



MATH 60638 – Méthodes de prévision

Projet de semestre – Partie 3

Présenté à Pre. Debbie Dupuis

Par l'Équipe G

Abderezak AMIMER – 11241872

Maryvonne ANGELO – 11286237

Franck BENICHOU – 11249092

Théo LEFEBVRE – 11314432

Joe YOUNES – 11285603

Note : 30,5/35

Note présentation orale : 20,2/25

Le 14 avril 2022 à Montréal

Objectif

La deuxième partie du projet consistait à évaluer les méthodes et modèles de lissage ainsi que les modèles de régression multiple à erreurs ARMA pour la prévision de la somme de la demande horaire pour jour $t+1$ dans les zones *North*, *North Central* et *East* du Texas selon la performance obtenue. Il a été conclu que TBATS est le plus approprié pour la région étudiée avec un MAPE de 5.35% et une couverture réelle de 94.07% pour un intervalle de prévision à 95%. Les paramètres du modèle étaient les suivants : $(0.772, \{1,2\}, -, \{<7,3>, <3625.25,2>\})$. Toutefois, le meilleur modèle de régression linéaire ajusté n'était pas adéquat et il a été corrigé pour cette troisième partie. Celui-ci possède maintenant la bonne structure d'erreur, c'est-à-dire, une erreur ARMA (5,2). Il inclut 13 variables explicatives dont le CDD_t , CDD_{t-1} , CDD_{t-2} , HDD_t , HDD_{t-1} , HDD_{t-2} , CDD_t , CDD_{t-1} , CDD_{t-2} , HDD_t , HDD_{t-1} , HD_t , le jour de la semaine, $Humidext$, CPT_t , les journées fériées, les journées qui précèdent et qui suivent un jour férié et, pour terminer, les mois de l'année. Le MAPE obtenu est de 4.11% avec une couverture réelle de 84.52% pour un intervalle de prévision à 95%. Cette performance a été obtenue en utilisant un « *moving window* » commençant le 1^{er} janvier 2012 et se terminant le 31 décembre 2019 avec un réentraînement quotidien. Pour cette dernière phase, il est question d'évaluer les modèles ARIMA/SARIMA et ARX. À titre de rappel, la série chronologique utilisée est divisée en trois segments : la plage d'entraînement, allant de 2012 à 2017, la plage de validation, allant de 2018 à 2019 et la plage de test, allant de 2020 à 2021 qui contient le confinement du 23 mars 2020 au 01 mai 2020 à cause de la crise Covid-19. Il faut donc garder l'impact de la COVID-19 en tête quand viendra le temps de mesurer la performance sur la plage de test.

Modèles ARIMA/SARIMA

Méthodologie – La stationnarité de la série chronologique est une des conditions importantes pour l'utilisation de ces modèles. L'étude en est donc faite. Le modèle ARIMA est écarté en

raison des saisonnalités hebdomadaire et annuelle qui ne sont pas prises en compte. Le meilleur modèle SARIMA sera recherché de façon itérative selon le respect des conditions. Les fenêtres *rolling* et une *expanding* sont employées pour la prévision. De même, nous comparerons un SARIMA avec entraînement unique versus un réentraînement journalier. Enfin, le test de Diebold-Mariano est fait à un niveau 5% afin de départager les ensembles de prévisions.

Stationnarité – La série chronologique de la somme de la demande journalière d'électricité n'est pas stationnaire car son graphique indique une variance et une espérance

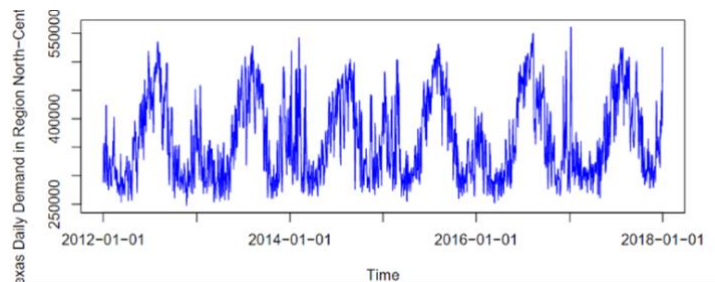


Figure 3 Série. Chronologique de la somme de la demande d'électricité journalière SOMME

qui varie en fonction du temps ainsi qu'une saisonnalité. De plus, l'ACF illustre Pics récurrents aux lags multiples de 7. La différenciation apparaît nécessaire afin

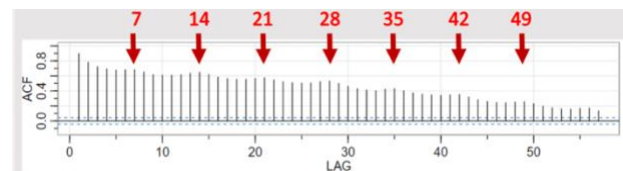


Figure 2 ACF de la série chronologique de la somme de la demande d'électricité journalière pour SOMME avant transformation

d'éliminer la composante saisonnière, mais avant de faire celle-ci la transformation Box-Cox est utilisée. La valeur de lambda est

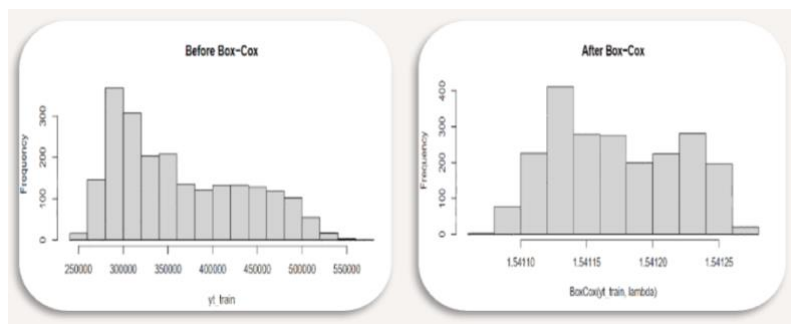


Figure 1 Série avant et après la transformation Box-Cox

de -0.65. Cette transformation rend la distribution en Figure 3 plus "normale" et améliore légèrement la stabilité de la variance. La série est ensuite différenciée

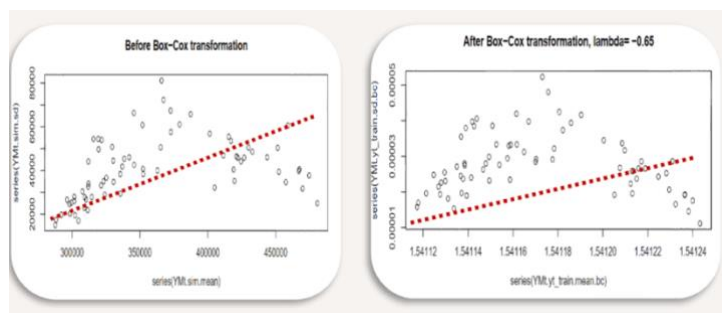


Figure 4 Série avant et après la transformation Box-Cox

d'ordre 1 à décalage 7, ce qui signifie des

paramètres saisonniers de SARIMA avec $D=1$ et $S=7$. Nous obtenons alors un graphique

ACF/PACF en Figure 5 bien plus convenable. La série chronologique différenciée de lag 7 en Figure 6 montre la disparition de la saisonnalité et une espérance nulle en tout temps. Malgré ce traitement, la variance n'est toujours pas constante comme l'illustre la figure 4. En opérant une procédure itérative, le modèle SARIMA le plus adéquat est le modèle SARIMA (1, 1, 2) (0, 1, 1)₇ avec une transformation Box-Cox ($\lambda = -0.65$).

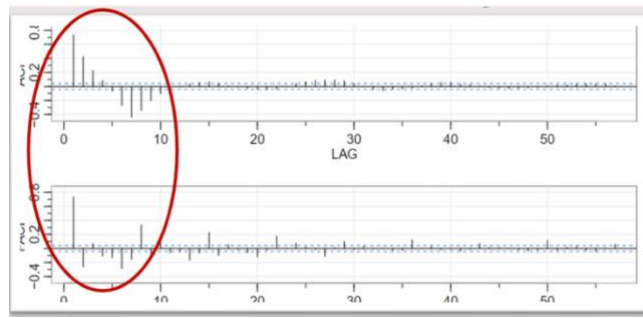


Figure 5 : ACF/ PACF de la série chronologique de la somme de la demande d'électricité journalière pour SOMME après différenciation

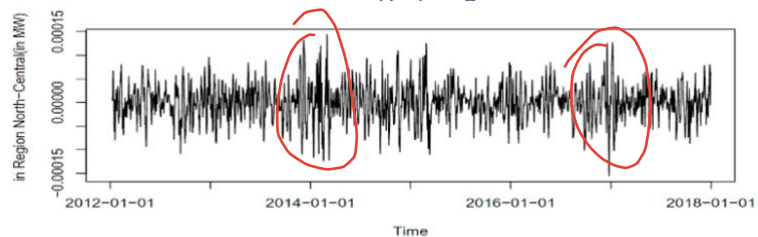


Figure 6 : Série chronologique de la somme de la demande journalière d'électricité pour SOMME différenciée de lag 7

Analyse des résultats – Les diagnostics graphiques nous indiquent une hétéroscédasticité durant les mois d'été et d'hiver avec de plus grandes erreurs de prévision comparés aux autres mois. De plus, le graphique quantile-quantile des résidus montre un comportement heavy

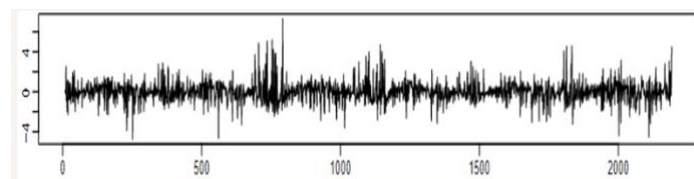


Figure 7 : Résidus standardisés

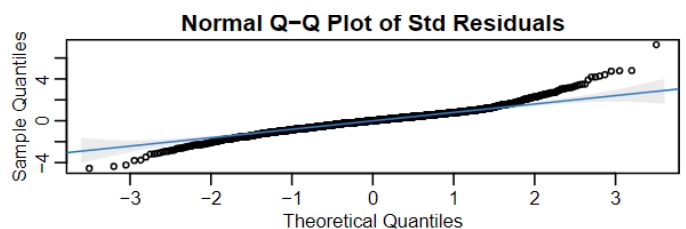


Figure 8 : QQ plot des résidus standardisés

tailed, ce qui ne satisfait pas la condition de normalité. En revanche, l'ACF des résidus est convenable car l'autocorrélation entre les résidus n'est pas observée. Enfin, le test de Ljung-Box nous montre que l'hypothèse nulle des erreurs indépendantes

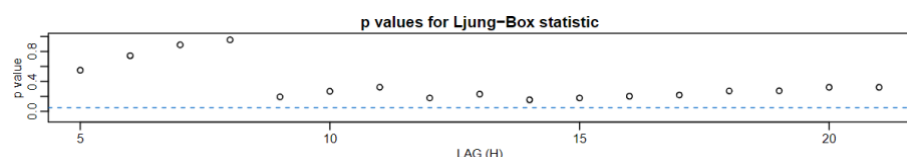


Figure 9 : Test Ljung-Box du modèle

dire plutôt ne peut pas être rejetée...

est valide. D'après cette analyse, on conclut que le modèle SARIMA ne respecte pas tous les postulats. En termes de prévisions, le tableau 1 montre que la performance du modèle SARIMA est similaire entre un entraînement unique et un entraînement journalier. Ici, une fenêtre *expanding* donne une performance similaire à une fenêtre *moving*. Le test bilatéral Diebold Mariano à un niveau de 5% donne une p-value de 0.8 : on conclut que l'erreur quadratique moyenne des prévisions est la même ~~quel que soit la fréquence d'entraînement~~ ^{plutôt pour les 2 fréquences considérées}. Le modèle choisi est le modèle SARIMA (1, 1, 2) (0, 1, 1)₇ entraîné une fois seulement car le réentraînement journalier exige beaucoup plus de temps sans aucune amélioration de la performance. Ce modèle sélectionné obtient un MAPE de 5,49%, ce qui correspond à une amélioration de 17,7% par rapport au benchmark *naïve no change* (MAPE de 6,34%). Pour finir, il a une couverture de 93.6% pour un intervalle de prévision à niveau 95%.

Entraînement	MAPE (%)	ME (MWh)
Unique	5.49	-434
Journalier	5.49	-421

Tableau 1 : Performance de SARIMA avec différents entraînements

Sur l'ensemble de validation, je suppose.

Modèles ARX

Méthodologie

La méthodologie pour les modèles ARX sera similaire à la régression linéaire avec erreurs

ARMA : une analyse des types de fenêtre d'entraînement, la fréquence de l'entraînement et

leur impact sur la performance. Les modèles ARX permettent d'accommoder les lags de la

demande journalière ainsi que les variables exogènes suivantes: journée de la semaine, journée fériée, avant journée fériée, après journée fériée, les variables humidex,

refroidissement éolien, HDD et CDD avec du bruit selon la même méthodologie utilisée dans

la partie 2 du rapport, les lags 1 et 2 des HDD et CDD, le temps t et les saisons. Une étude à

un seul lag de la demande et une à 7 lags de la demande ont été considérés.

Pourquoi seulement ces deux? Vous pouvez comparer plusieurs, non?

Analyse des résultats

Ceci laisse entendre plus que les $p=1$ et $p=7$ indiqués plus haut.

Plusieurs modèles avec combinaisons de variables exogènes ainsi qu'une analyse sur le nombre de lags de la demande à inclure a été effectuée. Au final, le **modèle ARX7** avec *moving window* et réentraînement journalier performe le mieux sur les données. Cependant, l'analyse des résidus du modèle démontre que ce dernier ne respecte pas tous les postulats requis. Le graphique des résidus montre une espérance constante autour de 0 et une variance constante à travers le temps. On remarque deux pics anormaux de résidus positifs qui sont expliqués par un froid inattendu qui se traduit par des valeurs élevées inattendues de HDD et CP au printemps.

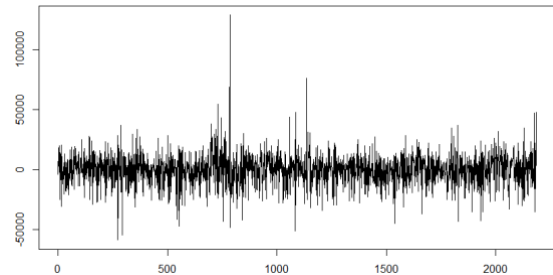


Figure 10 : Résidus du modèle ARX 7

Avez-vous comparé les résidus du ARX avec ceux du modèle de régression?

L'AFC et le PACF des résidus montre qu'il n'y a pas d'autocorrélation dans les résidus. Ce résultat



est plutôt surprenant puisqu'on aurait pu penser qu'il y aurait de l'autocorrélation à chaque 7^{ème} lag dû à la saisonnalité. En revanche, le

Mais vous avez des variables dummy pour les jours de la semaine (comme dans la régression avec erreurs ARMA où il n'avait plus de saisonnalités dans les résidus non plus).

diagramme quantile-quantile met en évidence que le postulat de normalité des résidus n'est pas respecté. En effet, on voit du côté droit un effet *heavy-tail* avec des résidus très positifs dû à une sous-estimation de la demande d'électricité.

Ce modèle ne respecte pas toutes les conditions sur les résidus, mais un ajustement sur l'échantillon de

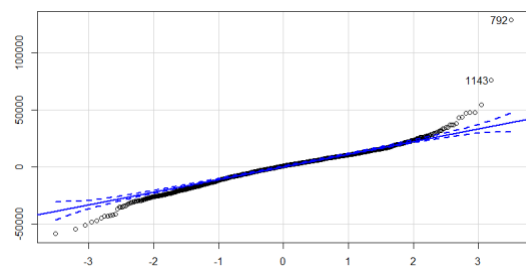


Figure 12 : QQ plot du modèle ARX7

validation a été effectué. La performance du modèle a été résumée dans le tableau suivant:

Le modèle choisit un MAPE de 3.94 et un taux de couverture de 85.3% pour un intervalle de prévision de 95%. Cela étant dit, la couverture du modèle n'est pas excellente, dû au fait qu'il ne respecte pas les postulats de normalité sur les résidus.

Ensuite, une analyse de la performance par trimestre a été effectuée. Le modèle performe mieux durant le dernier trimestre de l'année et moins bien durant les 3 premiers trimestres. Le biais obtenu est faible

	ME (MWh)	MPE	MAPE	Taux de couverture à 95%
Q1	426	-0.213	4.12	0.861
Q2	-1214	-0.57	4.1	0.836
Q3	2099	0.295	4.01	0.781
Q4	-1789	-0.689	3.54	0.935
Overall	-124	-0.295	3.94	0.853

Tableau 2 : Performance trimestrielle du modèle ARX7

comparés aux modèles ajustés précédemment et on observe un biais négatif durant le troisième trimestre qui indique que les prévisions sont basses durant cette période.

Comparaisons globales

Maintenant que les analyses des différentes méthodes et modèles à notre disposition ont été menées, les modèles candidats peuvent être comparés entre eux pour déterminer le meilleur. Au total, 5 modèles sont étudiés : le *naïve no change* qui est notre *benchmark*, le modèle TBATS parmi les méthodes de lissage, la régression linéaire avec erreurs ARMA (5,0,2) et les deux modèles ressortant des analyses faites durant cette partie, à savoir le SARIMA (1,1,2)*(0,1,1)₇ ainsi que le ARX(7). En comparant les performances précédemment citées, la régression linéaire avec erreurs ARMA (5,0,2) et le modèle ARX(7) performant le mieux avec un MAPE respectif de 4,11% et 3,94%. Un test de Diebold-Mariano au niveau de 5% est conduit pour nous permettre de choisir un modèle. Le test ne montre pas de différence significative entre la régression linéaire et ARX(7). Le modèle ARX(7) est choisi, non pas seulement de par son MAPE plus faible, mais également car le ME est beaucoup plus faible que celui de la régression linéaire. Le meilleur modèle choisit, il faut maintenant regarder la performance sur l'échantillon test. Le MAPE et le ME du modèle sur test sont supérieurs à ceux de validation, ce qui était attendu au vu des changements d'habitudes net en 2020-2021 dû au Covid-19. Le Q2 2020 performe particulièrement mal avec une MAPE de 5,1 car il correspond à la période de confinement dont nos régions ont fait l'expérience. Les trimestres suivants ont continué d'être impacté à cause de l'effet durable de la Covid-19. Le MPE a quant à lui diminué drastiquement

Nous rappeler si la performance de la régression était aussi uniforme.

Où? De combien? Un petit MPE est bon donc je n'arrive pas à suivre.

2
au Q2 2020, **signifiant de plus grandes surestimations** faites par notre modèle. Le secteur résidentiel ayant augmenté sa demande de manière imprévue avec le confinement, une hypothèse est qu'une forte diminution de la demande du secteur industriel soit la cause de ce résultat. Des données qui ne sont pas disponibles à ce moment seraient nécessaires pour affirmer cette hypothèse.

Conclusion et recommandation

En résumé, les diverses analyses mènent à la conclusion que **le modèle ARX(7) est celui**
recommandé afin d'effectuer la prévision à t+1 de la somme de la demande d'électricité
quotidienne dans les régions North, North Central et East. Néanmoins, un problème concernant les variables météorologiques se pose car **le niveau de bruit ajouté a un impact important sur la couverture.** Une analyse approfondie de la sensibilité du bruit serait donc opportune dans le futur pour trouver des valeurs reflétant au mieux les données réelles. Enfin, nous avons détecté des problèmes d'hétéroscédasticité. L'analyse par trimestre a permis de mettre en avant **une variance supérieure des prévisions sur le 1er trimestre ainsi que sur le 4ème.** Dans notre cas, le modèle choisi **performe le mieux sur tous les trimestres.** Dans le futur, la trouvaille d'une variable explicative manquante pourrait possiblement expliquer la non-constance de la variance des prévisions et donc amenuiser ce problème.

Je dirais plutôt que lorsque le modèle sera implanté, i.e. lorsque les prévisions météo pour le lendemain seront utilisées, on connaîtra mieux la performance réelle.

ou l'usage de modèles plus complexes qui permettent des variances non-constantes...

Ça prend vraiment un tableau qui montre les performances dans l'ensemble test, de préférence par année et par saison.

Vous n'avez pas donné la performance des IP dans l'ensemble test. À quoi peut s'attendre l'utilisateur?

Vous avez écarté le TBATS? Une couverture à 94,07% des IP à un niveau de confiance de 95% a sûrement du mérite.