

# Documentation technique sur l'outil d'aide à la détermination des prix spots de l'électricité.

L'outil développé est une aide à la détermination des prix spots de l'électricité de différents pays de l'Union Européenne, en fonction de la demande et de la production électrique. Dans notre cas, nous avons travaillé sur 3 pays : l'Allemagne, l'Espagne et les Pays-Bas.

Est-il possible de prédire le prix spot de l'électricité dans ces 3 pays en utilisant les données de consommation européenne, de prévision de production ainsi que les données temporelles associées ?

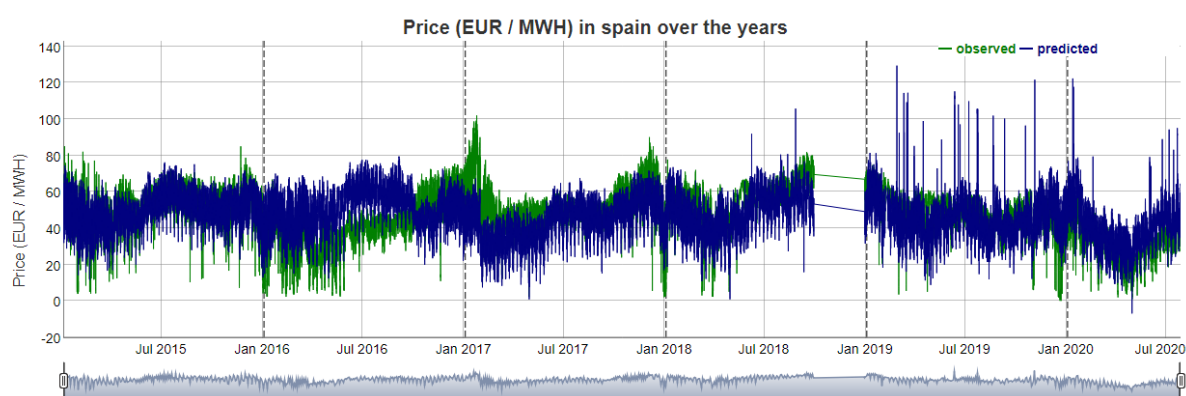
Voici le lien vers l'outil hébergé sur shinyapps.io :

[https://ichbinarnaud.shinyapps.io/shiny\\_app/](https://ichbinarnaud.shinyapps.io/shiny_app/)

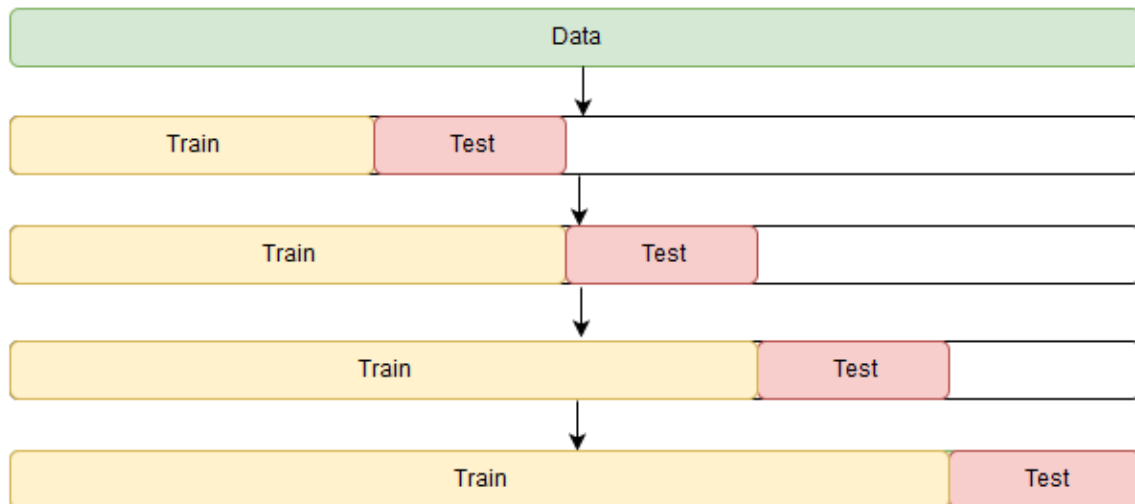
Ainsi que le code et les données du projet :

[https://github.com/ichbinarnaud/project\\_R\\_electricity\\_price/tree/main/project\\_R\\_electricity\\_price](https://github.com/ichbinarnaud/project_R_electricity_price/tree/main/project_R_electricity_price)

Chaque onglet représente un pays sauf le dernier, nommé « Correlation », et il serait assez facile d'en ajouter plus, mais nous avons décidé de ne nous concentrer que sur 3 pays.



Le premier élément affiché est le graphique temporel montrant l'évolution du prix spot sur plusieurs années. Il manque une partie des résultats de 2018. La courbe verte représente le prix spot observé que nous avons récupéré. La courbe bleue que nous considérons comme le prix spot « prédit » est issue de la prédiction effectuée sur la base de test en suivant ce schéma :



Ainsi, le prix prédit que nous voyons affiché de 2016 à 2017 a été prédit grâce au model entrainé sur les données de 2015 à 2016. Le prix prédit de 2017 à 2018 a été prédit grâce au model entrainé de 2015 à 2017. Et ainsi de suite jusqu'à la dernière année. La seule exception est la première année, où la base de test est la même que celle d'entrainement.

Il est par ailleurs possible de zoomer sur ce graphique afin de se concentrer sur une zone temporelle en particulier.

Le coefficient de determination est 0.6016


RMSE seconde annee : 18.21

RMSE troisieme annee : 12.77

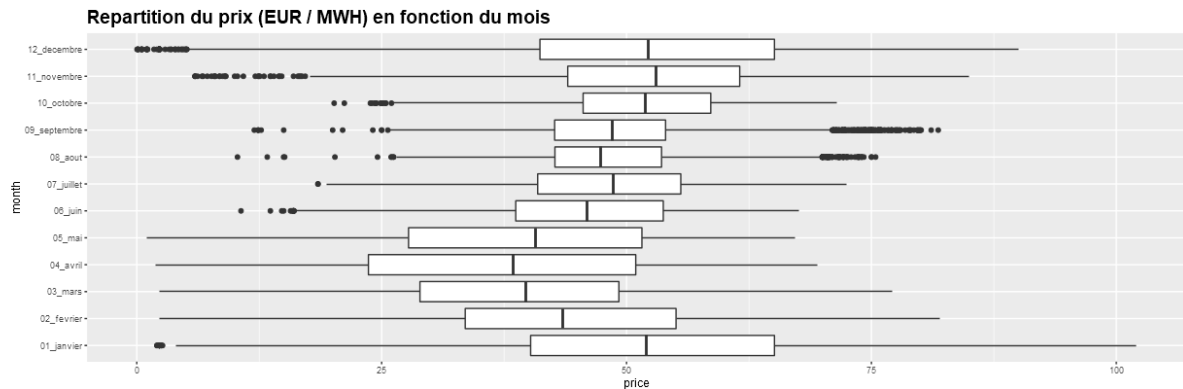
RMSE quatrieme annee : 11.51

RMSE cinquieme annee : 11.1

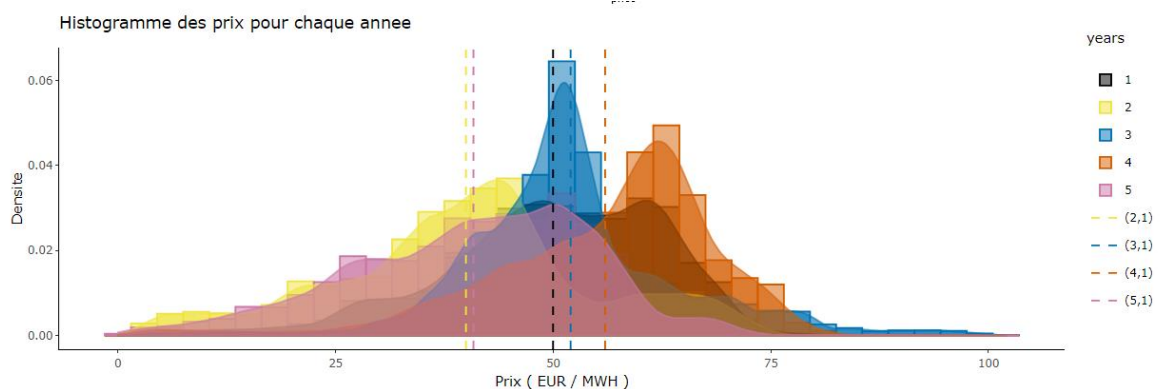
RMSE sixieme annee : 13.16

 Generer les donnees en csv

Les données affichées à gauche du graphique sont également liées à ce découpage. Le RMSE, racine de l'erreur quadratique moyenne, est calculé sur la base de test de chaque année pour donner une idée de la précision. Le coefficient de détermination est quant à lui le  $R^2$  ajusté calculé sur le modèle de régression linéaire entrainé sur toutes les données présentes. Bien que ce coefficient nous montre que l'utilisation de nos variables explicatives a un sens par rapport aux données que nous cherchons à prédire, sa valeur reste tout de même faible.



Les diagrammes en boîte représentent la répartition du prix spot en fonction des différents mois de l'année pour chaque pays. C'est intéressant lorsque l'on veut connaître une tendance générale. Dans notre cas, on peut apercevoir une hausse des prix lorsque les températures sont faibles, notamment en janvier et décembre, ce qui est plutôt logique.



Les histogrammes représentent la répartition des prix en fonction de l'année. Cela permet également de suivre la tendance générale à travers les années. Les traits verticaux représentent la médiane des prix de chaque année. Il est possible de cliquer sur la légende pour masquer un élément ou bien double cliquer pour l'isoler.

Plusieurs tableaux sont présents juste en dessous de cet histogramme. Ils nous donnent des informations sur l'impact des variables explicatives sur le prix spot.

Ils comprennent 4 colonnes :

- Estimate. C'est le coefficient de la variable explicative, déterminé par la régression linéaire.
- t-value. Cette valeur est une indication de l'importance de la variable dans notre modèle. Plus la valeur absolue de cette valeur est élevée, plus la variable est

importante. Il est possible de classer la t-value par ordre croissant ou décroissant dans notre tableau.

- Standard error. C'est l'estimateur de l'écart type du coefficient.

-  $\Pr(>|t|)$ . C'est la valeur-p du test de Student.

	Estimate	Std. Error	t value	Pr(> t )
	All	All	All	All
load_VE_austria	-0.00132829972125654	0.000117904318477823	-11.2659123805239	2.12837092208366e-29
load_VE_belgium	0.00260151217969415	0.000134528228577856	19.3380393631555	5.85087997585833e-83
load_VE_switzerland	-0.00158180916605219	0.0000792649809862679	-19.9559647447115	3.37858032713152e-88
load_VE_germany	0.000366435684498213	0.0000224689867229653	16.3085095476817	1.3036763100134e-59
load_VE_danemark	-0.00578477353907292	0.000246784800207846	-23.4405584711898	9.66493840722887e-121
load_VE_spain	0.00166691990181663	0.0000406987461938494	40.9575246833658	0
load_VE_france	0.000321079697649656	0.0000161510381251567	19.8798179511164	1.52392914727792e-87
load_VE_england	-0.0010262225654793	0.0000176285943540618	-58.2135220124828	0
load_VE_ireland	0.00911241040962787	0.000291608697224859	31.2487607411836	5.89505371855484e-212
load_VE_italy	0.000504404553996306	0.0000211473572755642	23.8518954129151	6.43350105308853e-125
load_VE_luxembourg	-0.00438142041488078	0.000656824834090203	-6.67060712000891	2.57864121458725e-11
load_VE_netherlands	-0.000054675040581433	0.0000536153937868278	-1.01976385362044	0.307846276354163
load_VE_norway	-0.000819468534908071	0.0000617349475213361	-13.2739812344516	3.93990044022808e-40
load_VE_portugal	0.000790135641553025	0.000180863898297668	4.36867528008609	0.0000125298203410825
load_VE_sweden	0.000371537303822945	0.0000519128384716618	7.15694450084367	8.38275146472444e-13

Showing 1 to 15 of 15 entries

Previous 1 Next

Nous pouvons alors connaître, grâce à ces éléments, quelles variables explicatives sont importantes si nous voulons déterminer le prix spot. De plus, il est possible de les importer dans un fichier excel pour faire plus de calculs.

Le premier tableau comprend toutes les variables de consommation des pays, qui commencent par « load ».

Le second comprend les variables de prévision de production, terminant par « generation ».

Le dernier comprend les variables temporelles : jours de la semaine, mois et heures.

Le prochain élément présent est le « calculateur ».

	fit	lwr	upr
1	57.5263620464291	56.3667135273876	58.6860105654707

Showing 1 to 1 of 1 entries

Showing 10 entries

Search:

Previous1Next

load_VE_austria	load_VE_belgium	load_VE_switzerland	load_VE_germany	load_VE_danemark	load_VE_spain	load_VE_france	load_VE_englan	
1	7025	9809	6687	55050	3717	28446	51678	355

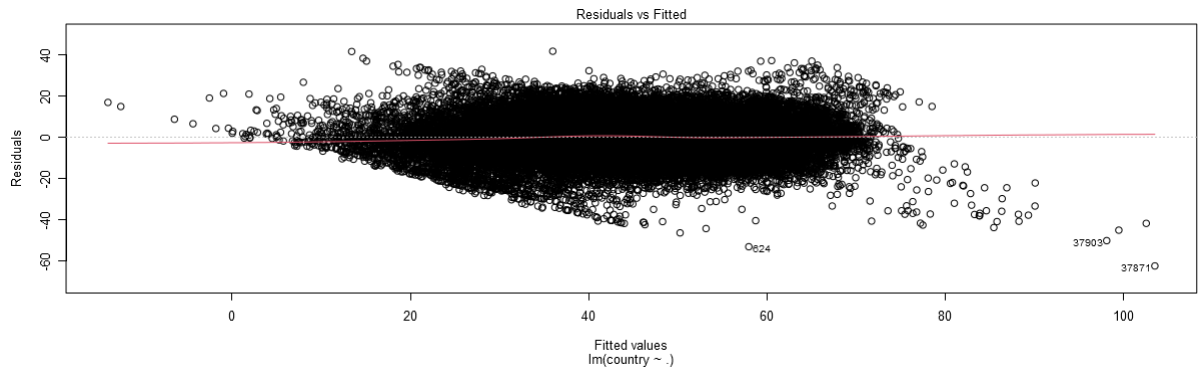
Showing 1 to 1 of 1 entries

Previous1Next

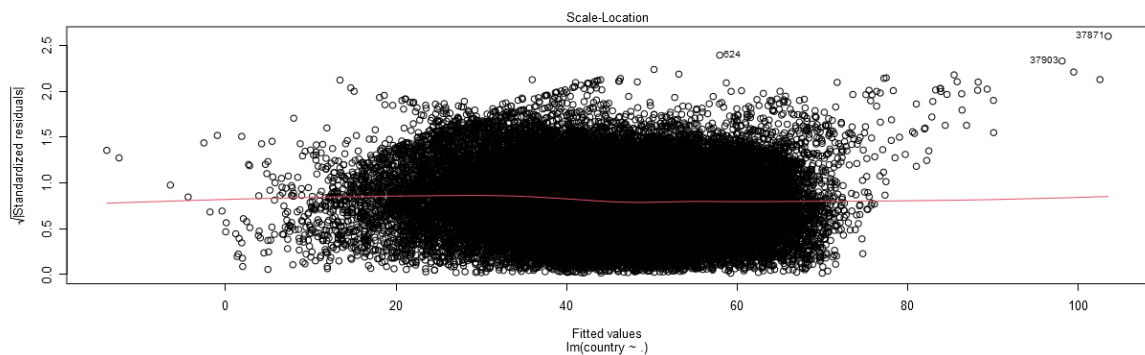
Il est possible d'entrer, dans le tableau du bas, les données que l'on veut pour chaque variable. Dans le tableau du haut s'affiche alors la valeur prédite par notre régression linéaire entraînée sur tout le modèle avec son intervalle de confiance à 99%.

Les derniers graphiques restant représentent la vérification des hypothèses de la régression linéaire :

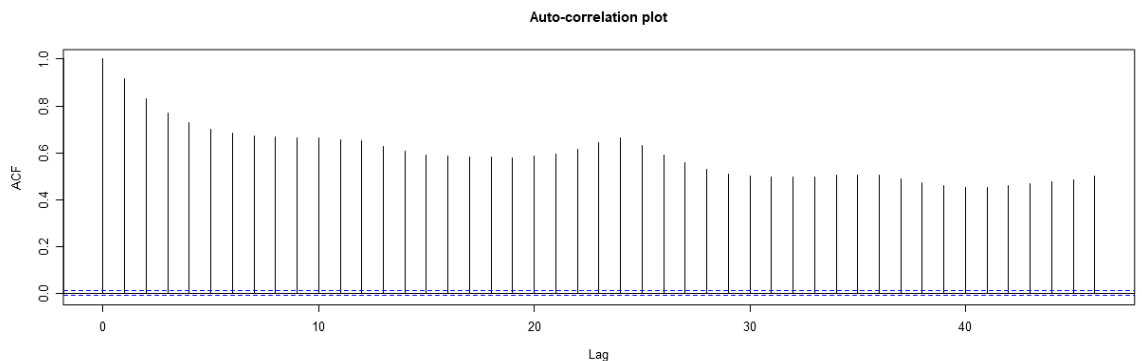
- Le premier vérifie que l'erreur est centrée, ce qui est vérifié par la ligne rouge du graphique horizontale et proche de 0.



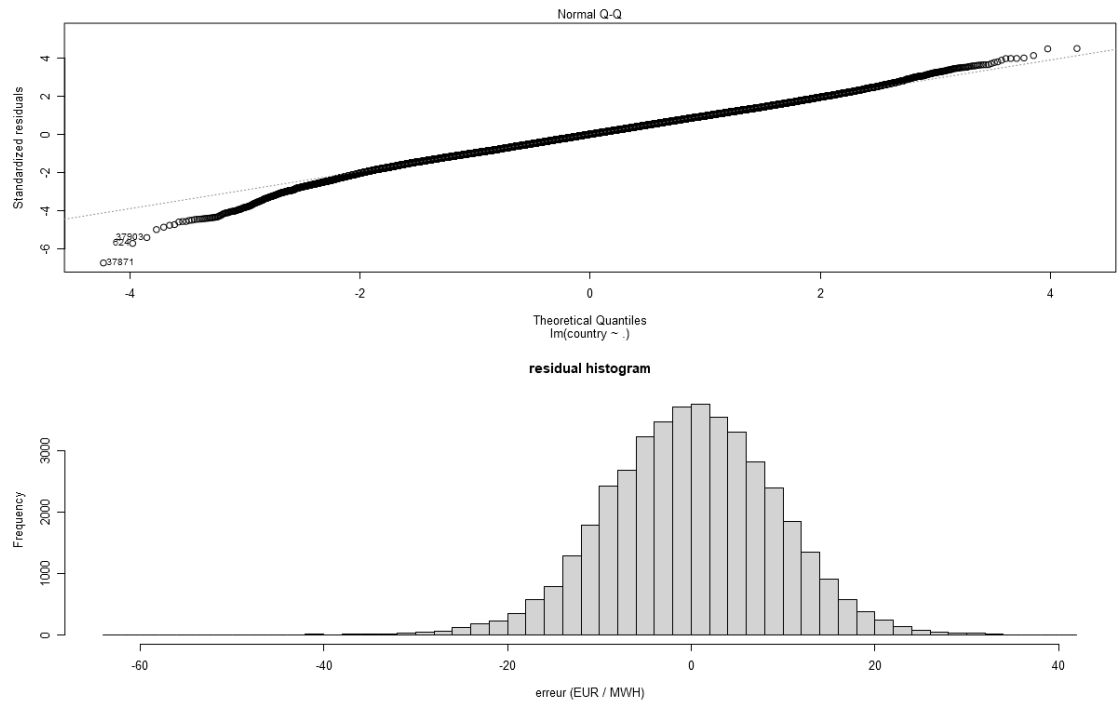
- Le second vérifie l'homoscédasticité, ce qui est vérifié par la ligne rouge horizontale et qui sépare de manière égale les points du graphique.



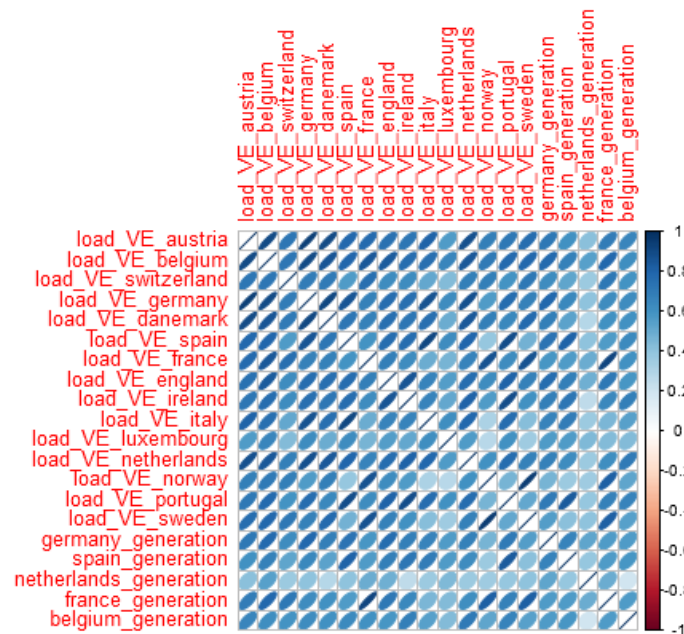
- Le troisième vérifie que les erreurs du modèle ne sont pas corrélées. Or ce n'est le cas pour aucun des pays. En effet, les barres horizontales sont toutes au-dessus de la ligne bleue, ce qui signifie que nous avons un problème d'autocorrélation.



- Les 2 derniers graphiques sont là pour vérifier que les résidus suivent une loi normale. Si les points sont alignés le long de la droite, ce postulat est validé. De plus, le second graphique permet de vérifier l'aspect visuel de la gaussienne des résidus.



Enfin, le dernier onglet, « Correlation », représente la corrélation entre les différentes variables explicatives utilisées, autres que les données temporelles (qui ne sont que des 0 ou 1), afin d'éviter les problèmes de multi-colinéarité. Dans notre cas, certaines données de consommation sont très corrélées et pourraient être enlevées mais nous n'avons pas constaté d'amélioration du modèle en les enlevant.





## Conclusion sur l'efficacité de l'outil

Le modèle utilisé ne permet pas de déterminer précisément le prix spot de l'électricité dans les 3 pays, son erreur étant élevée. Cela peut venir d'un manque de variables explicatives ou bien de la simplicité du modèle, qui n'est qu'une régression linéaire multiple. Nous avons cependant choisi de rester sur ce modèle simple car il permet d'avoir une meilleure idée de l'impact des variables explicatives sur le résultat prédit. En effet, même si notre précision est faible, la t-value des tableaux nous permet de connaître les éléments à surveiller car ayant un gros impact sur le prix spot de certains pays. Par exemple, une forte demande belge, couplée avec des prévisions de production hollandaise et française élevées, alors que l'Allemagne produit peu, devrait entraîner une forte hausse des prix en Allemagne. Afin d'être plus précis sur ces impacts, il faudrait alors retirer les variables explicatives trop corrélées, qui faussent la t-value et le coefficient, voire enlever les données temporelles, qui sont presque déjà « comprises » dans les données de consommation. Le modèle serait cependant moins précis pour la prédiction.