



MATH 60638 – Méthodes de prévision

**Projet de semestre – Partie 1**

Présenté à

Pre. Debbie Dupuis

Par l'Équipe D

Maryvonne ANGELO – 11286237

Christelle GEORGE – 11288106

Samy SENOUNE – 11290255

Joe YOUNES – 11285603

Le 12 février 2020

À Montréal

# Introduction

Aujourd'hui, près de 82% de la population vit en milieu urbain aux États-Unis (Banque mondiale, 2021). Cette urbanisation a favorisé les échanges à toutes les échelles et permet la croissance économique des villes. Ceci dit, un défi de taille s'impose, puisque cette croissance est tributaire d'un facteur clé : l'approvisionnement en énergie. Il s'agit de posséder des infrastructures adéquates pour supporter cet approvisionnement. Une interruption d'apport d'énergie, même brève, peut avoir des conséquences lourdes, tant sur le plan économique que social. Le 13 juillet 1677 à New York, à titre d'exemple, un orage a causé une panne de 25 heures, touchant plus de 10 millions de personnes avec des dommages reportés de l'ordre de 300 millions de dollars. La plus grande panne d'électricité, quant à elle, a touché l'Inde le 31 juillet 2012 : il s'agit de 670 millions d'indiens, soit 10% de la population mondiale (Inmesol, 2014). Ceci met en perspective les enjeux auxquels l'industrie énergétique doit faire face. Il est donc crucial d'assurer une bonne prévision de la consommation d'électricité, puisque des erreurs dans la prévision engendrent des coûts importants. Dans ce projet, notre étude porte sur Lincoln, une ville américaine capitale de l'État du Nebraska.

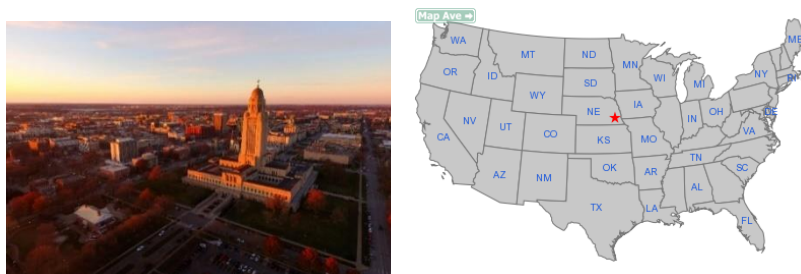


Figure 1 : Ville de Lincoln, Nebraska

Située dans le sud-est du Nebraska, cette ville possède une population de 287401 habitants, en date de 2018 (Bureau du recensement des États-Unis), pour une superficie de 242 km<sup>2</sup>. La ville

bénéficie d'un climat continental humide : les hivers sont froids et secs ; à contrario, les étés sont chauds et humides. D'autre part, l'activité de Lincoln est basée sur une économie de service : services administratifs, universités, banques, assurances et transport. Ainsi, les besoins en énergie de cette ville sont principalement divisés entre le secteur résidentiel et le secteur tertiaire.

Dans le cadre de ce projet, nous sommes amenés à prévoir le pic de la demande horaire d'électricité par jour pour la ville de Lincoln. Il s'agit d'une prévision à court-terme. Les données sont fournies par le Southern Power Pool (SPP), une compagnie qui gère le réseau électrique et le marché de l'électricité pour le centre des États-Unis. Ce premier rapport va s'articuler autour de trois points principaux. Dans un premier temps, une analyse exploratoire des données sera effectuée. Ensuite, plusieurs méthodes naïves seront évaluées en guise de benchmark. Finalement, une description des variables explicatives possibles sera menée afin de justifier le comportement observé.

## Analyse exploratoire des données

Les données fournies s'étalent du 1 janvier 2011, 1am au 31 décembre 2020, minuit. Dans un premier temps, il a fallu manipuler les données pour les adapter au bon fuseau horaire, de GMT à CST. Un graphique de la demande d'électricité horaire entre 2011 et 2021 dans la région de Lincoln a été tracé (Fig 2) afin d'avoir une idée générale de la variation de la demande d'électricité. Aucune valeur ne semble aberrante parmi les données : les statistiques sommaires indiquent un minimum de 223 MW et un maximum de 783 MW, ce qui coïncide avec la figure 2. Ceci dit, 7 valeurs

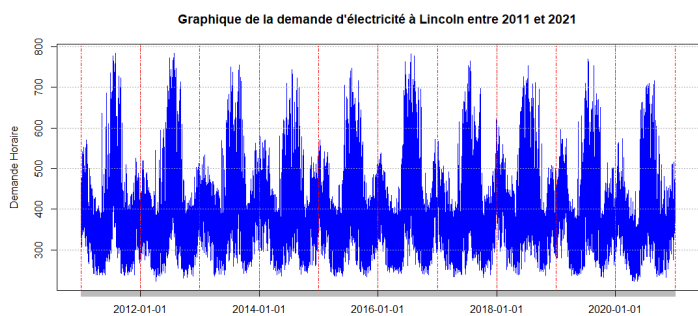


Figure 2 : Graphique de la demande d'électricité horaire à Lincoln

manquantes sont relevées parmi les données : la première en date du 27 septembre 2017 à 7pm, et

les autres le 12 décembre 2018 entre 2 et 7am. Si la valeur manquante associée à la première date reste inexpliquée, on voit que les autres valeurs manquantes correspondent à une plage horaire continue : une panne ou une réparation à la station pourrait expliquer ces valeurs manquantes. Ceci est plausible puisque le mois de décembre 2018 fut accompagné de fortes précipitations, avec plus de 6 cm au-dessus de la normale (Université de Nebraska, 2018). Toutes les données manquantes débutent un mercredi et nous avons choisi d'imputer ces valeurs par les données de la veille à la même heure. En effet, l'analyse exploratoire effectuée plus bas a mis en évidence que les demandes

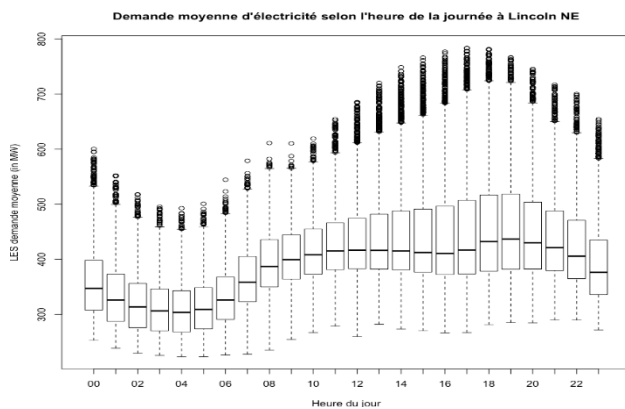


Figure 3 : Boxplot de la demande moyenne d'électricité horaire

d'électricité du mardi et du mercredi sont très similaires. De toute façon, l'imputation de ces valeurs aura peu d'impact pour la suite de notre étude. Comme l'indique la figure 3, la demande d'électricité est très faible durant la nuit (entre minuit et 6h, de 310 MW à 359 MW.) Elle semble ensuite suivre un plateau pour atteindre un maximum vers 18h, avec une valeur de 459 MW. Le pic de la demande est donc enregistré en fin d'après-midi : les valeurs imputées pour la plage horaire de 2 à 7am ne seront pas utilisées dans la série de la demande maximale journalière.

L'objectif de ce projet étant la prévision du pic de la demande horaire par jour, un graphique de la demande maximale journalière a également été produit (Fig 4). Les graphiques 2 et 4 illustrent le caractère saisonnier de la demande : la consommation d'électricité atteint un pic pendant la saison estivale à chaque année. Ce pic est précédé et suivi par des creux en termes de demande d'électricité. Enfin, un

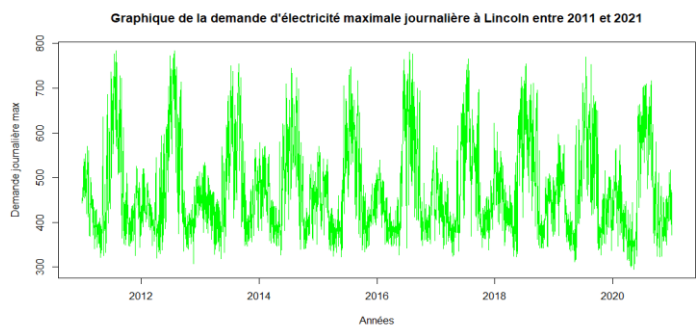


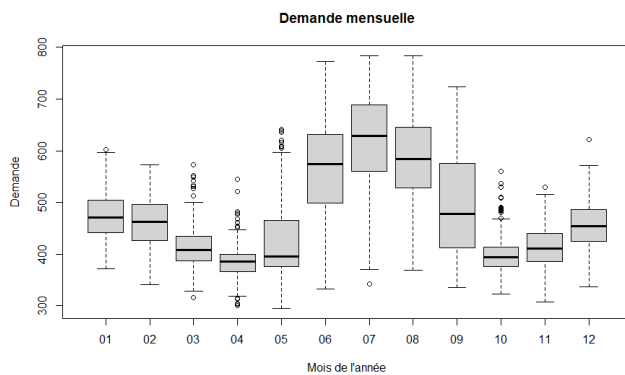
Figure 4 : Graphique de la demande maximale journalière à Lincoln

certain regain de demande d'électricité se manifeste en hiver. Ceci sera repris plus tard dans le rapport. Des statistiques sommaires de la demande maximale sont présentées dans le tableau suivant.

*Tableau 1 : Statistiques sommaires du pic de la demande d'électricité à Lincoln*

Min	1st Qu	Médiane	Moyenne	3rd Qu	Max
295.5	397.9	449.3	474.7	529.9	783.5

La demande moyenne du pic d'électricité mensuelle (Fig 5) confirme la tendance expliquée plus



*Figure 5 : Boxplot du pic de la demande moyenne mensuelle d'électricité à Lincoln*

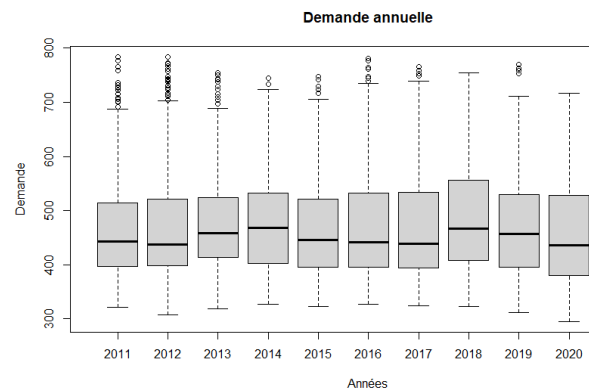
octobre (384 et 397 MW respectivement).

On s'intéresse à présent à la variation du pic de la demande annuelle en électricité (Fig 6) : il n'y a pas de variation significative entre les années. L'année 2018 représente l'année pour laquelle la demande

était, en moyenne, la plus grande (489 MW). Il s'agit aussi

de l'année qui a battu des records de chaleur à une échelle mondiale, particulièrement avec la canicule nord-américaine qui a certainement affecté Lincoln (Le Devoir, 2018). À contrario, l'année 2020 enregistre la plus petite demande (463 MW). Ceci serait peut être lié au confinement en réponse à la pandémie de la Covid-19, et sera exploré plus tard plus en détails.

haut : les mois de juin, juillet et août représentent la période où la consommation est la plus grande. Elle atteint un pic de moyenne 623 MW en juillet. À l'inverse, les deux mois où le pic de la consommation moyenne d'énergie est le plus faible sont en avril et



*Figure 6 : Boxplot du pic de la demande moyenne annuelle d'électricité à Lincoln*

D'autre part, il serait intéressant d'étudier la variation de la demande au cours de la semaine (Fig

7). Le graphique suivant indique une différence entre la demande d'énergie durant la semaine et la fin de semaine.

Le pic de la demande d'électricité chute en fin de semaine, passant de plus de 490 MW du lundi au vendredi

à 440 MW en moyenne le samedi et

dimanche. Encore une fois, ceci sera repris plus tard dans le rapport.

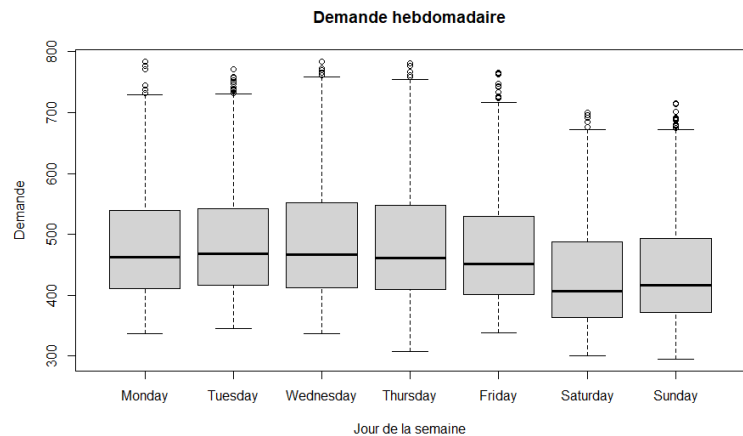


Figure 7 : Boxplot du pic de la demande moyenne hebdomadaire d'électricité à Lincoln

## Évaluation des méthodes naïves

Quatre méthodes naïves différentes ont été comparées. L'analyse exploratoire ayant mis en évidence une tendance similaire pour chaque année sans valeurs aberrantes, toutes les données ont donc été conservées. Ceci est important puisque les méthodes naïves se basent sur les données passées pour faire des prédictions futures. Ensuite, l'ensemble des observations est divisé en trois sous-groupes : l'échantillon d'entraînement (du 1er janvier 2011 au 31 décembre 2016), l'échantillon validation (du 1er janvier 2017 au 31 décembre 2018) et l'échantillon test (du 1er janvier 2019 au 31 décembre 2020). Cette séparation en sous-échantillons sera surtout utilisée dans les méthodes de la Partie 2 du rapport. Les méthodes naïves ne nécessitant pas de phase d'entraînement, elles ont été appliquées sur l'ensemble des observations. Cela étant dit, leurs mesures de performance sont évaluées sur l'échantillon de validation : ceci permettra de comparer

plusieurs méthodes sur ce même échantillon, et d'éliminer le biais - dans une certaine mesure – des méthodes.

Les quatre méthodes utilisées sont le *naïve no change*, le *naïve seasonal*, la moyenne mobile sur 3 jours, puis la moyenne mobile sur 7 jours. La méthode de *naïve no change* consiste à prendre la valeur observée du pic de la veille et l'utiliser comme prévision pour la journée même :  $\hat{y}_{T+1|T} = y_T$ . Cette méthode est optimale dans le cas de données de marche aléatoire. Ensuite, la méthode de *naïve seasonal*, dans notre cas, consiste à prendre la valeur observée de l'année passée pour prédire la valeur d'aujourd'hui :  $\hat{y}_{T+365|T} = y_T$ . Notre hypothèse est que cette méthode ne serait pas très performante, puisque comme l'a montré l'analyse exploratoire, le jour de la semaine (entre autres) est important pour estimer la demande d'électricité. De plus, les jours individuels ne sont pas les mêmes jours de la semaine d'une année à l'autre. Ainsi, des saisonnalités autres que la saisonnalité annuelle devraient être considérées. Puis, pour la moyenne mobile de 3 jours, nous utilisons la moyenne de valeurs observées des demandes d'électricité des 3 derniers jours comme prévision :  $\hat{y}_{T+1|T} = (y_T + y_{T-1} + y_{T-2})/3$ . C'est le même principe pour la moyenne mobile sur 7 jours. Ces deux méthodes nous assurent des prévisions qui varient moins d'un jour à l'autre, puisqu'elles regroupent une moyenne d'observation plutôt qu'une seule observation. Les résultats sont regroupés dans le tableau 2.

Nous observons que la méthode de *naïve no change* est la meilleure en termes de performance,

Tableau 2 : Performance des méthodes naïves

	<b>BIAIS</b>	<b>RMSE</b>	<b>MAE</b>	<b>BIAIS%</b>	<b>MAPE</b>
<b>Naïve No Change</b>	-0.103	53.989	38.971	0.564	8.081
<b>Naïve Seasonal</b>	-5.151	84.135	61.003	0.154	12.439
<b>Moyenne mobile 3 jours</b>	-0.16	61.77	46.38	1.02	9.73
<b>Moyenne mobile 7 jours</b>	-0.265	63.00	46.383	1.239	9.656

pour tous les indicateurs sauf le biais en pourcentage. En fait, les données de la demande d'électricité s'apparentent à des données de marche aléatoire par leurs longues périodes et les changements soudains présents. Les principales mesures à noter sont le RMSE et le MAPE : la méthode *naïve no change* possède respectivement une valeur de 53.989 et 8.081.

La méthode de *naïve seasonal* représente la pire méthode, ce qui confirme les raisons mentionnées plus haut. Finalement, les deux moyennes mobiles ont des résultats très similaires, mais la moyenne mobile sur 7 jours semble être plus biaisée : ceci s'explique par le fait qu'il lui faut plus de temps pour s'adapter à un changement que la moyenne mobile sur 3 jours. Somme toute, la méthode de *naïve no change* prend la première place, et la moyenne mobile de 3 jours puis de 7 jours

complètent le podium. Le *naïve seasonal* se retrouve en dernière place.

Le graphique illustre clairement tous les résultats énoncés, pour un

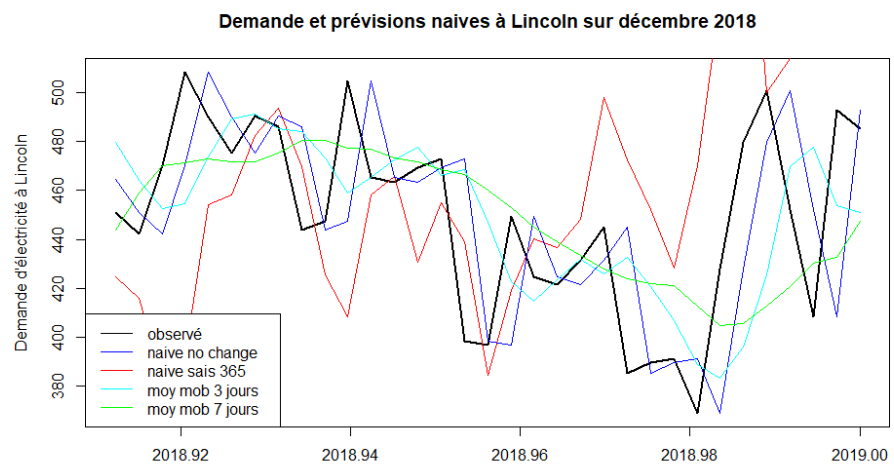


Figure 8 : Demande observée vs prédite naïvement pour le mois de décembre 2018

mois aléatoire (décembre 2018) parmi l'échantillon validation. Le *naïve no change* suit naturellement de très près les valeurs observées : on voit aussi que l'erreur sera plus petite pour des variations minimales de demande d'une journée à l'autre. Les méthodes naïves de moyennes mobiles ont toutes les deux des courbes bien plus lisses : elles ne sont pas en mesure de capturer correctement des pics et des variations abruptes de demande. Plus la durée de la moyenne augmente, moins la méthode prédit correctement. Finalement, le *naïve seasonal* ne semble pas avoir



de chance contre les autres modèles : sa courbe en rouge est beaucoup trop éloignée des valeurs observées en noir.

## Description des variables explicatives possibles

### Variables météorologiques

La météo a certainement un rôle à jouer quant à la prévision de la demande du pic journalier d'électricité. Les données météorologiques proviennent de National Centers for Environmental Information (NOAA). Les 7 températures manquantes ont été imputées avec leurs valeurs de la

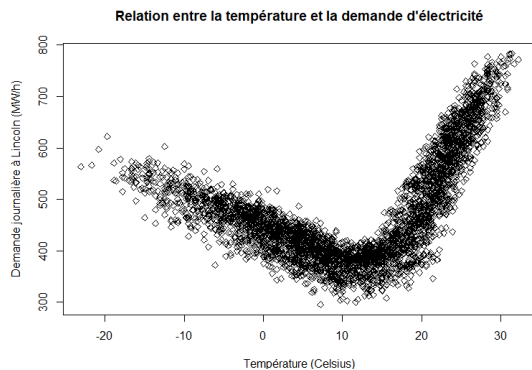


Figure 9 : Relation entre la température moyenne et la demande d'électricité à Lincoln

veille. Ensuite, la température moyenne a été obtenue avec la température minimale et maximale fournie. Le graphique 9 est révélateur : la demande en électricité augmente lorsque la température diminue à partir de 14°C, et elle augmente encore plus rapidement lorsque la température augmente à partir de 14°C : de hautes températures requièrent plus d'électricité que de basses températures. Pour approfondir cette analyse, deux variables additionnelles sont créées : le Heating degree-days HDD et le cooling degree-days CDD. Elles représentent la demande d'énergie pour chauffer ou refroidir les bâtiments, respectivement. La température de référence est fixée à 14°C (Fig 9) :

$$CDD = \begin{cases} T_{moy} - 14, & \text{si } T_{moy} > 14 \\ 0 & \text{sinon} \end{cases} \quad HDD = \begin{cases} 14 - T_{moy}, & \text{si } T_{moy} < 14 \\ 0 & \text{sinon} \end{cases}$$

La relation entre le CDD et la demande est linéaire, celle entre le HDD et la demande semble quadratique. Deux différenciations successives de décalage 1 permettent de retrouver une tendance linéaire (Fig 10).

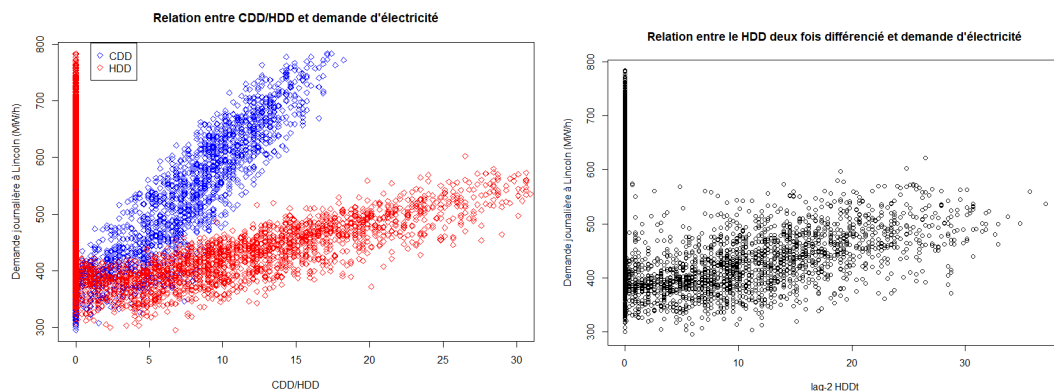


Figure 10 : Relation entre CDD/HDD et la demande (GAUCHE) et graphique du HDD différencié deux fois (DROITE)

Somme toute, les différentes saisons de l'année illustrent ces différences de températures. Ainsi, une variable catégorielle est créée pour indiquer la saison de l'année ; les données de la demande sont analysées (Fig 11). La demande est maximale pendant l'été, avec une moyenne de 582 MW et un pic de 623 MW atteint au mois de juillet plus particulièrement. L'hiver représente la deuxième saison où la demande est élevée, mais dans une moindre mesure : en moyenne, 454 MW. Finalement, l'automne et le printemps sont deux saisons où la demande est minimale. Le graphique de gauche met en évidence que la demande estivale est bien plus grande que la demande hivernale. Ceci vient confirmer ce qui a été dit dans le paragraphe précédent : l'utilisation de l'air climatisé en été pour lutter contre les températures chaudes demande bien plus d'électricité que le chauffage en hiver contre le froid.

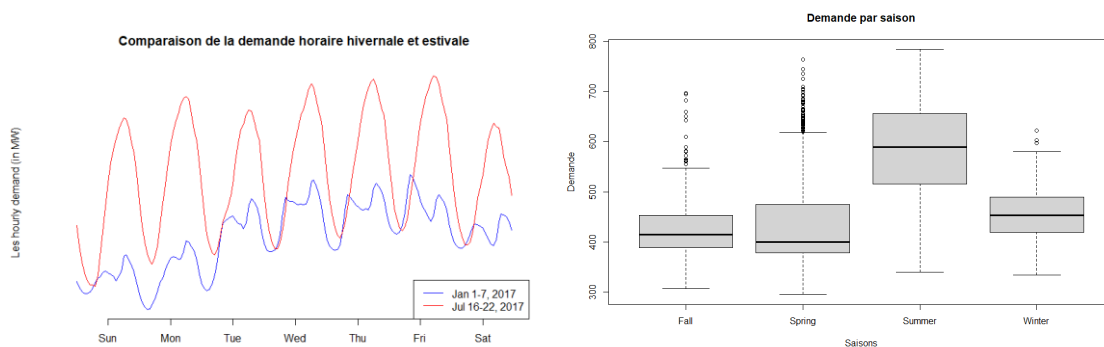


Figure 11 : Comparaison du pic de la demande horaire hivernale et estivale (GAUCHE) et boxplot du pic de la demande d'électricité par saison (DROITE)

## Variables chronologiques

D'autre part, l'analyse exploratoire a mis en évidence l'influence du temps sur le pic de la demande d'électricité. Dans cette section, nous nous intéressons au temps comme construction sociale tel que défini par l'Homme :

- L'heure de la journée : « hour » allant de 0 (minuit) à 23
- Le jour de la semaine : « weekday » allant de 1 (lundi) à 7 (dimanche)
- Le mois de l'année : « month » allant de 1 (janvier) à 12 (décembre)

Ces variables sont créées directement à partir de la date des données d'électricité fournies : aucune source n'était nécessaire. Comme découvert lors de l'imputation des valeurs manquantes au début de l'analyse exploratoire, l'heure de la journée influe sur la demande horaire avec un pic recensé à 18h (Fig 3). Cette variable est uniquement mentionnée par soucis de complétude : elle ne sera pas approfondie puisque la demande maximale journalière est à l'étude dans ce rapport.

En ce qui concerne le jour de la semaine,

la figure 7 indiquait que la demande d'électricité était plutôt constante pendant les jours de semaine et diminuait

pendant les jours de fin de semaine. Une

analyse supplémentaire indique que, bien

que la demande augmente

considérablement en été, cette tendance se

manifeste quelle que soit la saison (Fig 12). Cette chute de demande coïncide avec le moment où

la plupart de la population est en congé lors de la fin de semaine : ce sujet sera approfondi dans la

suite.

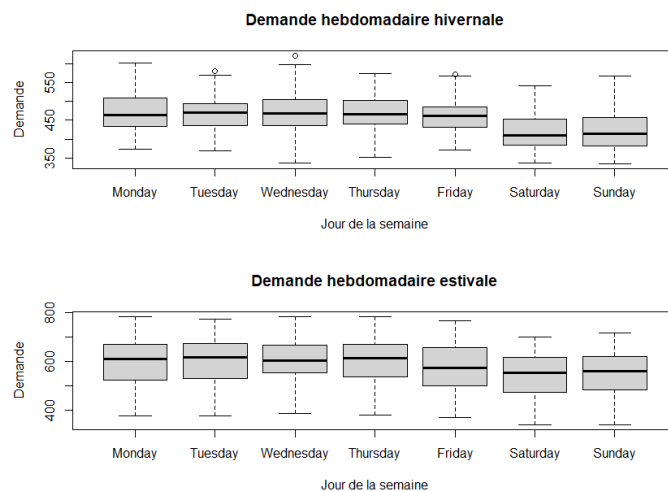


Figure 12 : Comparaison du pic de la demande hebdomadaire moyenne en hiver et en été

Finalement, l'influence du mois de l'année sur la demande est visualisée dans la figure 5 plus haut. Ceci est repris sur la figure 13 : quelle que soit l'année (2011, 2015 ou 2020 sur le graphique), la demande est maximale pendant le mois de juillet (623 MW) et minimale en avril (384 MW). La tendance est identique et dépend donc du mois.

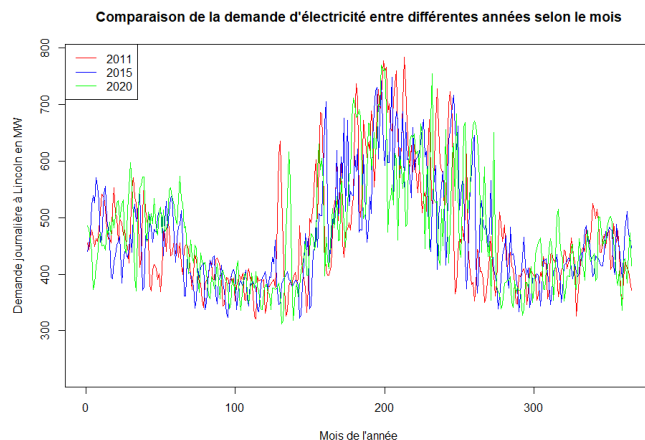


Figure 13 : Comparaison du pic de la demande en fonction de l'année selon le mois

## Jour de congés

Une variable binaire ayant pour but de préciser un congé a été créée dans le jeu de données. Elle regroupe deux conditions : il s'agit d'un jour férié ou d'un jour de fin de semaine. L'analyse a été menée sur les trois premières années des données, soit entre 2011 et 2013. Les jours fériés regroupent : le premier jour de l'an, le jour de Martin Luther King Junior, le jour du Président, le Memorial Day, l'indépendance, la journée du travail, Columbus Day, la journée des anciens combattants, l'action de grâce et Noël. Cette variable s'est montrée pertinente puisque comme l'indique la figure 13, la demande d'électricité

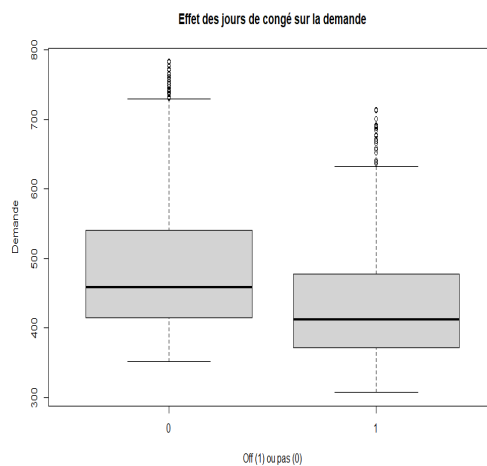


Figure 14 : Boxplot du pic de la demande moyenne d'électricité en fonction d'un jour de congé (1) ou non (0)

est plus faible pendant les jours de congé. Ceci serait dû au fait que, bien que les individus consomment de l'électricité à la maison pendant les jours fériés, les entreprises et les bureaux sont fermés, et cette fermeture entraîne une diminution significative de la demande d'électricité. Finalement, la question de l'effet du confinement sur la demande en électricité s'est posée mais a

rapidement été écartée : le Nebraska fait partie des cinq états parmi les 50 états américains qui n'a pas imposé de confinement (La Presse, 2020).

## Autres

Nous avons finalement exploré une possible corrélation entre certaines variables économiques (revenu moyen par habitant, taux de chômage) et démographiques avec la demande en électricité. Ceci dit, la période à l'étude étant assez courte, aucun lien n'a pu être explicité à ce stade : la ville de Lincoln n'a pas subi de changements majeurs entre 2011 et 2020.

## Références

- Banque mondiale, 2020. *Population urbaine*. Tiré de <https://donnees.banquemondiale.org/indicateur/SP.URB.TOTL.IN.ZS>
- Inmesol, 2014. *Les plus grosses pannes d'électricité de l'histoire*. Tiré de <http://www.inmesol.fr/blog/les-plus-grosses-pannes-deelectricite-de-lhistoire>
- Wikipédia, 2020. *Lincoln (Nebraska)*. Tiré de [https://fr.wikipedia.org/wiki/Lincoln\\_\(Nebraska\)](https://fr.wikipedia.org/wiki/Lincoln_(Nebraska))
- Wikipédia, 2021. *Électricité aux États-Unis*. Tiré de [https://fr.wikipedia.org/wiki/%C3%89lectricit%C3%A9\\_aux\\_%C3%89tats-Unis](https://fr.wikipedia.org/wiki/%C3%89lectricit%C3%A9_aux_%C3%89tats-Unis)
- Lincoln weather and Climate, 2018. *Decembre 2018 Data*. Tiré de <https://lincolnweather.unl.edu/december-2018-data>
- Le devoir, 2018. *Des records de chaleur dans tout l'hémisphère nord*. Tiré de <https://www.ledevoir.com/societe/environnement/533211/des-records-de-chaleur-dans-tout-l-hemisphere-nord>
- NOAA, 2021. *Climate Data Online Search*. Tiré de <https://www.ncdc.noaa.gov/cdo-web/search>
- La Presse, 2021. *Des États rouges refusent toujours le confinement*. Tiré de <https://www.lapresse.ca/international/etats-unis/2020-04-10/des-etats-rouges-refusent-toujours-le-confinement>
- Sunset Bld, 2020. *Jours fériés et fêtes aux USA*. Tiré de [https://www.sunsetbld.com/preparer-voyage-usa/jours-feries-fetes/#:~:text=12%20f%C3%A9vrier%203A%20Lincoln's%20Day%20\(Colorado,aussi%20le%20premier%20pr%C3%A9sident%20r%C3%A9publicain.](https://www.sunsetbld.com/preparer-voyage-usa/jours-feries-fetes/#:~:text=12%20f%C3%A9vrier%203A%20Lincoln's%20Day%20(Colorado,aussi%20le%20premier%20pr%C3%A9sident%20r%C3%A9publicain.)