

Consommation et prix de l'électricité

Vadim BERTRAND, Kassim KONE

Table des matières

Introduction	3
1. Consommation électrique	3
1.1. Analyse descriptive	3
1.2. Analyse des composantes déterministes : tendance et saisonnalité(s)	5
1.3. Comparaison de différents modèles d'estimation	6
1.3.1. Court terme - moyen terme - long terme	6
1.3.2. Multi saisonnier	7
2. Prix de l'électricité	8
2.1. Analyse descriptive	8
2.2. Analyse des composantes déterministes : tendance et saisonnalité(s)	8
2.3. Comparaison de différents modèles d'estimation	9
2.3.1. Court terme - moyen terme - long terme	9
2.3.2. Multi saisonnier	10
3. Liens entre consommation électrique et prix de l'électricité	11
Conclusion	12

Introduction

Nous avons choisi d'observer le prix du mégawattheure d'électricité sur le marché boursier français en euros (€/MWh) et la consommation électrique en mégawattheure (MWh) en France entre 2015 et 2021. Ces données sont mises à disposition par RTE (le gestionnaire français du Réseau de Transport d'Electricité) au pas horaire.

Nous avons exclu les années 2020 et 2021 de l'étude en raison de la crise du COVID-19 qui a entraîné une forte chute de la consommation en 2020, probablement en raison du ralentissement économique ; et de la crise énergétique dans laquelle nous nous trouvons actuellement, démarrée en 2021, qui a engendré une hausse spectaculaire du prix de l'électricité. La *Figure 1* met en avant ces variations après un lissage des données sur une année glissante.

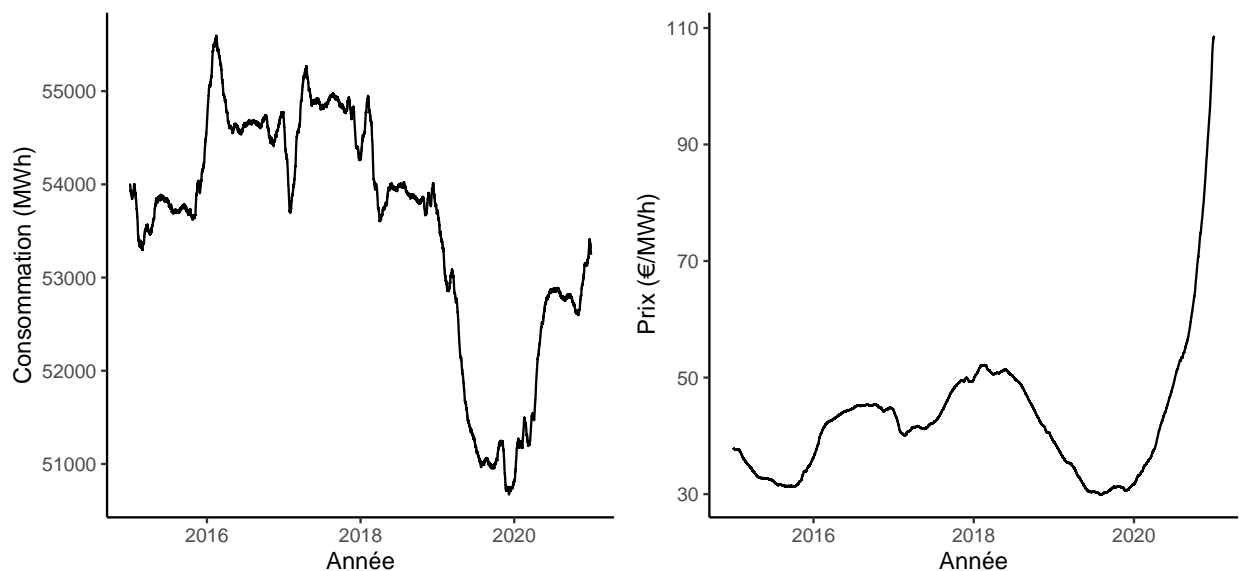


Figure 1: Variations lissées de la consommation électrique et du prix de l'électricité entre 2015 et 2021

Pour chacune des deux séries temporelles, notre étude s'intéressera tout d'abord à la détection d'une tendance et des saisonnalités, puis au choix d'un modèle d'estimation et de prédiction convenable. Enfin nous essaierons d'identifier un lien entre les deux séries.

1. Consommation électrique

1.1. Analyse descriptive

Nous remarquons sur la *Figure 2* que la tendance de la consommation électrique semble très légèrement décroissante, et qu'une saisonnalité de variance constante se dégage : pendant les mois d'hiver la consommation augmente.

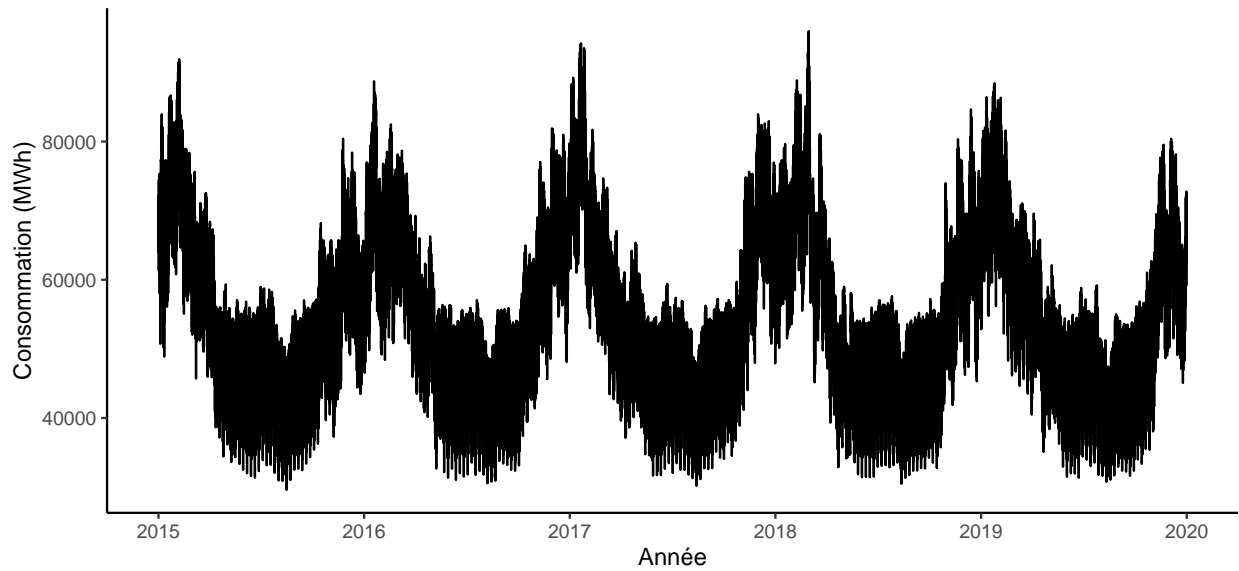


Figure 2: Variations de la consommation électrique au pas horaire

En plus de la saisonnalité annuelle, nous pouvons voir sur la *Figure 3* représentant l'ACF des séries à différentes fréquences (horaire, journalière et mensuelle) que deux autres saisonnalités peuvent être explorées : journalière (nous consommons moins la nuit) et hebdomadaire (l'activité industrielle est moins importante le weekend).

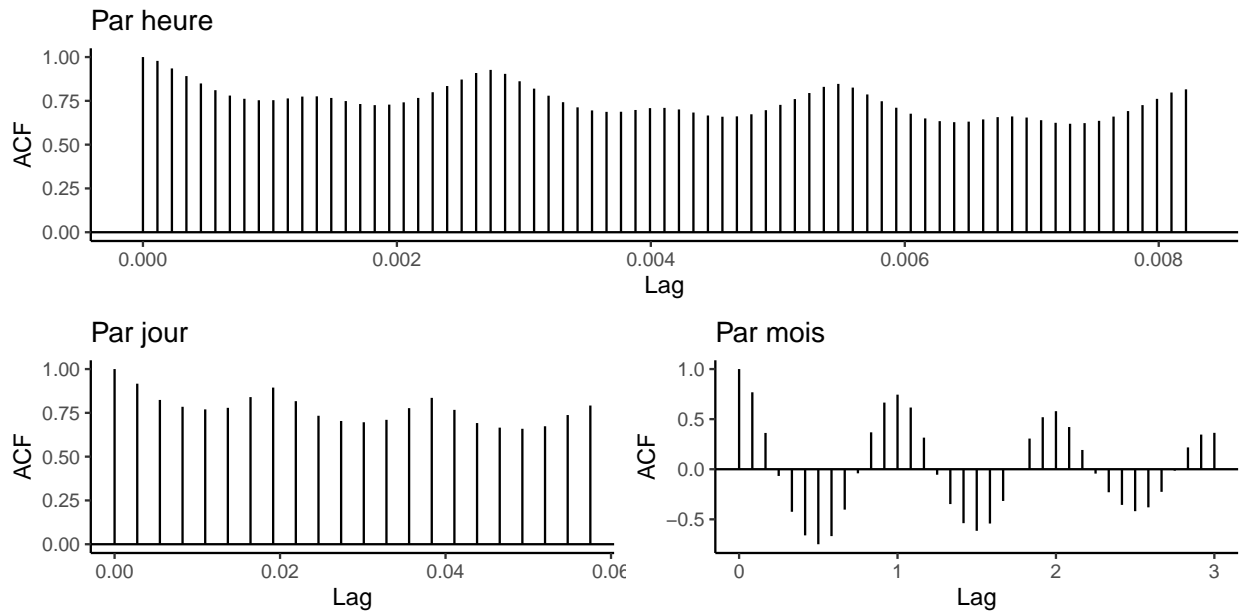


Figure 3: ACF de la consommation électrique avec une fréquence horaire, journalière et mensuelle

1.2. Analyse des composantes déterministes : tendance et saisonnalité(s)

Afin de considérer les saisonnalités journalière et hebdomadaire en plus de la saisonnalité annuelle nous avons appliqué la décomposition STL estimant successivement plusieurs composantes saisonnales en utilisant la méthode de régression *LOESS* [1]. *LOESS* estime dans un premier temps chaque point par une régression linéaire ou polynomiale faite sur une fenêtre glissante centrée et pénalisant les points utilisés dans la régression en fonction de leur distance temporelle au point d'estimation. L'opération est répétée en pénalisant cette fois-ci également par la distance entre la valeur initiale et la première estimation, ce qui permet de réduire la sensibilité aux valeurs anormales.

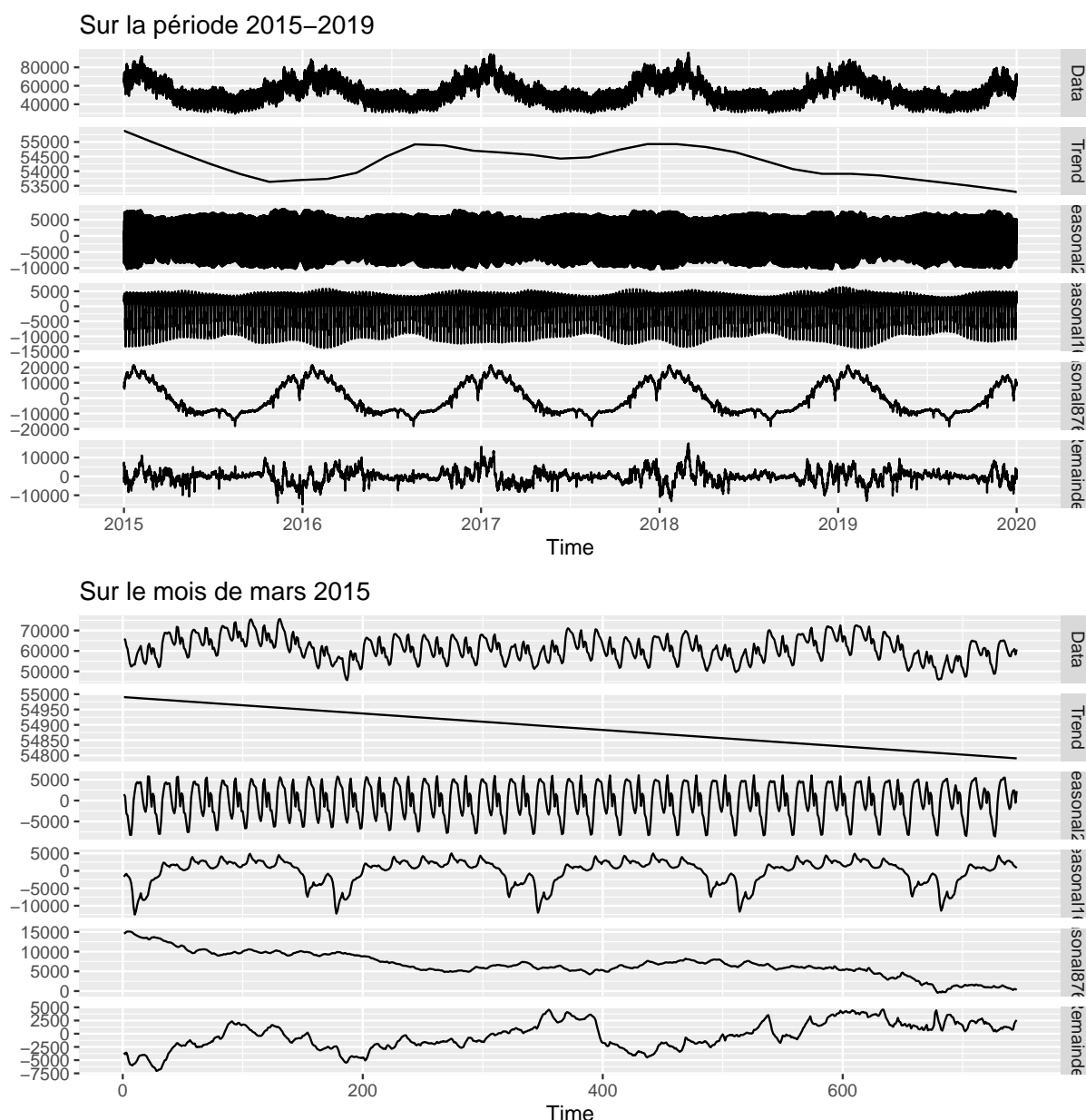


Figure 4: Décomposition STL de la consommation électrique

La *Figure 4* présente deux décompositions : sur la période 2015-2019 nous observons bien la saisonnalité annuelle et la tendance légèrement décroissante, en revanche les saisonnalités journalières et hebdomadaire ne

sont pas visibles en raison de leur fréquence trop grande ; nous avons donc zoomé sur la période de mars 2015 qui les met en évidence.

1.3. Comparaison de différents modèles d'estimation

Maintenant que nous avons caractérisé la partie déterministe du signal, nous allons expérimenter plusieurs modèles d'estimation. Pour ce faire nous avons divisé notre série en un jeu d'entraînement et un jeu de test sur lequel nous avons évalué la performance des modèles.

Nous nous intéresserons à quatre types de modèles différents :

- *court terme* (CT) prenant en compte la saisonnalité journalière
- *moyen terme* (MT) : saisonnalité hebdomadaire
- *long terme* (LT) : saisonnalité annuelle
- *multi saisonnier* intégrant les trois saisonnalités

1.3.1. Court terme - moyen terme - long terme

Nous avons utilisé trois jeux de données différents en fonction de l'horizon et du pas de prédiction souhaité :

- *CT* : données horaires, du 2 au 5 décembre 2019 pour l'entraînement et le 6 décembre 2019 pour le test
- *MT* : données journalières, du 26 août au 22 septembre 2019 pour l'entraînement et du 23 septembre au 29 septembre 2019 pour le test
- *LT* : données mensuelles, de début 2015 à fin 2018 pour l'entraînement et 2019 pour le test

Pour chacun des horizons, nous avons expérimenté trois modèles : Buys-Ballot (BB), Holt-Winters (HW) et SARIMA que nous comparons sur la base du carré de l'erreur moyenne quadratique (RMSE) et de l'erreur absolue moyenne exprimée en pourcentage (MAPE) réalisée sur le jeu de test.

La *Table 1* présentant les erreurs induit par ces modèles révèle que SARIMA est globalement le plus performant, tant sur l'estimation que la prédiction. Nous pouvons noter que les deux autres modèles ont donné des résultats similaires, mais en nécessitant l'estimation de plus de paramètres. Il est surprenant de constater que Holt-Winters a souvent moins bien performé que Buys-Ballot, peut-être en raison du caractère stationnaire de la série.

Sur la *Figure 5* nous pouvons constater que le profil des prédictions est, lui aussi, semblable pour tous les modèles.

En nous basant sur le test du porte-manteau de Ljung et l'analyse de l'ACF nous avons vérifié que les résidus obtenus par ces modèles correspondent tous à un bruit-blanc.

Table 1: RMSE et MAPE des modèles CT, MT et LT de la consommation électrique

	Entrainement	Test
CT BB	1551.12 (1.83%)	4401.86 (5.15%)
CT HW	697.88 (0.59%)	3791.57 (4.02%)
CT SARIMA	457.14 (0.41%)	1834.36 (1.80%)
MT BB	491.38 (0.83%)	577.31 (1.10%)
MT HW	745.67 (1.16%)	1396.21 (2.85%)
MT SARIMA	455.43 (0.65%)	512.51 (1.02%)
LT BB	2337.51 (2.78%)	2289.55 (3.60%)
LT HW	3503.90 (3.94%)	2586.19 (3.78%)
LT SARIMA	2988.61 (3.21%)	2165.37 (3.08%)

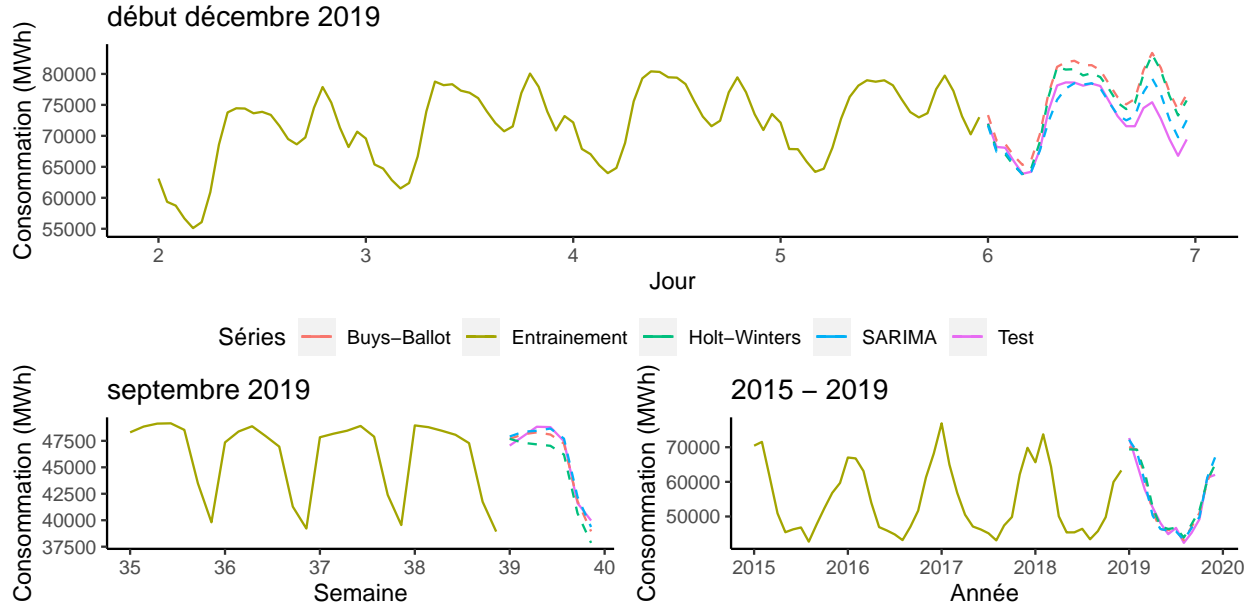


Figure 5: Prédiction de la consommation électrique pour différents horizons et pas

1.3.2. Multi saisonnier

TBATS est un modèle très répandu pour la prédiction de séries temporelles ayant plusieurs saisonnalités [2]. Malheureusement TBATS construit et évalue plusieurs modèles potentiels pour choisir le meilleur en se basant sur l'AIC et cette sélection de modèles implique un coût de calcul important.

Une alternative est d'utiliser un modèle ARIMA, non saisonnier, mais en spécifiant des régresseurs permettant de modéliser les saisonnalités [3]. Les régresseurs de chaque saisonnalité sont les termes d'une série de Fourier approchant la saisonnalité en question. Ici encore, déterminer le nombre de termes de Fourier à considérer pour modéliser les saisonnalités est coûteux.

Une approche plus naïve et moins gourmande peut être mise en oeuvre : utiliser la décomposition STL (voir Figure 4) pour re-saisonnaliser la prédiction faite sur la tendance.

Nous avons mis en oeuvre les trois approches en utilisant les données horaires de 2015 à 2018 pour l'entraînement et de 2019 pour le test et en considérant trois saisonnalités : journalière, hebdomadaire et annuelle.

Comme nous pouvons le voir sur la Table 2 et la Figure 6 les trois modèles proposent des résultats assez proches, et de manière surprenante c'est STL, le modèle le moins coûteux, qui fait légèrement mieux, nous le privilégierons donc fortement. Globalement la prédiction est très bonne : à peine plus de 5% d'erreur moyenne absolue avec un horizon d'un an et un pas horaire !

Contrairement aux modèles uni saisonniers, les résidus ne peuvent être apparentés à un bruit-blanc aussi il pourrait être envisagé de tenter d'améliorer les prédictions en les modélisant par des processus (S)AR(I)MA.

Table 2: RMSE et MAPE des modèles prédictifs multi saisonnier de la consommation électrique

	Entrainement	Test
Multi STL	228.53 (0.30%)	4445.78 (6.26%)
ARIMA + Fourier regresseurs	1343.02 (1.97%)	4645.31 (6.45%)
TBATS	981.98 (1.34%)	4982.87 (6.90%)

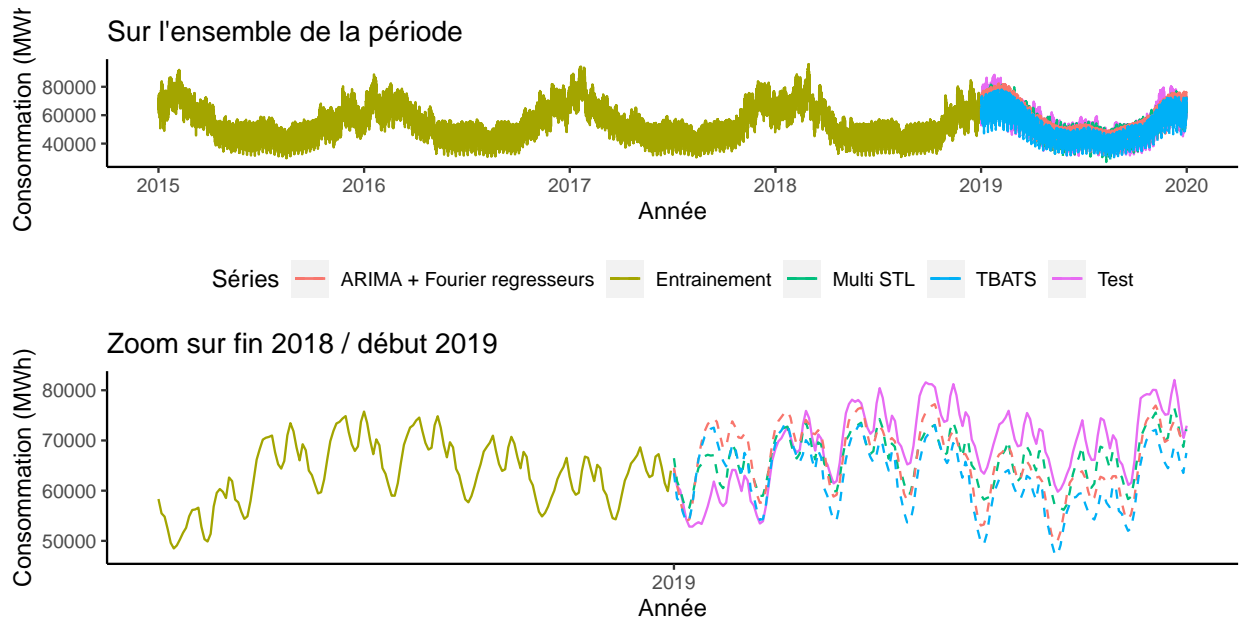


Figure 6: Prédictions multi saisonnières de la consommation électrique

2. Prix de l'électricité

Nous avons reproduit les démarches explicitées dans la première partie de ce rapport sur les prix de l'électricité sur le marché boursier français. Afin d'éviter trop de redondance nous passerons plus rapidement sur la mise en oeuvre pour nous concentrer sur les résultats.

2.1. Analyse descriptive

Via l'analyse des ACF aux fréquences horaire, journalière et mensuelle nous avons relevé les mêmes saisonnalités que pour la consommation : journalière, hebdomadaire et annuelle. Cependant, la saisonnalité annuelle est moins marquée ici.

2.2. Analyse des composantes déterministes : tendance et saisonnalité(s)

La décomposition STL permet de bien extraire les composantes saisonnières journalière et hebdomadaire. Cependant, probablement en raison de fortes fluctuations pouvant s'apparenter à du bruit, la composante annuelle est moins claire que celle de la consommation d'électricité. Nous avons souhaité mettre en lumière ces points via la *Figure 7*.

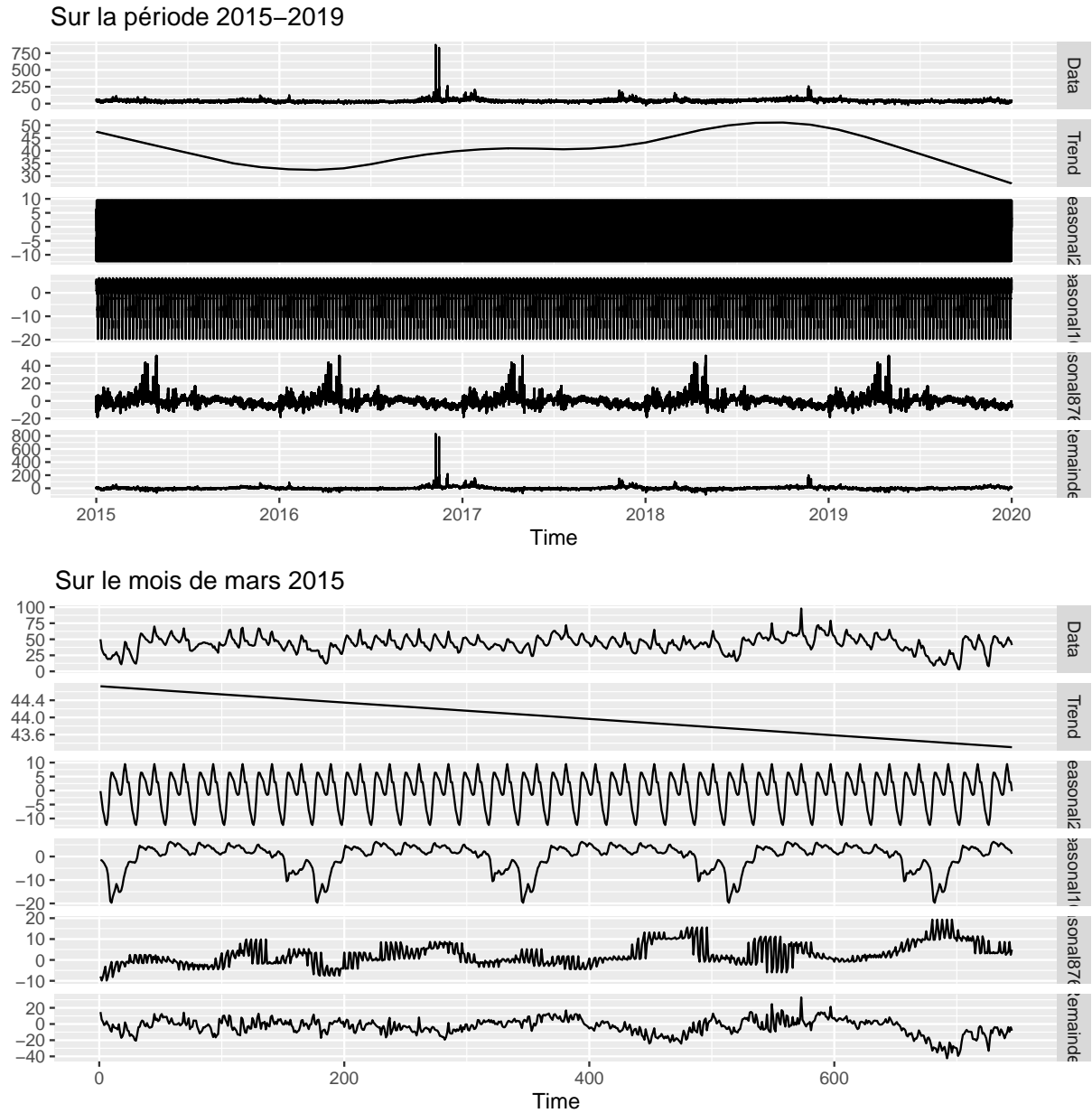


Figure 7: Décomposition STL du prix de l'électricité

2.3. Comparaison de différents modèles d'estimation

2.3.1. Court terme - moyen terme - long terme

Afin de comparer les résultats des prédictions obtenus sur le prix de l'électricité avec celles sur la consommation, nous avons conservé les mêmes périodes et les mêmes types de modèles.

Comme nous pouvons le voir sur la *Table 3* et la *Figure 8* les prédictions du prix de l'électricité sont nettement moins bonnes que celles de la consommation électrique. Le prix semble avoir un caractère plus ératique.

Table 3: RMSE et MAPE des modèles CT, MT et LT du prix de l'électricité

	Entrainement	Test
CT BB	3.40 (4.44%)	15.40 (27.20%)
CT HW	3.28 (4.23%)	16.36 (27.38%)
CT SARIMA	2.92 (3.38%)	11.26 (16.73%)
MT BB	3.96 (8.97%)	4.63 (12.23%)
MT HW	4.56 (10.31%)	8.54 (20.87%)
MT SARIMA	4.27 (9.16%)	4.97 (14.06%)
LT BB	7.92 (15.54%)	15.95 (37.96%)
LT HW	7.89 (12.78%)	13.90 (33.98%)
LT SARIMA	8.15 (14.83%)	18.43 (48.42%)

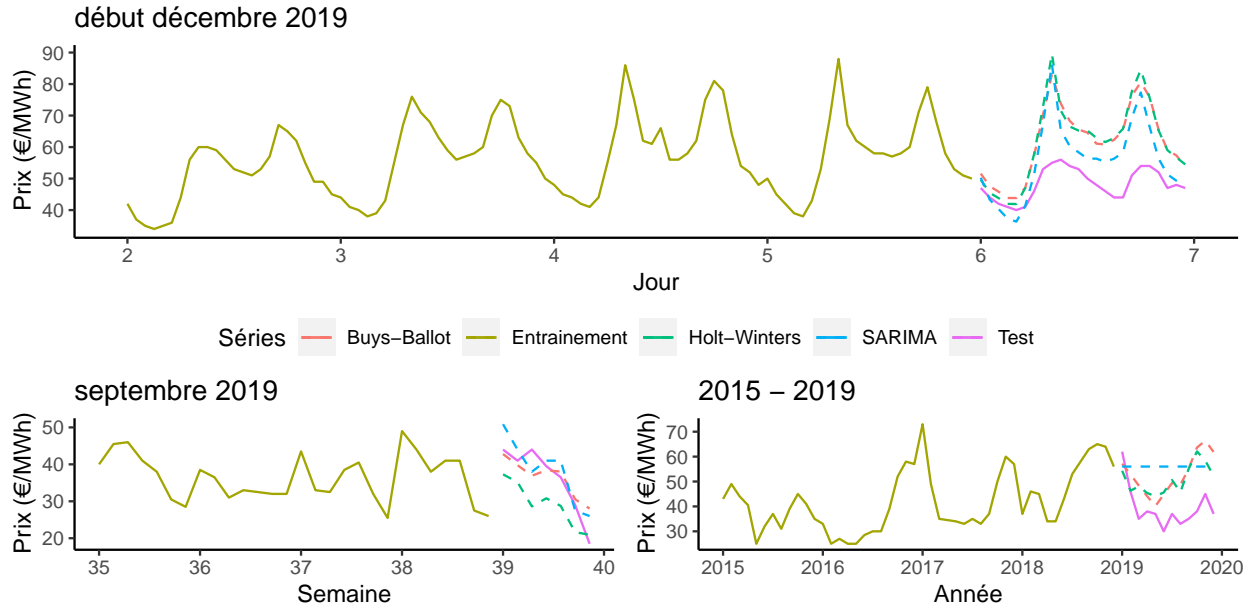


Figure 8: Prédictions du prix de l'électricité pour différents horizons et pas

2.3.2. Multi saisonnier

La Table 4 donne la RMSE des modèles multi saisonnier, nous n'avons pas pu calculer la MAPE car le prix est parfois nul ou très proche de zéro. En moyenne, le meilleur modèle (ARIMA + Fourier régresseurs) performe aux alentours de 33% d'erreur, ce qui équivalent aux modèles uni saisonniers, mais nettement inférieur aux modèles prédictifs de la consommation.

Table 4: RMSE des modèles prédictifs multi saisonnier du prix de l'électricité

	Entrainement	Test
Multi STL	5.24	16.98
ARIMA + Fourier regresseurs	8.99	13.73
TBATS	8.93	16.82

Sur la *Figure 9* nous pouvons voir que les modèles multi STL et TBATS ont capturé la forte hausse apparaissant souvent en fin d'année, mais que celle-ci ne s'est pas produite en 2019, ce qui a généré une erreur importante à cette période.

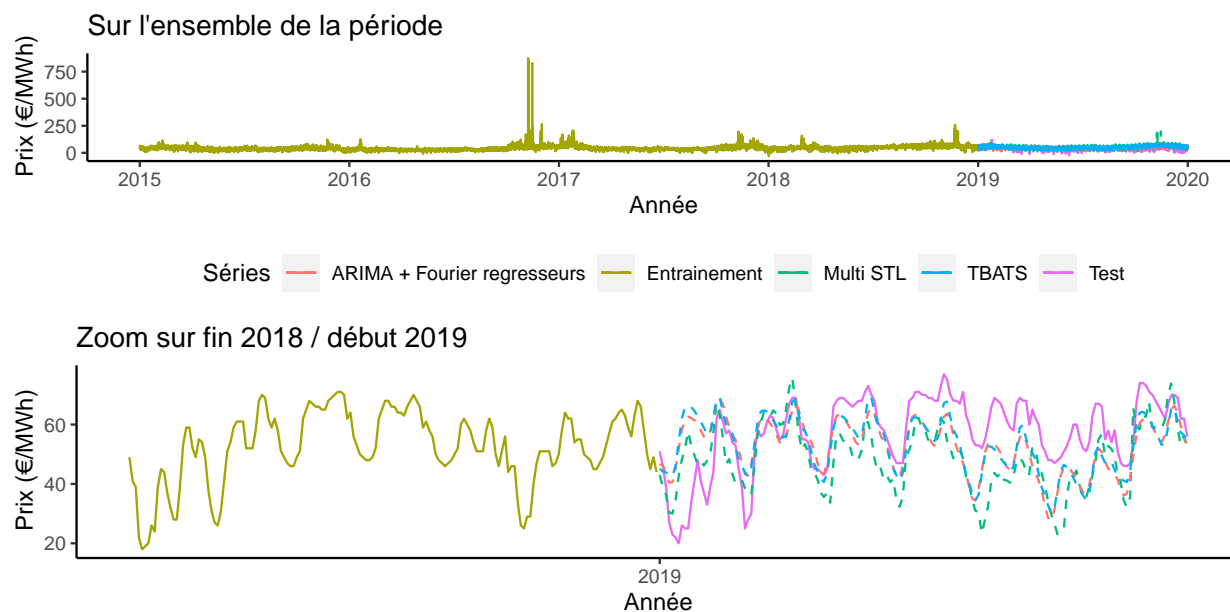


Figure 9: Prédictions multi saisonnières du prix de l'électricité

3. Liens entre consommation électrique et prix de l'électricité

Nous avons vu qu'il est possible de faire de très bonnes prédictions de la consommation électrique, mais que c'est bien plus compliqué avec le prix de l'électricité.

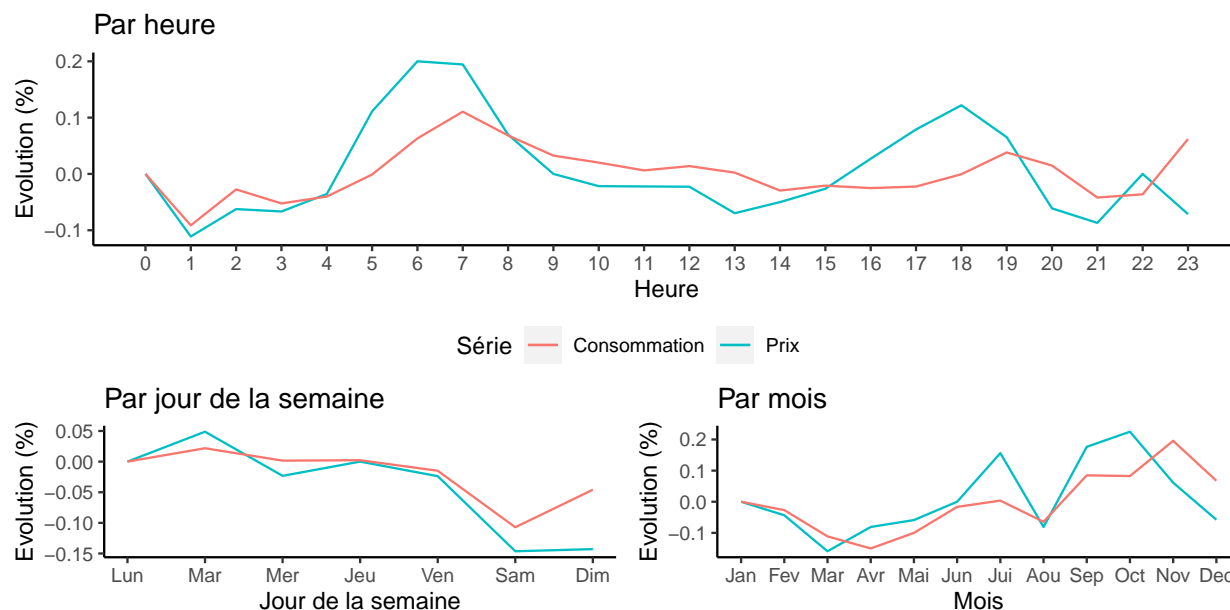


Figure 10: Evolution jointe du prix et de la consommation d'électricité

Cependant, la *Figure 10* suggère un lien entre les évolutions de la consommation et du prix. Nous pourrions donc être intéressé par la prédiction du prix de l’électricité en se basant sur la consommation ou la prédiction de la consommation.

Pour nous assurer de l’existence de ce lien, nous pouvons calculer la distance entre les deux séries en se basant sur leur cross-corrélation, afin de s’intéresser aux variations plutôt qu’aux valeurs prises. Un test de causalité de Granger [4], dont l’hypothèse nulle est que le modèle $y \sim Lags(y, k) + Lags(x, k)$ ne donne pas plus de puissance que le modèle $y \sim Lags(y, k)$ (donc qu’il n’y a pas de causalité) pourrait également nous servir.

Table 5: Métriques pour évaluer la causalité entre prix et consommation d’électricité

	Métrique
Cross-corrélation distance	0.03
p-value du test de Granger	0.00

La *Table 5* montre que la distance est faible et que la p-value du test de Granger conduit à rejeter avec une forte certitude l’hypothèse d’absence de causalité.

Il semblerait donc très pertinent de construire un modèle prédictif pour le prix de l’électricité en s’appuyant sur la consommation d’énergie.

Conclusion

Nous avons vu que la consommation électrique peut être prédite très correctement, y compris à un horizon lointain, mais qu’il n’en va pas de même avec la prédiction du prix de l’électricité.

La mise en lumière d’une causalité entre consommation et prix nous permet cependant d’envisager d’autres modèles prédictifs pour le prix qui se baseraient également sur la consommation ou son estimation.

Ces résultats encourageants sont néanmoins à nuancer : en effet, nous avons vu en introduction que des événements peuvent fortement impacter la consommation ou le prix de l’électricité (cf *Figure 1*). Les prédictions long terme seraient alors peu pertinentes et se contenter de modèles court ou moyen terme (multi saisonniers ou non) pourrait s’avérer suffisant et même préférable.

References

- [1] Jean E. McRae Robert B. Cleveland William S. Cleveland and Irma Terpenning. “STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)”. In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73.
- [2] Rob J. Hyndman Alysha M. De Livera and Ralph D. Snyder. “Forecasting time series with complex seasonal patterns using exponential smoothing”. In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1513–1527. DOI: <https://doi.org/10.1198/jasa.2011.tm09771>.
- [3] Wlodek Tych Peter C. Young Diego J. Pedregal. “Dynamic harmonic regression”. In: *Journal of Forecasting* 18.6 (1999), pp. 369–394. DOI: [https://doi.org/10.1002/\(SICI\)1099-131X\(199911\)18:6%3C369::AID-FOR748%3E3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-131X(199911)18:6%3C369::AID-FOR748%3E3.0.CO;2-K).
- [4] C. W. J. Granger. “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”. In: *Econometrica* 37.3 (1969), pp. 424–438.