



MATH 60638 – Méthodes de prévision

**Projet de semestre – Partie 3**

*Évaluation des méthodes ARIMA / SARIMA, modèle ARX,  
comparaisons globales et recommandations*

Présenté à

Pre. Debbie Dupuis

Par l'Équipe D

Christelle GEORGE – 11288106

Samy SENOUNE – 11290255

Le 16 avril 2021

À Montréal

## Objectif

Dans la partie 2 du projet, il était question d'évaluer les méthodes de lissage et de régression pour la prévision du pic de la demande horaire pour jour  $t + 1$  à Lincoln. Le modèle TBATS  $(0, \{4,0\}, -, \{< 7,3 >, < 365.25,5 >\})$  représente notre candidat pour les méthodes de lissage, avec un MAPE obtenu de 6.5% sur la fenêtre expanding. Pour les méthodes de régression, le modèle de régression linéaire aux erreurs AR(1) est sélectionné, avec un MAPE de 4.1% obtenu par entraînement unique. Le but de la partie 3 est d'évaluer les méthodes ARIMA/SARIMA, d'effectuer des comparaisons globales afin d'émettre des recommandations et de conclure. À titre de rappel, l'ensemble des observations est divisé en trois groupes : l'échantillon d'entraînement (du 1er janvier 2011 au 31 décembre 2016), de validation (du 1er janvier 2017 au 31 décembre 2018) et de test (du 1er janvier 2019 au 31 décembre 2020).

## Méthodes ARIMA/SARIMA

**Méthodologie** – Une étude de la stationnarité de la série chronologique du pic de la demande est effectuée : il s'agit d'une condition importante pour l'application de ces méthodes. La famille ARIMA est écartée en raison de la saisonnalité apparente de la série du pic de la demande. Ensuite, le modèle sera recherché de façon itérative. Deux types de fenêtres seront employées pour la prévision: une rolling et une expanding. D'autre part, on comparera l'approche où le modèle est entraîné une unique fois avec celle où le modèle est réentraîné à chaque pas de temps (réentraînement journalier). Finalement, le test de Diebold-Mariano est employé à un niveau 5% afin de départager des ensembles de prévisions entre eux.

**Stationnarité** – La série du pic de la demande n'est pas une série stationnaire : son graphique indique une espérance et une variance qui varie en fonction du temps, particulièrement en été, ce

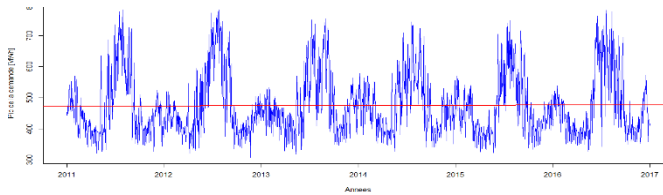


Figure 1 : Graphique de la série du pic de la demande

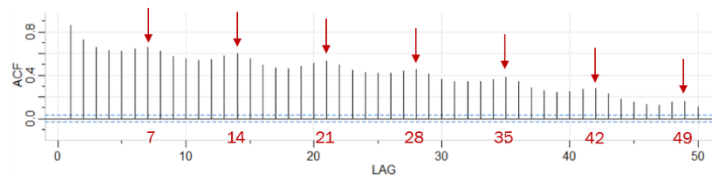


Figure 2 : ACF de la série du pic de la demande

qui viole les propriétés de la stationnarité. De plus, la saisonnalité est évidente. L'allure de l'ACF témoigne de sa non-stationnarité, avec des pics récurrents aux décalages 7 : la différenciation semble nécessaire pour enlever la composante saisonnière. Avant de procéder à la différenciation, on a recours à une transformation BoxCox. Avec un lambda obtenu de -0.57, cette transformation accorde une allure plus « normale » aux valeurs de la série et permet d'améliorer la stabilité de la variance, comme l'indique la figure 3. La série est ensuite différenciée d'ordre 1 à décalage 7, ce qui se traduit par des paramètres saisonniers de SARIMA de  $D=1$ ,  $S=7$ . Ce faisant, on cherche à éliminer l'impact de la saisonnalité hebdomadaire. Cette

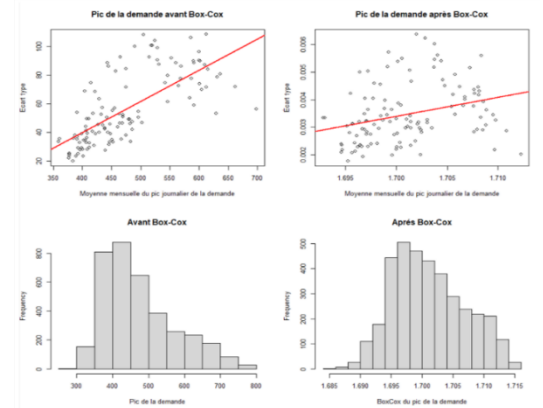


Figure 3 : Impact de la transformation BoxCox sur la série du pic de la demande

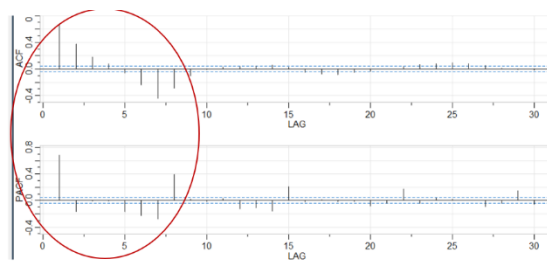


Figure 4 : ACF/PACF de la série du pic de la demande différenciée à lag 7

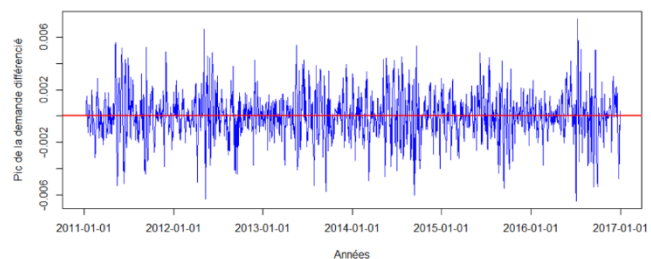


Figure 5 : Graphique de la série du pic de la demande différenciée à lag 7

différenciation permet d'obtenir un graphique ACF/PACF bien plus convenable. Aussi, la figure 5 indique que l'aspect saisonnier est éliminé et que l'espérance de la série est nulle en tout temps. Cependant, la variance n'est toujours pas constante, quel que soit le traitement effectué. Une procédure itérative est poursuivie par la suite afin de trouver le modèle SARIMA le plus adéquat possible : le modèle SARIMA  $(2,0,3) * (0,1,1)_7$ , appliqué sur la série différenciée à lag 7 avec

une transformation BoxCox, permet d'obtenir les diagnostics graphiques les plus convenables.

**Analyse des résultats** – Les diagnostics graphiques des résidus indiquent une hétéroscédasticité, particulièrement lors des mois d'été, associés à de plus grandes erreurs de prévision. D'autre part,

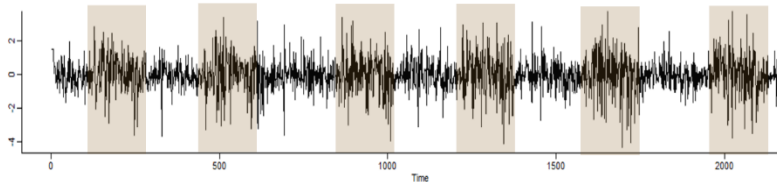


Figure 6 : Graphique des résidus standardisés

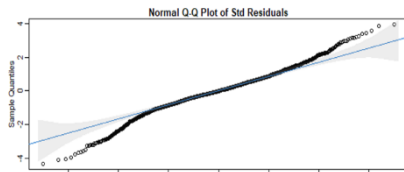


Figure 7 : Diagramme QQ des résidus du modèle

le graphique quantile-quantile des résidus adopte un comportement « heavy tailed », ce qui ne respecte pas la condition de normalité. Il traduit une trop grande proportion de grands résidus, positifs et négatifs. Ceci dit, l'ACF des résidus est convenable : il n'y a pas d'autocorrélation dans les résidus. Finalement, le test de Ljung-Box

obtient des valeurs convenables de p-values jusqu'au lag 14, ce qui appuie l'hypothèse nulle des erreurs indépendantes. Somme toute, le modèle SARIMA

ne parvient pas à bien capter les différentes dynamiques des saisons, notamment celle de la saison estivale. Ceci dit,

il s'agit du meilleur modèle obtenu suite aux nombreuses

itérations effectuées. En terme de prévisions, le tableau 1 indique

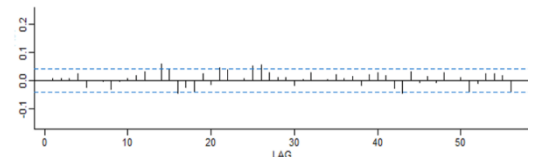


Figure 8 : ACF des résidus du modèle

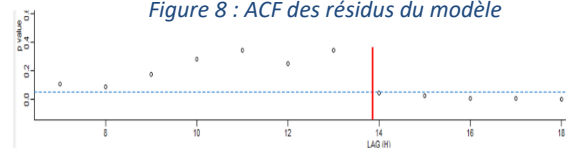


Figure 9 : Test Ljung-Box du modèle

les performances du modèle obtenues avec un entraînement unique ainsi qu'un réentraînement journalier, sur une fenêtre expanding. Le test bilatéral de Diebold Mariano à niveau 5% donne une

Tableau 1 : Performance du modèle SARIMA sur l'ensemble de validation

| Entraînement | RMSE   | MAPE  | ME     |
|--------------|--------|-------|--------|
| Unique       | 46.486 | 6.658 | -3.360 |
| Journalier   | 46.575 | 6.664 | -3.225 |

p-value de 0,112 : l'erreur quadratique moyenne des prévisions est la même, quelle que soit la fréquence

du réentraînement. Les observations de l'ensemble de validation ne rajoutent donc pas de pouvoir prédictif au modèle. Ainsi, le modèle choisi est le modèle **SARIMA (2, 0, 3) \* (0, 1, 1)<sub>7</sub>** entraîné une unique fois, puisque le réentraînement journalier requiert bien plus de temps sans améliorer la performance. Le modèle choisi permet d'obtenir un **MAPE de 6.66%**, ce qui correspond à une

amélioration de 17.7% par rapport à notre benchmark naïve no change (MAPE de 8.09%). En dernier lieu, 53 observations sur les 730 de validation, soit 7,3%, ne se trouvent pas dans l'intervalle de prédiction à niveau 95% : l'ajustement du modèle a mis en évidence le non-respect de la condition de normalité avec le comportement « heavy-tailed » observé. Il s'agit cependant d'une valeur acceptable. Il est important de mentionner que les résultats de la fenêtre rolling ne sont pas discutés dans ce rapport car ils sont sensiblement les mêmes : le type de fenêtre n'a pas d'impact sur la performance de la prévision du pic de la demande avec le modèle SARIMA.

## **Modèles ARX**

**Méthodologie** – La méthodologie adoptée est similaire à celle des méthodes SARIMA : il sera question d'évaluer le type de fenêtre (rolling et expanding) ainsi que la fréquence de l'entraînement (unique et journalier). D'autre part, les variables exogènes choisies sont les suivantes : HDD, CDD, leurs lags 1 et 2, le refroidissement du vent CP, les « dummy » pour les jours de la semaine, les jours fériés, veille et lendemain des jours fériés ainsi que des termes autorégressifs du lag de la demande. Deux stratégies sont considérées pour ces termes autorégressifs : d'une part, on considère uniquement le lag1 de la demande (modèle ARX1). D'autre part, on considère toute la semaine passée, soit les lags 1 à 7 du pic de la demande (modèle ARX7). Les « dummy » pour les mois ont été considérés initialement mais suite à votre remarque sur leur pertinence (en raison de leur corrélation avec les variables de température), ils ont été enlevés de l'analyse : leur retrait du modèle a montré une légère amélioration de la performance. Finalement, du bruit est ajouté aux variables de températures moyennes et de vents selon la même méthodologie qu'énoncée dans la partie 2 : bruit de loi normale  $N(0, 0.5)$  °C pour la température et  $N(0, 1.5)$  km/h pour le vent.

**Analyse des résultats** – Le **modèle ARX7** correspond au meilleur modèle obtenu, en terme de performance. Ceci dit, l'ACF obtenu des résidus n'est pas satisfaisant : on note des pics récurrents

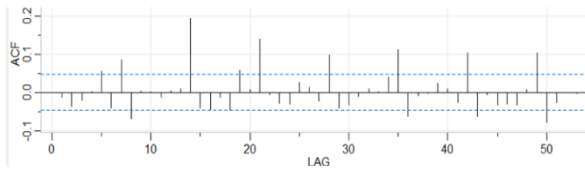


Figure 10 : ACF des résidus du modèle ARX7

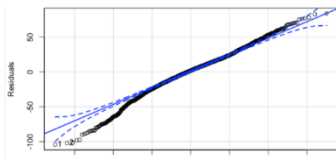


Figure 11 : Diagramme QQ des résidus du modèle ARX7

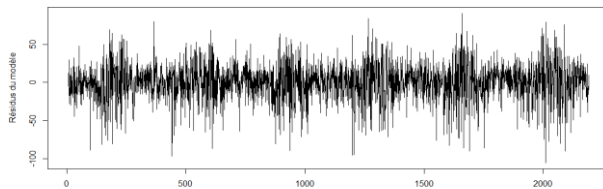


Figure 12 : Résidus du modèle ARX7

le cadre de la prévision du pic de la demande. Sa performance est résumée dans le tableau 2. La

fenêtre rolling permet l'obtention d'un meilleur MAPE que la fenêtre expanding dans le cas d'un

Tableau 2 : Performance du modèle ARX7 sur l'ensemble de validation

| Réentraînement         | RMSE   | MAPE  | ME     |
|------------------------|--------|-------|--------|
| Unique initial         | 28.413 | 4.410 | -1.545 |
| Journalier – EXPANDING | 28.378 | 4.786 | -1.390 |
| Journalier – ROLLING   | 28.404 | 4.454 | -1.011 |

modèle ARX7 sur les 6 années de l'échantillon d'entraînement initial : le **MAPE atteint vaut**

**4.41%** et représente une amélioration significative de 45% par rapport à notre benchmark naïf !

Somme toute, les observations de validation n'accordent pas davantage de pouvoir prédictif au

modèle. En dernier lieu, 67

observations de validation – soit

9% de l'échantillon – se retrouvent

en dehors de l'intervalle de

prévision à niveau 95%. La bonne

performance de ce modèle est

perceptible dans la figure 13 : on

voit que la courbe rouge des prédictions ARX 7 est

aux décalages 7 : la saisonnalité persiste dans la

structure des résidus et induit de l'autocorrélation dans

les résidus. Le diagramme quantile-quantile indique

que le postulat de normalité n'est pas respecté : il met en évidence une

grande proportion de grands résidus négatifs, soit de surestimation du

pic de la demande. En dernier lieu, le graphique des résidus du modèle

traduit le problème d'hétéroscédasticité bien marqué

pendant la saison estivale. Bien que ce modèle ne

respecte pas les conditions de départ, il est utilisé dans

réentraînement journalier du modèle. Ceci dit, la meilleure

performance est obtenue pour un entraînement unique du

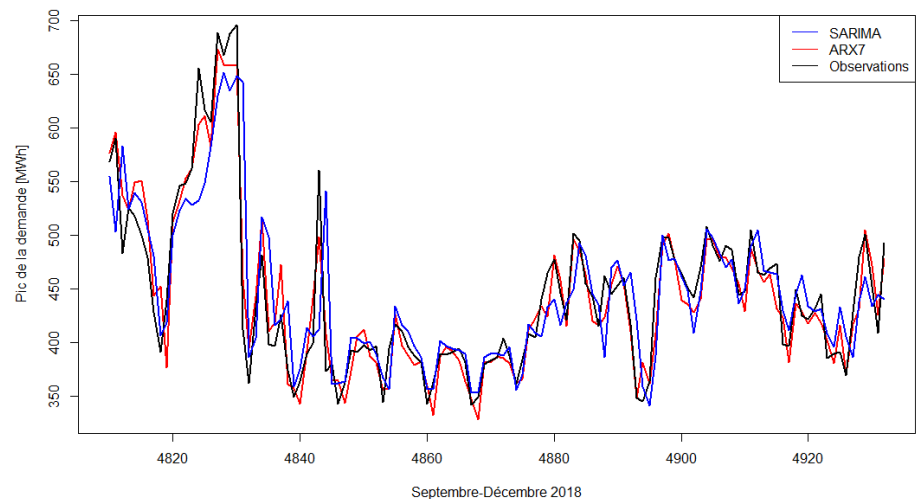


Figure 13 : Comparaison de la prévision du pic de la demande avec le modèle SARIMA vs le modèle ARX7

bien plus proche de la courbe noire des observations que la courbe bleu du modèle SARIMA, entre septembre et décembre 2018.

## Comparaisons globales

Les modèles candidats sont à présents comparés entre eux pour déterminer le meilleur. À titre de rappel, il y a 5 modèles à l'étude. Le benchmark est le modèle naïf no change. Le modèle TBATS est le modèle retenu quant aux méthodes de lissage. Pour la famille des méthodes de régression linéaire, le modèle candidat est celui aux erreurs AR(1). Finalement, les deux modèles vu dans ce rapport sont le SARIMA  $(2,0,3) * (0,1,1)_7$  et le modèle ARX7 précédemment décrits. Plusieurs

tests successifs de Diebold-Mariano unilatéral (alt = less ou greater) permettent de déterminer le

Tableau 3 : Performance de tous les modèles candidats sur l'ensemble de validation

| Classement | MODÈLES                               | MAPE (%) | ME (MWh) |
|------------|---------------------------------------|----------|----------|
| 1          | Régression linéaire aux erreurs AR(1) | 4.09     | 1.61     |
| 2          | Modèle ARX 7                          | 4.43     | -1.57    |
| 3          | TBATS                                 | 6.45     | 2.68     |
| 4          | SARIMA $(2, 0, 3) * (0, 1, 1)_7$      | 6.66     | -3.36    |
| 5          | Naïf no change (benchmark)            | 8.09     | -0.11    |

classement de ces méthodes : c'est la **régression**

**linéaire aux erreurs AR(1)** qui se hisse en

première position, avec un MAPE de 4.1% sur la

validation. Afin d'éliminer la mesure de performance de tout biais et obtenir le MAPE le plus représentatif de l'erreur de prévision, la performance est désormais évaluée sur les deux années 2019-2020 de l'échantillon test : le MAPE obtenu est de 5,06%. Ceci dit, nous portons un regard

Tableau 4 : Performance du modèle de régression sur l'ensemble de test

|          | Validation | Test 2019-2020 | Test 2019 | Test 2020 |
|----------|------------|----------------|-----------|-----------|
| MAPE (%) | 4.09       | 5.06           | 4.38      | 5.72      |
| ME (MWh) | 1.61       | 12.45          | 10.11     | 14.75     |

critique sur cette valeur qui semble assez grande,

d'autant plus que l'année 2020 représente une année

particulière. Une analyse plus poussée révèle que le mois de mars 2020 amorce un déclin significatif du MAPE pour le reste de 2020. En d'autres termes, la mauvaise performance sur 2020 s'explique par le confinement instauré à Lincoln en

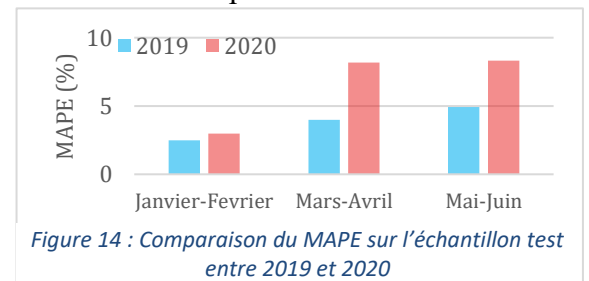


Figure 14 : Comparaison du MAPE sur l'échantillon test entre 2019 et 2020

mars, qui a changé les habitudes de vie de la population : la pandémie de Covid-19 a eu un impact significatif sur le pic de la demande d'électricité. Ainsi, l'échantillon test est révisé et sera constitué

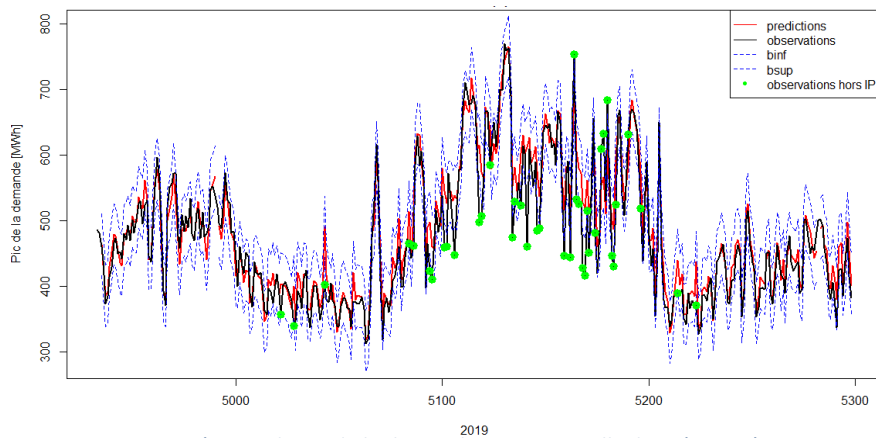


Figure 14 : Prédiction du pic de la demande avec intervalle de prévision à niveau 95% avec le modèle gagnant de régression linéaire aux erreurs AR(1) sur l'année 2019 de l'échantillon test

acceptable. La figure 14 met en exergue, une fois de plus, la dynamique de la saison estivale difficilement capturée par le modèle, même le plus performant : les points verts indiquent l'ensemble des observations situées hors de l'intervalle de prévision à niveau 95%.

## Conclusion et recommandation

*In fine*, bien que le **modèle de régression linéaire aux erreurs AR(1)** ne respecte pas les conditions de normalité ni d'homoscédasticité, il s'agit du modèle qui performe le mieux quant à la prévision du pic de la demande horaire d'électricité à Lincoln. D'autre part, la **température** est un **facteur très important** à considérer : les deux modèles considérant la température (régression linéaire et ARX) obtiennent une performance significativement plus élevée que les autres modèles, avec un MAPE respectif de 4,1% et 4,4% (Tableau 3). La dynamique du pic de la demande est fortement liée à la dynamique des saisons : considérer la température rajoute un grand pouvoir prédictif au modèle. Finalement, un **problème d'hétéroscédasticité** récurrent est relevé, particulièrement en été : les températures chaudes engendrent un besoin de climatisation important qui se traduit par de plus grandes erreurs dans la prévision du pic de la demande. Il serait pertinent d'envisager une **méthode d'ensemble** en créant un **modèle par saison**, afin de capturer la dynamique adéquate et d'améliorer la performance.