
DÉMARCHE STATISTIQUE POUR LA DÉTERMINATION DU PRIX D'ÉQUILIBRE DU MARCHÉ AFRR

par

Valentin Dorseuil, Théo Gaboriaud

Résumé. — Nous présentons dans ce rapport notre travail de prédiction du prix d'enchère du marché aFRR dans le cadre du cours de *machine learning* pour la prévision. Ce travail fut l'occasion de mettre en oeuvre les principes vus en cours dans un cadre concret pour lequel nous avons pensé une métrique *métier* que nous avons tâché de minimiser en déployant divers algorithmes de prédiction statistique. Le problème de prédiction s'est avéré particulièrement challengeant, notamment car il nous fut relativement difficile de battre notre baseline en raison du peu d'information de corrélation temporelle présent dans les données au delà de la veille du jour de la prédiction. Nous avons du pallier ce problème en mobilisant une autre source de données.



Table des matières

1. Introduction.....	2
1.1. Contexte.....	2
1.2. Objectifs.....	2
1.3. Démarche.....	2
1.4. Présentation du jeu de données.....	2
2. Métriques et baseline.....	3
2.1. Protocole d'évaluation.....	3
2.2. Métriques classiques.....	3
2.3. Métrique métier.....	3
2.4. Baseline.....	4
3. Exploration.....	4
3.1. Dépendances temporelles de la série à prédire.....	4
3.2. Pertinence d'autres données.....	5
4. Comparaison de modèles statistiques.....	6
4.1. Estimation par méthodes linéaires.....	6
4.2. Estimation par méthodes d'arbres.....	7
4.3. Boosting.....	8
4.4. GAM.....	8
4.5. Agrégation d'experts.....	10
5. Discussion.....	11
Conclusion.....	11

1. Introduction

1.1. Contexte. — Le marché de l'électricité repose sur un équilibre permanent entre l'offre et la demande. Pour garantir cette stabilité, différents mécanismes sont mis en place, parmi lesquels figure l'Automatic Frequency Restoration Reserve (aFRR). L'aFRR est un service de régulation permettant d'ajuster automatiquement la production ou la consommation d'électricité afin de maintenir la fréquence du réseau à son niveau nominal. Le prix d'achat et de vente des réserves d'électricité pour l'aFRR est défini selon un système d'enchère qui se tient quotidiennement.

Notre projet s'inscrit dans ce contexte. Nous avons à notre disposition un jeu de données contenant le prix d'équilibre des du marché d'enchère sur les neuf premiers mois de l'année 2024, ainsi que quelques statistiques sur les enchères. Notre projet consiste à prédire ce prix d'équilibre pour deux pays, l'Allemagne et l'Autriche.

1.2. Objectifs. — L'objectif principal de ce projet est donc de développer un modèle de prévision probabiliste permettant d'anticiper, à $J+1$, la valeur du prix d'équilibre des enchères. La prévision des prix pourra ainsi tirer parti de plusieurs sources d'informations :

- L'historique des prix observés sur ces enchères.
- Les tendances globales du marché
- Des variables explicatives externes (température, vent, ...)

1.3. Démarche. — Pour parvenir à ces objectifs, nous commençons par établir des métriques et un modèle de référence. Nous explorons ensuite les données à notre disposition puis mettons en œuvre divers méthodes de prédiction à la lumière de la compréhension progressive du problème tirée par les résultats de nos expérimentations. Dans un souci de simplicité, nos prédictions se concentrent sur un seul créneau d'enchère (il y en a 12 par jour au total).

1.4. Présentation du jeu de données. — Notre jeu de données contient le prix d'équilibre du jour ainsi que la demande en import ou export des deux pays. Les autres variables contenues dans le jeu sont constantes au fil du temps et n'apportent pas d'information. Les données sont explorées plus en détail en partie 3.

2. Métriques et baseline

Comme expliqué en 1.2, nous cherchons à réaliser des prédictions sur la série temporelle du prix d'équilibre du marché AFRR pour l'Allemagne et l'Autriche. Afin de comparer nos modèles entre eux, nous avons besoin de définir des métriques et un protocole d'évaluation unifié. Nous définissons aussi une *baseline*, qui correspond à un modèle naïf qui nous donne un score de base.

2.1. Protocole d'évaluation. — Nous adoptons un protocole d'évaluation basé sur une validation croisée en fenêtre extensible (*extending window*). À chaque itération, le modèle est entraîné sur les t premiers points de la série temporelle, puis évalué sur le point $t + 1$. La fenêtre d'entraînement s'élargit progressivement : elle débute avec l'intervalle $[0, 100]$, puis s'étend à $[0, 101]$, $[0, 102]$, et ainsi de suite. Cette approche permet d'exploiter un historique de données de plus en plus riche, reflétant ainsi un scénario réaliste où de nouvelles observations s'ajoutent continuellement. Elle favorise une meilleure adaptation du modèle aux tendances et aux variations de la série temporelle tout en assurant une évaluation rigoureuse et progressive de sa capacité de généralisation.

2.2. Métriques classiques. — L'évaluation des prédictions est tout d'abord réalisée à l'aide de l'erreur absolue moyenne (MAE), calculée uniquement à partir de $t = 100$, lorsque les premières prédictions commencent. Nous privilégions cette métrique à l'erreur quadratique moyenne (MSE) car les pertes monétaires sont linéaires par rapport aux erreurs de prédiction.

2.3. Métrique métier. — Pour contextualiser notre analyse d'une série temporelle portant sur le prix d'équilibre d'une enchère, nous avons défini une seconde métrique spécifique à notre étude, distincte de l'erreur absolue moyenne. L'objectif est de créer une métrique qui reflète le coût financier engendré par des prédictions erronées. Notre approche s'appuie sur un modèle simplifié d'enchères, offrant une estimation indicative du coût à des fins théoriques, sans viser l'exactitude. Ce choix, bien que perfectible, a selon nous le mérite de nous permettre d'adapter notre analyse à un objectif précis.

Nous considérons le modèle d'enchère suivant, où nous nous plaçons du point de vue d'un acteur souhaitant déterminer le prix de vente de ses produits sur le marché aFRR :

- Nous possédons n produits à vendre, chacun ayant un coût minimal de vente p_i (les coûts différents sont représentatifs du marché de l'électricité où le coût de production varie selon la méthode de production).
- On suppose que notre acteur veut maximiser son gain. Étant donné une prédiction p_{pred} du prix d'équilibre, il met chaque produit en vente au prix p_{pred} si $p_i < p_{pred}$ et le laisse à p_i sinon. Il espère donc réaliser un profit $(p_{pred} - p_i)_+$.
- Les résultats de l'enchère déterminent un prix d'équilibre p_{true} .

Nous analysons l'impact de la qualité de la prédiction faite modèle comme suit :

- Tous les produits dont le coût de vente est supérieur à p_{true} ne peuvent être vendus, ne réalisant ainsi ni profit ni perte (par perte, nous entendons ici le manque à gagner par rapport à une prédiction parfaite).
- Lorsque nous sur-estimons p_{pred} et fixons tous nos prix à p_{pred} , nous ne vendons rien, entraînant ainsi un coût fixe important en perte, indépendamment de la valeur de p_{pred} .
- Lorsque nous sous-estimons p_{pred} et fixons les prix p_i , inférieurs à p_{pred} , à la valeur p_{pred} , nous réalisons une plus-value mais aussi des pertes potentielles, proportionnelles à $(p - p_{pred})$.

Cette analyse est illustrée par le schéma en figure 1. Elle, nous conduit à choisir la métrique métier suivante, c'est-à-dire le coût de perte potentielle à minimiser, tracé en figure 2 :

$$\text{Cost}(p_{pred}, p_{true}) = a(p_{true} - p_{pred})_+ + C\mathbb{1}_{p_{pred} \geq p_{true}}$$

Ce coût donne une idée de la perte potentielle en argent par jour, avec un coût fixe important si p_{pred} dépasse p_{true} et un coût proportionnel sinon. Comme p est exprimé en €/MW/h, le paramètre a correspond à la puissance moyenne d'électricité échangée multipliée par la durée de la plage horaire (ici 4 heures). Pour coller à la réalité, a est calculé à partir de nos données. La

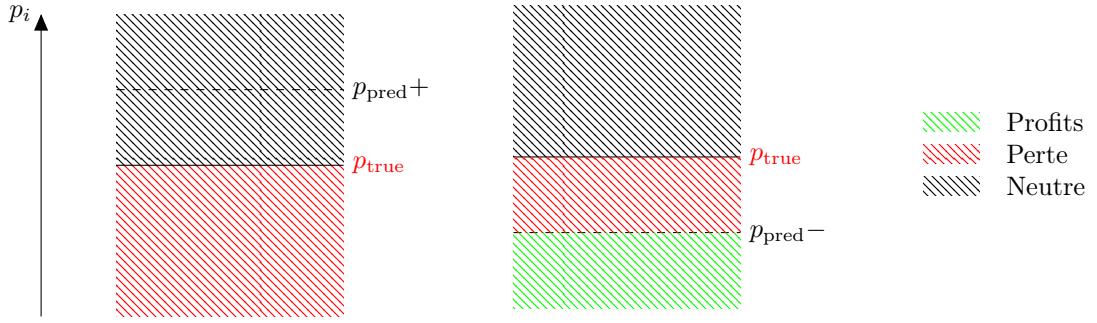


FIGURE 1. Représentation schématique du modèle d'enchère. À gauche, le prix d'équilibre est sur-évalué. À droite, le prix d'équilibre est sous-évalué.

condition $\text{Cost}(p_{pred} = 0) = \text{Cost}(p_{pred} = \infty)$ nous donne $C = a \times p_{true}$. Nous l'appellerons MML pour *Mean Monetary Loss*.

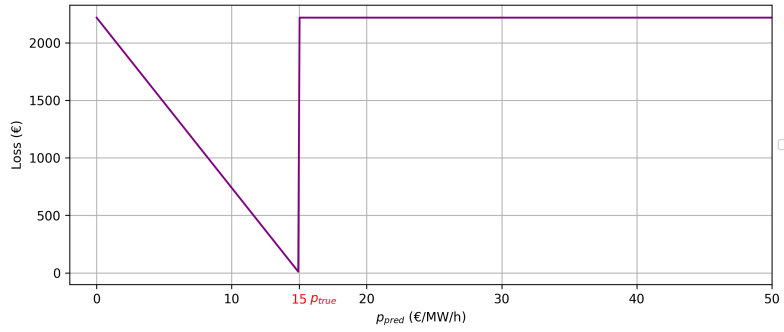


FIGURE 2. Graphique de la métrique de coût

2.4. Baseline. — Pour mesurer l'efficacité de nos futures modèles, nous avons choisi d'utiliser une approche simple, que nous appelons *baseline*, afin de disposer d'un point de référence pour évaluer les performances de nos modèles.

Pour la série temporelle (y_t) que nous souhaitons prédire, notre *baseline* est définie comme (y_{t-1}) . Cela signifie que notre baseline prévoit la valeur actuelle en utilisant la valeur du jour précédent. C'est un choix qui nous semble cohérent au vue de la corrélation très positive entre la valeur du jour et celle du lendemain (cf figure 4).

3. Exploration

Nous nous employons dans cette partie à tirer des informations qualitatives sur les données dans le but de pouvoir raffiner notre stratégie de prédiction.

3.1. Dépendances temporelles de la série à prédire. — Nous nous doutons que la série que nous tentons de prédire comporte de fortes dépendances temporelles, voici ce à quoi elle ressemble :

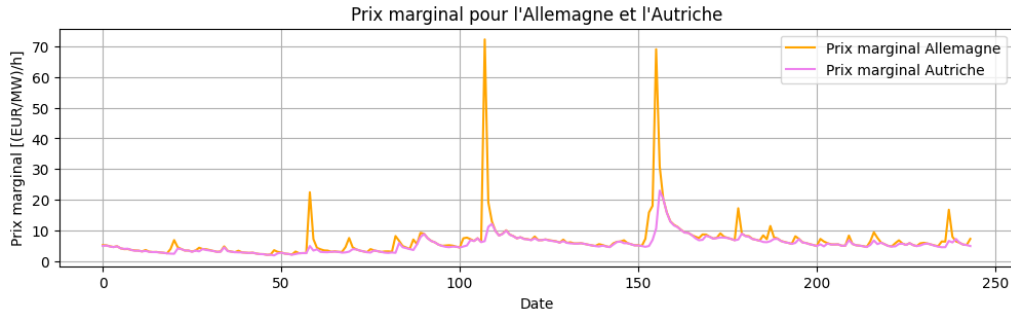


FIGURE 3. Série temporelle du prix marginal.

Sans pour l'instant chercher à les stationnariser, nous affichons les autocorrélogrammes des séries pour comprendre dans quelle mesure leurs dépendances temporelles nous permettront de prédire leurs prochaines valeurs :

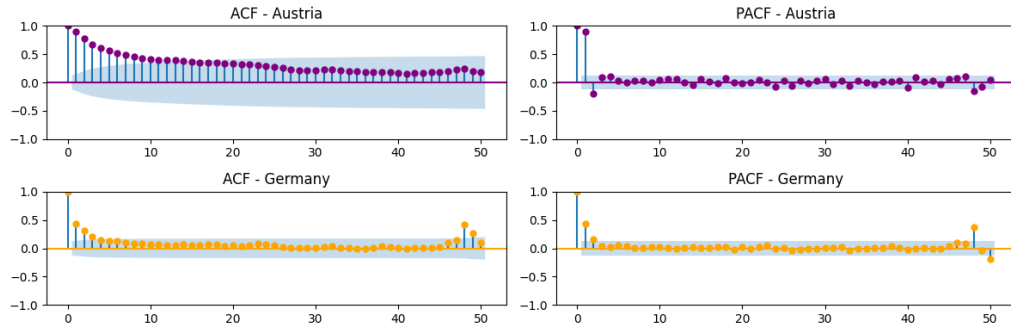


FIGURE 4. Autocorrélogrammes et autocorrélogrammes partiels des séries

Sans pour l'instant choisir une façon d'exploiter cette dépendance pour la prédiction, nous observons donc que cette série présente des grandes corrélations temporelles et comprenons donc que l'exploitation des valeurs passées représentera une cause déterminante du succès des prédictions.

3.2. Pertinence d'autres données. — Nous voulons également savoir si d'autres données pourraient nous être utiles pour la prédiction. Nous nous intéressons en particulier aux données météo, puisque nous supposons que la météo peut à la fois affecter la demande d'électricité (un temps nuageux pousse à rester chez soi) et l'offre d'électricité renouvelable (le vent et les éoliennes, l'ensoleillement et les panneaux solaires...), lesquels influenceront le prix d'équilibre sur le marché aFRR. Nous récupérons donc les données pertinentes que nous trouvons sur météo stat et affichons leurs corrélations avec la valeur d'intérêt :

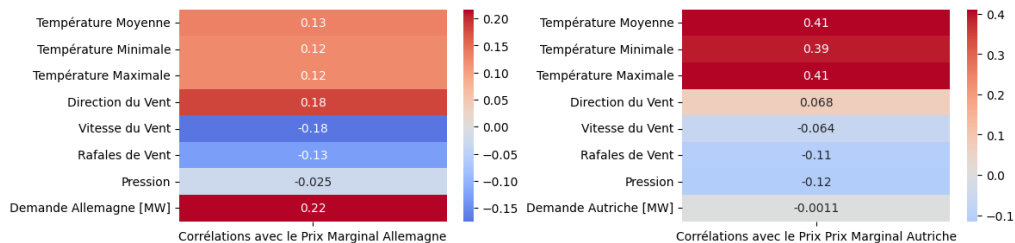


FIGURE 5. Corrélations entre les données météo et les variables d'intérêt

On observe que toutes les variables sont toutes deux corrélées avec au moins le prix d'enchère pour un des deux pays. Nous précisons d'ailleurs que ces données doivent être sues à l'avance pour

la prédiction, mais que nous faisons l'hypothèse que les outils de prédiction météo sont suffisamment précis du jour pour le lendemain pour que nous puissions négliger l'erreur de prédiction.

Conclusion. — L'exploration des données nous montre qu'à la fois les dépendances temporelles de la série et sa corrélation avec des variables auxiliaires liées à la météo sont essentielles pour faire le meilleur prédicteur possible. Il nous faudra donc penser des algorithmiques permettant d'exploiter simultanément ces deux sources d'information.

4. Comparaison de modèles statistiques

Dans cette section, nous mettons en oeuvre la démarche énoncée en introduction. Du plus simple au plus complexe, nous utilisons plusieurs modèles statistiques pour aboutir vers le prédicteur le plus performant possible.

4.1. Estimation par méthodes linéaires. — On va dans un premier temps s'intéresser à la modélisation suivante :

$$(1) \quad Y_{t+1} = \mathbb{E}[Y_{t+1}|\mathcal{F}_t] + \epsilon_{t+1}(T_{t+1})$$

où $\mathcal{F}_t = \sigma((Y_k)_{k \leq t})$ et le vecteur T_t désigne l'ensemble des variables auxiliaires à notre disposition. On peut d'abord remarquer que l'on ne prend pas en compte l'influence causale des termes $T_k, k \leq t$. Même si on pourrait argumenter que la météo deux jours avant le jour d'intérêt peut influencer la disponibilité de certaines énergies (comme l'énergie hydraulique), nous pensons que l'apport de cette dernière (privée de l'information contenue dans la météo de la veille) est négligeable.

Cette séparation additive permet donc d'essayer de déterminer d'abord $\mathbb{E}[Y_{t+1}|\mathcal{F}_t]$ puis de faire de la prédiction *stationnaire* sur $\hat{\epsilon}_{t+1}(T_{t+1}) = \hat{Y}_{t+1} - \hat{\mathbb{E}}[Y_{t+1}|\mathcal{F}_t] = g(T_{t+1})$ où g ne dépend pas de t .

Nous nous restreignons à chercher des **fonctions linéaires** pour exprimer $\hat{\mathbb{E}}[Y_{t+1}|\mathcal{F}_t]$ en fonction de \mathcal{F}_t , ce qui nous permet d'utiliser tous les outils type "ARMA" vus en cours.

Nous commençons par trouver une expression pour $\hat{\mathbb{E}}[\cdot|\mathcal{F}]$, et nous essayons ensuite de proposer des fonctions g .

4.1.1. Stationarisation. — La première étape pour déployer les modèles ARMA est de stationnariser la série. Dans notre cas, une simple différenciation $\tilde{y}_t = y_t - y_{t-1}$ suffit à passer le test Dickey-Fuller (hypothèse nulle (la série possède une racine unitaire) rejetée au seuil de 1% pour les deux séries Allemagne et Autriche). Voici les autocorrélogramme après différenciation :

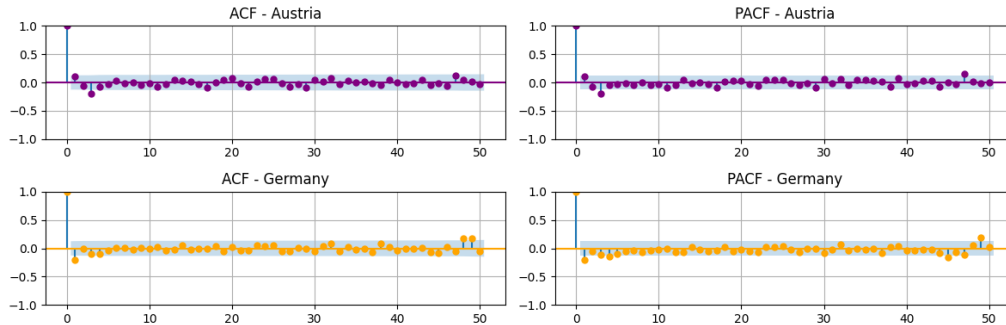


FIGURE 6. Autocorrélogrammes et autocorrélogrammes partiels des séries après différenciation

Difficile donc d'avoir des doutes sur la stationnarité de la série.

4.1.2. Modèle ARMA. — Les autocorrélogrammes et autocorrélogrammes partiels invitent à tester les $ARMA(p, q)$ pour $p, q \in \{1, \dots, 4\}^2$. Cependant, aucun modèle ne donne de résultat probant au regard des erreurs absolues et quadratiques sur les ensembles de test (ils obtiennent les mêmes erreurs que de prédire 0 à chaque pas de temps).

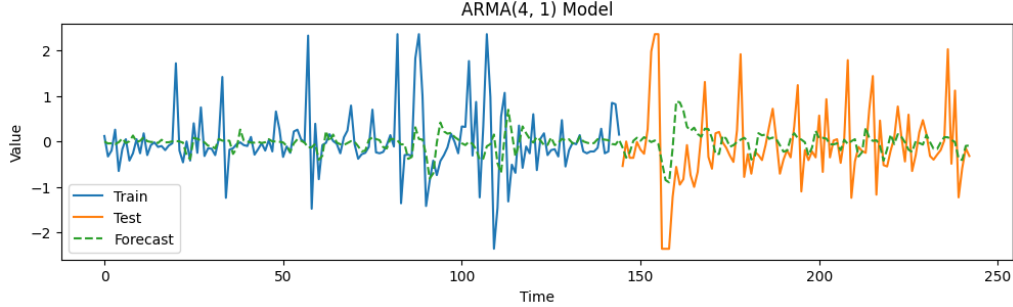


FIGURE 7. Exemple de prédicteur pour la série différenciée Autriche

Conclusion. — Les méthodes ARMA et le modèle (1) sont pertinents pour une structure de corrélation suffisamment riche pour que la solidité de la méthodologie d'estimation des coefficients des modèles ARMA compense la restriction à une contribution additive des variables extérieures. Or ici la structure d'autocorrélation est élémentaire : il est inutile d'aller plus loin dans la démarche.

4.1.3. Régressions linéaires. — À ce stade, puisqu'il est inutile de séparer l'estimation, on peut donc essayer de faire une regression linéaire sur toutes les variables présentes. La démarche habituelle pour cela consiste à d'abord sélectionner les variables d'intérêt avec un Lasso, puis à calculer les meilleurs coefficients avec un Ridge en trouvant la meilleur pénalisation par cross-validation. Dans notre cas, et comme vu dans le schéma ci-dessous :

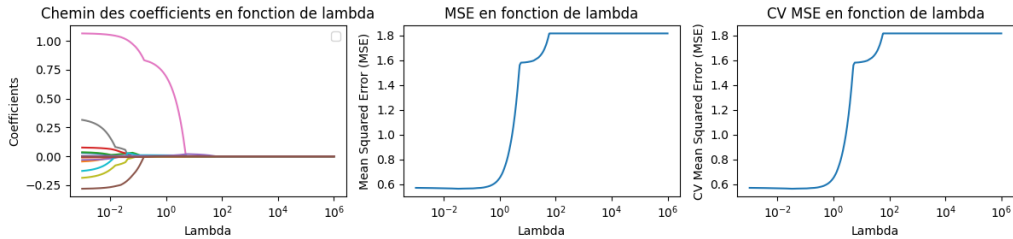


FIGURE 8. Estimateur Lasso pour l'Autriche

Le Lasso ne sélectionne que la dernière valeur y_{t-1} (seule valeur significative sur le schéma pour le minimum d'erreur de généralisation). On peut calculer la meilleure regression linéaire à l'aide d'un ridge pour obtenir la meilleure prédiction linéaire (valeurs numériques pour l'Autriche) : $\hat{y}_t \approx 0.8\hat{y}_{t-1} + 0.8$. Pour l'Allemagne, il se passe la même chose. Nous reportons les scores dans le tableau 1.

4.2. Estimation par méthodes d'arbres. — Jusque là, nous avons donc vu que les données météo peuvent être utiles à la prédiction (par leur corrélation avec la variable d'intérêt), mais que leur donner un poids linéaire ne convenait pas. Nous nous tournons donc vers des méthodes non linéaires. Pour commencer, nous nous intéressons aux méthodes utilisant des arbres. Contrairement aux méthodes vues précédemment, les méthodes à arbres ne sont initialement pas conçues pour traiter des séries temporelles mais peuvent être adaptées, notamment dans notre cas où nous possédons aussi des *features* externes. L'étude de la *PACF* en figure 4, nous incite à ne rajouter que y_{t-1} comme *feature* dans notre jeu de données (en plus des données externes).

4.2.1. Méthode CART. — Le paramètre le plus important à déterminer lorsqu'on réalise un arbre CART est la profondeur de ce dernier. En raison de notre faible nombre de données et de la présence de certaines valeurs extrêmes, la méthode par validation croisée k -fold aléatoire, ne fonctionne pas (la profondeur est trop sensible au découpage du k -fold). Nous utilisons une méthode de k -fold non aléatoire avec des *sliding window* de taille 10.

Nous calculons donc la profondeur optimale et trouvons pour les deux pays une profondeur optimale de 3. En appliquant le protocole défini en 2.1, nous calculons nos métriques sur les modèles CART, avec la profondeur trouvée. Les résultats ne sont pas très probants, car les modèles ont du mal à gérer les extrêmes, du fait d'une profondeur assez faible et d'un modèle assez simple.

4.2.2. Méthode CART avec régression Ridge. — Pour pallier à ce problème, nous implémentons un modèle de CART avec une régression Ridge au sein des feuilles dans l'idée de complexifier le modèle. Un modèle Ridge est donc entraîné pour chaque feuille. Le modèle performe un peu moins bien que le CART classique.

De manière générale, les mauvais résultats des méthodes CART s'expliquent probablement par le fait que chaque feuille d'un arbre ne contient qu'une seule valeur, alors que nos données dépendent fortement de y_{t-1} , une valeur très volatile.

4.2.3. Random Forest. — Il est donc naturel de s'intéresser aux modèles de Random Forest, pour complexifier nos modèles d'arbres qui semblent trop simples. Nous appliquons des techniques de Bootstrap afin de créer des arbres les moins corrélés possibles. À l'aide d'un *grid search*, nous trouvons une profondeur idéale de 10 pour l'Autriche et 3 pour l'Allemagne, ainsi qu'un nombre d'estimateurs de 60 pour les deux.

Comme précédemment, le modèle a du mal à prédire les valeurs extrêmes. Cela est particulièrement accentué par le faible nombre de données que nous possédons. Les résultats sont tout de même meilleur que pour les méthodes CART (cf table 1) et s'approchent du *baseline*. Les graphiques sont représentés en figure 9.

Conclusion. — Les méthodes à arbres comme CART et Random Forest montrent des performances limitées dans notre cas, *i.e.* celui d'une série temporelle avec peu de données. Ces méthodes semblent peiner à capturer parfaitement les dépendances temporelles, ce qui est nécessaire pour améliorer la baseline. Bien que le Random Forest améliore la précision, le manque de diversité dans les données empêche sûrement de générer des graphes suffisamment divers. De plus, le modèle étant assez opaque, son interprétation reste difficile.

4.3. Boosting. — Nous testons également l'algorithme XGBoost sur le problème. Nous constatons qu'il n'est pas plus efficace de prédire le résidu ($X_t - X_{t-1}$) que la valeur directement. Les résultats sont reportés dans 1.

4.4. GAM. — Enfin, nous nous intéressons aux GAMs (Generalized Additives Models), qui étendent les modèles linéaires en permettant au prédicteur linéaire d'être une somme de fonctions lisses (souvent des splines) des variables explicatives. Cela permet de capturer des relations non linéaires entre la variable réponse et les prédicteurs.

Il existe plusieurs méthodes pour adapter les GAMs pour les séries temporelles et nous avons décidé de nous concentrer sur les filtres de Kalman, étant donné que la majorité de notre information semble être contenu dans la partie temporelle de notre série (plus que dans les features). Les filtres de Kalman s'intègrent de la manière suivante dans un modèle GAM :

$$\mathbb{E}[y_t] = \beta_0 + \sum_{j=1}^p f_j(x_{j,t}) + \epsilon_t,$$

$$\epsilon_t = A\epsilon_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, Q),$$

où y_t est la valeur observée à l'instant t , β_0 est l'ordonnée à l'origine, $f_j(x_{j,t})$ sont les fonctions lisses appliquées aux prédicteurs, ϵ_t représente le résidu à l'instant t qui évolue selon une dynamique

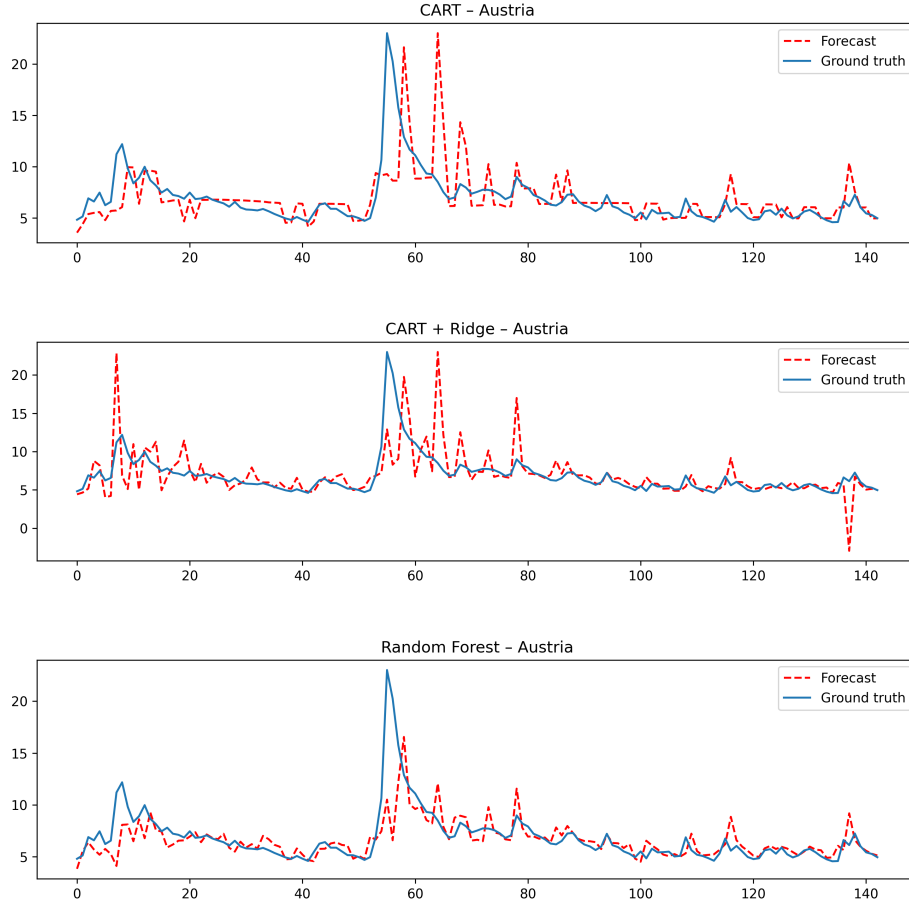


FIGURE 9. Comparaison des modèles CART, CART avec régression Ridge, et Random Forest.

linéaire, A est la matrice de transition et w_t est le bruit de processus, supposé gaussien avec une covariance Q .

Le filtre de Kalman intervient ici pour estimer de manière récursive ϵ_t à partir des observations, permettant ainsi de corriger en temps réel les prédictions du GAM en tenant compte de la dynamique des erreurs.

Après de nombreuses expériences, nous avons trouvé que les modèles les plus performants étaient ceux avec des faibles DOF (Degree of Freedom), *a priori* en raison du faible nombre de données. Pour l'Allemagne, le modèle le plus performant utilisait des spleens de la série à prédire, décalée d'un pas de temps, ainsi que de la demande d'import-export de l'Allemagne. Pour l'Autriche il s'agissait d'un modèle avec la température moyenne du jour et la série à prédire (décalée aussi d'un pas de temps). Pour chaque pays nous avons calculé plusieurs paramètres de régularisation différents (λ) afin d'obtenir les meilleures performances comme représenté en figure 10. Pour l'Autriche le meilleur modèle fut obtenu pour $\lambda = 10$ ce qui correspond à un degré de liberté de 9.8. Pour l'Allemagne on obtient un $\lambda = 500$ pour un degré de liberté de 7.

Conclusion. — Les GAMs sont des modèles puissants qui permettent de capturer des relations non linéaires entre les variables de manière flexible. Ils surpassent nos autres modèles pour notre métrique métier MML, car ils semblent sous estimer plus souvent la valeur de p . Cependant, leur optimisation fut délicate, notamment en présence de peu de données et parfois d'un nombre élevé de degrés de liberté, ce qui a rendu l'ajustement des paramètres plus difficile.

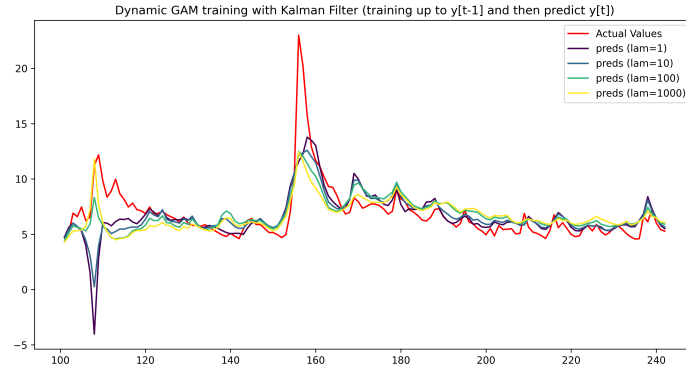


FIGURE 10. Comparaison de modèles de GAM avec filtre de Kalman pour différentes valeurs de λ pour l’Autriche.

4.5. Agrégation d’experts. — Pour finir, nous agrégeons nos modèles en utilisant un algorithme d’agrégation d’expert et en le monitorant ou bien par notre métrique métier ou bien par la MAE. Pour ceci nous avons utilisé l’algorithme SLSQP (*Sequential Least Squares Quadratic Programming*) qui considère notre problème comme un problème linéaire d’optimisation sous contraintes. Pour chaque donnée à prédire, nous calculons les meilleurs poids sur les modèles pour minimiser l’erreur sur le jeu d’entraînement, puis nous utilisons ces poids pour prédire la prochaine valeur.

Assez paradoxalement, l’agrégation d’expert donne de meilleurs résultats sans le GAM, qui était pourtant le modèle donnant la meilleure MML pour nos deux pays. Nous mentionnons également que notre métrique d’évaluation en *extending window* ne garantit pas de stabilité des poids avec le temps.

Cette méthode nous a permis d’améliorer nos résultats et ainsi d’obtenir des valeurs plus basses pour la métrique métier, confirmant le fait que nos modèles ne sont pas tous parfaitement corrélés. Les résultats détaillés pour chaque modèle, pays et métrique sont présentés en table 1. Toutefois, il nous semble important de préciser que l’agrégation d’expert que nous avons implémenté nous semble naïve puisqu’elle minimise simplement la perte sur le jeu d’entraînement. Dès lors, nous pensons qu’une voie très sérieuse d’amélioration serait la pénalisation de l’agrégation pour moins *overfitter*. Cela reste néanmoins un challenge puisque notre métrique *métier* n’est pas différentiable (et la lisser rendrait quand même son optimisation différentielle fastidieuse).

Nom de la méthode	MAE Aus	MAE Ger	MML Aus	MML Ger
Baseline	0.71	2.65	1144	1165
Méthodes linéaires	0.38	1.36	850	936
CART	1.37	4.17	1090	1020
CART + Ridge	2.42	5.1	1274	1228
Random Forest	0.97	3.54	1007	1203
XGBoost	1.12	4.04	1182	1132
GAM	1.01	1.99	774	860
Aggrégation d’experts MML	-	-	732	938
Aggrégation d’experts MAE	0.83	3.11	-	-

TABLE 1. Comparaison des méthodes (Aus pour Autriche et Ger pour Allemagne). Les quatre meilleurs résultats sont en gras.

5. Discussion

Un point qui peut sembler étonnant au premier regard des résultats, est que l’Autriche a une MAE bien moins élevée que l’Allemagne (en raison d’une série temporelle avec moins de valeurs extrêmes) mais une MML similaire. Cela est dû au fait qu’en moyenne, l’Autriche exporte plus d’électricité que l’Allemagne sur le marché aFRR sur ce créneau. Le paramètre a que nous avons calculé pour notre coût dans la MML est donc plus élevé pour l’Autriche que pour l’Allemagne.

De manière générale, nos résultats suggèrent que pour notre jeu de données, qui est relativement petit et présente une forte dépendance temporelle à $t - 1$ avec peu de corrélation avec les *features*, des modèles simples tels que les méthodes linéaires donnent des résultats déjà presque optimaux. Cette tendance peut s’expliquer par le fait que les autres modèles testés, très efficaces pour les données tabulaires, ne sont pas spécifiquement adaptés aux séries temporelles. Toutefois, des approches comme les GAMs et les Random Forests parviennent également à produire de très bons résultats sur notre jeu de données.

Comme on pouvait s’y attendre, l’agrégation des données permet d’améliorer les résultats, en augmentant la performance globale du modèle. Surtout, l’agrégation est la seule étape de notre démarche où l’on peut directement maximiser la performance de notre modèle selon notre métrique. Cependant, cette amélioration s’accompagne d’une perte de transparence et d’interprétabilité. En effet, plus les données sont agrégées, plus il devient difficile de comprendre précisément comment chaque facteur contribue aux prédictions du modèle.

Nous pourrions également tester, (nous l’avons notamment fait pour le XGBoost et cela ne fonctionnait pas dans notre cas) si séparer la prédiction temporelle de la prédiction basée sur les *features* apporte réellement une amélioration en termes de performance. De plus, il serait intéressant de tester d’autres types de modèles, tels que les réseaux de neurones (*deep learning*) dans le cas où plus de données seraient disponibles ou encore des modèles de séries temporelles comme les LSTMs.

Pour finir, un petit point pratique sur notre métrique métier, juste pour valider nos estimations. Avec la MML (notre métrique), on estime un manque à gagner moyen de 750€ par jour pour le meilleur modèle pour l’Autriche. Si l’on multiplie ça par 12 créneaux et 365 jours par an, on arrive à plus de 3 millions d’euros de perte annuelle. Voilà pourquoi il est crucial d’avoir des modèles qui fonctionnent bien et qui aident à optimiser tout ça !

Conclusion

En conclusion, ce projet fut l’occasion de mettre en œuvre les principes vus en cours dans le cadre d’un exercice concret. Il mit en évidence la diversité des options en termes de modélisation et de stratégie d’estimation, options qui deviennent d’autant plus riches lorsqu’elles sont orientées par une proximité avec les données et un objectif de prédiction précis.