

Mémoire présenté
pour l'obtention du Diplôme de
Master mention ÉCONOMETRIE, STATISTIQUES
Parcours *Économie Quantitative pour la DÉcision*

Le 19 septembre 2025

Par : GHASSIR Meryam

Titre : Analyse et quantification des risques de marché et de Flex d'un portefeuille de consommation dans le secteur de l'électricité.

Confidentialité : Non

Tuteur universitaire : M. RULLIERE Jean-Louis

Entreprise / Institution de recherche : Union des Producteurs Locaux d'Électricité (Énergie d'Ici)

Tuteur Entreprise : M. GARCIER Antoine

FORMATIONS

(ACTUARIAT) (ECONOMETRIE & STATISTIQUES) (CONTINUE & VAE) (DOCTORALE)

Remerciements

Je tiens à remercier chaleureusement Antoine GARCIER, directeur général d’Energie d’Ici de m’avoir offert l’opportunité de réaliser mon stage dans le secteur de l’électricité, domaine que je ne connaissais pas mais qui a suscité un fort intérêt au fur et à mesure de l’avancée de mon stage. Je remercie grandement Jodie REASA, la responsable du pôle modélisation pour son suivi, son écoute et sa disponibilité. Elle m’a permis d’en apprendre davantage sur le fonctionnement du secteur, ses spécificités et les possibilités immenses de modélisation dans divers pans de ce secteur, tout cela en respectant mon rythme de travail et en me laissant une certaine liberté de travailler sur divers sujets. Elle m’a également partagé ses expériences, ses connaissances et ses conseils qui m’ont permis de formaliser ce mémoire qu’elle a volontiers relu, corrigé et enrichi, et je la remercie une seconde fois pour cela !

Je tiens également à remercier l’ensemble de mes collègues du service comptabilité et financier, au service adhérent en passant par les autres membres du service support et de la direction pour leur joie de vivre, leur ouverture et leur accueil qui m’ont permis de m’intégrer au sein de cette entreprise.

J’adresse également mes remerciements à l’université Lyon 1 pour leur accompagnement scolaire et spécifiquement à l’ISFA et aux enseignants qui m’ont permis de monter en compétences et m’ont apporté les connaissances nécessaires au métier tout en cultivant une certaine curiosité pour continuer à s’améliorer et à apprendre afin de continuellement évoluer. Je remercie spécifiquement Monsieur Jean-Louis RULLIERE de m’avoir offert l’opportunité d’intégrer le master EQUADE et d’avoir cru en ma future réussite dans ce parcours.

J’aimerais également remercier l’équipe administrative et surtout Juliette SUDON pour son écoute, son dynamisme et sa volonté d’intégrer tous les étudiants dans la vie de l’ISFA en nous partageant continuellement des bonnes informations et des opportunités et sans qui je n’aurais pas participé à l’organisation du forum de l’ISFA et au concours d’éloquence. Ces expériences m’ont permis de mieux m’intégrer auprès des autres étudiants, de rencontrer des professionnels de divers secteurs, de parler en public sans appréhension, ce qui m’a permis de m’enrichir personnellement et professionnellement.

Je remercie mes camarades de classe et mes camarades des autres parcours, devenus amis pour certains et avec qui cette année n’aurait pas eu le même impact. Nous avons partagé ensemble nos connaissances et expériences, nos sessions révision et rédaction de mémoire mais aussi des sorties amusantes et des moments de rires et de bonne humeur.

Je remercie également mes parents qui m’ont insufflé la volonté d’étudier, d’être curieuse intellectuellement et de persévérer dans tous les projets que j’entreprenais.

Ce mémoire clos le chapitre de ma vie d’étudiante parsemé d’apprentissage, de recherche, de questionnement, de doute et aussi de bonne rencontre, de bons souvenirs, de découverte et de développement de ma personne. Il ouvre un nouveau chapitre sur une vie de jeune active que je l’espère aussi enrichissant et plein de nouvelles aventures.

Résumé

Ce mémoire analyse les principaux risques auxquels sont confrontés les fournisseurs d'électricité en se concentrant plus précisément sur les aléas liés à la consommation du portefeuille clients d'Énergie d'Ici. Dans un contexte post-crise énergétique marqué par une volatilité persistante des prix et une évolution réglementaire continue, l'objectif est d'identifier, d'évaluer l'impact et de quantifier ces risques.

La démarche repose d'abord sur une analyse exploratoire des profils et comportements de consommation puis sur le calcul et la modélisation de primes de risque, avec projection à l'horizon 2026. L'étude met en évidence la variabilité des écarts entre consommation prédite et réelle influencée à la fois par des facteurs exogènes (conditions météorologiques, calendrier) et par des comportements de consommation divers et spécifiques. L'équilibre permanent nécessaire entre consommation et production impose de compenser ces écarts sur le marché spot, ce qui peut générer des coûts difficiles à anticiper dans un environnement de prix extrêmement volatil.

La quantification de ces risques a permis de mesurer concrètement leur impact économique et de constituer une base chiffrée pour appuyer les recommandations tarifaires. Ce travail se distingue par l'articulation entre une analyse exploratoire, des modélisations statistiques et une projection prospective, offrant ainsi des pistes d'action adaptées aux spécificités d'Énergie d'Ici.

Mots clés : énergie, électricité, gestion des risques, risque de volume, risque de Shape, risque de flexibilité, risque de marché, équilibre offre-demande, consommation d'électricité, volatilité des prix, optimisation tarifaire, primes de risque, exposition financière, gestion de portefeuille

Abstract

This academic work analyses the main risks faced by electricity suppliers, with a particular focus on the consumption uncertainties of Énergie d'Ici's customer portfolio. In a post-crisis context marked by a persistent high volatility, and continuous regulatory changes, the purpose of this study is to identify, evaluate and quantify these risks.

The approach first relies on an exploratory analysis of consumption profiles and behaviours, followed by the calculation and modelling of risks premia with a projection in 2026. This dissertation highlights the variability of gaps between predicted and real consumption influenced by both exogenous factors (such as weather conditions and calendar effects) and by diverse and specific consumer behaviours. The permanent balance required between consumption and production forces the need to compensate for these gaps on the spot market, which can generate costs that can be difficult to anticipate in an extremely volatile price environment.

The quantification of these risks made it possible to concretely assess their economic impact and to establish a numerical basis to support the pricing recommendations. This work stands out through its combination of exploratory analysis, statistical modelling, and forward-looking projections, thereby providing actionable insights tailored to the specificities of Énergie d'Ici.

Keywords: energy, power energy, risk management, volume risk, Shape risk, flexibility risk, market risk, supply-demand balance, electricity consumption, price volatility, pricing optimization, (risk) premiums / premia, financial exposure, portfolio management

Table des matières

Remerciements	2
Résumé	3
Abstract	4
Table des matières	5
Introduction	8
Présentation de l'entreprise	9
1. Contexte, enjeux et défis du marché de l'électricité	11
1.1. Le marché de l'électricité	11
1.1.1. Historique	11
1.1.2. Fonctionnement	12
1.1.3. Spécificités	15
1.2. Les facteurs de volatilité	16
1.2.1. Impact conjoncturel	16
1.2.2. Les EnR comme catalyseur de volatilité	17
1.2.3. Transformation du mix énergétique	18
1.3. Offre et Demande en électricité	19
1.3.1. La demande (Consommateurs)	19
1.3.2. L'offre (Producteurs et Fournisseurs)	20
1.3.3. L'équilibre offre/demande, un impératif systémique (Responsable d'équilibre)	21
1.4. Tarification et cadre réglementaire	22
1.4.1. La formation des prix	23
1.4.2. La composition globale du prix de l'électricité	24
1.4.3. Tarifs spécifiques et nouveaux enjeux tarifaires	25
1.5. Cartographie globale des risques	25
1.5.1. Le risque de marché	26
1.5.2. La constitution du risque de Flex	27
1.5.3. Les autres risques	29
1.5.4. Stratégies de gestion et Hedging	31
2. Analyse des risques	33
2.1. Présentation du portefeuille clients et des variables	33
2.1.1. Segmentation clients	33
2.1.2. Variables clés et besoins analytiques	34
2.2. Données	38
2.2.1. Collecte et structuration	39

2.2.2.	Prétraitement et nettoyage	40
2.3.	Analyse exploratoire des données	42
2.3.1.	Statistiques descriptives	42
2.3.2.	Analyse des résidus	45
2.4.	Cartographie des groupes de risque	50
2.4.1.	Clustering	51
2.4.2.	Répartition des risques	54
3.	Calculs et modélisations	57
3.1.	Calculs et quantification des primes de risque	57
3.1.1.	Prime de risque de marché	57
3.1.2.	Prime de risque de Flex	60
3.1.3.	Effet de foisonnement	65
3.2.	Simulation des primes	68
3.2.1.	Simulation Monte Carlo : méthode directe	68
3.2.2.	Simulation Monte Carlo : méthode des variations	70
3.2.3.	Simulation avec le processus Ornstein-Uhlenbeck	72
3.3.	Modélisations avancées	76
3.3.1.	Variables explicatives	76
3.3.2.	Modèles testés	78
3.3.3.	Comparaison des performances et résultats pour chaque risque	82
3.4.	Prédiction des primes pour l'année 2026	84
3.4.1.	Features utilisées	84
3.4.2.	Risque de marché	86
3.4.3.	Risque de Flex	87
4.	Recommandations stratégiques et limites de l'étude	90
4.1.	Recommandations concrètes	90
4.2.	Limites générales	91
4.3.	Perspectives	92
	Conclusion	95
	Bibliographie	96
	Annexe 1	100
	Annexe 2	100
	Annexe 3	103
	Annexe 4	104
	Annexe 5	105

Annexe 6	107
Annexe 7	113
Annexe 8	113
Annexe 9	116
Annexe 10	117
Annexe 11.....	118
Annexe 12	120
Annexe 13	122
Annexe 14	123

Introduction

Ces dernières années, le secteur de l'électricité a été profondément bouleversé par des transformations majeures : des événements climatiques extrêmes et de plus en plus fréquents tels que les vagues de chaleur et les périodes de grand froid, qui alternent avec des crises économiques et des tensions géopolitiques. Loin d'être isolés, ces phénomènes s'inscrivent dans un contexte global marqué par l'accélération de la transition énergétique et l'intégration européenne des marchés de l'énergie. Les systèmes électriques nationaux, historiquement régulés et centralisés, doivent aujourd'hui composer avec des contraintes croissantes : un besoin accru de flexibilité de l'offre et de la demande, une volatilité importante des prix et un environnement incertain dans lequel évolue le secteur du fait de ses spécificités mais aussi par les réglementations et l'actualité géopolitique dans le monde.

La non-stockabilité¹ intrinsèque de l'électricité complique la gestion de ce marché. En effet, pour garantir une alimentation fiable et en continue, l'offre doit s'ajuster à la demande en temps réel. Cette contrainte impose aux acteurs d'anticiper, avec une précision maximale, la consommation. C'est un défi qui est rendu complexe par les variations autant de production que de consommation quotidiennes, hebdomadaires et saisonnières, mais aussi par l'impact de la météo, des jours fériés et des événements imprévus.

Le marché électrique, désormais libéralisé et fortement interconnecté au niveau européen, est caractérisé par une forte instabilité des prix notamment de très court terme appelé prix spot pouvant évoluer jusqu'à 15 minutes avant la livraison de l'énergie avec parfois des fluctuations extrêmes. Cette volatilité est accentuée par le développement rapide des énergies renouvelables notamment solaire et éolien, qui en raison de leur intermittence, introduisent des incertitudes supplémentaires dans la production.

Dans ce contexte, la gestion des risques devient cruciale pour les fournisseurs d'énergie et les autres acteurs. Il s'agit pour eux de maîtriser le risque lié à l'écart entre la consommation réellement observée et les prévisions effectuées, le risque découlant de la fluctuation des prix sur le marché et celui associé à la solvabilité des clients. Cette gestion fine des risques doit permettre d'ajuster la tarification pour éviter les pertes financières tout en assurant une certaine compétitivité et équité tarifaire. L'enjeu est donc double : anticiper des besoins fluctuants difficiles à modéliser et établir des prix qui reflètent au mieux la réalité du marché et des risques associés. De ce fait, ce mémoire cherchera à déterminer comment implémenter une gestion du risque basée sur l'analyse de la consommation électrique et sur des prévisions robustes, afin de déterminer la tarification des risques tout en minimisant l'exposition globale à ces derniers.

Pour répondre à cette problématique, le mémoire s'organisera en trois principales parties présentant premièrement le contexte et les enjeux du secteur de l'électricité pour comprendre son fonctionnement et les risques qui lui sont rattachés. Puis, en deuxième partie, l'analyse du portefeuille de clients qui permettra d'entamer dans la troisième partie les calculs de prime de risque qui constitue une première approche à la quantification des risques. Cela servira de base pour les modélisations et prévisions effectuées qui vont suivre pour la quantification future du risque. Enfin, la conclusion permettra de synthétiser cette étude en présentant les recommandations concrètes, les limites et les perspectives d'amélioration.

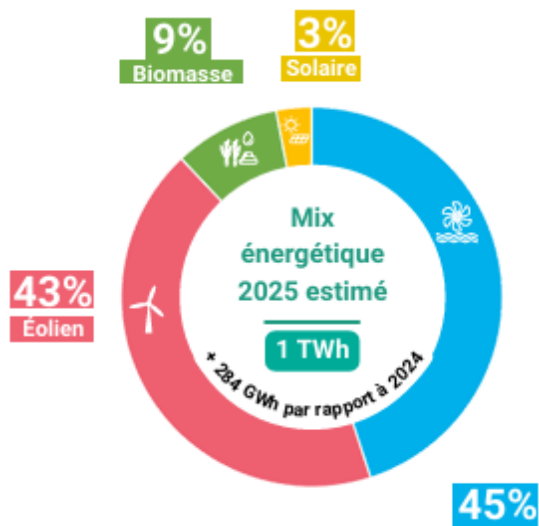
¹ L'énergie électrique est un bien primaire économique spécifique du fait qu'il doit être consommé quasiment de suite et ne peut être stocker pour des échanges ultérieurs ce qui accroît l'instabilité du marché.

Présentation de l'entreprise

Ce mémoire a été réalisé dans le cadre de mon stage au sein de l'entreprise Énergie d'Ici, fournisseur d'électricité locale et renouvelable en France.

Fondée en 2016 par plusieurs producteurs associés qui souhaitent, grâce à la libéralisation progressive du marché de l'électricité, vendre directement une énergie renouvelable produite localement au consommateur final.

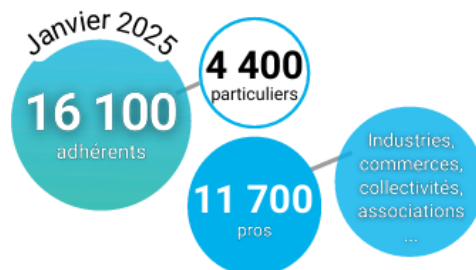
Comprenant initialement que des centrales hydrauliques, le portefeuille de production s'est étoffé en intégrant les premières centrales solaires en 2019 puis les centrales éoliennes en 2020 et biomasses en 2022. Aujourd'hui, Énergie d'Ici fournit de l'électricité via près de 230 centrales hydrauliques (135 centrales), éoliennes (23 parcs), solaires (69 parcs) et biomasses (2 centrales) constituant la base de son mix énergétique avec 1 TWh d'électricité renouvelable agrégée et répartie de la manière suivante :



Où l'hydraulique, énergie fondatrice du fournisseur, garde sa place historique de principale source de production (45% du mix énergétique) mais, est, en 2025, suivit de près, en proportion de production, par les parcs éoliens pour 43%. Se retrouve ensuite, dans une proportion moindre mais non négligeable du portefeuille, les centrales biomasses et parcs solaires pour respectivement 9% et 3% de la part de production.²

Le nombre de centrales évolue chaque année en fonction des besoins des adhérents avec une proportion des sources d'énergie qui change d'une année à l'autre.

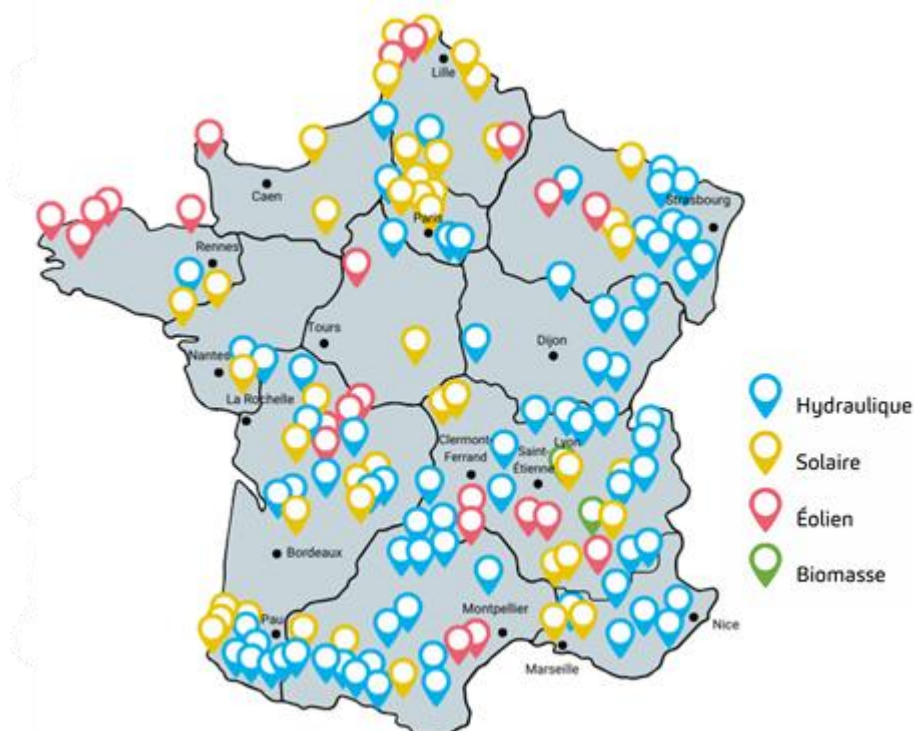
Plus qu'un fournisseur, Énergie d'Ici propose d'accompagner ses plus de 16 000 adhérents³ dans la gestion et l'optimisation de leur consommation d'énergie. Des clients en grande partie professionnels de tout secteur et de toute taille : de l'industrie au tertiaire en passant par l'artisanat et le commerce ; également de nombreuses collectivités locales, associations et services publics mais aussi des adhérents particuliers dit résidentiels représentant près d'un quart du portefeuille de consommation.



² Données issues du dossier de presse de l'entreprise Energie d'Ici présent sur leur site Internet, version 2025, [Dossier de presse 2025](#)

³ Les clients sont des adhérents à des formules d'achat d'électricité renouvelable et locale

Les adhérents sont présents dans tout le territoire français et pour préserver la localité de la production, le portefeuille de production comprend des centrales réparties dans toute la France comme le montre la carte suivante afin d'avoir une production au plus proche du consommateur :



4

Cette proximité entre producteurs et consommateurs fait partie de l'une des valeurs de l'entreprise garantissant un circuit court de l'énergie et un ancrage local conséquent démontrant l'attachement au territoire du fournisseur qui contribue également au développement socio-économique des territoires. La pérennité de l'activité que procure les énergies renouvelables permettent de contribuer favorablement à la transition énergétique, enjeu majeur de ces dernières années et de celles à venir pour la France.

Énergie d'Ici est un acteur important en matière d'investissement notamment pour la rénovation et la pérennisation des actifs productifs avec plus de 100 millions d'euros investis depuis sa création.

La réalisation de ces missions est possible grâce aux 30 collaborateurs⁵, répartis à Lyon et à Pau travaillant au développement d'Énergie d'Ici via la relation avec les adhérents, les partenariats avec les producteurs, la facturation, les prévisions de la consommation et de la production et d'autres services essentiels au bon fonctionnement.

Énergie d'Ici combine ainsi production 100% renouvelable et fourniture d'électricité de proximité en proposant des offres locales et durables. Grâce à un portefeuille diversifié de moyens de production, à ses investissements, à ses valeurs et engagements, le fournisseur se positionne comme un acteur clé de la transition énergétique en France et dans le marché de l'électricité nationale.

⁴ Carte issue du dossier de presse précédemment mentionnée, version 2025

⁵ Nombre de salariés en mai 2025, donnée issue du « Blog d'Ici » du site Internet de l'entreprise, article du 22 mai 2025, [Énergie d'ici franchit le cap de 1 TWh d'électricité renouvelable agréée](#)

1. Contexte, enjeux et défis du marché de l'électricité

1.1. Le marché de l'électricité

1.1.1. Historique

Le secteur de l'électricité en France s'est développé à partir de la fin du XIX^e siècle avec l'émergence des premiers usages industriels, de l'éclairage public électrifié dans les grandes villes puis progressivement dans les foyers. Jusqu'à dans les années 40, les services de production, de transport et de distribution étaient assurés par de nombreuses compagnies régionales, souvent municipales ou privées, entraînant une électrification inégale selon les territoires.

La situation change radicalement après la Seconde Guerre mondiale. La loi du 8 avril 1946 nationalise l'ensemble des entreprises privées du secteur électrique et gazier, donnant naissance à Électricité de France (EDF), entreprise publique à monopole intégré. Cette centralisation visait à unifier le réseau, harmoniser les tarifs, et garantir un accès égal à l'électricité sur tout le territoire, conformément aux missions de service public.

Dans les années 1950-1960, l'accès à l'électricité se généralise rapidement avec le développement de capacité de production supplémentaire et une croissance de l'économie mondiale et de la France majeure de 5,3% par an poussant la consommation en énergie à la hausse. Mais, un tournant majeur intervient lors de la crise pétrolière de 1973 qui révèle la dépendance énergétique de la France aux énergies fossiles importées. En réponse, le gouvernement lance le Plan Messmer (1974), visant à développer massivement l'énergie nucléaire pour assurer l'indépendance énergétique et stabiliser les coûts de production, constituant encore aujourd'hui la base du mix énergétique français.

Pour comprendre l'impact de cette première réforme et les suivantes, il est nécessaire de visualiser la chaîne de valeur de l'électricité. Historiquement intégrée verticalement au sein d'EDF, elle se compose de quatre segments principaux :

- La production qui consiste à la génération d'électricité à partir de différentes sources d'énergie (nucléaire, hydraulique, éolien, solaire, ...) et à sa vente sur le marché ;
- Le transport qui permet l'acheminement de l'électricité via un réseau de haute et très haute tension (HTA) depuis les sites de production jusqu'aux points de connexion avec le réseau de distribution ;
- La distribution permet ensuite la transmission de l'électricité à travers un réseau de moyenne et basse tension depuis les points de connexion avec le réseau de transport jusqu'aux consommateurs finaux ;
- La fourniture consiste en la gestion des relations commerciales avec les clients / consommateurs finaux, incluant la facturation, le service client et le suivi des contrats. Cela peut également inclure un accompagnement au suivi de la consommation d'électricité aux clients.

A partir de la fin des années 1990, sous l'impulsion des directives européennes, le secteur de l'électricité va connaître une ouverture progressive à la concurrence. En premier lieu, la directive 96/92/CE du Parlement européen et du Conseil, adoptée le 19 décembre 1996 et transposée en France en 2000, introduit les premiers fondements du marché intérieur de l'électricité en imposant une séparation entre

les activités de production, de transport, de distribution et de fourniture, afin de prévenir les conflits d'intérêts et les pratiques anti-concurrentielles.

Cette dynamique se poursuit avec la directive 2003/54/CE du 26 juin 2003, qui approfondit la libéralisation en imposant des séparations juridique et comptable renforcées ainsi qu'un élargissement des catégories de consommateurs éligibles à l'ouverture du marché. Enfin, dans le cadre du « troisième paquet énergie », la directive 2009/72/CE du 13 juillet 2009 renforce l'indépendance des gestionnaires de réseaux, clarifie les conditions d'accès aux infrastructures et renforce la protection des consommateurs.

En France, ces directives ont été traduites progressivement dans la législation nationale : dès février 1999, les sites consommant plus de 100 GWh/an sont autorisés à entrer sur le marché libre ; en juin 2000, ce seuil est abaissé à 16 GWh/an ; en février 2003, à 7 GWh/an ; puis, en juillet 2004, l'accès est étendu à toutes les entreprises et collectivités, avant d'être ouvert à tous les consommateurs, y compris résidentiels, en juillet 2007, leur permettant d'avoir la possibilité de sortir du tarif réglementé de la vente d'électricité proposé historiquement par EDF.

Pour encadrer cette ouverture progressive du marché, la Commission de régulation de l'Énergie (CRE) a été créée le 24 mars 2000, en application de la loi du 10 février 2000 relative à la modernisation et au développement du service public de l'électricité. Sa mission est de veiller au bon fonctionnement des marchés de l'électricité et du gaz en France et de garantir la transparence et la concurrence loyale entre acteurs.

Enfin, en 2010, la loi NOME (Nouvelle Organisation du Marché de l'Électricité) introduit l'Accès Régulé à l'Énergie Nucléaire Historique (ARENH), permettant aux fournisseurs alternatifs d'acheter jusqu'à 100 TWh/an d'électricité nucléaire d'EDF à un tarif fixé par la CRE (40 €/MWh puis 42 €/MWh). Cette mesure vise à favoriser la concurrence tout en reflétant les coûts réels de production, les coûts d'exploitation, les charges liées aux installations (de sécurité et de maintenance notamment) et les coûts d'investissement.

Ces réformes ont permis de mettre fin au monopole historique d'EDF sur la production et la fourniture d'énergie. Elles ont permis à des producteurs indépendants et de nouveaux producteurs d'émerger en facilitant le développement et l'intégration des énergies renouvelables tels que l'éolien et le solaire via la création d'un marché de gros qui permet aux producteurs de vendre leur production. Elles ont également permis le développement de divers fournisseurs proposant des offres diverses et variées en électricité aux consommateurs collant mieux à leurs besoins.

1.1.2. Fonctionnement

Seules la production et la fourniture sont ouvertes à la concurrence, le transport et la distribution restant des monopoles naturels publics. Ce sont des activités qui nécessitent une unicité de réseau, sûr, efficace et équitablement réparti dans tout le territoire. Elles sont donc confiées à des gestionnaires spécifiques :

- Pour le transport d'électricité, RTE (Réseau de Transport d'Électricité) gère plus de 100 000 km de lignes de haute et très haute tension de manière indépendante⁶.

⁶ Données issues de RTE, [Le transport d'électricité, comment ça fonctionne ? | RTE](#), version 13/08/2025

- Pour la distribution d'électricité, l'acteur principal couvrant 95% du territoire français avec un réseau de près de 1,4 million de km est Enedis⁷. Les 5% restants sont gérés par des entreprises locales de distribution⁸. Ces distributeurs s'occupent du réseau électrique en moyenne et basse tension reliant directement les consommateurs finaux.

Ces acteurs vont permettre d'acheminer l'énergie produite par les producteurs dans lesquels on distingue trois types d'acteurs :

- Les producteurs historiques tel que EDF en France qui dispose de parc diversifié avec des centrales nucléaires, de centrales hydrauliques, de centrales thermiques, ...
- Les producteurs indépendants qui exploitent des centrales ou des parcs renouvelables (hydraulique, éolien, solaire ou biomasse)
- Les producteurs occasionnels ou décentralisés, souvent de petite taille, qui vont injecter leur surplus de production sur le réseau (par exemple des entreprises qui ont des installations photovoltaïques sur le toit de leurs locaux)

L'énergie produite sera ensuite achetée par les fournisseurs pour la revendre aux consommateurs finaux de différents types (particuliers, entreprises, industries, collectivités). Ils doivent garantir à chacun de leur client un approvisionnement contenu et adapté à leurs besoins au bon moment et en quantité suffisante. Certains fournisseurs disposent de moyens de production propres ou via des contrats de gré à gré avec des producteurs, tandis que d'autres, dits « pur acheteurs », s'approvisionnent uniquement via les marchés de gros.

Il existe plusieurs types de marché :

- Les bourses de l'électricité, offrent un mécanisme de commercialisation anonyme et transparent, en permettant de faire coïncider l'offre et la demande et de déterminer des prix publics de gros. Parmi elles, l'European Power Exchange (EpeX Spot) couvre la France, l'Allemagne, l'Autriche et la Suisse, représentant près de 40 % de la consommation européenne. D'autres plateformes existent en fonction de la proximité des frontières entre les pays de l'Europe.
- Les marchés de gré à gré (OTC, Over The Counter), qui organisent des transactions bilatérales entre producteurs et fournisseurs, généralement via des contrats cadres négociés avec l'aide de courtiers. Ces marchés ne disposent pas de centralisation comparable à une bourse, mais permettent une grande flexibilité pour échanger différents types de produits d'électricité.
- Les marchés OTC organisés, qui combinent la structure d'une bourse avec la flexibilité d'un marché bilatéral. Les acteurs y formulent offres et demandes sur une plateforme centralisée, avec un équilibrage continu de l'offre et de la demande, permettant la conclusion de contrats en temps quasi réel.

Ces différents types de marchés servent de base pour la formation des prix et la planification de l'approvisionnement des fournisseurs mais l'électricité étant un produit non-stockable, il est nécessaire de gérer l'équilibre entre production et consommation en temps réel. D'où l'intervention, au sein de la bourse de l'électricité, d'un marché au comptant et de marchés dérivés :

⁷ Données issues d'Enedis, [Le marché de l'électricité | Enedis](#), version du 13/08/2025

⁸ Ce sont « des distributeurs non nationalisés dans la loi du 8 août 1946 » qui sont en gestion d'une zone spécifique locale, [Que signifie ELD ? | Enedis](#), version du 13/08/2025

Les marchés dérivés ou aussi appelé à terme ou d'options permettent de sécuriser les prix et les volumes sur des périodes futures via l'achat de produits standards (blocs de kWh à diverses échéances : hebdomadaire, mensuelle, trimestrielle, semestrielle, annuelle) :

- Le marché Forward permet aux acteurs de négocier aujourd'hui des volumes d'électricité pour livraison à une date future, sur une période et un profil spécifiques. Il s'agit d'un contrat bilatéral entre producteurs et fournisseurs pour se couvrir contre les variations de prix. Le prix va donc refléter en général l'anticipation du marché sur le prix moyen spot futur d'un pays donné.
- Le marché Future est similaire au Forward mais standardisé et négocié sur des bourses spécialisées. Il couvre plusieurs maturités (à court terme (quotidien, hebdomadaire), moyen terme (mensuel, trimestriel) et long terme (annuel), offre une liquidité plus importante et des mécanismes de compensation par une chambre de compensation, réduisant le risque de contrepartie.
- Le marché au comptant appelé marché spot permet de gérer les transactions de court terme, pour le lendemain (marché Day-Ahead) ou durant la journée (marché Intra-Day) :
 - ⇒ Le marché Day-Ahead (à livraison le lendemain), où les volumes d'électricité sont négociés pour chaque heure de la journée suivante. Les prix sont fixés en fonction de l'offre et de la demande anticipées, permettant aux fournisseurs et producteurs de planifier leurs opérations et de sécuriser leurs positions.
 - ⇒ Le marché Intra-Day permet d'ajuster les positions en cours de journée pour tenir compte des écarts entre prévisions et consommations réelles. Les transactions peuvent se faire toutes les 15 à 30 minutes, offrant une flexibilité essentielle pour répondre à des variations imprévues de la demande ou de la production, notamment liée aux énergies renouvelables.

L'ensemble de ces marchés gèrent deux types de flux :

- Flux contractuels, correspondant aux volumes achetés et vendus sur les différents marchés (Future, Forward, Day-Ahead, Intra-Day ou OTC).
- Flux physiques, qui reflètent l'électricité effectivement injectée ou soutirée sur le réseau.

Les marchés à terme et le marché spot sont complétés par le marché Balancing, qui joue un rôle clé dans la stabilité du système électrique. Même lorsque producteurs et fournisseurs se sont coordonnés sur les volumes à livrer via les marchés Futures, Forwards ou Spot, des écarts peuvent survenir entre les flux contractuels et les flux physiques réellement injectés ou consommés sur le réseau. Le marché Balancing permet de corriger ces déséquilibres en temps réel, en faisant intervenir les responsables d'équilibre, désignés pour maintenir l'équilibre instantané entre production et consommation. Lorsqu'un écart apparaît, un mécanisme d'ajustement est déclenché sur le marché Balancing : producteurs et fournisseurs qui sont en surplus ou en déficit ajustent leurs positions pour corriger ces déséquilibres et éviter toute perturbation du réseau.

Le fonctionnement du marché de l'électricité est complexe et spécifique par rapport aux autres marchés primaires. Certaines particularités propres à la France, liées à la régulation, aux infrastructures et aux sources de production, influencent directement sa gestion et sa stabilité, et seront détaillées dans la section suivante.

1.1.3. Spécificités

Le marché de l'électricité en général mais aussi propre à la France présente plusieurs particularités qui le distinguent des autres marchés primaires et influencent sa régulation, sa stabilité et la gestion des risques pour les acteurs.

L'électricité est un bien économique pas comme les autres du fait de sa nature non-stockable à grande échelle. Contrairement à d'autres matières premières, elle doit être consommée dès sa production impliquant un équilibre permanent entre offre et demande. Cette contrainte rend cruciale la bonne gestion et la planification des volumes produits et achetés et des écarts entre la consommation (volume soutiré) et la production (volume injecté) par les responsables d'équilibre. L'équilibre doit être constant car même de faibles déséquilibres peuvent entraîner des perturbations sur le réseau ou des coûts de rééquilibrage pour les producteurs et les fournisseurs.

L'enjeu est donc d'avoir un équilibre continu entre offre et demande qui est dépendant de plusieurs facteurs notamment les conditions météorologiques. En effet, la demande en électricité et la production notamment renouvelable peut fortement varier selon les conditions climatiques. Les périodes hivernales sont particulièrement critiques, avec des pics de consommation élevés, tandis que les vagues de chaleur estivales peuvent également générer des contraintes sur le réseau. Ce phénomène est caractérisé par la sensibilité thermique ou thermo-sensibilité des consommateurs qui va se mesurer via des indicateurs tels que les HDD (Heating Degree Days) et les CDD (Cooling Degree Days) qui permettent d'ajuster les prévisions de consommation et d'anticiper les besoins de production.

Une autre spécificité propre au marché français impact la nature de ce dernier : l'ARENH⁹. Ce mécanisme permet aux fournisseurs alternatifs d'acheter une partie de l'électricité produite par les centrales nucléaires, toutes détenues par EDF, à un prix fixé par la CRE. Il vise à stimuler la concurrence sur le marché de détail tout en garantissant un prix plancher réaliste basé sur les coûts historiques de production nucléaire. Cependant, ce mécanisme prendra fin en 2025 et sera remplacé par un mécanisme prenant en compte le coût des investissements futures en termes d'accroissement du parc nucléaire français et du plan de maintenance et de prolongation de durée de vie des centrales existantes.

Au-delà des spécificités françaises, le marché est également encadré par les directives européennes, qui imposent la séparation des activités et la transparence des marchés. L'intégration européenne du marché de l'électricité, via des plateformes comme EPEX Spot, influence la formation des prix et l'arbitrage entre différents hubs, tout en maintenant une harmonisation progressive entre les pays membres. Le développement des interconnexions aux frontières entre pays permet également de fluidifier les échanges sur les marchés en augmentant les capacités de transport d'énergie.

Toutes ces caractéristiques spécifiques contribuent à la volatilité des prix et des volumes mais d'autres facteurs accroissent ces effets et ont des impacts directs sur le marché.

⁹ Accès Régulé à l'Énergie Nucléaire Historique

1.2. Les facteurs de volatilité

1.2.1. Impact conjoncturel

Le marché de l'électricité est particulièrement sensible à de nombreux facteurs conjoncturels d'ordre économique, géopolitique ou réglementaire. Les fluctuations de prix observées au cours des deux dernières décennies illustrent l'ampleur de cette volatilité importante et croissante.

Dès les débuts de la libéralisation, le marché a été impacté par la crise financière de 2008 entraînant une hausse spectaculaire des prix spot de l'électricité, atteignant des niveaux historiques depuis la création du marché. Cette flambée des prix est due à une combinaison de facteurs : une demande encore soutenue malgré le ralentissement économique, une dépendance aux prix élevés des énergies fossiles et une offre de production réduite sur certaines périodes. Cela met en évidence la vulnérabilité du marché face aux chocs macroéconomiques et aux contraintes structurelles de production (un niveau de capacité production non encore suffisante pour ne plus être fortement dépendant des énergies fossiles).

A contrario, la pandémie de Covid-19 a provoqué un effondrement temporaire de la demande industrielle et commerciale, couplée au développement des capacités de production des énergies renouvelables, entraînant une baisse significative des prix spot. Les changements soudains dans les habitudes de consommation causés par le confinement ont notamment accentué la volatilité horaire, ce qui illustre ici la sensibilité du marché aux comportements des consommateurs et aux perturbations économiques rapides.

A la sortie de la pandémie, la reprise rapide de l'activité accompagnée d'une forte hausse des prix du gaz naturel et aggravée par la baisse des approvisionnements russes en gaz après l'invasion de l'Ukraine en 2022, et une disponibilité des moyens de production notamment nucléaire moins importante due au décalage des programmes de maintenance des centrales initialement prévus en 2020, ont engendré une forte instabilité des prix sur les marchés européens (accentuée par l'influence des interconnexions européennes) avec des hausses de prix spectaculaires sur le marché spot et les marchés à terme.

Au-delà des crises, en France, la Programmation Pluriannuelle de l'Énergie (PPE) et la loi Énergie-Climat fixent des objectifs contraignants de réduction des émissions et de développement des EnR influençant la structure des prix et la disponibilité des capacités pilotables. Le mécanisme ARENH joue également un rôle positif dans la formation des prix, en garantissant un accès régulé à l'électricité nucléaire à un prix fixe (42€/MWh) permettant aux fournisseurs alternatifs d'accéder à une énergie propre à coût stable et limitant leur exposition à la volatilité du marché.

La France est d'ailleurs l'un des pays européens les plus exportateurs d'énergie électrique propre faciliter via le développement des interconnexions électriques européennes entre les pays membres qui permet de fluidifier les échanges et optimiser les ressources. Ces interconnexions peuvent lisser les chocs locaux : par exemple, un déficit de production hydraulique en France peut être compensé par des importations d'énergie photovoltaïque d'Espagne ou éolienne d'Allemagne. Mais, elles peuvent aussi propager la volatilité comme cela a été le cas lors de la crise énergétique de 2022 où les tensions sur le marché allemand notamment liées au gaz ont entraîné une contagion des prix élevées vers la France via les interconnexions, malgré une situation locale moins tendue.

Ces impacts conjoncturels influencent fortement les prix et la disponibilité de l'électricité à court terme mais les énergies renouvelables aussi contribuent à l'importante volatilité du marché électrique.

1.2.2. Les EnR comme catalyseur de volatilité

Les énergies renouvelables mentionnées sous le sigle EnR ont fortement bénéficié de la libéralisation du marché et du consensus entre les pays membres de l'Union Européenne pour développer ces technologies, dans le cadre de la transition écologique et de la décarbonation du mix énergétique. Cet objectif de développement s'inscrit dans les engagements climatiques européens¹⁰ et vont se traduire par des politiques de soutien ciblées, un cadre réglementaire incitatif favorisant les investissements et facilitant la vente d'énergie, et un accès facilité aux réseaux de transport et de distribution.

Il existe quatre grands types de technologies dites renouvelables :

- Les centrales hydrauliques utilisent la force de l'eau, soit stockée dans des barrages, soit issue des cours d'eau, pour produire de l'électricité via des turbines. Leur production dépend des cycles hydrologiques et de la disponibilité en eau, ce qui rend cette source sensible aux conditions climatiques (sécheresses, précipitations abondantes).
- Les parcs éoliens terrestres ou maritimes (aussi appelés offshore) produisent de l'électricité grâce à la force du vent. La production d'énergie est intermittente car elle est dépendante des régimes de vent qui peuvent varier fortement d'une saison ou d'un jour à l'autre. La quantité de production va également dépendre de la vitesse et de la constance du vent.
- Les parcs photovoltaïques permettent de transformer directement la lumière du soleil en électricité. La production dépend de l'ensoleillement ce qui la rend très variable à une échelle journalière et saisonnière avec une absence totale de production la nuit et une baisse de production en période hivernale et en cas de couverture nuageuse importante.
- Les centrales à biomasse exploitent la combustion ou la méthanisation de matières organiques tels que le bois, l'agriculture ou les déchets¹¹ pour produire de l'électricité et/ou de la chaleur. Cette énergie est plus prévisible que l'éolien et le solaire (qui sont fortement dépendants des conditions météorologiques) mais elle reste dépendante des cycles de production et de l'approvisionnement en matières premières.

Les centrales hydrauliques et biomasse, étant en partie pilotables, offrent une production plus stable et ajustable, tandis que l'éolien et le solaire restent des sources intermittentes.

Le déploiement de ces technologies a été soutenu par deux principaux mécanismes en France, qui vont influencer indirectement la volatilité des prix sur le marché :

- Obligation d'Achat (OA) : Mécanisme permettant de garantir un revenu stable au producteur, indépendamment de la fluctuation des prix de gros via la vente de la production d'électricité à un prix fixé, généralement supérieur au prix de marché, à un acheteur désigné (fournisseur historique comme EDF ou alternatif). Cependant, ce dispositif peut conduire à injecter sur le réseau électrique même en période de prix très bas ou négatifs, les producteurs étant insensible aux fluctuations des prix du marché, ce qui accentue la volatilité à court terme.
- Complément de Rémunération (CR) : Système incitant davantage les producteurs à optimiser leur injection de production sur le réseau où la production est directement vendue sur le marché.

¹⁰ Sous l'impulsion du Paquet Énergie Propre (Clean Energy Package) et plus précisément la directive européenne 2018/2001 encadrant le développement des EnR dans l'Union Européenne.

¹¹ Les déchets vont comprendre les biodéchets ménagers, les déchets de la restauration, des industries agroalimentaires, ...

Les producteurs recevront une prime « complémentaire de revenus » qui sera calculée comme la différence entre un prix de référence (déterminé lors d'appels d'offres contractuellement) et le prix moyen du marché. De ce fait, la prime peut être négative en fonction des périodes rendant les producteurs plus sensibles aux périodes de prix négatifs notamment ce qui devrait logiquement faire arrêter leur production lors de ces périodes limitant l'accentuation de la volatilité et des prix négatifs.

Si les EnR réduisent la dépendance aux énergies fossiles, elles introduisent une intermittence structurelle dans la production d'électricité. Par exemple, lors des périodes de forte production renouvelable et de faible demande, l'excédent peut provoquer une baisse brutale des prix, voire des prix négatifs. A l'inverse, lors des périodes de forte consommation mais de faible production renouvelable (par exemple, lors d'un hiver froid sans soleil ni vent) cela peut créer d'importante tension sur les prix avec des prix atteignant des pics extrêmes (d'autant plus durant une période de crise).

Ces phénomènes sont amplifiés par l'effet de cannibalisation : lorsque de nombreuses installations produisent en même temps, elles se concurrencent sur le marché spot, ce qui réduit le prix de vente moyen pour tous les producteurs d'énergie, avec une demande globale stable.

Les EnR transforment la dynamique des marchés de l'électricité en réduisant le coût marginal de production sur certaines périodes mais accroissant également l'amplitude des variations de prix ce qui rend plus complexe les prévisions, la gestion des risques plus exigeante et influence profondément la composition du mix énergétique.

1.2.3. Transformation du mix énergétique

Le mix énergétique français est historiquement dominé par l'énergie nucléaire qui représente en 2024, environ 67%¹² de la production totale. Cependant, cette structure vise à être modifier progressivement dans le cadre des objectifs climatiques de l'Union Européenne et du Plan de Programmation de l'Energie actant le projet de réduire la part de l'énergie nucléaire du mix à 50% d'ici 2035 impliquant le développement rapide d'autres sources d'énergie propres.

En effet, même si le parc nucléaire français garantit une production stable, faiblement émettrice de CO₂ et relativement indépendante des fluctuations des prix des énergies fossiles, l'âge avancé du parc (avec des réacteurs qui ont plus de 30 ans d'activité pour la majorité) impose un programme massif de maintenance couteux poussant encore une fois à la réduction de la part de cette énergie dans le mix français.

D'autres sources d'énergie ont connu une croissance importante et soutenue telles que l'éolien terrestre avec une puissance installée passant de 6,1 GW en 2010 à 22,9 GW en 2024, et le solaire avec une puissance installée photovoltaïque qui est passée de 0,8 GW en 2010 à 24,3 GW en 2024. Une croissance facilitée par les mécanismes de soutien qui seront précédemment présentés.

Cependant, la nature non-stockable de l'électricité et l'intermittence de ces sources rend moins stables les prévisions de production. À l'inverse, l'hydroélectricité et la biomasse offrent des productions plus régulières et pilotables, constituant des leviers de flexibilité essentiels pour équilibrer le réseau.

¹² Données issues du Bilan électrique 2024, RTE

Pour répondre à cette variabilité croissante, plusieurs solutions de flexibilité se développent depuis quelques années :

- De stockage : batteries en lithium-ion et l'hydrogène pour stocker les excédents de production
- D'élasticité volontaire de la demande consistant à la modulation volontaire de la consommation en modifiant les plages d'heures creuses par exemple ou en incitant les industriels à fortement consommer lors d'importante production des EnR.
- D'optimisation des outils de pilotage de la production en recourant à la prévision météo et en implémentant des algorithmes ou modèles d'intelligence artificielle pour optimiser les flux.

Par ailleurs, la transition énergétique s'inscrit dans un contexte de sobriété énergétique et de réindustrialisation. En effet, la France vise à réduire sa consommation globale via différentes actions (rénovation des bâtiments, régulation des usages, prévention des habitudes quotidiennes de sobriété, ...) tout en développant et en relocalisant les industries de secteurs énergivores tels que la production de batteries, la construction de data centers, l'hydrogène vert et les composants électroniques. Cette double dynamique de sobriété et de réindustrialisation crée un paradoxe : Réduire la consommation tout en augmentant la production industrielle. Cela implique alors une planification énergétique fine, conciliant efficacité, flexibilité et faible pollution carbone.

Cette transition implique une plus grande variabilité des prix horaires, une dépendance accrue aux conditions météorologiques, et donc un besoin renforcé en gestion des risques (via des instruments de couverture et de prévisions plus rigoureux de la consommation et de la production). Il est alors nécessaire pour les acteurs du marché d'adapter leurs stratégies pour composer avec ce système de plus en plus volatil.

1.3. Offre et Demande en électricité

Ces facteurs de volatilité influencent directement l'équilibre entre l'offre et la demande d'électricité. La demande et l'offre, en plus d'être régulées par les volumes physiques, dépendant également des comportements des consommateurs, de la saisonnalité, des prix et des conditions climatiques. Cette section se concentre sur la compréhension détaillée de la demande et de l'offre, ainsi que sur le rôle central des responsables d'équilibre.

1.3.1. La demande (Consommateurs)

La demande en France est très hétérogène due à la diversité des consommateurs finaux comprenant les particuliers, les petites et moyennes entreprises, les industries lourdes et les collectivités. Cette demande va être représentée par une courbe de charge qui va illustrer la variation au pas horaire de la consommation sur une journée, une semaine, un mois ou une année. Ces courbes vont permettre de prévoir la consommation globale future et d'anticiper les périodes de pointe et de creux de consommation. Elles sont essentielles aux producteurs pour adapter leur production et aux fournisseurs pour prévoir leurs achats.

Les profils de consommation permettent d'affiner la définition des comportements de consommation en distinguant les grandes catégories d'utilisateurs. On va retrouver les ménages, catégorisés en RES (pour résidentiels), les petites et moyennes entreprises, catégorisées en PRO et les grandes entreprises du tertiaire et industries, catégorisées en ENT. Ces profils vont consommer différemment : par exemple,

une usine métallurgique va avoir besoin d'énergie en continue tandis que les ménages fonctionnent en pics et creux de consommation.

Les consommateurs sont séparés en deux groupes en fonction du type de mesure de la consommation :

- Les clients télé-relevés qui comprennent les segments C2 et C4 et pour qui les données sont collectées automatiquement via le compteur par pas de 5 minutes en temps réel ce qui permet de générer une courbe de charge mesurée pour chaque client afin de déterminer leurs profils de consommation réels à une échelle de temps fine. Cela permet au fournisseur de prévoir la consommation à venir des clients de manière précise et anticiper l'achat nécessaire de bloc d'énergie pour couvrir la demande.
- Les clients profilés qui correspondent au segment C5 et pour qui la reconstitution des flux est faite via la courbe de charge profilée. Contrairement aux clients équipés de compteurs communicants mesurant la consommation en continu, leurs données ne sont pas relevées heure par heure. Elles sont estimées à partir de profils types établis par le gestionnaire du réseau, reposant sur des moyennes statistiques et segmentées selon des périodes horo-saisonnnières auxquelles est multiplié un facteur d'usage.

Ces périodes distinguent notamment :

- ⇒ PTE (Pointe) : novembre à avril, jours ouvrés, de 9 h à 10 h et de 18 h à 19 h ;
- ⇒ HPH (Heures Pleines Hiver) : novembre à avril, jours ouvrés, de 6 h à 22 h ;
- ⇒ HCH (Heures Creuses Hiver) : novembre à avril, toutes les heures des jours fériés et week-ends, ainsi que de 22 h à 6 h les jours ouvrés ;
- ⇒ HPE (Heures Pleines Été) : avril à octobre, jours ouvrés, de 6 h à 22 h ;
- ⇒ HCE (Heures Creuses Été) : avril à octobre, toutes les heures des jours fériés et week-ends, ainsi que de 22 h à 6 h les jours ouvrés.

Cette méthode de profilage pour le segment C5, bien que moins précise pour détecter des variations fines ou des événements ponctuels de consommation, reste adaptée pour la facturation et l'analyse agrégée, tout en permettant de limiter les coûts liés aux relèves et au stockage de données détaillées.

Plusieurs facteurs vont influencer la demande : la température extérieure, les prix de l'électricité, les habitudes de consommation, les événements exceptionnels (comme le confinement ou les matchs de foot important) et les incitations tarifaires (avec les tarifs heures creuses / heures pleines). La saisonnalité va également jouer un rôle déterminant où l'hiver va se caractériser par des tensions accrues sur le système avec l'usage important de chauffage électrique (d'autant plus lorsque les hivers sont plus froids que la normale) tandis que l'été va présenter des pointes de consommation liées à la climatisation lors de périodes de fortes chaleurs / canicule.

La demande traduit donc les besoins, influencés par divers facteurs, des consommateurs qui ne peut être satisfaite que dans la limite des capacités de production disponibles. L'étude de l'offre s'impose donc en complément.

1.3.2. L'offre (Producteurs et Fournisseurs)

L'offre d'électricité repose sur l'action combinée des producteurs, fournisseurs et gestionnaires de réseaux, complémentaires entre eux. Les producteurs génèrent l'électricité à partir de différentes

technologies renouvelables (hydraulique, éolien, solaire, biomasse) ou non (nucléaire, fossile, gaz) et mettent cette énergie à disposition sur le marché. Les fournisseurs, quant à eux, achètent cette électricité pour la revendre aux consommateurs finaux (particuliers, entreprises, collectivités) et peuvent être à la fois producteurs et commerçants.

Pour structurer leurs approvisionnements, les fournisseurs s'appuient sur différents types de produits d'électricité :

- Le produit BASELOAD correspond à une puissance constante fournie sur l'ensemble de la période contractuelle assurant la demande minimale d'électricité nécessaire sur cette période indépendamment des variations quotidiennes. Les centrales fonctionnant en base sont généralement les réacteurs nucléaires et certaines centrales hydrauliques qui sont capables de fournir une production stable et continue dans le temps et ne dépendant pas ou très peu des aléas climatiques.
- Le produit PEAKLOAD qui correspond aux périodes de forte demande notamment lors des heures de pointe sur les jours ouvrés (du lundi au vendredi, de 8h à 20h). Ces pics de consommation vont faire appel aux centrales à gaz, à certaines installations hydrauliques, à des groupes thermiques qui vont être mobilisés pour y faire face. Ces produits permettent d'ajuster l'offre à la demande instantanée en assurant la stabilité du réseau en période de consommation élevée.
- Le produit OFFPEAK est l'inverse du produit PEAKLOAD et va donc correspondre aux périodes où la demande en électricité est faible. Il va inclure les nuits, les week-ends et les jours fériés, des périodes caractérisées par une consommation plus stable et moins intense permettant aux centrales de réguler leur production sans mobiliser des moyens de pointe coûteux. Ce produit permet de maximiser l'utilisation des capacités existantes à moindre coût tout en optimisant la planification du portefeuille de production du fournisseur.

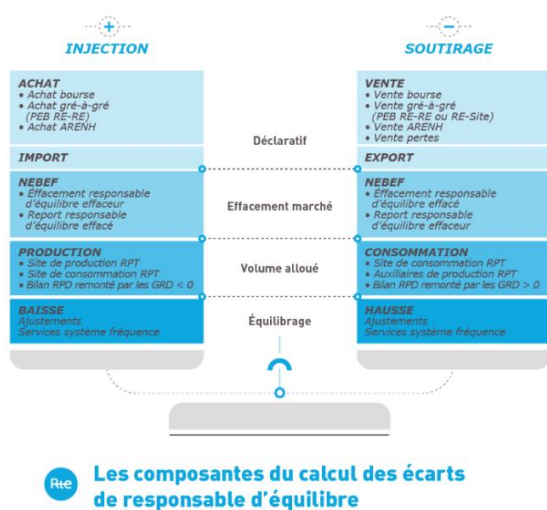
Les fournisseurs vont acheter des blocs d'énergie en fonction de ces produits afin de répondre à la demande de leur portefeuille adaptée aux profils de consommation de leurs clients et à l'obligation de couverture continue de la demande. Cette distinction va permettre également d'aligner la production sur la demande en limitant les risques financiers liés aux écarts entre volumes achetés et consommés en fonction des périodes et des heures de la journée, ces derniers seront traités par les responsables d'équilibre, dernier acteur essentiel pour la régularisation finale et le respect strict de l'équilibre entre la consommation et la production à l'instant t.

1.3.3. L'équilibre offre/demande, un impératif systémique (Responsable d'équilibre)

La synchronisation continue entre production et consommation est un impératif pour le système électrique. Contrairement à d'autres biens, l'électricité ne peut pas être stockée massivement et de manière économique, ce qui rend l'équilibre offre/demande particulièrement sensible. Le moindre déséquilibre peut entraîner des effets immédiats tels qu'une dégradation de la qualité du courant et, dans les cas extrêmes, des coupures d'électricité. Face à ce constat, différents mécanismes de régulation ont été mis en place, confiés aux responsables d'équilibre, acteurs contractuellement désignés pour assurer ce suivi au nom des producteurs et des fournisseurs.

Ces écarts sont scrupuleusement surveillés par le gestionnaire de réseaux, RTE en France, chargé de l'équilibre physique global du réseau, qui confie la bonne gestion et l'adéquation entre les injections

(production) et les soutirages (consommation) à des responsables d'équilibre au sein de leur périmètre d'équilibre¹³.



Les écarts sont calculés sur la base de la reconstitution des flux d'injection et de soutirage grâce à la remontée de données du réseau de transport et du réseau de distribution.¹⁴

Les écarts nécessitant l'intervention de la RTE seront calculés a posteriori et facturés. Cela correspond au coût d'équilibrage qui sera répercuté entre les responsables d'équilibre qui présente effectivement ces écarts, et qui eux-mêmes feront payer le coût de ces déséquilibres à leurs clients (fournisseurs principalement).

Les responsables d'équilibre doivent donc surveiller en temps réel les écarts entre les volumes contractuels et les flux physiques réels observés sur le réseau. Pour corriger les déséquilibres qui apparaissent, ils peuvent activer plusieurs options correctives : achat ou vente d'énergie supplémentaire sur le marché Infra-Day, ajustement des centrales pilotables (arrêt ou démarrage en cas, respectivement, de forte production lors de période de faible consommation ou de forte consommation pour laquelle la production à l'instant t n'est pas suffisante), de recours à d'autres solutions de flexibilité (usages de batteries, encore peu développés, de capacité de réserve disponible sur certaines centrales hydrauliques ou d'incitation à la consommation auprès d'entreprise lors de période de forte production et de faible consommation par exemple). Ces interventions permettent de limiter les écarts et réduire les coûts d'ajustement pouvant être significativement coûteux pour les fournisseurs.

Si, toutefois, le déséquilibre persiste et que les acteurs en charge ne respectent pas leurs engagements d'équilibre, le cadre réglementaire prévoit également des sanctions financières, ce qui encourage une planification rigoureuse de la consommation et de la production de chaque acteur et une gestion proactive des risques.

1.4. Tarification et cadre réglementaire

En général, la tarification est déterminée par la confrontation de l'offre et de la demande jusqu'à un point d'équilibre permettant l'établissement d'un prix. Cependant, l'électricité devant être produite et consommée quasiment immédiatement, avec une demande aléatoire et soumise à des « effets de pointe », un prix du MWh variant au gré des saisons sur la base du comportement de consommation et un besoin d'équilibre strict en permanence, sa tarification ne repose pas simplement sur le principe offre-demande classique.

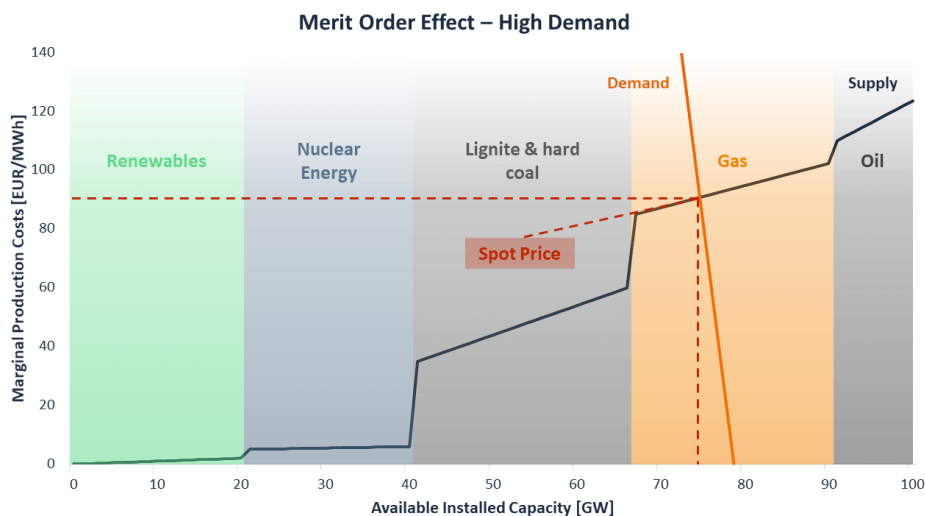
¹³ Un périmètre comprend les producteurs et les consommateurs en gestion des fournisseurs qui sont responsable d'équilibre ou ont désigné un responsable d'équilibre à qui ils ont délégué la responsabilité d'équilibre.

¹⁴ Schéma et explications de la méthode de calcul issus de l'article « Le prix de règlement des écarts de responsable d'équilibre », publié par RTE, version 2025

1.4.1. La formation des prix

Les prix de l'électricité vont se former à plusieurs périodes, en fonction des marchés de gros sur lesquels sont échangés les volumes d'énergie soit en livraison immédiate (sur le marché spot) soit future (marché à terme).

Le système de formation des prix sur les marchés de gros repose sur un principe économique fondamental : le Merit Order. Ce mécanisme consiste à classer les moyens de production disponibles par ordre croissant de leur coût marginal, c'est-à-dire le coût de production d'un MWh supplémentaire induit par la dernière unité produite (hors coûts fixes déjà engagés). Le Merit Order ne prend en compte que les coûts variables (les coûts de fonctionnement et du combustible).



En pratique, les centrales produisant de l'électricité dite « fatale » sont appelées en premier (hydraulique au fil de l'eau, éolien et solaire). Le coût marginal de ces énergies est quasi nul. Est appelé ensuite, en fonction des besoins de la demande si les énergies intermittentes ne suffisent pas à y répondre, les centrales nucléaires qui ont des coûts fixes élevés mais un coût marginal faible. Enfin, viennent les moyens de production thermiques à combustibles fossiles qui sont plus coûteux à exploiter : les centrales au gaz naturel avec un coût marginal qui va être fortement dépendant des prix du gaz sur les marchés mondiaux, et les centrales à charbon ou au fioul, également sensibles au prix du combustible et aux quotas d'émissions de CO₂¹⁵ au niveau européen.

La demande d'électricité étant quasi inélastique à court terme, le prix de marché est fixé par la centrale marginale, c'est-à-dire la dernière unité de production nécessaire pour satisfaire la demande à un instant donné. En période hivernale, lors de pics importants de consommation et avec une faible production des EnR, par exemple, la dernière centrale appelée est souvent thermique fossile.

Ce mécanisme explique la forte sensibilité des prix de l'électricité au coût du gaz et des quotas CO₂ notamment depuis la crise énergétique de 2022 avec la flambée du prix du gaz qui a directement provoqué une envolée historique des prix spot de l'électricité en Europe.

¹⁵ Droit à payer pour obtenir l'autorisation d'émettre une tonne d'équivalent de dioxyde de carbone. C'est un outil politique utilisé pour réduire les émissions de gaz à effet de serre.

Le Merit Order illustre bien le rôle des marchés de gros dans la détermination des prix de l'énergie via la confrontation entre l'offre et la demande mais aussi leur forte sensibilité aux variables exogènes tels que le prix des combustibles, la météo ou encore la disponibilité du parc nucléaire. Cette logique s'applique aussi bien au marché spot qu'aux marchés à terme où les prix futurs sont fixés sur la base des anticipations de coûts marginaux.

1.4.2. La composition globale du prix de l'électricité

Une fois le prix de l'énergie déterminé sur le marché de gros, celui-ci n'est qu'une partie du montant payé par le consommateur final. La facture d'électricité est composée de plusieurs éléments qui reflètent à la fois les coûts de production, les frais d'acheminement, les taxes et les marges des différents acteurs.

On a premièrement le coût de la fourniture de l'énergie qui va dépendre du type d'offre souscrite par le consommateur. Il y a les offres indexées qui suivent l'évolution des prix sur le marché de gros avec parfois un décalage temporel, et les offres à prix fixe où le fournisseur achète à l'avance (ou se couvre via des contrats à terme) et inclut dans le prix une prime de risque couvrant les fluctuations futures des prix. Cette part variable du prix va également comprendre les coûts de contribution au mécanisme de capacité et aux économies d'énergie (CEE)¹⁶.

On a ensuite le coût de l'abonnement fixe qui va dépendre de la puissance souscrite et qui sert à financer les coûts d'accès et de maintenance du réseau, les coûts de commercialisation des fournisseurs (publicité, marketing, gestion clientèle) et la gestion administrative des contrats

Vient ensuite les taxes et contributions :

- On a précédemment cité les contributions au mécanisme de capacité et aux économies d'énergie qui sont variables ;
- Il y a également la taxe TURPE (Tarif d'Utilisation des Réseaux Publics d'Électricité) qui couvre l'entretien et le développement des réseaux de transport (RTE) et de distribution (Enedis, entreprises locales) ;
- L'Accise sur l'électricité (anciennement CSPE (Contribution au Service Public de l'Électricité)) qui finance notamment le soutien aux énergies renouvelables et les tarifs sociaux ;
- La CTA (Contribution Tarifaire d'Acheminement) qui finance les retraites des employés des industries électriques et gazières ;
- Et la TVA qui est à 20% sur la consommation et sur l'abonnement¹⁷.

On a enfin, dans le prix final, le coût du risque : une marge intégrée par le fournisseur pour se couvrir de la volatilité des prix, des aléas météorologiques et de l'intermittence de certaines énergies exacerbant la dispersion des prix.

¹⁶ Le mécanisme de capacité permet de garantir la sécurité d'approvisionnement électrique en France et se présente comme un surcoût sur la facture d'électricité. Les CEE sont des obligations pour les fournisseurs d'énergie à financer des actions de maîtrise de la demande d'énergie qui sont inclus dans la part variable du prix.

¹⁷ La TVA sur l'abonnement est passée de 5,5% à 20% le 1^{er} août 2025 dans le cadre de l'uniformisation de la TVA en France demandée par l'Union Européenne.

1.4.3. Tarifs spécifiques et nouveaux enjeux tarifaires

Historiquement, le marché français reposait sur les tarifs réglementés fixés par les pouvoirs publics mais depuis la libéralisation, de nouvelles offres plus diversifiées ont émergées.

Les TRV ou Tarifs Réglementés de Vente sont proposés par EDF et se distingue par une couleur en fonction du niveau de puissance souscrit. Il y avait le tarif Bleu (puissance souscrite de 3 à 36kVA), le tarif Jaune (de 36kVA à 250kVA) et le tarif Vert (supérieure à 250kVA). Ces deux derniers tarifs proposés aux professionnels ont été supprimés en 2016 dans le cadre de la loi NOME.

On a d'autres types de tarification spécifiques basés sur :

- Les heures pleines et les heures creuses où le tarif est différencié selon des plages horaires prédéfinies (16 heures pleines et 8 heures creuses), incitant à consommer lorsque la demande est plus faible et les moyens de production moins couteux ;
- Les produits Peak-Load et Off-Peak qui sont destinés aux industriels et qui correspondent à des blocs horaires standardisés comme nous l'avons présenté au 1.2.3. ;
- Le tarif base avec un prix constant tout au long de la journée sans différenciation d'heures pleines et d'heures creuses.

La montée en puissance des EnR et l'interconnexion croissante des réseaux en Europe imposent une adaptation des modèles tarifaires auxquels une commission d'enquête du Sénat en juillet 2024 a émis des recommandations sur le prix de l'électricité à l'horizon 2035 et 2050. Est suggéré la flexibilité de la demande qui est actuellement inélastique à court terme en introduisant des tarifs incitatifs ou la modification des plages d'heures creuses et heures pleines avec une plage supplémentaire d'heures creuses de 14h à 17h au printemps et en été (période de forte production solaire).

On pourrait également avoir une flexibilité contractuelle sur la fixation des prix de l'énergie qui se ferait en fonction des conditions de marché et non pas juste annuellement. Ou encore, proposer une tarification dynamique où le coût de l'énergie serait basée sur les prix du marché spot ajustée heure par heure, permettant ainsi aux consommateurs de bénéficier de prix bas en période de forte production EnR (ce qui n'est pas le cas aujourd'hui).

À terme, l'enjeu est de faire converger la tarification vers un système capable de refléter en temps réel les conditions de production et de consommation. Cela suppose un déploiement massif des compteurs communicants, des outils d'analyse des données et une pédagogie tarifaire renforcée.

1.5. Cartographie globale des risques

La présentation des facteurs de volatilité et la définition de chaque composante de la tarification nous permette de comprendre l'importance de la constitution des primes de risque pour un fournisseur. La gestion d'un portefeuille de consommation repose sur la compréhension et la quantification des différents types de risques auxquels fait face le fournisseur. Ces risques se manifestent à plusieurs niveaux : au niveau de la consommation réelle par rapport aux prévisions, des variations de prix sur le marché, ou encore de la flexibilité et du comportement des clients. Il est donc essentiel de cartographier clairement ces risques afin d'identifier les leviers d'action et de prioriser les stratégies de couverture ou d'atténuation.

1.5.1. Le risque de marché

Le risque de marché reflète l'exposition financière du fournisseur aux fluctuations imprévisibles des prix de l'électricité sur les marchés spot et à terme.

Ce risque résulte directement de l'écart entre le prix contractuel, c'est-à-dire le prix auquel le fournisseur a sécurisé l'électricité pour un volume donné via des contrats à terme, des blocs d'achat (CAL), ou des contrats de gré à gré avec les producteurs, et le prix spot, correspondant au prix réel sur le marché horaire. L'exposition au risque de marché est concentrée sur les volumes qui n'ont pas été préalablement couverts par des contrats fixes, c'est-à-dire les écarts de volumes entre consommation réelle et consommation prédite qui doivent être achetés sur le marché spot pour répondre à la consommation réelle dans un souci de respect de l'équilibre. La formule générale de quantification du risque de marché s'écrit ainsi :

$$\text{Risque marché} = \sum_t V_t^{\text{exposé}} * (P_{\text{spot},t} - P_{\text{terme}})$$

Où :

- $V_t^{\text{exposé}}$ (en MWh) représente le volume d'électricité non couvert par un contrat à terme ou un contrat bilatéral et donc directement exposé aux fluctuations du marché,
- $P_{\text{spot},t}$ (en €/MWh) est le prix de l'électricité sur le marché spot à l'instant t ,
- P_{terme} (en €/MWh) est le prix convenu sur le volume contractuel.

Cette approche permet d'évaluer et de quantifier le coût horaire de l'exposition au marché des écarts à la consommation, traduisant directement la sensibilité du fournisseur aux fluctuations du marché. Mais certains points de vigilance sont à prendre en compte. En effet, les variations des prix sur le marché électrique sont fortement influencées par la saisonnalité, la conjoncture énergétique, la disponibilité des centrales et les conditions météorologiques. Ainsi, les pics de demande, les aléas liés à la production d'énergies renouvelables ou les fluctuations des coûts des combustibles fossiles peuvent provoquer des écarts importants entre le prix spot et le prix contractuel, générant des pertes ou gains significatifs selon la couverture mise en place.

De ce fait, on peut également utiliser la méthode de Value-at-Risk (VaR) qui permet d'estimer la perte maximale probable sur un horizon de temps donnée à un niveau de confiance α . On reprend la formule précédente de calcul de risque où on définit la perte horaire de la manière suivante :

$$Loss_t = V_t^{\text{exposé}} * (P_{\text{spot},t} - P_{\text{terme}})$$

Et qui se définit sur une période T comme suit :

$$L_T = \sum_{t=1}^T L_t = \sum_{t=1}^T V_t^{\text{exposé}} * (P_{\text{spot},t} - P_{\text{terme}})$$

Avec $Loss_t = L_t$

La VaR au niveau de confiance α s'écrit alors :

$$VaR_{\alpha}(L_T) = \inf \{l \in \mathbb{R} : P(L_T \leq l) \geq \alpha\}$$

Les prix spot étant extrêmement volatils et pouvant atteindre des pics de prix conséquents autant positifs que négatifs, le modèle CVaR (Conditional Value-at-Risk) permet de mieux appréhender le risque en tenant compte des queues de distribution. Il permet d'estimer la perte moyenne en cas de pertes extrêmes :

$$CVaR_{\alpha}(L_T) = \mathbb{E}[L_T | L_T \geq VaR_{\alpha}(L_T)]$$

Une fois le risque marché couvert sur les volumes non sécurisés, le fournisseur doit considérer la différence entre la consommation réelle et la consommation prévue pour chaque client, ainsi que l'impact du profil horaire en fonction du prix auquel il vend son énergie. C'est là qu'intervient le risque Flex ou risque de flexibilité, qui valorise les écarts entre prévision et consommation effective au prix spot, mais en prenant en compte le prix facturé au client, combinant ainsi l'effet du volume total et du Shape (ou profil en français) horaire.

1.5.2. La constitution du risque de Flex

Le risque de flexibilité (ou risque Flex) représente le risque financier auquel un fournisseur est exposé lorsque la consommation réelle des clients diffère des prévisions établies pour un portefeuille donné, en tenant compte à la fois du volume total et du profil horaire de consommation. Il peut être considéré comme une conséquence financière de la combinaison de ces deux risques sur les écarts de volume et les écarts de Shape (profil), valorisés par la différence entre le prix de vente et le prix d'achat au spot.

1.5.2.1. Le risque de volume

Le risque de volume correspond à la différence entre la consommation totale prévue et la consommation totale réellement observée sur une période donnée pour un client ou un portefeuille de clients. Formellement, pour un client i sur une période T :

$$\Delta V_i^{total} = \sum_{t=1}^T (C_{i,t}^{réelle} - C_{i,t}^{prévision})$$

Où :

- $C_{i,t}^{réelle}$ est la consommation réelle du client i à l'instant t ,
- $C_{i,t}^{prévision}$ est la consommation prévue du client i à l'instant t ,
- T est la période considérée.

Pour obtenir la valorisation financière du risque de volume, cet écart est multiplié par la différence entre le prix contractuel d'achat ou de couverture de l'énergie (CAL, produits à terme Base ou Peak) au client et le prix spot du marché :

$$L_i^{volume} = \Delta V_i^{total} * (P_{spot,t} - P_{terme})$$

On calcule donc ici l'exposition du fournisseur sur ses approvisionnements indépendamment de la marge commerciale avec les clients.

1.5.2.2. Le risque de Shape

Le risque de Shape ou risque de profil horaire correspond aux écarts horaires entre le profil réel de consommation et le profil sur lequel le fournisseur a couvert sa consommation, indépendamment du volume total. Cela correspond donc au volume réellement disponible à chaque heure de la journée et à la répartition par heure du bloc de volume total. On le définit comme :

$$\Delta S_{i,t} = C_{i,t}^{réelle} - C_{i,t}^{profil_couvert}$$

Où $C_{i,t}^{profil_couvert}$ correspond à la consommation sur laquelle le fournisseur a structuré ses achats ou sa couverture (Base-Load ou Peak-Load par exemple).

La valorisation financière se calcule de manière similaire à la valorisation du risque de volume :

$$L_i^{shape} = \Delta S_{i,t} * (P_{spot,t} - P_{terme})$$

Quand, pour le risque de volume, on va comparer les consommations totales journalières ou mensuelles en regardant l'écart global à cette maille ; on va comparer la répartition horaire dans le risque de Shape, c'est-à-dire qu'on calcule l'écart intra-journalier entre profil couvert et profil réellement consommé.

Les écarts de Shape entraînent des ajustements coûteux si la consommation se concentre sur les heures de pointe ou creuses, entraînant un rachat ou une revente à des prix défavorables.

1.5.2.3. Le risque de Flex

Le risque de Flex combine volume et Shape, et mesure l'impact économique net de tous les écarts (totaux et horaires) valorisés au prix spot par rapport au prix de vente au client :

$$L_i^{Flex} = \sum_{t=1}^T (C_{i,t}^{réelle} - C_{i,t}^{prévision}) * (P_{spot,t} - P_{vente,i})$$

Où :

- $P_{vente,i}$ correspond au prix contractuel de vente au client i ,
- $P_{spot,t}$ correspond au prix spot horaire,
- $C_{i,t}^{réelle} - C_{i,t}^{prévision}$ est la différence de consommation qui inclut l'effet volume total et l'effet Shape horaire (puisque les écarts sont calculés heure par heure)

Plusieurs scénarii se présente :

- Si $C_{i,t}^{réelle} > C_{i,t}^{prévision}$ et $P_{vente} > P_{spot,t}$ alors le fournisseur a sous-estimé la consommation réelle et doit donc recourir au marché pour combler le manque à un coût plus avantageux au spot qui est inférieur au prix de vente de l'énergie au client permettant de potentiellement dégagée un gain pour le fournisseur.
- Si $C_{i,t}^{réelle} > C_{i,t}^{prévision}$ et $P_{vente} < P_{spot,t}$ alors le fournisseur doit acheter le manque d'énergie sur le marché spot à un prix supérieur à celui du prix de vente donc il est en perte.

- Si $C_{i,t}^{réelle} < C_{i,t}^{prévision}$ et $P_{vente} > P_{spot,t}$ alors le fournisseur a surévalué la consommation et doit revendre le surplus d'énergie au spot à un prix inférieur au prix de vente ce qui constitue potentiellement une perte sur marge pour le fournisseur.
- Si $C_{i,t}^{réelle} < C_{i,t}^{prévision}$ et $P_{vente} < P_{spot,t}$ alors le fournisseur revend le surplus d'énergie plus cher sur le marché spot que s'il l'avait vendu au client ce qui lui permet d'engendrer potentiellement des gains.

Le risque de Flex va tenir compte à la fois des écarts de volume total et des écarts de profil horaire, et valorise l'ensemble selon la différence entre le prix de vente facturé au client i et le prix spot à l'instant t .

La formalisation de cette prime permet de quantifier précisément les pertes potentielles liées à l'incertitude de consommation, d'ajuster en conséquence la couverture (aussi appelée Hedging) et de calibrer les primes de risque pour différents segments ou groupes de clients.

1.5.3. Les autres risques

En complément des risques directement liés au volume, au profil de consommation et à la flexibilité, un fournisseur d'électricité est également exposé à d'autres catégories de risque. Ces derniers, bien que parfois moins directement quantifiables, peuvent avoir un impact significatif sur la rentabilité et la pérennité d'un fournisseur.

Le risque de crédit, risque financier auquel chaque entreprise fait face indépendamment du secteur, correspond au risque de non-paiement, de retard de paiement ou de défaut de paiement des clients finaux. Le fournisseur se retrouve face à ce risque lorsqu'il fournit de l'énergie à crédit ce qui est le cas dans le secteur de l'électricité. Ce risque va se formaliser par la perte attendue de la façon suivante :

$$L^{crédit} = \sum_{i=1}^N (EAD_i * PD_i * LGD_i)$$

Où :

- EAD_i (Exposure At Default) correspond à l'exposition du fournisseur vis-à-vis du client i , c'est-à-dire le montant facturé et non encore réglé (en €),
- PD_i (Probability of Default) représente la probabilité de défaut du client i estimée à partir de sa solvabilité (score de crédit, historique de paiement, notation interne),
- LGD_i (Loss Given Default) correspond au taux de perte en cas de défaut, défini comme la proportion non recouvrable après procédures de recouvrement ou saisie de garanties = $\frac{\text{Montant non récupéré}}{\text{Montant total dû}}$.

Cette valorisation est rapportée au MWh livré au client qui correspond à sa consommation totale sur une année afin d'exprimer le coût du risque unitairement en prime en €/MWh.

Le risque d'imbalance correspond au coût engendré par les écarts entre les programmes notifiés par le fournisseur à son responsable d'équilibre (prévisions de consommation et d'approvisionnement) et les

consommations réelles mesurées sur son périmètre. Ces écarts sont réglés auprès du gestionnaire de réseau (RTE en France) aux prix de règlement des écarts, sur la base du coût de l'énergie d'équilibrage.

Pour un client i (ou un portefeuille de clients) à l'instant t , les écarts se définissent :

$$\Delta V_{i,t}^{imb} = C_{i,t}^{réelle} - C_{i,t}^{notifiée}$$

Où :

- $C_{i,t}^{réelle}$ correspond au volume soutiré (ou injecté) physiquement mesuré
- $C_{i,t}^{notifiée}$ représente les valeurs de volume de consommation prévue ou programmée (soutirage ou injection) transmises au responsable d'équilibre (après ajustements sur le marché Intra-Day)

Cet écart sera valorisé au prix du règlement d'écart avec l'application d'un prix dual¹⁸ :

- PRE_t^\uparrow (Prix d'imbalance dit « upward ») appliqué lorsque le périmètre consomme/soutire plus que programmé ($\Delta V_{i,t}^{imb} > 0$). Le système est donc à la hausse et nécessite des achats d'énergie d'équilibrage au prix PRE_t^\uparrow .
- PRE_t^\downarrow (Prix d'imbalance dit « downward ») appliqué lorsque le périmètre consomme/soutire moins que programmé ($\Delta V_{i,t}^{imb} < 0$). Le système est alors à la baisse et nécessite de « revendre » l'excédent d'énergie au prix PRE_t^\downarrow .

Le coût de règlement sur l'horizon T est donc de :

$$L_i^{imb} = \sum_{t=1}^T [\max(\Delta V_{i,t}^{imb}, 0) * PRE_t^\uparrow + \min(\Delta V_{i,t}^{imb}, 0) * PRE_t^\downarrow]$$

Le risque réglementaire désigne l'impact financier potentiel des changements législatifs ou réglementaires imposés par les autorités nationales ou européennes. Ces changements peuvent concerner :

- Les tarifs d'accès aux réseaux (TURPE en France),
- Les taxes et contributions (TVA, accises),
- Les mécanismes de capacité ou d'obligation de certificats,
- Les régulations spécifiques en cas de crise énergétique (plafonnement des prix, ARENH, etc.).

Les deux premiers changements concernant les taxes sont directement supportés par le consommateur final. Les deux derniers sont eux supportés par le fournisseur car ils impliquent une gestion différente de leurs achats / ventes auprès des producteurs et du marché pouvant impliquer des risques d'hausse des coûts de l'énergie notamment.

¹⁸ Un prix qui représente la valeur marginale du coût associé à une contrainte du problème d'optimisation correspondant au respect de l'équilibre strict entre offre et demande à un instant donné. Il permet de connaître la valeur supplémentaire générée ou économisée en augmentant la quantité d'électricité consommée ou produite pour obtenir l'équilibre.

On peut modéliser l'exposition au risque réglementaire comme une variation du coût unitaire :

$$L^{reg} = \sum_{t=1}^T \Delta P_t^{reg} * C_t^{réelle}$$

Avec :

- ΔP_t^{reg} qui correspond à la variation du coût induite par le changement réglementaire par MWh à l'instant t,
- $C_t^{réelle}$ qui représente la consommation réelle livrée.

D'autres risques existent tels que le risque d'illiquidité du marché, le risque opérationnel, le risque de modélisation, le risque d'appel de marge, le risque de faux profil de consommation mais vont impacter les coûts ou la rentabilité de l'entreprise indirectement et ont une probabilité de survenance moins importantes que les risques mentionnés.

Pour la suite de ce mémoire, les risques de marché et de Flex seront approfondis avec des approches pratiques pour les quantifier et les modéliser à des fins de prédictions.

1.5.4. Stratégies de gestion et Hedging

Ces risques induisent une gestion efficace et consciencieuse du portefeuille de consommation pour lesquels, après les avoir identifiés, des stratégies de couverture seront mise en place pour réduire l'exposition financière du portefeuille à ces risques.

Les risques peuvent être minimiser via des stratégies pures de management et de gestion via la diversification des profils clients, la détection des clients avant souscription présentant un risque de crédit par exemple ou encore la répartition géographique de la production sur tout le territoire national afin de capter l'ensemble des conditions météorologiques favorables aux productions des différentes technologies EnR et réduisant le risque de non-production sur certaines périodes.

Ils peuvent être également modéliser via des méthodes dites de « Hedging » permettant de couvrir en volume les écarts ou limiter l'impact des variations des prix du marché en valeur. Ils consistent en deux stratégies : le Hedging en volume et le Hedging en valeur qu'on présente ci-dessous.

Le Hedging en volume consiste à couvrir les écarts de consommation prévus par rapport à la consommation réelle afin de neutraliser le risque financier associé à la quantité d'électricité à fournir. Pour un client i sur une période T, le fournisseur peut définir un volume couvert $V_i^{couvert}$ proche de la consommation prévue $\sum_{t=1}^T C_{i,t}^{prévision}$.

L'exposition au risque en volume s'écrira donc :

$$\Delta V_i = \sum_{t=1}^T (C_{i,t}^{réelle} - C_i^{couvert})$$

Le Hedging en volume va consister à acheter ou vendre sur le marché à terme le volume ΔV_i pour neutraliser ou réduire fortement l'exposition. Le coût ou gain associé à cette stratégie est :

$$L_i^{Hedge\ vol} = \Delta V_i * (P^{contrat} - P_i^{spot})$$

Où :

- $P^{contrat}$ est le prix fixé par le contrat d'achat ou de vente à terme,
- P_i^{spot} est le prix spot réel observé sur le marché

Ainsi, si $\Delta V_i > 0$ (consommation réelle supérieure au volume couvert), le fournisseur achètera le complément au prix spot. A l'inverse, si $\Delta V_i < 0$, il revendra le surplus sur le marché.

Le Hedging en valeur vise à limiter l'impact des fluctuations de prix sur la rentabilité du fournisseur. Contrairement au Hedging en volume, l'objectif est de sécuriser la marge financière, indépendamment des écarts de consommation. On valorise l'exposition aux prix par :

$$L_i^{Hedge\ valeur} = \sum_{t=1}^T (C_{i,t}^{réelle} * (P_t^{spot} - P^{contrat}))$$

Où :

- $C_{i,t}^{réelle}$ est la consommation réelle à l'instant t,
- P_t^{spot} est le prix spot observé à l'instant t.

Dans ce cas, le fournisseur peut se couvrir via des :

- Contrats à terme (Forward ou Futures) : volume fixe d'électricité acheté ou vendu à un prix déterminé pour neutraliser la sensibilité aux fluctuations du marché,
- Options sur l'électricité : droit d'acheter ou de vendre à un prix donné, limitant les pertes en cas de mouvements extrêmes du prix,
- Swaps : échange de flux variables contre des flux fixes, stabilisant la marge.

En combinant ces deux stratégies de Hedging en volume et en valeur, le fournisseur pourra réduire la volatilité financière du portefeuille de clients, ajuster la couverture nécessaire selon les prévisions de consommation faites et prioriser la couverture sur des segments de clients ou des périodes jugées plus à risque ou celles les plus exposées au risque. Il peut également inclure les coûts des pénalités d'imbalance ou réglementaires dans la stratégie globale.

Mathématiquement, la combinaison des deux stratégies se définit, pour un portefeuille de N clients :

$$L^{hedge} = \sum_{i=1}^N \sum_{t=1}^T [(C_{i,t}^{réelle} - C_i^{couvert}) * (P^{contrat} - P_t^{spot})]$$

Cette formule synthétise les couvertures en volume et en valeur et représente la base du calcul de la prime de risque du portefeuille.

2. Analyse des risques

Pour arriver à calculer les primes de risque, il est essentiel de comprendre la composition du portefeuille de consommation et son comportement dans le temps. La notion de temps est prépondérante dans cette étude car la majorité des données manipulées ont une dépendance temporelle qui doit être prise en compte pour la suite de l'analyse et des calculs qui seront effectués.

L'analyse de données se base sur l'année 2024 ce qui permet d'avoir un historique pour chaque variable sur une année entière le plus récent possible.

2.1. Présentation du portefeuille clients et des variables

Énergie d'Ici comprend plus de 16 000 points de livraison¹⁹ d'électricité répartis dans des segments et avec des profils variés de consommation qui seront détaillés dans cette partie.

2.1.1. Segmentation clients

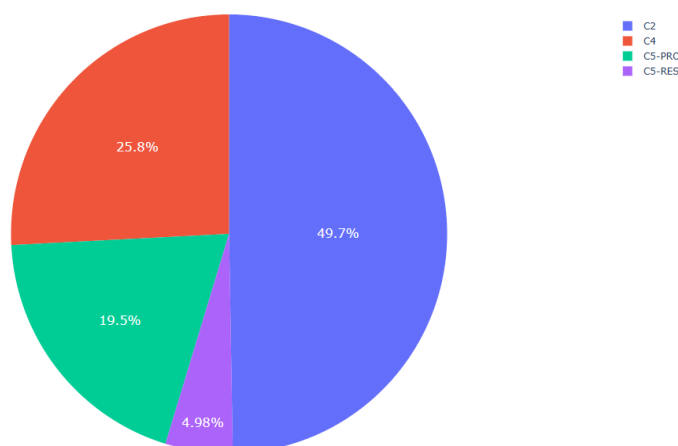
Le portefeuille de consommation (comprenant donc l'ensemble des points de livraison) sont répartis par segment. Les critères de segmentation sont définis par Enedis et vont être principalement liée à la puissance souscrite, au point de connexion et au profil de consommation²⁰. On retrouve les segments :

- **C2** qui correspond aux clients qui sont raccordés en HTA et qui disposent d'une puissance souscrite supérieure à 250 kVA. Ce sont de très grands consommateurs qui vont principalement correspondre à des clients industriels (industrie chimique, cimenteries, ...) et tertiaires lourds (grandes plateformes logistiques, grands centres hospitaliers, ...). Ces clients ont un profil type de consommation importante et continue, sur une grande surface ou sur plusieurs sites.
- **C4** qui correspond aux clients qui sont raccordés en basse tension avec une puissance souscrite comprise entre 36 kVA et 250 kVA. On va retrouver des clients d'industrie à taille intermédiaire (menuiserie industrielle, imprimeries, ...), des établissements scolaires ou universités de grande taille, des collectivités locales et des grandes surfaces commerciales. Ils ont un profil type de consommation soutenue mais moins constante et plus influencée par les heures de pointe et/ou la saisonnalité.
- **C5** qui comprend les clients dits « petits » professionnels et les clients particuliers appelés « résidentiels » qui sont raccordés en basse tension avec une puissance souscrite inférieure à 36 kVA. Ces clients ont un profil type de consommation plus faible et souvent concentré sur les heures ouvrées, les jours ouvrés avec une forte variabilité selon l'activité ou les périodes du calendrier.

¹⁹ Point de livraison ou en abrégé PDL fait référence au local ou lieu d'habitation pour lequel un contrat a été souscrit et où l'électricité est acheminée. Le terme de client n'est pas utilisé ici car un client peut représenter plusieurs PDL.

²⁰ Ce sont des critères standards utilisés par Enedis et que les fournisseurs utilisent pour segmenter leur portefeuille de consommation, [A quoi correspondent les points de connexion C1 à C5 ? | Enedis, 12/08/2025](#)

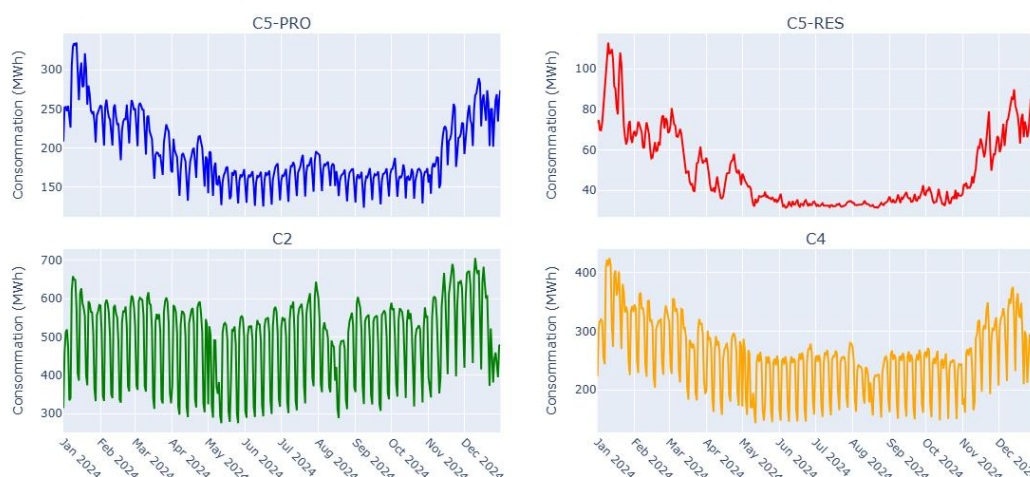
Proportion des consommations par segment



Ces segments sont répartis dans le portefeuille en proportion de volume total consommé avec une part majoritaire de clients industriels et professionnels représentant 95% du volume de consommation du portefeuille et les résidentiels représentant le volume restant, soit 5%.

Les profils type de consommation se caractérisent bien sur le graphique ci-dessous qui représente la courbe de charge²¹ sur l'année 2024 au pas journalier²² de chaque segment avec des volumes de consommation différents et des formes de courbe représentant l'activité plus ou moins soutenue des clients du segment. Une différenciation dans le segment C5 entre les clients professionnels et résidentiels permet de voir qu'ils n'ont pas la même forme de courbe de consommation.

Consommation journalière par segment en 2024



Le volume de consommation est un résultat influencé par le comportement de consommation des clients qui dépend de variables que nous allons présenter dans le prochain paragraphe.

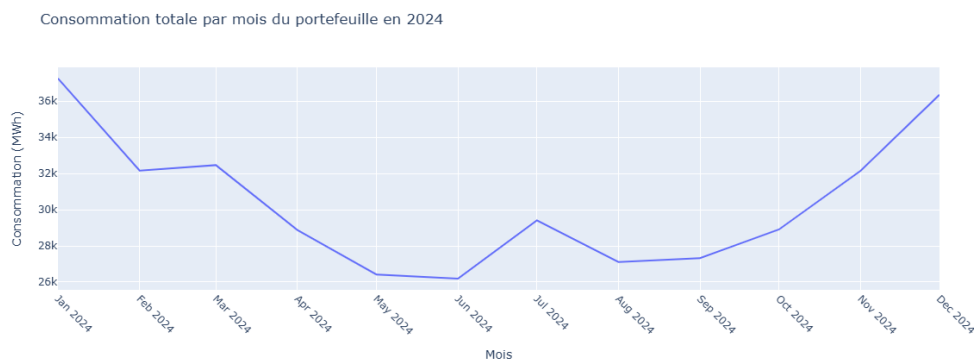
2.1.2. Variables clés et besoins analytiques

La demande en électricité est influencée par les saisons et la temporalité comme nous l'avons précédemment présenté en partie 1.

²¹ Une courbe de charge est un graphique représentant la consommation d'électricité en fonction d'une période définie.

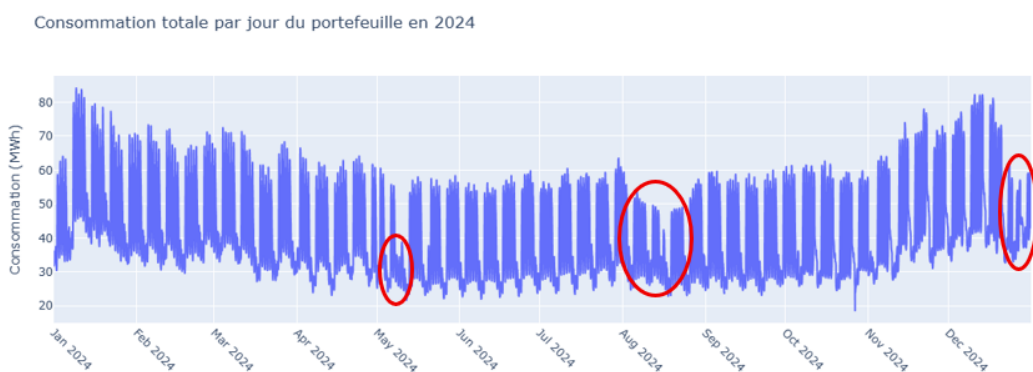
²² Les données sont initialement au pas horaire ou pas 5 minutes pour les segments C2 et C4, c'est-à-dire qu'il y a des données de consommation chaque heure ou chaque 5 minutes qu'on a agrégé et sommé pour avoir une consommation par jour qu'on appelle « au pas journalier » pour une meilleure visibilité des courbes de charge.

C'est bien le cas dans le portefeuille de clients qui présente une saisonnalité annuelle en fonction des saisons été/hiver comme le montre le graphique ci-dessous :



Avec un pic de consommation sur la période hivernale de novembre à février plus une décroissance jusqu'à la période estivale de juillet à août avec les départs en vacances et une activité moins soutenue des professionnels puis une consommation à la hausse dès septembre caractérisant une reprise de l'activité économique.

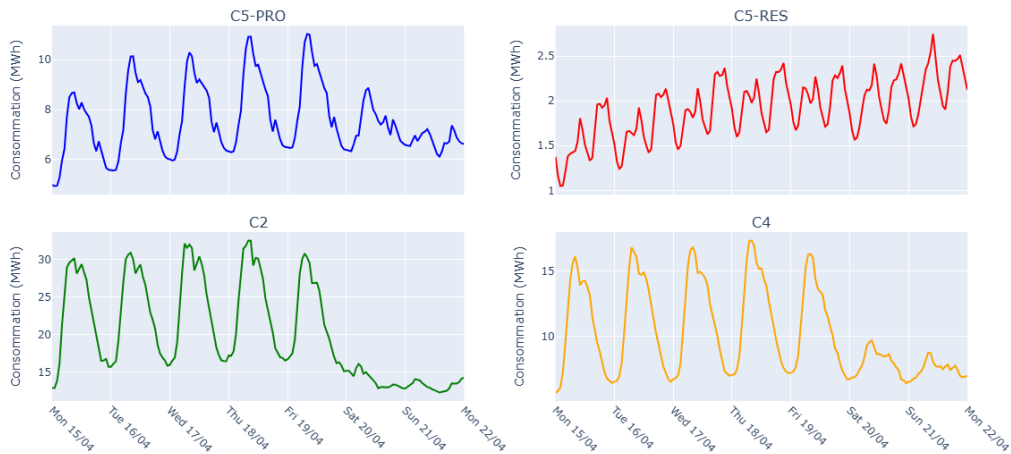
Au pas horaire, sur une saisonnalité annuelle, le prochain graphique nous présente également l'influence des jours fériés et jours de vacances sur la consommation d'énergie où la demande des clients professionnels et industriels diminuent tandis que celle des résidentiels augmente dans une moindre proportion par rapport à la baisse des clients professionnels :



On a une baisse de la consommation bien visible au mois de mai avec 4 jours fériés puis au mois d'août, lors des vacances d'été où l'activité des professionnels est en partie ralentie ou arrêtée pour congés estivaux et on a les vacances de Noël qui représente, en pleine période estivale, une part de consommation plus faible.

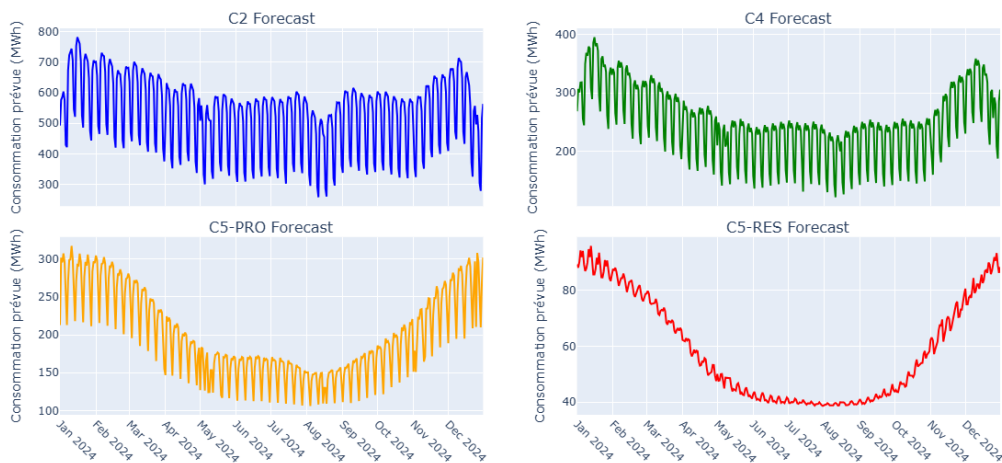
On a également une saisonnalité hebdomadaire avec une consommation moindre les jours de week-end par rapport aux jours de semaines et une saisonnalité journalière avec des périodes de faibles et hautes consommations en fonction du moment de la journée comme nous pouvons le constater sur le graphique suivant :

Consommation horaire par segment - Semaine du 15/04/2024



C'est sur ces données historiques que va se baser la modélisation du volume de consommation à prédire correspondant à la consommation prédite qui sera utile pour le calcul des écarts de consommation pour la détermination et la quantification des primes de risques. Le graphique ci-dessous présente les prévisions journalières de consommation de chaque segment :

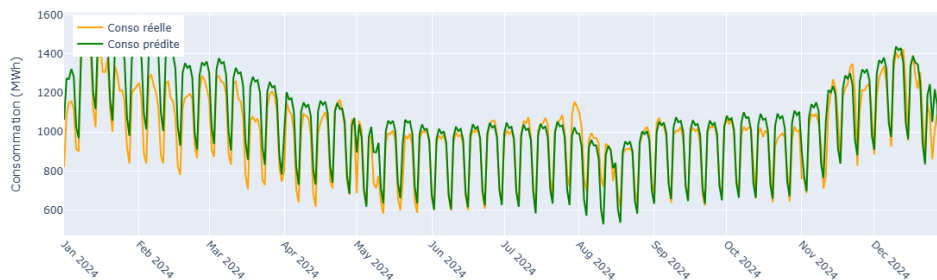
Prévisions journalières par segment en 2024



Ces écarts à la consommation qu'on définit comme la déviation de volume se calcule en soustrayant la consommation prédite au pas horaire à la consommation réelle au pas horaire :

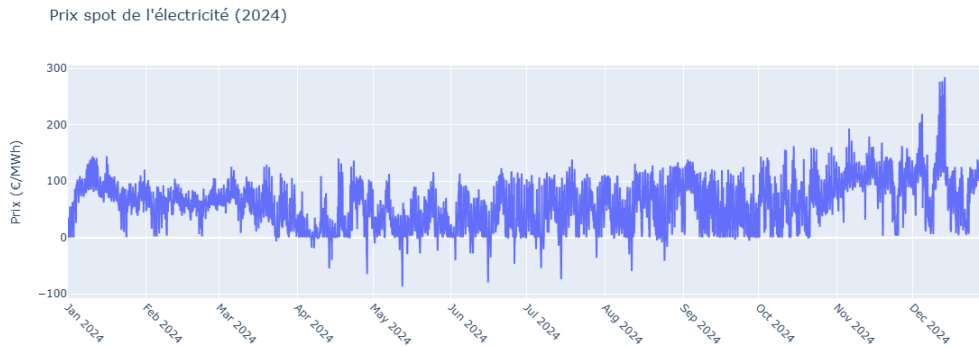
$$\Delta Volume = C_{réelle} - C_{prédite}$$

Comparaison de la consommation totale du portefeuille : réelle vs prédite (journalier, 2024)



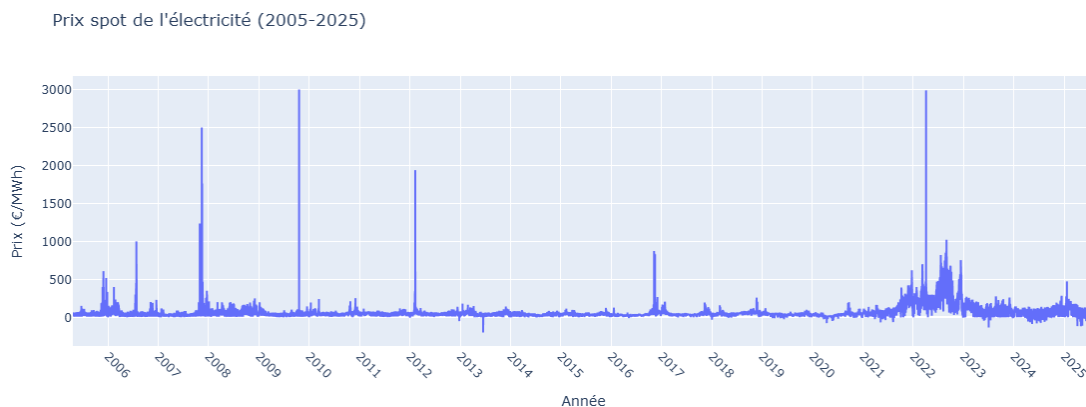
La demande est également influencée par la météo et les températures extérieures notamment pour le chauffage et, avec le réchauffement climatique et la hausse des températures, la climatisation qui exerce une influence sur les prix et le coût du risque.

Nous avons d'abord les prix des marchés notamment les prix de court terme appelé prix spot présentant également une saisonnalité similaire à celle de la demande :



On a des pics de prix en période hivernale, là où la demande est la plus importante et des pics inverse, vers le bas, lorsque la production est plus importante que la consommation, notamment en période estivale avec la forte production des parcs solaires couplée à une faible consommation.

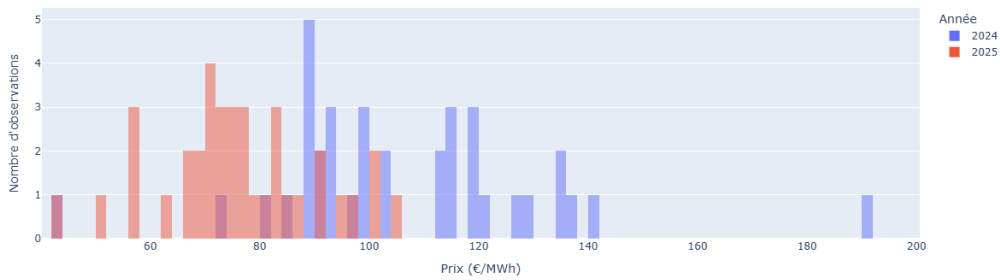
Les prix spot vont finalement suivre la demande et la production qui est mise en face pour y répondre en fonction de l'activité des consommateurs et de leurs besoins mais vont également refléter les crises économiques ou énergétiques que traverse le marché comme on peut le voir avec le graphique des données de prix de 2005 à 2025 :



Il y a également les prix long terme, appelé prix Future, acheté en valeur CAL²³ par le fournisseur, qui vont dépendre aussi de la demande prévisionnelle, des capacités de production disponibles mais dépend plus fortement, comme le prix spot, des tensions et crises énergétiques qu'il y a eu ou est en cours ou aura. On voit bien la répartition des prix CAL par maturité dans le graphique suivant avec en 2024 des prix plus élevés faisant suite encore à la crise que 2025, où les marchés sont moins tendus et les prix continuent leur décroissance :

²³ CAL pour Calendar qui correspond au prix d'achat de bloc d'énergie pour une année

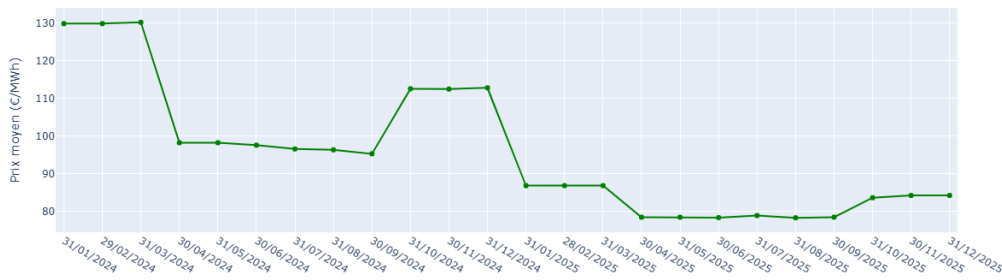
Histogrammes superposés des prix CAL Marché pour 2024 et 2025



Ce prix Future traduit le sentiment des producteurs et acheteurs sur l'évolution des prix du marché de l'électricité sur les années à venir en prenant en compte le contexte économique et géopolitique.

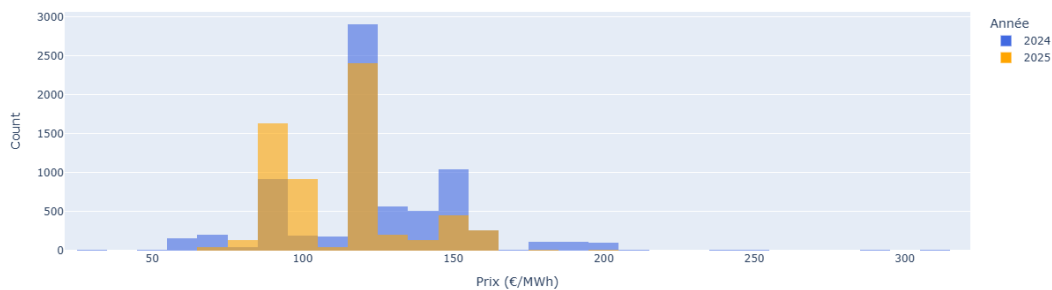
On voit sur ce graphique l'évolution des prix CAL marché mais nous avons également à disposition, étant donné la nature d'Énergie d'Ici qui est fournisseur et producteur, les prix CAL producteurs, des prix issus de contrats négociés en amont avec les producteurs pour l'achat de bloc d'énergie sur une période donnée, à l'année ou au trimestre. Les prix se répartissent sur les années 2024 et 2025 de la façon suivante avec une nette décroissance constatée de la valeur du prix du MWh :

Prix moyen par mois des produits Future producteurs (2024-2025)



Enfin, nous avons les prix de contrat, les montants tarifés aux clients pour la part d'énergie consommée sur la base des prévisions influencer par le coût d'achat de l'énergie et le contexte économique et géopolitique avec des prix encore élevés en 2024 comme nous le présente le graphique ci-dessous :

Distribution des prix de contrat en 2024 et 2025



2.2. Données

Ces données sont issues de diverses sources, internes et externes, en libre accès ou payante, et permettent de réaliser les calculs nécessaires notamment pour la définition des primes de risque. La qualité de ces données est donc essentielle pour la robustesse des analyses, la pertinence des calculs et la fiabilité des

conclusions. L'hétérogénéité des sources, des formats et des granularités temporelles des données impose un travail méthodique de collecte, de structure et de traitement avant de pouvoir les exploiter.

2.2.1. Collecte et structuration

Les données exploitées proviennent de plusieurs sources, à la fois interne et externe. Elles sont mesurées sur des pas de temps variables et ont des caractéristiques spécifiques.

On va avoir besoin de collecter les données de consommation fournies par Enedis pour les clients télé-relevés, pour lesquels les fichiers comprennent directement les courbes de charge, et pour les clients profilés, sous format de bloc de volume total consommé pour le segment C5 notamment, un fichier qui devra être retraité et travaillé pour avoir les courbes de charges des clients C5. Chacun de ces fichiers pour les clients télé-relevés sont à la granularité temporelle 15 minutes. Les fichiers des clients profilés sont au pas mensuel voire trimestriel pour certains clients qui sont encore avec un compteur de relève manuelle (et non télé-communicants comme les compteurs Linky).

De ces données, vont être modéliser la consommation future des segments afin de déterminer les achats futurs d'énergie. Les résultats de ces modélisations vont servir à constituer le fichier de prévision de la consommation par segment permettant ensuite de construire le fichier des écarts de consommation entre réelle et prévisionnelle, au pas horaire, pour chaque segment et même chaque PDL (pour les segments C2 et C4).

Est mis à disposition en interne, le prix d'achat des blocs d'énergie sur 3 ans depuis 2024 auprès du marché et des producteurs (en gré à gré) et la tarification des clients en €/MWh pour les segments C2 et C4.

On va également récupérer les prix des produits à terme (Future / Forward) et des prix spot de chaque année depuis 2022 pour le marché à terme et 2005 pour le marché spot, via le serveur d'EEX, bourse de l'énergie en Europe.

Pour les modélisations avancées, on va récupérer les données externes suivantes :

- Les données météo (température, force du vent, humidité et pression atmosphérique) récupérées automatiquement via un lien avec l'API de Météo France. Les données sont stockées sur Google Big Query.
- Les calendriers des jours fériés et des vacances scolaires sont téléchargés directement sur le site data.gouv.fr

La diversité de ces sources induit une hétérogénéité des formats. En effet, on va avoir des fichiers en format CSV ou Excel issus des téléchargements et fichiers créés, qui eux-mêmes peuvent avoir des formats spécifiques (séparation des données avec des virgules ou des points-virgules, indexation de la date ou non, ...). Il y a également les fichiers en format JSON pour les flux de données issus d'APIs puis les bases de données relationnelles internes (Google Cloud Plateform ou BigQuery) accessibles via requête SQL. Il faut donc uniformiser tous ces types de données pour ne travailler qu'avec un seul format dans les scripts de code. Dans le cadre de cette étude, le format CSV a été choisi pour la simplicité d'import et d'export des tableaux de données dans les scripts.

Afin de traiter cette diversité de format, l'extraction des données se fait par automatisation via des scripts Python en utilisant les bibliothèques pandas pour le traitement tabulaire (fichier en format CSV et Excel), requests et json pour l'appel et la gestion des données issues des APIs ainsi que de google. cloud pour les requêtes BigQuery. Certaines données sont sous plusieurs fichiers qu'il faut concaténer en un seul ce qui est permis par l'automatisation. Ce processus permet d'économiser du temps de traitement et de limiter les oublis ou erreurs.

Ces scripts automatisés permettent également de récupérer une volumétrie importante de données. En effet, un seul client peut générer plusieurs dizaines de milliers de points de mesure annuellement lorsque la granularité est au pas quinze minutes (à titre d'exemple, pour l'année 2024, il y a 8784 heures soit 8784 lignes au pas horaire, mais, au pas 15 minutes ce nombre monte à $8784 * 4 = 35\ 136$ lignes, et ceci uniquement pour un seul client !). Ainsi, à l'échelle d'un portefeuille, ce volume se chiffre rapidement en plusieurs millions voire dizaine de millions de lignes. Un volume de données qui ne peut plus être gérés ni traités ni analysés correctement uniquement via les outils simples comme Excel et nécessite donc l'usage de BigQuery (Google Cloud) pour le stockage des données et l'exécution de requête via les scripts Python.

La récupération, le traitement et l'usage des données pour analyse et modélisation nécessite également une bonne documentation des codes créés et utilisées afin de garantir la bonne traçabilité des scripts et de permettre la mise à jour des codes et leur réutilisation facilement. De ce fait, chaque étape de collecte, de transformation et de modélisation des données a été documenté dans le cadre de ce mémoire. Les scripts sont organisés par modules correspondant aux étapes du pipeline (import, prétraitement, enrichissement, stockage des données) et de notebooks commentés pour accompagner l'exécution des scripts modulaires. Cette démarche permet de structurer le travail et de garantir une bonne lisibilité de la logique de programmation pour faciliter les potentielles réutilisations.

2.2.2. Prétraitement et nettoyage

Une fois les données collectées, elles doivent être préparées²⁴ pour les analyses et modélisations à faire. Une étape essentielle pour détecter et corriger les incohérences, combler les manquantes et mettre les séries de données dans un format homogène et pratique pour les calculs.

Pour utiliser les données dans les codes on doit les convertir en DataFrame en s'assurant qu'elles soient toutes uniformisées, c'est-à-dire avec la même unité de mesure. Pour la consommation par exemple, les consommations réelles sont en Wh et les consommations prédites en MWh, pour plus de praticité, on convertit les consommations réelles en MWh. Les prix spot et à terme sont déjà en €/MWh et pour la météo, on ne garde que les données de température, qui sont plus utiles dans notre cas pour la consommation. Pour la consommation, on doit également harmoniser l'horodatage. En effet, les données de prix spot et à terme sont en UTC alors que celles des consommations sont au fuseau horaire de Paris. On convertit donc les données de prix en fuseau horaire des consommations pour éviter des décalages temporels.

On va ensuite vérifier, dans les données, la présence de doublons de date ou de PDL qui peuvent apparaître notamment avec le changement horaire et le passage en heure d'été perturbant la continuité

²⁴ Les codes de récupération des données sont disponibles dans le répertoire personnel GitHub suivant : <https://github.com/Meryam5/Memoire>. Les étapes de prétraitement sont faites en amont des calculs et modélisations dans les scripts partie2.ipynb et partie3.ipynb

logique de la série temporelle qui n'a plus une journée de 24 heures mais se retrouve avec une journée spécifique à 25 heures. Pour ce faire, on somme, pour la date de passage à l'heure d'hier, à 2 heures, les consommations pour n'avoir qu'une seule ligne à 2 heures cette journée. Si on se retrouve avec des PDL identiques dans les données, on supprime la colonne ou la ligne en doublon.

Pour le changement d'horaire, on aura également une problématique pour le passage à l'heure d'été où on « perd » une heure ce qui résulte à une journée de 23 heures perturbant également la continuité temporelle. Pour ce faire, on ajoute une ligne à 2 heures de la journée du changement d'heure avec un volume à 0 MWh de consommation.

On peut également avoir des données manquantes dans les données de consommation ce qui est le cas pour le segment C5 où nous n'avons les données qu'au pas mensuel voir trimestriel et où il faut reconstituer les données pour obtenir une courbe de charge horaire exploitable. Cette reconstitution se fait en extrapolant les données profilées en données télé-relevées. Pour ce faire, on applique une méthode de reconstitution basée sur les profils standard fournis par Enedis. Ces profils traduisent la forme type de la consommation pour chaque demi-heure d'une journée donnée. Ces coefficients sont calculés à l'aide d'un échantillon représentatif d'une population suivant un ensemble de critères permettant de regrouper les consommateurs en strates. On divise ensuite chaque point demi-horaire de cette courbe par sa moyenne pour obtenir les coefficients du profil.

On normalise ensuite la courbe type afin d'obtenir un profil horaire unitaire : $\sum_{t \in j} Profil^{(j)}(t) = 1$. On multiplie ce profil par un facteur d'usage FU_{ij} qui assure la cohérence entre l'énergie effectivement consommée par le client i sur le poste horo-saisonnier j .

On a ainsi la courbe de charge reconstituée qui est défini par :

$$Cdc_i(t) = FU_{i,j} * Profil^{(j)}(t), \quad \forall t \in j$$

Avec :

- $Profil^{(j)}(t)$ qui correspond au profil horaire fourni par Enedis, normalisé sur le poste j
- FU_{ij} qui est le facteur d'usage, défini comme l'énergie du client sur le poste j de la façon suivante : $FU_{i,j} = E_{i,j}$

Si on ne dispose pas des consommations détaillées par poste, ce qui est le cas dans notre étude, et qu'on n'a uniquement l'énergie annuelle E_i , on répartit cette énergie entre les postes via des coefficients $\alpha_{i,j}$ également fournis par Enedis, de sorte que :

$$FU_{i,j} = \alpha_{i,j} * E_i \quad \text{avec} \quad \sum_j \alpha_{i,j} = 1$$

Cela nous permet de vérifier si la somme annuelle reconstituée est bien égale à la consommation observée tout en restituant la dynamique horaire de la consommation des clients profilés avec les profils fournis par Enedis.

Il peut également avoir des données manquantes de consommation dans quelques cellules précises qu'on comble par interpolation linéaire suivant la continuité de la série temporelle.

Enfin, on va commencer un travail d'analyse en vérifiant la nature des valeurs extrêmes (ou outliers). On va chercher à savoir si ces outliers sont « normaux » et correspondent à des pics de consommation dus à des événements spécifiques (journée froide, hausse de l'activité résultant d'une hausse importante de la consommation, ...), ou bien s'ils proviennent d'une anomalie de mesure ou d'une défaillance de remontée des données. Une valeur aberrante apparaît constamment dans les données et correspond à l'heure de passage à l'heure d'été où 2 heures du matin a une consommation nulle ajoutée pour garder la continuité des données. Lorsqu'apparaît un « trou dans les données » comme cela a été le cas dans cette étude dans les prix spot pour la date du 17 juillet où une indisponibilité technique des données n'a pas permis de récupérer les prix de ce jour, on effectue une interpolation linéaire simple se basant sur la moyenne des jours du mois de juillet correspondant à un jeudi en attente de correction technique.

Au terme du prétraitement, les données sont consolidées dans des fichiers et bases propres. Les données sont continues, exemptes de « trous » majeurs, horodatées de manière uniforme et correcte, et alignées entre variables internes et externes. Les valeurs aberrantes sont soit corrigées, soit identifiées comme événements particuliers pouvant renforcer la volatilité. Ces bases de données ainsi nettoyées nous permettent alors de pouvoir effectuer les analyses, calculs et modélisations dans la suite de ce mémoire.

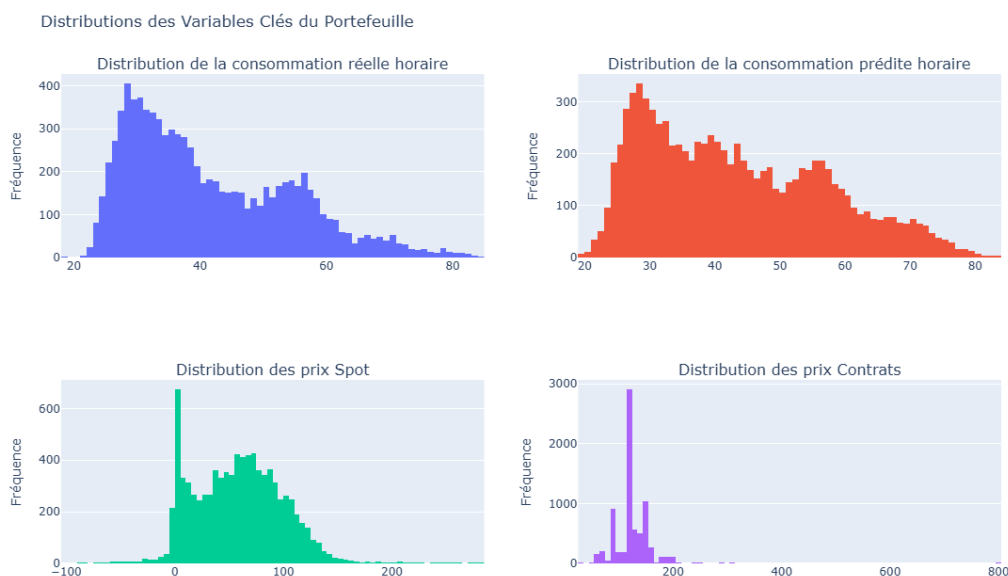
2.3. Analyse exploratoire des données

Cette analyse exploratoire permet de pousser plus loin l'observation des précédents graphiques des données et variables utilisées afin de les décrire de manière statistique.

Cette analyse sera différente en fonction du choix de la maille/échelle de travail qui influence les résultats obtenus notamment à cause de l'effet de foisonnement qui sera présenté dans la troisième partie. De même, si on choisit de travailler au pas horaire, journalier ou mensuel, les résultats seront différents.

2.3.1. Statistiques descriptives

On commence par la répartition de chacune des variables présentées précédemment pour visualiser la distribution des données sur la totalité du portefeuille :

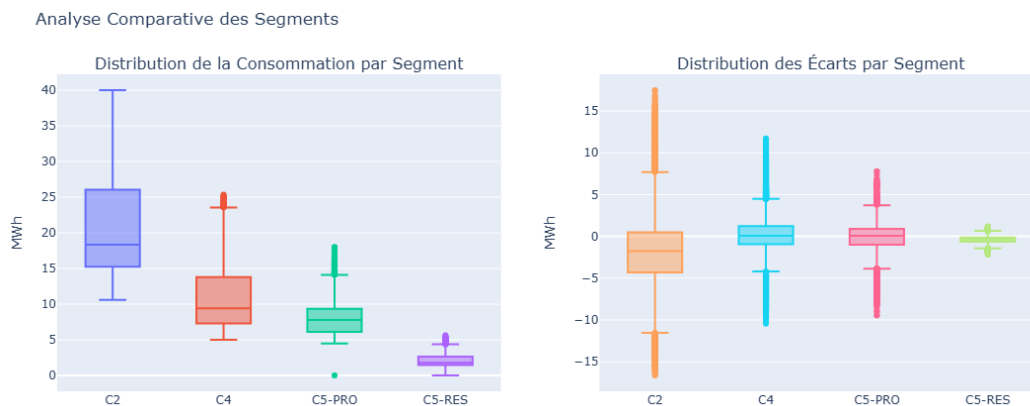


Les prix Spot présentent une asymétrie marquée vers la droite avec la majorité des prix compris entre 0 et 100 €/MWh. Mais il existe une longue queue épaisse de prix élevés avec un maximum à 284,23€/MWh et une queue moins longue mais non négligeable de prix négatifs avec un minimum à -87,29€/MWh pour une moyenne sur l'année 2024 de 87,73€/MWh avec un écart-type de 40,68€/MWh traduisant la forte volatilité du prix Spot.

Pour les prix des contrats (uniquement des segments C2 et C4), qui correspondent aux prix auxquels le fournisseur vend le MWh aux clients, ils sont plus condensés autour de la moyenne (124,61€/MWh) et reflètent les prix encore élevés dus à la crise énergétique de 2022 où ces contrats ont été négociés en amont pour garantir un prix fixe d'achat aux clients et une vente à un prix fixe pour le fournisseur. Avec un minimum à 28,56€/MWh, les prix de contrat sont allés jusqu'à un maximum de 800€/MWh.

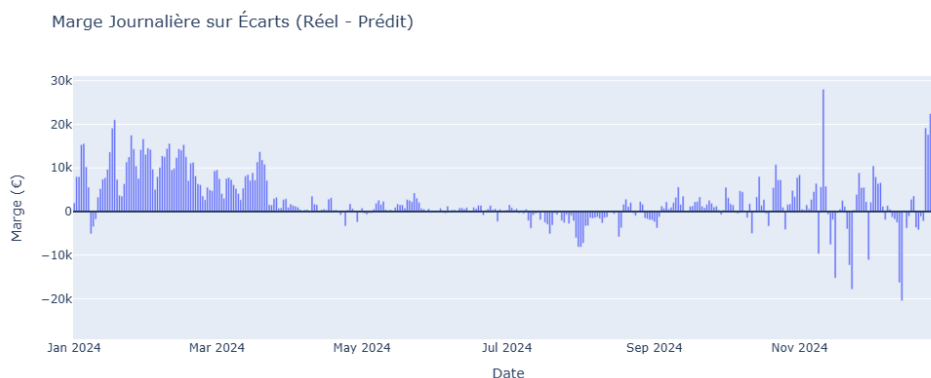
La consommation réelle au pas horaire présente une forme bimodale, c'est-à-dire avec deux pics qui suggèrent l'idée de deux régimes de consommation différencié dans la fréquence horaire du niveau de consommation. Cela pourrait correspondre aux différences entre les heures de forte activité (les heures de pics au sein d'une journée, les jours ouvrés, les heures lors de la période hivernale) et de faible activité (durant la nuit, les week-ends, la période estivale). On a une consommation moyenne de 41,5 MWh par heure avec un minimum de consommation à 18,6 MWh et un maximum de consommation à 84,1 MWh.

Enfin, la consommation prédite présente également deux pics mais moi distincts que la consommation réelle avec une moyenne de consommation par heure de 43,4 MWh. Elle présente un minimum par heure de consommation à 19,4 MWh et un maximum à 83,7 MWh. On peut déjà imaginer des écarts potentiels de consommation entre la réelle et celle prédite avec une tendance à la surestimation de la consommation lors de la prédiction. Ces écarts sont effectivement observables sur le graphique suivant en comparaison de la consommation globale par segment :



On voit, par rapport à la consommation de chaque segment, que les écarts sont plus ou moins importants. Pour le segment C2 par exemple, l'amplitude des écarts est grande même si la médiane est négative (on a une surestimation de la consommation où il faut donc revendre le surplus), on a tout de même des pics d'écarts à plus de 17 MWh. On a, dans une moindre proportion, le même type de disparité pour les segments C4 et C5-PRO mais une très faible dispersion pour le segment C5.

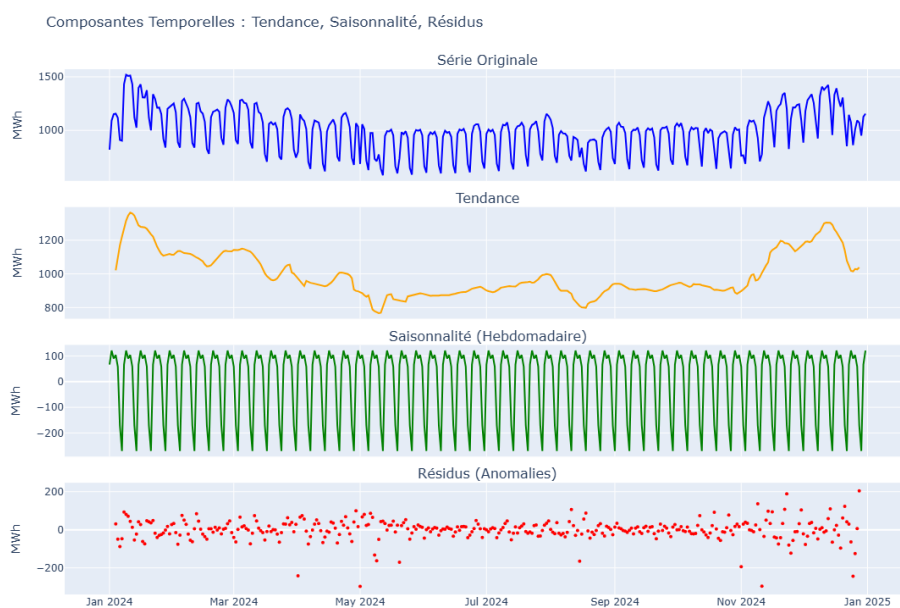
On peut déjà chercher à observer la marge journalière générée par ces écarts entre la consommation réelle et la consommation prédite sur l'année 2024 avec le graphe ci-dessous :



Toutes les barres au-dessus de 0 représentent un gain potentiel qui se produit lorsque la consommation réelle est inférieure à la consommation prédite où l'énergie excédentaire est alors vendue sur le marché spot. A l'inverse, les barres négatives représentent un coût dans le sens où le fournisseur devra acheter l'énergie manquante sur le marché spot à un coût non-connu à l'avance.

Sur l'année, on voit des pics plus importants de janvier à avril puis d'octobre à décembre et une période estivale plutôt stable. Cela met en évidence une performance variable du modèle de prévision actuellement utilisé en fonction des saisons avec des biais de sous et surestimation pouvant être conséquents. Mais globalement, la gestion des écarts a été profitable sur la période avec une baisse des prix Spot sur l'année 2024 avantageuse pour l'achat d'énergie, moins pour la revente mais qui a lieu lors de moment de pics de demande en énergie. Cette saisonnalité des écarts se retrouve généralement dans la saisonnalité de la consommation, une composante de la série temporelle que caractérise les données de consommation électrique.

La forme de la série temporelle nous permet d'avoir une idée de la forme de chacune de ses composantes comme le montre le graphique suivant :



On a une tendance saisonnière annuelle classique pour la consommation électrique avec une consommation élevée en période hivernale (début et fin d'année) et un creux de consommation pendant l'été.

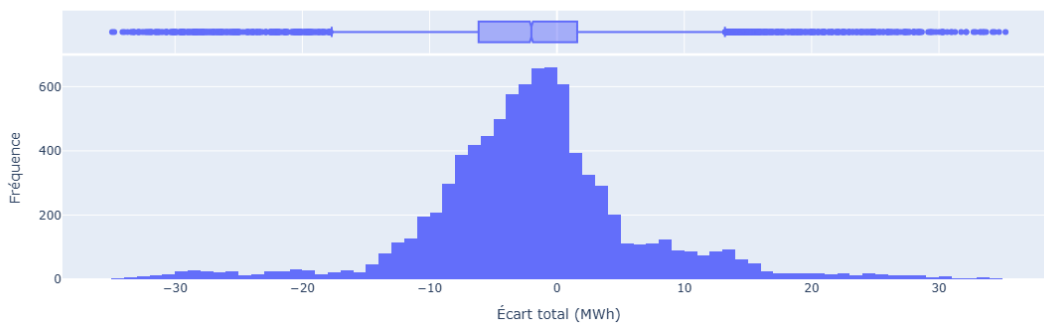
On a un cycle hebdomadaire très net de la saisonnalité (qu'on voit bien grâce à la mise à l'échelle des données au pas journalier. Un choix de visualisation au pas journalier a été fait car la saisonnalité, au pas horaire, est bruitée et ne nous permet pas de visualiser une saisonnalité). On voit bien une saisonnalité élevée pendant les jours de la semaine et une chute significative durant le week-end reflétant la différence d'activité entre les jours ouvrés et les jours de repos.

Enfin, les résidus correspondent aux variations inexplicables (appelées également bruits) après avoir retiré la tendance et la saisonnalité. Si une grande partie des points se retrouvent autour de 0, il y en a plusieurs qui s'écartent fortement de 0 (positivement et négativement) signalant des jours où la consommation a été anormalement élevée ou faible par rapport au modèle. Ces anomalies peuvent correspondre à des jours fériés, des événements climatiques exceptionnels ou à d'autres éléments non-récurrents tels qu'une crise dans le secteur de l'énergie par exemple. L'analyse de ces résidus est donc essentielle pour identifier les événements imprévisibles.

2.3.2. Analyse des résidus

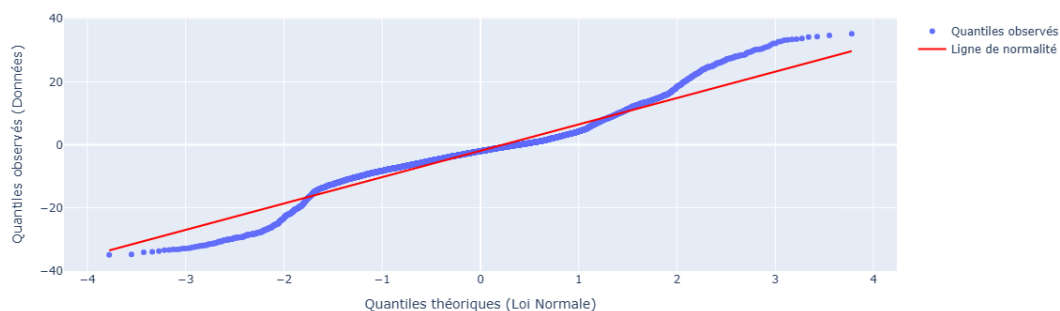
Même si on a une forte proportion d'écart autour de 0, on a une asymétrie marquée à gauche avec un nombre d'écarts négatifs plus important mais une dispersion des écarts majeures caractérisée par des queues épaisses et longues des deux côtés.

Distribution des écarts totaux du portefeuille (2024)



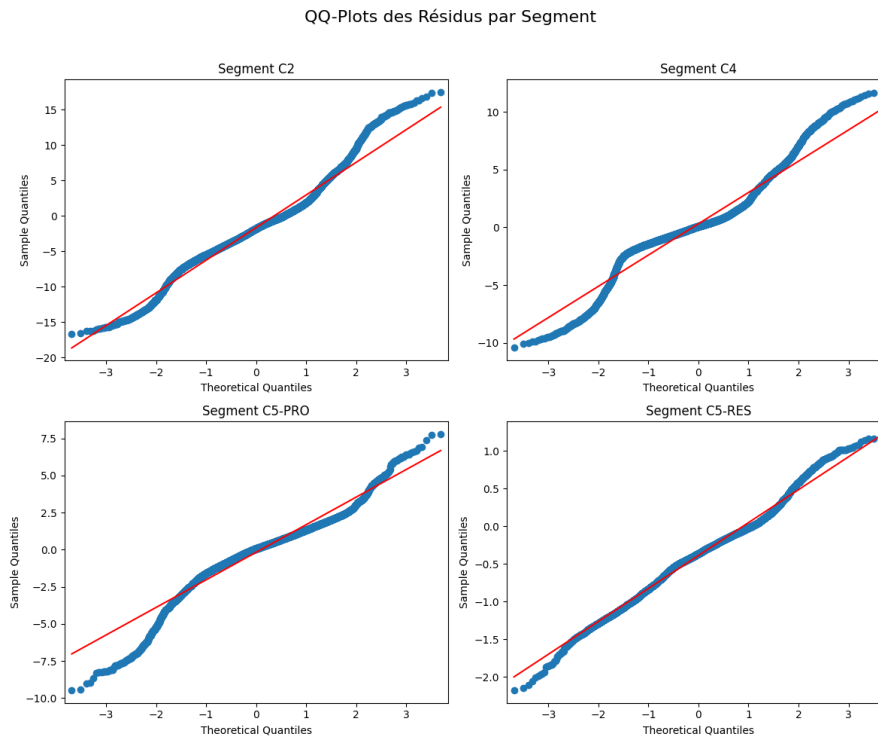
L'objectif est alors de déterminer la forme de ces écarts totaux et par segment, et leur loi de distribution pour au mieux les appréhender en commençant par comparer la distribution des écarts par une distribution normale (représentée par la ligne rouge dans le graphique suivant) :

QQ-plot des écarts totaux du portefeuille



Sur la totalité du portefeuille, les quantiles au centre sont assez proches de la ligne de normalité signifiant que la majorité des écarts (compris entre les quantiles -2 et 2) se comportent similairement à une loi

normale. Mais aux extrémités, on observe une déviation à gauche où les points sont systématiquement en-dessous de la ligne rouge, et à droite où les points sont tous au-dessus de la ligne de normalité. La distribution présente des queues épaisses qui ne sont pas correctement modélisé par une loi normale. On a le même constat à la maille des segments comme on peut le voir sur le graphique suivant :



Ce constat reste visuel. Il faut le tester statistiquement pour rejeter ou accepter la normalité des résidus. Pour ce faire, on applique le test de normalité Jarque-Bera qui est plus robuste pour les grands échantillons de données, moins sensible aux écarts minimes de distribution et une alternative au test de Shapiro-Wilk. Les résultats par segment et sur le portefeuille sont donnés dans le tableau suivant qui nous confirme le rejet de l'hypothèse de normalité à 5% pour l'ensemble des cas :

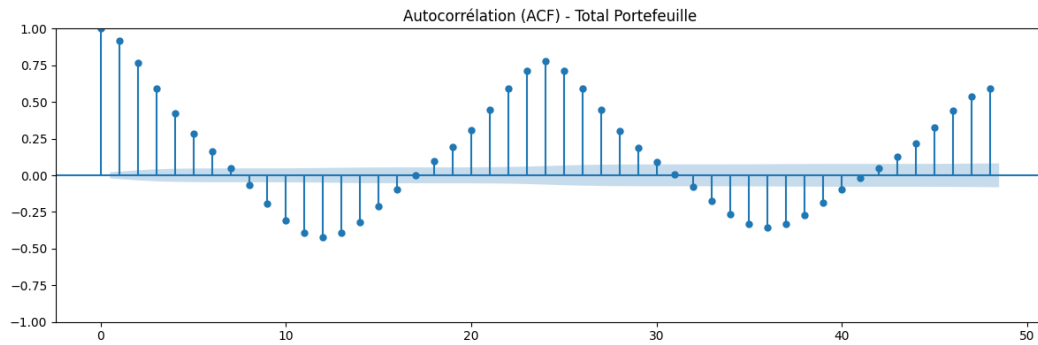
	count	mean	std	min	25%	50%	75%	max	skew	kurtosis	jarque_bera_stat	jarque_bera_pvalue
C2	8,784.00	-1.64	4.61	-16.64	-4.32	-1.76	0.49	17.50	0.42	1.72	1,346.52	0.00
C4	8,784.00	0.31	2.71	-10.44	-0.94	0.08	1.23	11.71	0.26	2.87	3,112.24	0.00
C5-PRO	8,784.00	-0.17	1.86	-9.46	-1.01	0.08	0.90	7.78	-0.90	3.10	4,689.72	0.00
C5-RES	8,784.00	-0.39	0.44	-2.17	-0.65	-0.36	-0.12	1.18	-0.04	0.70	181.02	0.00
Total	8,784.00	-1.89	8.61	-34.94	-6.16	-2.01	1.58	35.23	0.12	2.67	2,624.20	0.00

Ce test nous permet également de décrire la forme des queues de distribution (via la mesure statistique de kurtosis) qui est assez marquée pour le segment C5-PRO avec un kurtosis supérieur à 3, typique d'une distribution plus épaisse ou lourde que celle d'une loi normale ayant pour conséquence la présence plus fréquente de valeurs extrêmes. Les autres cas présentent un kurtosis inférieur à 3 mais proche de 3, signifiant des queues moins épaisses ou lourdes mais la présence tout de même de valeurs extrêmes (bien moins fréquemment pour le segment C5-RES).

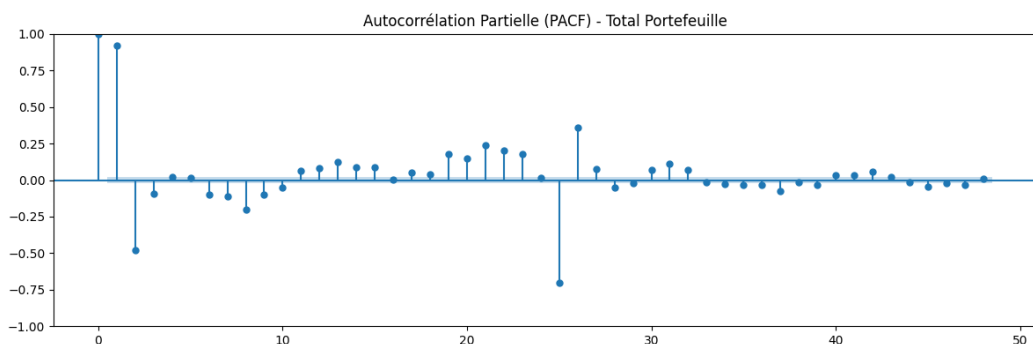
La loi normale n'étant pas la loi appropriée pour au mieux modéliser la distribution des résidus, d'autres lois ont été testées pour déterminer laquelle peut au mieux l'approcher. La loi Student et le mélange de deux lois normales semblent les plus appropriées comme détaillée en annexe 1.

Après avoir identifié, par les lois testées, la loi la plus adaptée pour représenter statiquement la distribution des résidus, il est également nécessaire d'examiner la dépendance temporelle entre les résidus. En effet, si les erreurs ne sont pas aléatoires et présente une dépendance entre elles, l'usage de modèles de série temporelle pour mieux capter la dynamique du risque que présente ces résidus est plus approprié.

Pour ce faire, on réalise les graphiques ACF et PACF ci-dessous :



Ce premier graphique présente une forme sinusoïdale claire des résidus avec des pics positifs significatifs aux lags temporels 24 et 48 révélant une saisonnalité journalière. On observe également des barres qui dépassent largement la zone de confiance en bleue (signifiant alors que les résidus ne sont pas aléatoires). Cela indique donc une forte corrélation entre les erreurs.



Ce second graphique présente des pics initiaux très significatifs sur les premiers lags qui suggèrent qu'une partie de l'erreur peut être expliquée directement par les erreurs des heures précédentes, puis un pic saisonnier significatif à 24 comme observer sur l'ACF.

Les résidus / écarts présentent une structure temporelle avec une forte dépendance aux erreurs passées et une saisonnalité journalière évidente nécessitant l'usage d'un modèle temporel pour au mieux capturer leur dynamique.

Avant de modéliser la volatilité de ces résidus, on utilise un modèle SARIMA, plus adapté pour les séries temporelles présentant une forte saisonnalité, afin de capturer l'information temporelle résiduelle pour pouvoir correctement modéliser la volatilité des données avec un modèle GARCH. Pour pouvoir appliquer ces modèles, on doit s'assurer de certaines hypothèses de validité notamment la stationnarité de la série temporelle pour le modèle SARIMA et la présence d'hétéroscédasticité dans les données pour le modèle GARCH.

La présentation des méthodes et résultats d'un modèle GARCH simple et d'un modèle GARCH avec application d'un modèle SARIMA sur les résidus en amont est faite en Annexe 2.

Les séries temporelles modélisant la consommation électrique n'ont pas de tendance. En effet, il n'y a pas de croissance ni décroissance continue dans la représentation des données, la consommation suivant plutôt un cycle saisonnier qui recommence chaque année et dépend surtout notamment des conditions météorologiques et des tensions politiques ou techniques pouvant influencer l'offre ou la demande.

C'est pour cela que nous pouvons nous baser sur l'historique de consommation de chaque client afin de pouvoir modéliser et prédire la consommation future. Le test de Dickey-Fuller nous présente des séries stationnaires ce qui nous permet d'utiliser les modèles cités.

Les données sont stationnaires sur l'ensemble du portefeuille et sur les segments C2 et C4. Cependant, elles ne sont pas stationnaires pour les segments C5 (PRO et RES) ce qui peut s'expliquer par une hausse de la consommation via une hausse du nombre d'adhérents et un hiver plus froid que la normale ce qui fait augmenter la consommation d'énergie pour le chauffage notamment.

Pour la suite de l'analyse, les modèles seront appliqués sur l'ensemble du portefeuille qui présente une série stationnaire.

Une dernière étape avant d'utiliser le modèle GARCH consiste à vérifier si les données sont homoscédastiques. Si les résidus sont hétéroscédastiques, il sera utile d'utiliser un modèle GARCH pour modéliser la variance changeante.

Pour déterminer la nature homoscédastique ou non des données, on applique le test ARCH d'Engle qui nous permet de tester l'hypothèse H_0 « La variance est constante, les données sont homoscédastiques » il n'y a donc pas de volatilité si on garde cette hypothèse. A l'inverse, si cette hypothèse est rejetée, les variances sont différentes entre elles, donc les données sont hétéroscédastiques à 5% ce qui signifie qu'à 95% de confiance, la volatilité dans les données n'est pas constante et l'usage d'un modèle GARCH est alors pertinent pour approcher et modéliser la volatilité des résidus. Dans notre cas, les variances ne sont pas constantes donc les données sont hétéroscédastiques et l'usage d'un modèle GARCH est pertinent.

Les hypothèses de base étant vérifiées, on va pouvoir modéliser les résidus avec une approche SARIMA pour laquelle la recherche des meilleurs paramètres a été automatisé pour trouver la meilleure combinaison qui minimise le score AIC qu'on détaille en Annexe 3.

Une fois ces paramètres définis, on ajuste le modèle SARIMA et on récupère les résidus pour analyser leur corrélation entre eux comme précédemment. Ils ne présentent pas, cette fois-ci, de pics significatifs. Les résidus du modèle se comportent maintenant comme un bruit blanc ce qui signifie qu'ils sont aléatoires et non corrélés dans le temps. Cela nous permettant ainsi de pouvoir ajuster un modèle GARCH pour lequel les résidus doivent être non-corrélés. On vérifie les résidus en effectuant l'ACF et PACF qu'on retrouve en Annexe 4.

Ainsi, l'application du modèle SARIMA nous a permis d'avoir des résidus proches d'un bruit blanc permettant au modèle GARCH de correctement capter les volatilités les plus importantes sur l'année.

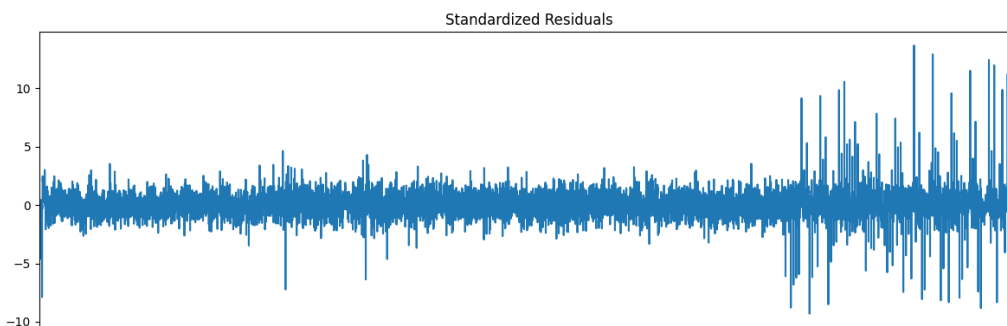
On voit d'ailleurs, dans le tableau de résultat, que le modèle SARIMA a correctement « nettoyé » les résidus de tout biais ou autocorrélation avec un μ (mu) statistiquement non-significatif traduisant que la moyenne des résidus est statistiquement nulle, condition essentielle pour appliquer le modèle GARCH.

L'ensemble des coefficients sont significatifs (avec une p-value largement inférieure à 5%). On a :

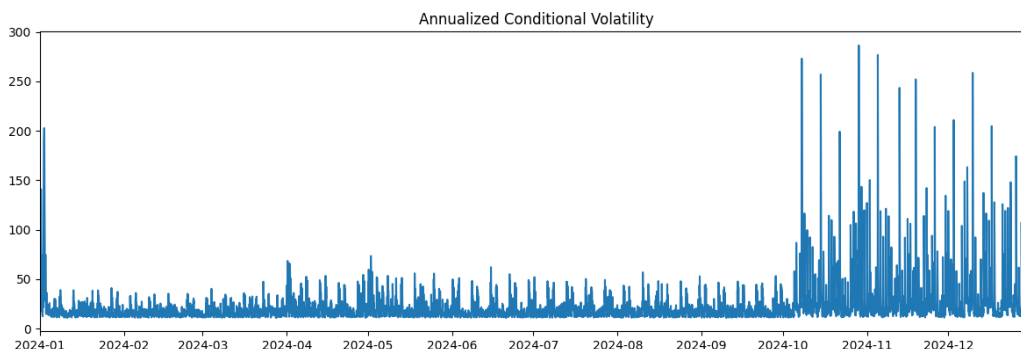
- Omega (coefficient = 0,2674) qui représente la variance de base caractérisant la volatilité à long-terme ;
- Alpha (coefficient = 0,5655) qui mesure l'impact d'un choc passé sur la volatilité d'aujourd'hui. La valeur de 0,57 indique que les chocs ont un impact fort et immédiat sur la volatilité ;
- Beta (coefficient = 0,4345) qui mesure la persistance de la volatilité. Une valeur de 0,43 montre que la volatilité est persistante c'est-à-dire qu'une journée à forte volatilité a tendance à être suivie par une autre journée de forte volatilité.

La somme des deux derniers coefficients est égale à 1 ce qui indique que les chocs de volatilité sont extrêmement persistants et mettent beaucoup de temps à se dissiper. En d'autres termes, la volatilité est très réactive et persistante dans le sens où le risque d'aujourd'hui est très fortement lié au risque d'hier.

On pourrait alors conclure que le modèle permettrait en l'état de simuler les résidus futurs mais le prochain graphique représentant les résidus standardisés montre certes des résidus pour une grande partie autour de 0 mais les erreurs à partir d'octobre 2024 ont une amplitude plus importante signifiant qu'il pourrait y avoir un changement dans la demande ou un facteur externe non modélisé et qui est important pour capter ces résidus. Cela est très probablement dû à la baisse des températures durant cette période augmentant la demande en énergie pour couvrir les besoins en chauffage.



Cependant, le modèle nous permet de détecter et quantifier les changements dans les régimes de volatilité. Il permet de déterminer les périodes à risques faibles de celle à risques plus élevés comme le montre le graphique suivant :

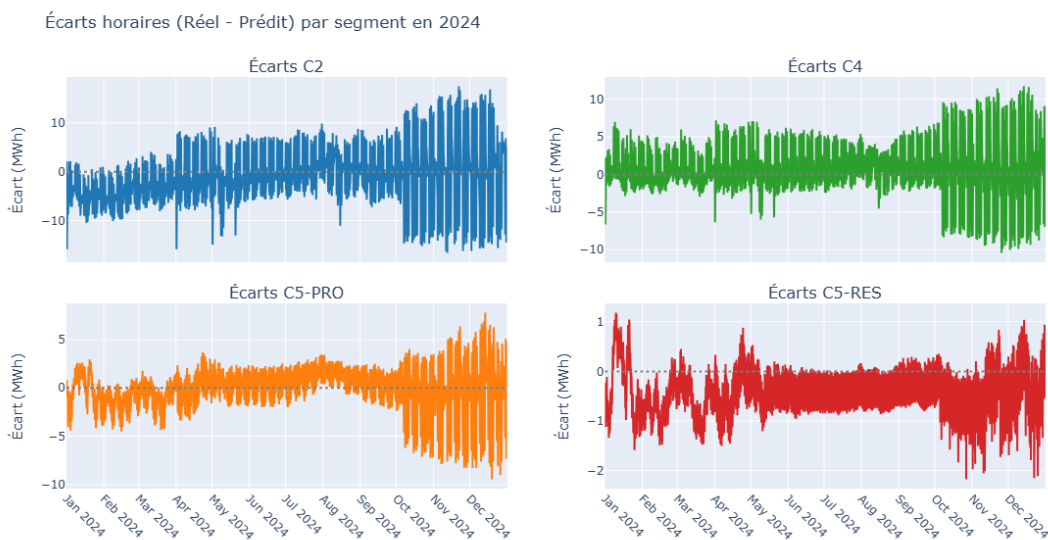


L'analyse des résidus et leur ajustement via un modèle GARCH ont permis de caractériser la volatilité intrinsèque de la série temporelle sur les écarts de consommation du portefeuille. La section suivante consiste à déterminer des groupes à risque similaire permettant de finaliser l'analyse des risques avant calculs et modélisations.

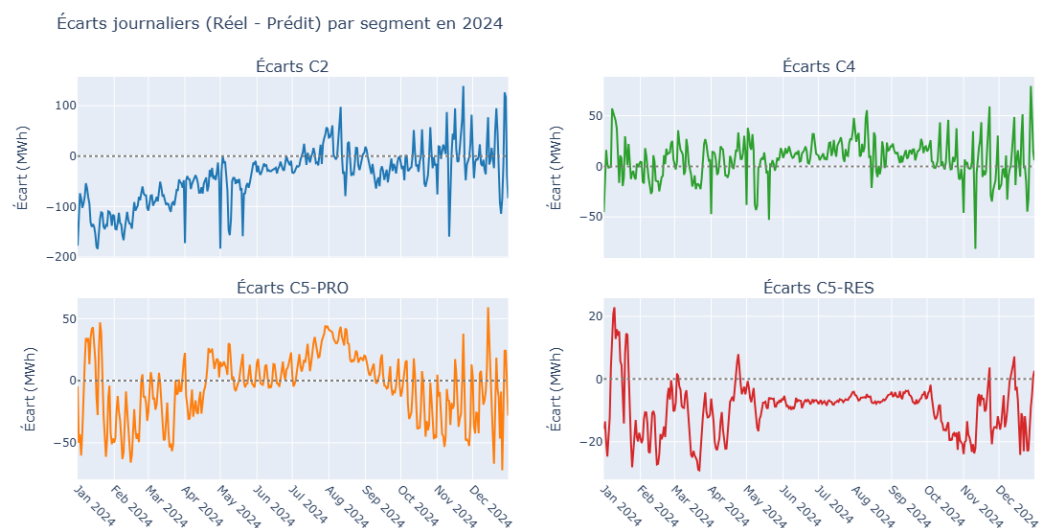
2.4. Cartographie des groupes de risque

Les facteurs de risque vont principalement, dans le cadre du risque de marché et de flexibilité, se concentrer dans la proportion de consommation, le volume d'écarts et les profils de consommation.

On a des écarts positifs ou négatifs (achat ou vente d'énergie nécessaire) qui dépendent de la saisonnalité et des profils de consommation comme on peut le voir sur le graphique suivant au pas horaire :



On voit la même quantification des écarts cumulés sur la journée mais avec l'effet de foisonnement qui permet d'éliminer les écarts qui se compense mutuellement et ne présente que les écarts à effectivement couvrir sur une journée :



Les écarts par segment permettent de voir d'une manière globale le risque mais l'analyse des données précédentes a montré des écarts extrêmes minimaux et maximaux suggérant des valeurs extrêmes et des individus comportant un risque plus important avec un profil de consommation prédit qui diffère fortement du profil réel ou des courbes de charge à comportement différent au sein des segments. Il est donc intéressant de constituer des sous-groupes en fonction de la consommation totale, des écarts et de la forme de la courbe de charge au sein des segments via la méthode de clustering avec K-means qui est détaillée en Annexe 5.

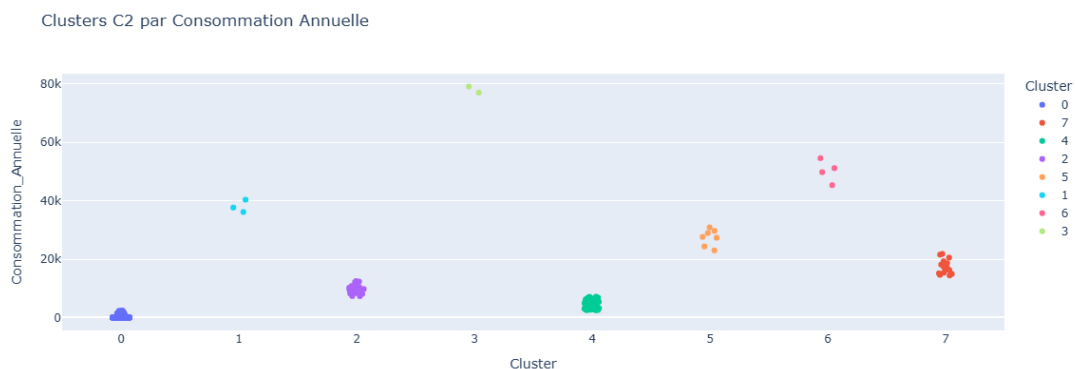
A noter qu'il n'est possible d'effectuer un clustering uniquement pour les segments de clients C2 et C4 (représentant près de 3/4 du portefeuille en termes de consommation) pour lesquels nous pouvons avoir les données de consommation au pas horaire pour chaque client composant le segment.

2.4.1. Clustering

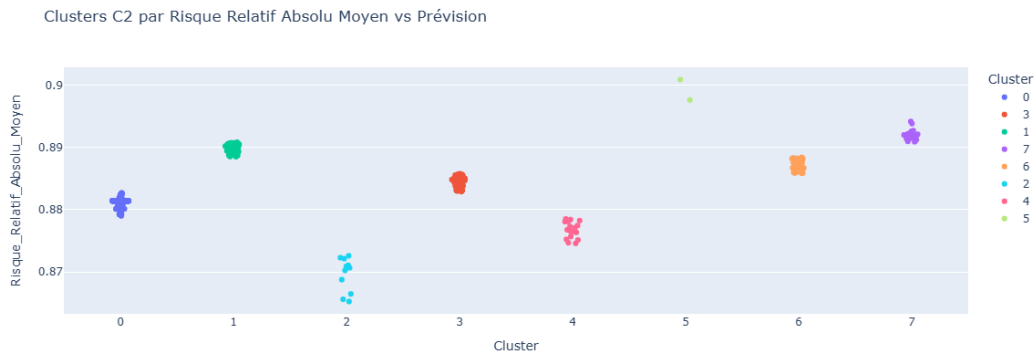
Avant de constituer les clusters, on détermine le nombre de groupe en recherchant le k optimal (nombre de groupe) via la méthode du coude pour chacun des deux segments. On applique ensuite ce nombre pour construire les clusters en fonction des critères d'attribution à chaque groupe : en fonction de la consommation totale (8 et 9 clusters pour les segments C2 et C4 respectivement), en fonction du risque relatif (même nombre de clusters que précédemment) et en fonction des profils de consommation (avec 12 et 20 clusters pour les segments C2 et C4 respectivement).

La décomposition en sous-groupe présente un premier groupe avec une grande partie des individus dans le premier cluster (plus de 50% voire 70% du nombre total d'individus dans le segment concerné en fonction des conditions utilisées) puis des plus petits groupes en fonction des segments (le détail de la composition de chaque cluster est en Annexe 6).

Visuellement, nous avons les clusters répartis, en fonction des conditions, de la façon suivante (nous présentons dans cette section seulement les résultats du segment C2 mais les résultats du segment C4 seront également présentés en Annexe 7) :



On a une répartition bien nette de chaque cluster et des individus rattachés en fonction de leur consommation annuelle avec des groupes présentant de faible consommation et d'autres, une consommation totale plus élevée.



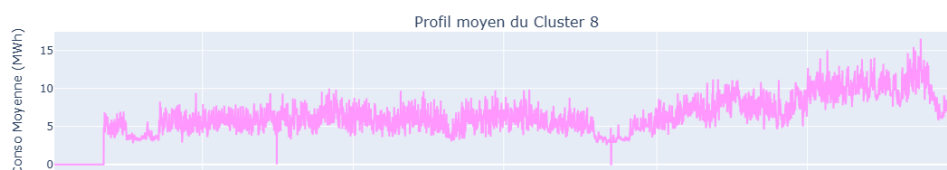
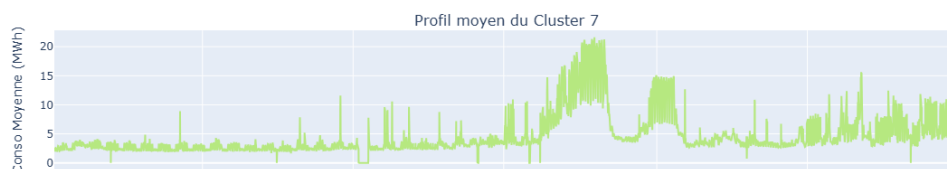
On a le même constat pour les risques relatifs moyens avec des groupes à risque plus faible d'écart que d'autres.

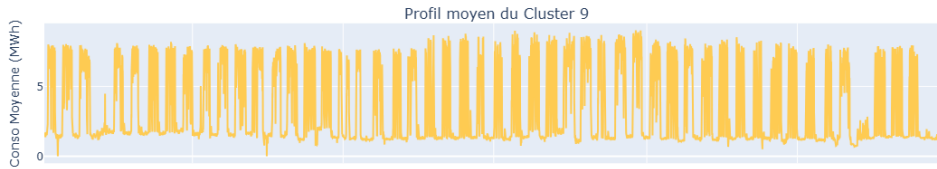
La méthode du risque relatif absolu moyen s'appuie sur le calcul, pour chaque PDL, de l'erreur de prévision moyenne en pourcentage par rapport à sa consommation réelle. Techniquement, elle se calcule comme la méthode de MAPE (Mean Absolute Percentage Error) de la façon suivante :

- On va calculer l'écart horaire entre la consommation réelle et la consommation prédite auquel on va prendre sa valeur absolue pour ignorer si l'erreur est positive ou négative et garder la somme totale des écarts ;
- On normalise cet écart absolu en le divisant par la consommation réelle de la même heure pour obtenir alors une erreur en pourcentage pour chaque heure ;
- Enfin, on calcule la moyenne de toutes ces erreurs sur l'ensemble de l'année permettant d'avoir un score unique par PDL représentant son imprévisibilité relative.

Cette méthode nous permet de normaliser le risque par rapport à la taille du client c'est-à-dire que l'on va purement regarder le pourcentage d'écart et non la corrélation de l'écart par rapport au volume de consommation nous permettant, par exemple, de considérer un faible consommateur avec une erreur de prévision de 50% comme plus risqué qu'un gros consommateur avec une erreur de 5%.

Enfin, la dernière méthode de clustering permet de segmenter en fonction de profil moyen de consommation. Un exemple de trois clusters sur les douze (les résultats sont visibles dans leur intégralité en Annexe 8) nous montre bien la diversité des profils de consommation moyen au sein du segment C2 avec des pics ou creux de consommation différents, des consommations continues indépendamment des variations saisonnières et des consommations sensibles aux fluctuations météorologiques.

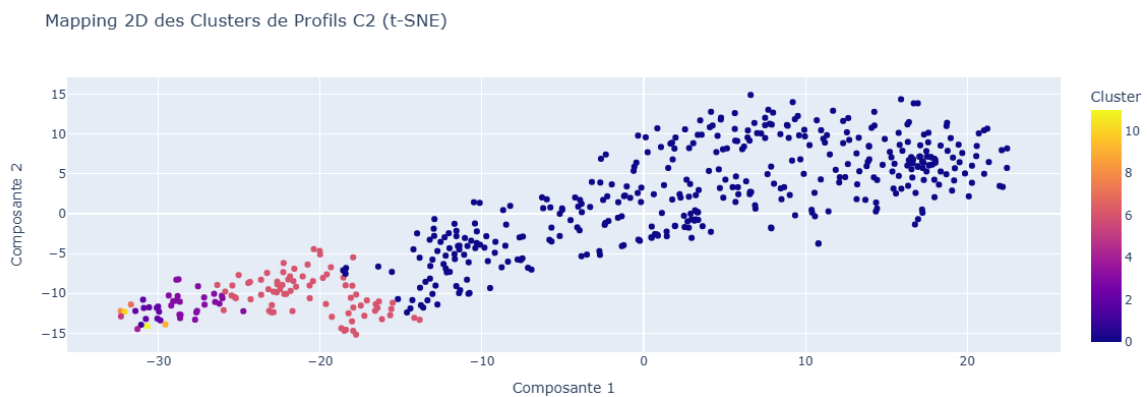




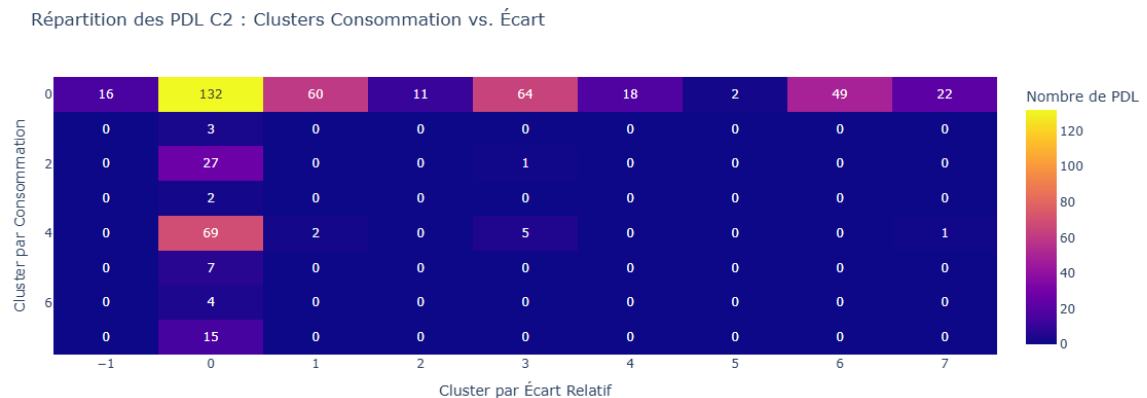
Le prochain graphique nous permet de mieux visualiser la répartition de ces clusters entre eux en réduisant la dimensionnalité des données (8760 heures pour chaque PDL où on arrive très rapidement à un nombre de dimension trop important pour pouvoir visualiser les individus et leur cluster d'attribution).

On utilise pour ce faire la méthode t-SNE (t-Distributed Stochastic Neighbor Embedding) qui permet d'aplatir les données en préservant les relations de voisinage. En effet, si deux PDL ont des profils de consommation très similaires dans l'espace à 8760 dimensions, cette méthode fera en sorte qu'ils soient très proches l'un de l'autre sur la carte en deux dimensions.

On voit donc sur le graphique ci-dessous la répartition des PDL dans chaque cluster ce qui permet de juger visuellement de leur bonne assignation. On a des groupes plutôt distincts et cohérents où on observe un « super-cluster » en bleu foncé représentant une grande famille de profils de consommation similaire.



Il est également intéressant de faire une analyse croisée des clusters avec un même nombre de groupe et une base de variable commune (consommation totale vs risque relatif absolu moyen) pour observer des similitudes entre certains clusters et potentiellement subdivisé en sous-groupe pour analyser plus finement les risques :



Par exemple, il y a 132 points de livraison (PDL) qui appartiennent simultanément au cluster de consommation n°0 (axe vertical) et au cluster d'écart relatif n°0 (axe horizontal). En d'autres termes, c'est le nombre de clients qui se trouvent à l'intersection de ces deux groupes spécifiques : ils ont une consommation totale et un risque relatif similaires.

La définition de ces clusters nous permet ensuite de répartir les risques en fonction de ces sous-groupes.

2.4.2. Répartition des risques

Pour la répartition des risques, on garde la définition du risque relatif absolu moyen permettant de représenter le risque de sous-estimation et de surestimation des segments C2 et C4.

On la visualise sous forme de distribution violon sur le graphique ci-dessous :



Ce graphique présente deux violons de forme très différente entre les deux segments. En effet, le segment C2 présente un violon très large comprenant plusieurs bosses sur les côtés signifiant que le segment n'est pas homogène en termes de risque et confirme l'approche précédente de la présence de plusieurs sous-groupes de clients avec des profils de risque distincts. Le segment C4 a une forme bien plus fine au centre mais a des queues beaucoup plus longues. Une grande proportion des clients de ce segment présente un risque similaire et concentré autour de 0,86 mais il existe des clients extrêmes ou outliers dont le risque est soit plus faible que la moyenne soit plus élevé.

On observe également que le cœur de la distribution du segment C2 est plus haut que celui du segment C4 suggérant, qu'en moyenne, les clients C2 sont plus risqués ou plus difficiles à prédire.

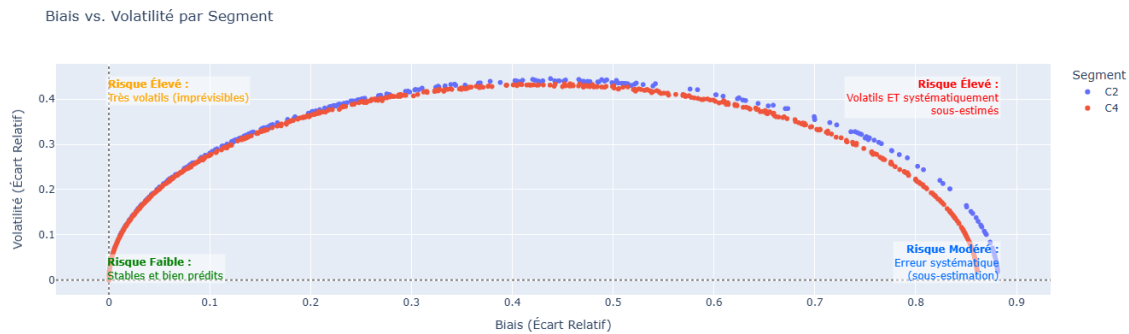
Globalement, le segment C2 est plus risqué et plus hétérogène que le segment C4 qui lui est plus homogène mais présente un risque d'outliers plus marqué. Le segment C2 nécessite d'utiliser les sous-groupes de risque pour au mieux l'approcher et la gestion des risques pour le segment C4 peut se concentrer sur l'identification et l'isolation des clients outliers.

A noter que les pourcentages de risque sont très élevés dans ce graphique (compris entre 0,79 et 0,90) car il prend en compte les écarts absolus (qui comprend les écarts négatifs et positifs de consommation).

Chaque client présente un risque d'écart plutôt concentré dans les deux segments (aux outliers près) mais ils n'ont pas tous la même volatilité de consommation. En effet, un client peut avoir un risque d'écart certain élevé mais une volatilité faible le rendant plus prévisible contrairement à un client

similaire mais avec une volatilité de consommation bien plus élevée. La prise en compte de la volatilité, en plus des écarts permet alors d'approfondir la notion de risque que chaque client porte.

Le graphique ci-dessous illustre cette dualité en positionnant les clients selon leur écart moyen (caractérisé par le biais qui est la moyenne de l'écart relatif ici) et leur volatilité de consommation ce qui permet d'identifier les profils les plus risqués :



Un biais proche de 0 signifie, qu'en moyenne, les erreurs positives et négatives s'annulent mais s'il est positif, il y a de la sous-estimation constante plus ou moins importante dans la consommation du client. Pour la volatilité, plus elle est élevée, plus l'erreur est imprévisible et instable rendant le client difficile à anticiper.

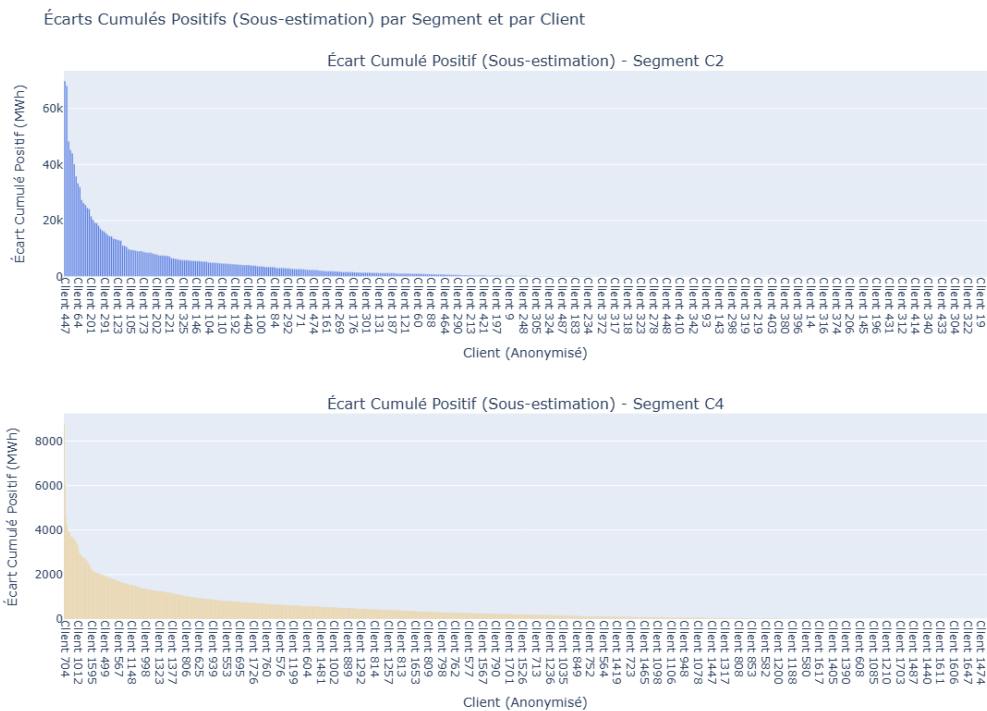
On a une vision bidimensionnelle nous permettant de déterminer 4 catégories de risque :

- Risque faible (en bas, à gauche) avec un biais et une volatilité faible qui correspond aux clients prédictibles avec un comportement stable
- Risque modéré (en bas, à droite) avec un biais élevé mais une volatilité faible correspondant aux clients mal prédits mais qui ne présentent pas de volatilité propre à son profil (les écarts peuvent être dus à une hausse de la production pérenne ou saisonnière, à des actions pour réduire sa consommation, aux renouvellements de matériels moins énergivores, ...). On a une erreur grande mais constante qui peut potentiellement être corrigée.
- Risque élevé (en haut, à gauche) avec un biais faible mais une volatilité élevée où les clients ont une consommation moyenne proche des prévisions mais un profil horaire imprévisible pour lesquels on ne peut pas anticiper leurs pics ou creux de consommation dans la journée.
- Risque très élevé (en haut, à droite) avec un biais élevé et une volatilité élevée qui va représenter les clients les plus risqués du portefeuille présentent un comportement erratique et une difficulté à prédire systématiquement leur consommation.

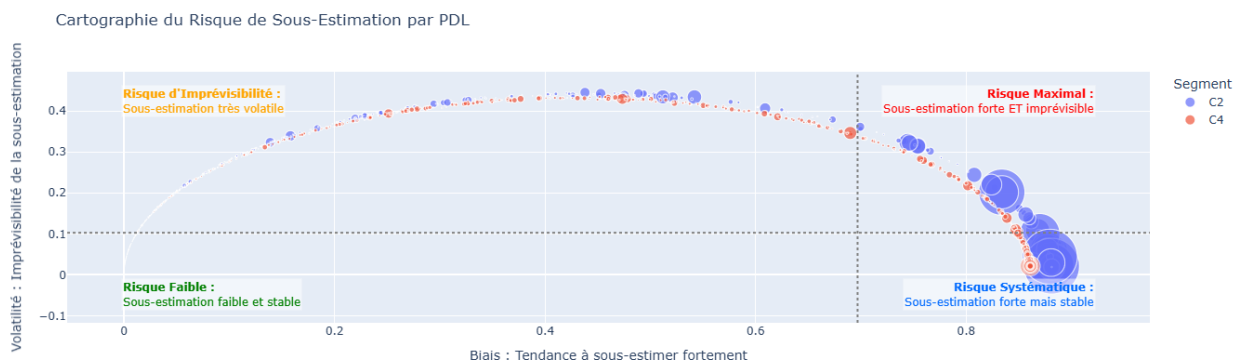
Cette démarche permet de visualiser les écarts relatifs en fonction de la volatilité, tout écart confondu. Or, le fournisseur subi des pertes lors de la sous-estimation des écarts, donc lorsque l'écart entre la consommation réelle et la consommation prédite est positif. De ce fait, il est pertinent d'adapter cette analyse uniquement sur les écarts positifs permettant d'évaluer en plus de la volatilité propre à chaque client, leur impact potentiel sur le portefeuille global.

L'idée est donc de savoir si les écarts cumulés positifs par client, en fonction de leur volatilité de prédiction au pas horaire, a un impact plus ou moins important par rapport aux autres clients du segment dans le portefeuille global. En effet, on peut visualiser des écarts cumulés positifs plus important pour

certain clients par segment comme on peut le voir sur les graphiques suivant mais cela ne signifie pas forcément que ce seront effectivement les clients les plus à risque du portefeuille :

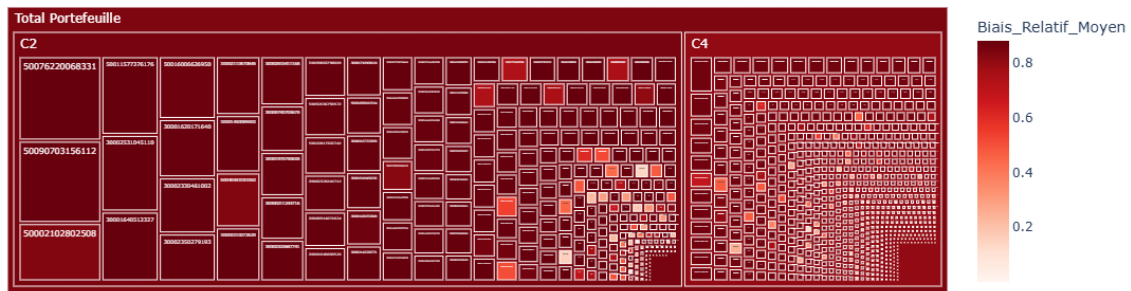


On reprojette les clients de chaque segment (C2 et C4) en fonction de leur biais et de leur volatilité de sous-estimation, en ajoutant leur impact dans le portefeuille :



On voit donc que les clients à plus fort impact dans le portefeuille ont tendance à avoir un risque modéré dit systématique où la sous-estimation est forte mais stable c'est-à-dire que ces clients sont effectivement mal prédits par les modèles utilisés pour la prévision mais il ne présente pas de profil instable de consommation au pas horaire.

Le but est de déterminer quels sont ces clients, ce que nous montre la Heatmap suivante, par segment, à l'échelle du portefeuille, où on a un risque plus important avec les clients du segment C2 que ceux du segment C4 présentant surtout des clients mal-prédits sans forte volatilité dans leur profil :



Cette analyse de risque détaillée va nous permettre de guider et d'évaluer les calculs des primes de risque afin de juger de leur vraisemblance avec cette analyse et de quantifier en €/MWh le coût de ces risques.

3. Calculs et modélisations

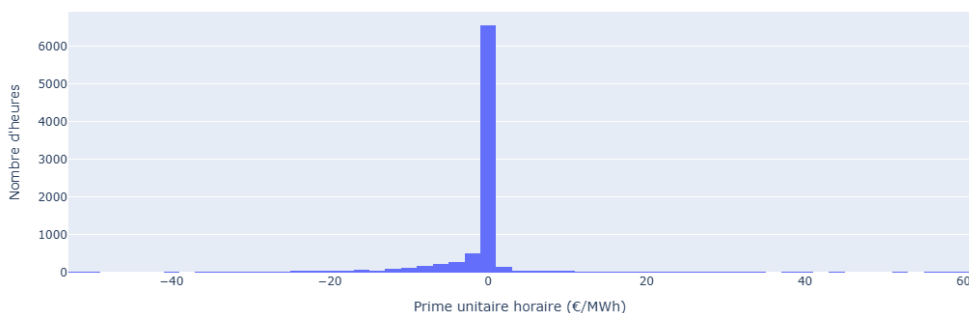
3.1. Calculs et quantification des primes de risque

3.1.1. Prime de risque de marché

La prime de risque de marché va se baser sur les écarts de la consommation positif ($C^{réelle} - C^{prévisionnelle} > 0$), c'est-à-dire qu'on va calculer la prime uniquement sur le volume exposé au marché spot. Cela correspond au volume manquant non-prévu mais qui doit être acheter à très court terme pour faire face à la demande plus élevée que les prévisions l'avaient estimée. Cet écart va donc générer des coûts pour le fournisseur qui doit le quantifier et le prévoir pour au mieux l'intégrer dans sa tarification.

On commence par calculer les primes par heure sur une année pour laquelle on dispose de toutes les données d'historique pour quantifier le risque passé qui nous servira de base pour les modélisations avancées.

Distribution de la prime unitaire horaire de risque marché (2024)



Le calcul nous présente, sur la distribution des primes unitaires, une forte proportion de prime nulle ou très proche de 0 avec une asymétrie à gauche de la distribution des primes. On a une prime unitaire moyenne de -0,99€/MWh²⁵ qui signifie qu'en moyenne le prix spot était plus avantageux que le prix à terme et a permis une économie moyenne de 0,99€ par MWh acheté en plus pour assurer l'équilibre et

²⁵ La prime unitaire a été calculé en divisant la prime par le volume totale de consommation du portefeuille (et non juste le volume exposé) pour obtenir une prime en €/MWh. C'est un choix qui permet de lisser la prime sur l'ensemble de la consommation.

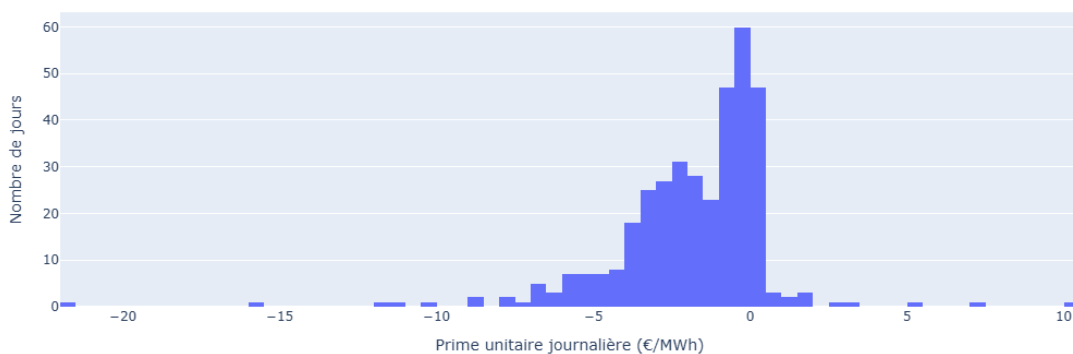
comblent le manque d'énergie sollicitée par la demande. On a cependant une dispersion marquée des primes due à la volatilité des prix spot avec un écart-type de la prime unitaire à 5,58€/MWh où on a une prime minimale à -51,35€/MWh et une prime maximale à 60,09€/MWh. Cela correspond à des pics et creux importants dans les prix spot par rapport au prix à terme.

De plus, on a au moins 75% des primes calculées qui sont nulles ce qui s'explique par un nombre d'heure conséquent non-exposées au spot, ou négatives où l'achat du manquant d'énergie sur le marché spot était plus avantageux par rapport au coût d'achat sur le marché à terme. On a un intervalle de confiance à 90% comprenant les valeurs suivantes : [-11,58 ; 0,87] €/MWh où il y a 5% d'heures avec une prime qui dépasse 0,87€/MWh caractérisant un coût pour le fournisseur où l'achat d'énergie à court terme était plus coûteux que l'achat à long terme.

Au pas horaire, le calcul de la prime permet de détecter les pics de risque et les heures critiques mais il est intéressant d'avoir une vision à la journée où la demande est fluctuante en fonction des jours.

On retrouve cette même logique avec le calcul des primes au pas journalier à la différence que les primes sont moins concentrées en 0 et où la temporalité journalière a permis de lisser les fluctuations horaires :

Distribution de la prime unitaire journalière de risque marché (2024)



En effet, la prime de risque moyenne est de -1,77€/MWh, plus « faible » au pas horaire. En moyenne, chaque MWh de consommation achetée sur le marché spot a permis un gain de 1,77€ par rapport au prix à terme. On a d'ailleurs une variabilité moins élevée mais toujours conséquente avec un écart-type à 2,58€/MWh et des valeurs extrêmes moins élevées avec une prime minimale à -21,55€/MWh et une prime maximale de 10,19€/MWh²⁶. On a un intervalle de confiance à 90% moins éparse : [-5,82 ; 0,24]€/MWh où 90% du temps les primes sont comprises dans cet intervalle. Les 5% des pires jours ont une prime positive (donc un coût pour le fournisseur) supérieure à 0,24€/MWh.

On effectue les mêmes calculs mais pour avoir une prime annuelle par segment qu'on compare au portefeuille en Annexe 9.

La prime négative s'explique du fait des prix Future (CAL) encore élevés fin 2023 (pour CAL Base Year 2024) avec des prix spot en baisse courant 2024 et moins élevés en moyenne que le prix CAL 2024 : Prix spot moyen 2024 à 57,71€/MWh contre CAL 2024 à 81,27€/MWh.

²⁶ Le jour où les prix spot ont été les plus élevés a coûté 10,19 €/MWh acheté sur le marché et le jour où le prix spot était le plus faible a rapporté, par rapport au prix à terme, 21,55 €/MWh.

Cette baisse des prix spot, en plus de la conjoncture économique, a d'autres causes telles que l'augmentation de la capacité globale de production en France mais aussi le doublement du nombre d'heure en prix négatifs.

De ce fait, recourir au marché pour compenser les écarts était moins coûteux globalement que l'achat de bloc de base en amont. Donc, la prime de risque est négative (marge potentielle à soustraire de la tarification du client en 2024).

Ce phénomène est une "décompensation" de la crise énergétique de 2021-2022 où le marché n'est plus en tension, les prix sont à la baisse mais ils n'ont pas retrouvé leur niveau d'avant crise. Les prix spot gardent d'ailleurs un certain niveau de volatilité encore élevé.

La volatilité estimée par RTE en 2024 est de 66% contre 56% en 2023 et une moyenne de 30% entre 2004 et 2023, ce qui reste encore élevée. On peut supposer que la volatilité du marché spot ne reviendra pas à un niveau avant crise (28% en 2019) notamment avec la hausse de la part de production d'électricité d'origine renouvelable comme l'éolien et le solaire qui viennent cannibaliser le marché avec une augmentation du nombre d'heure à prix spot négatifs et rend les prix spot plus volatils sur l'année. Ceci se traduit, en termes de risque, à des primes plus dispersées à l'échelle d'un portefeuille ou d'un segment mais aussi à une granularité temporelle différente comme on le voit dans le tableau suivant qui représente les primes unitaires moyennes par trimestre et par segment :

	T1	T2	T3	T4	Annuel
C2	-0,30	-2,21	-2,68	-0,33	-1,33
C4	-3,16	-6,08	-5,42	-0,39	-3,59
C5-PRO	-1,27	-4,69	-6,24	-0,06	-2,75
C5-RES	-1,14	-0,64	-0,36	-0,35	-0,69
Portefeuille	-1,34	-3,56	-3,92	-0,30	-2,16

Tableau 1 : Primes unitaires moyennes du risque de marché en €/MWh

On pourrait donc, de ce fait, chercher à définir des indicateurs de risque plus précis avec notamment le calcul de la VaR et CVaR qui permet d'estimer la perte maximale probable, mais le risque de marché en l'état n'étant pas caractérisé pour l'année 2024, les indicateurs ne sont pas interprétables dans le sens correct de leur usage. En effet, avec les résultats obtenus, on a une probabilité de 5 % que la variation journalière de la prime de risque soit inférieure à -5,82 €/MWh (VaR), et une moyenne des 5 % pires scénarios de -8,92 €/MWh (CVaR). Le signe négatif traduit non pas une perte mais un gain attendu pour le client, ce qui inverse l'interprétation habituelle de la VaR et de la CVaR. Ces indicateurs restent pertinents, mais leur lecture doit être adaptée au contexte : en 2024, le recours au marché spot était en moyenne²⁷ avantageux par rapport à l'achat à terme.

On effectue les mêmes calculs pour la prime de risque Flex dans la section suivante avec la formule adaptée au risque de Flex.

²⁷ Il est important de préciser en moyenne car, heure par heure, notamment lors des pics de consommation en période hivernale, le risque est bien présent avec un prix spot supérieur au prix à terme provoquant un surcout en cas d'écart à équilibrer via l'achat sur le marché spot.

3.1.2. Prime de risque de Flex

Le calcul de cette prime ne se fera que sur les segments C2 et C4 pour lesquels nous avons à disposition des données de prix de contrat fixe et à jour pour l'ensemble des PDL. Pour le segment de clients C5, la tarification est variable pour chaque client en fonction de la catégorisation « PRO » ou « RES » et des prix Heures Creuses et Heures Pleines avec une différenciation tarifaire en saison Haute (du 1^{er} novembre au 31 mars) et en saison Basse (du 1^{er} avril au 31 octobre). N'ayant pas le détail de la consommation par PDL, nous ne pouvons pas reconstituer via le calcul la tarification propre à chaque client et il n'y a pas de fichier existant à ce jour comprenant un listing des montants des factures payés par chaque client du segment C5.

Pour le calcul de la prime de risque de Flex, on va considérer l'ensemble des écarts de consommation qu'ils soient positifs ou négatifs où les pertes pour le fournisseur se caractérise dans deux cas :

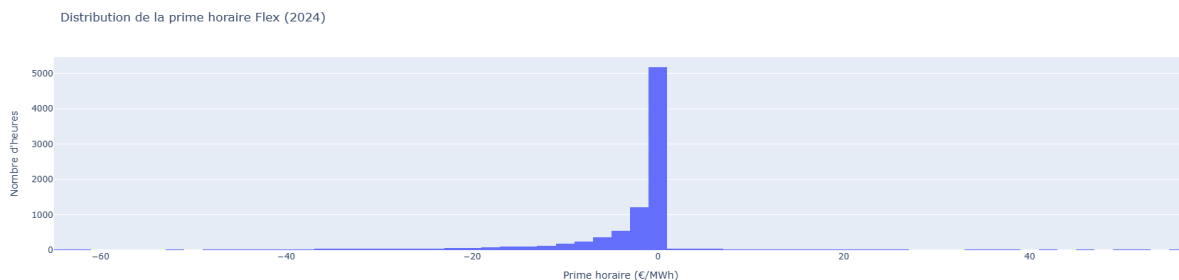
- $C_{réelle} - C_{prévisionnelle} > 0$ et $P_{spot} > P_{vente}$: il y a un manque d'énergie pour répondre à la consommation effective par rapport aux prévisions qui doit être acheter sur le marché spot à un prix supérieur au prix de vente.
- $C_{réelle} - C_{prévisionnelle} < 0$ et $P_{spot} < P_{vente}$: il y a un surplus d'énergie prévue par rapport à la consommation effective qu'il faut revendre sur le marché spot pour le respect de l'équilibre à un coût inférieur au prix de vente payé par le client

Ces deux scénarios génèrent des coûts pour le fournisseur et vont se caractériser par une prime positive.

Cependant, le surplus d'énergie à vendre ne constitue pas une perte réelle, car cette énergie n'a pas été achetée à un coût supplémentaire pour le client. En réalité, elle aurait été vendue au prix spot, ce qui relève plutôt d'une perte d'opportunité que d'une perte effective. La notion de perte effective s'applique plutôt dans le cas où un manque d'énergie doit être comblé via des achats à court terme au prix spot, générant un surcoût réel. Néanmoins, le coût d'opportunité peut devenir un coût effectif par rapport au prix à terme.

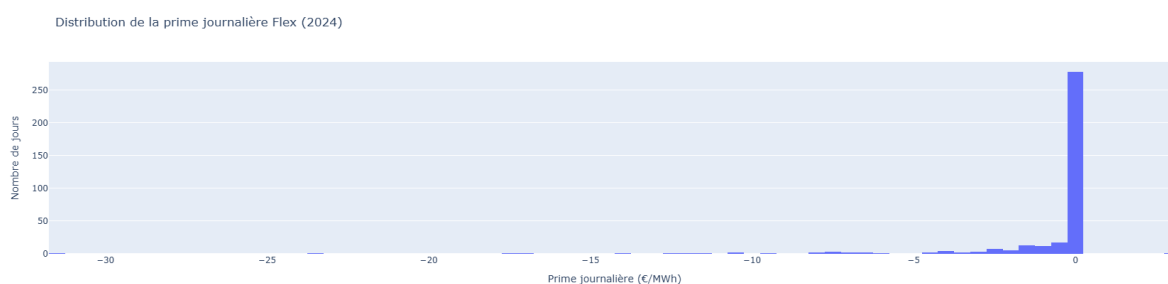
Il est pertinent de continuer l'étude sur les deux scénarios en les différenciant pour bien capter l'effet de chacun. Le nombre d'occurrence des écarts positifs et des écarts négatifs n'est pas strictement égal. On a en effet plus d'écarts négatifs (sous-consommation par rapport aux prévisions) que d'écarts positifs (surconsommation par rapport aux prévisions). Il y a 5760 écarts négatifs (65%) sur l'année 2024 et 3064 positifs (35%). Une disproportion qui pourrait influencer la valeur des primes en prenant le calcul de l'ensemble des écarts sans différenciation.

On commence par le premier scénario sur les écarts positifs avec le calcul des primes au pas horaire présenté sous forme de distribution ci-dessous :



On a une prime moyenne de $-3,14\text{€/MWh}$ qui indique que chaque MWh acheté pour couvrir la consommation est en moyenne moins cher que ce qui aurait été facturé au client. La prime négative n'est donc pas une perte réelle mais une protection de la marge dans le sens où le fournisseur achète de l'énergie à un coût inférieur au prix de vente, préservant ainsi le bénéfice. Dans 90% des cas, la prime est négative ou nulle et n'a pas impliqué de perte pour le fournisseur : $[-13,02 ; 0] \text{€/MWh}$. On a un écart-type assez élevé s'établissant à $7,28\text{€/MWh}$ et témoignant d'une forte variabilité au pas horaire des écarts et des prix spot. Cette volatilité s'observe bien dans les valeurs extrêmes où la prime minimale est de $-64,88\text{€/MWh}$ et la prime maximale est de $55,54\text{€/MWh}$.

Comme pour les primes de risque de marché, le calcul au pas journalier permet de lisser les effets horaires et de réduire la valeur des primes de risque Flex à cette échelle ce qui résulte d'une asymétrie très marquée à gauche vers les primes négatives (donc un gain pour le fournisseur s'il achète l'énergie manquante à un coût inférieur au prix de vente au client lui permettant de préserver sa marge) :



On a, en effet, une prime moyenne de $-0,96\text{€/MWh}$ avec une variabilité des primes moins importante de $3,10\text{€/MWh}$. 90% des valeurs des primes sont d'ailleurs comprises dans un intervalle à 90% moins dispersé : $[-6,58 ; 0] \text{€/MWh}$. On a des valeurs extrêmes lissées et moins conséquentes avec une prime minimale de $-31,39\text{€/MWh}$ et une prime maximale de $2,78\text{€/MWh}$.

Le calcul des indicateurs de risque VaR et CVaR ont été effectué pour le risque de Flex mais mène aussi aux mêmes conclusions que pour le risque de marché : on a une probabilité de 5 % que le gain unitaire soit supérieur à $6,58\text{€/MWh}$ (VaR = $-6,58\text{€/MWh}$), et dans les 5% des meilleurs cas, le gain moyen attendu est de $12,04\text{€/MWh}$ (CVar = $-12,04\text{€/MWh}$).

On retrouve le détail des primes annuelles par segment en Annexe 10.

On va constater une différence temporelle de valeur des primes au pas trimestriel et à l'échelle des segments dans le tableau suivant qui présente les résultats de calcul :

	T1	T2	T3	T4	Annuel
C2	-1,15	-6,46	-6,88	-6,90	-5,34
C4	-1,90	-12,91	-14,71	-13,79	-10,29
Portefeuille	-1,36	-4,91	-4,94	-4,68	-3,91

Tableau 2 : Primes unitaires moyennes du risque de Flex en €/MWh

Au pas trimestriel, chaque prime est calculée indépendamment des autres trimestres avec des prix spot propres à cette période et des volumes de consommation différents en fonction des saisons. On peut donc avoir un trimestre avec des prix très bas et des écarts élevés générant une prime unitaire fortement négative et créant une grande dispersion des valeurs des primes. A l'inverse, au pas annuel, tous les écarts positifs sont agrégés. En effet, les fluctuations trimestrielles vont se compenser partiellement : les

trimestres défavorables sont contrebalancés par d'autres plus favorables, réduisant ainsi l'écart-type et rendant la prime annuelle moyenne, en fonction des périodes et des segments moins négatives (ou moins positives dans le cas d'une année où les prix spot sont moins avantageux dans la durée et leur occurrence). Le pas annuel va donc fournir une vision plus lissée et représentative du risque global en prenant en compte les compensations possibles entre les risques.

Pour couvrir la totalité du risque Flex, on va maintenant étudier le second scénario de perte où la consommation réelle est inférieure à la consommation prévisionnelle. Dans le respect de l'équilibre constant, ce surplus doit être revendu sur le marché au prix spot qui peut être soit supérieur au prix de vente proposé au client et dans ce cas, le fournisseur est gagnant, soit inférieur au prix de vente et dans ce cas on va parler de coût d'opportunité et quantifier si ce coût d'opportunité correspond à une perte de marge uniquement ou à une perte effective si le prix spot est en plus d'être inférieur au prix de vente, inférieur au prix à terme (prix auquel le producteur achète l'énergie).

Le scénario de surplus d'énergie va être traité en trois étapes :

- 1- Le calcul du coût d'opportunité en reprenant la formule de la prime de risque Flex qu'on applique sur les écarts de consommation négatifs :

$$Flex_t = \sum_{t=1}^T \Delta V_t^- * (Prix_t^{spot} - Prix_t^{vente})$$

$$\text{Avec } \Delta V_t^- = (C_t^{réelle} - C_t^{prévue}) < 0$$

Si la prime est positive, c'est-à-dire que $(Prix_t^{spot} - Prix_t^{vente}) < 0$ (le prix spot est inférieur au prix de vente), le fournisseur fait face à un coût d'opportunité. En d'autres termes, il revend moins cher sur le marché que ce qu'il aurait pu vendre auprès du client, ce qui détériore sa marge mais n'implique pas forcément une perte effective. En effet, le prix spot peut être supérieur au prix à terme et ne pas créer de perte au fournisseur d'où la notion de coût d'opportunité. L'écart entre le prix spot et le prix de vente est désavantageux pour le fournisseur mais l'écart entre le prix spot et le prix à terme constitue toujours un gain relatif (le prix spot est supérieur au prix à terme). Cependant, si le prix spot est aussi inférieur au prix à terme alors le fournisseur enregistre une perte effective correspondant à l'écart entre le prix spot de vente et le prix à terme d'achat de l'énergie. Pour quantifier cette perte, il est nécessaire d'effectuer une seconde étape où on utilise le calcul de la prime de risque de marché sur les écarts négatifs.

- 2- Le calcul du risque de marché sur écarts négatifs comme élément permettant de différencier la perte d'opportunité de la perte effective. On l'applique sur les écarts négatifs comme suit :

$$Marché_t = \sum_{t=1}^T \Delta V_t^- * (Prix_t^{spot} - Prix_t^{terme})$$

Avec ΔV_t^- variable identique à celle de la prime $Flex_t$

On distingue deux scénarios possibles :

- ⇒ Si $Marché_t > 0$ alors $Prix_t^{spot} - Prix_t^{terme} < 0$ (le prix spot est inférieur au prix d'achat de l'énergie) donc la prime est positive et on a une perte lors de la vente du surplus

⇒ Si $Marché_t < 0$ alors $Prix_t^{spot} - Prix_t^{terme} > 0$ (le prix spot est supérieur au prix d'achat) donc la prime est négative ce qui implique un gain relatif pour le fournisseur lors de la vente du surplus d'énergie sur le marché

3- Une fois ces deux étapes réalisées, on doit distinguer les différents scénarios qui se présentent. On commence par se positionner sur la prise en compte du risque Flex sur écarts négatifs que lorsque la prime est positive. On va donc se concentrer uniquement sur le coût d'opportunité avéré. Ensuite, on utilise le calcul de la prime de risque de marché comme boussole d'identification entre coût d'opportunité pur et coût effectif. On va distinguer trois cas :

⇒ Cas n°1 : $Flex_t > Marché_t$ (la prime Flex est strictement supérieure à la prime de marché)

➔ Si $Marché_t > 0$ alors le fournisseur comptabilise une perte effective égale à $Marché_t$

➔ Si $Marché_t < 0$ alors le fournisseur n'est pas en cas de perte et enregistre une perte d'opportunité égale à $Flex_t$

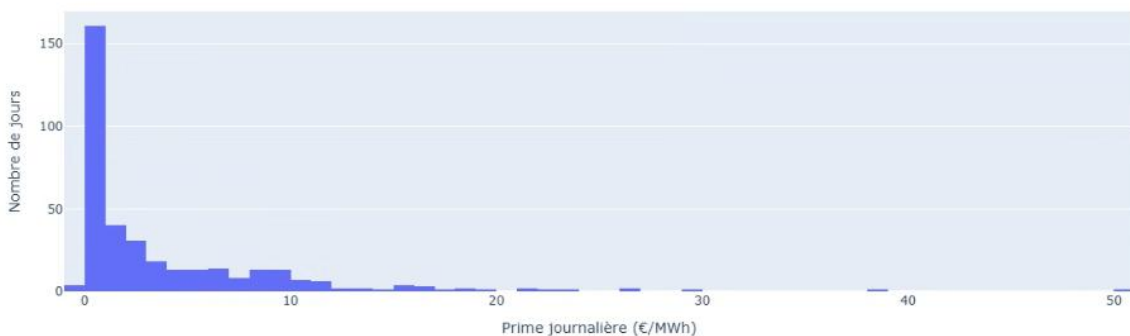
⇒ Cas n°2 : $Flex_t = Marché_t$: On est dans le cas où le prix de vente est égal au prix d'achat de l'énergie ce qui n'est pas réaliste pour une entreprise à but lucratif (sauf campagne promotionnelle de vente à prix coutant pour attirer de nouveaux clients)

⇒ Cas n°3 : $Flex_t < Marché_t$: On suppose que pour que le fournisseur soit rentable $Prix_t^{vente} > Prix_t^{terme}$ or si nous sommes dans ce cas, on a $Prix_t^{vente} < Prix_t^{terme}$ ce qui n'est pas possible dans le cadre de la supposition initiale. Donc, on part du principe que ce cas n'est pas cohérent et n'apparaîtra pas dans nos calculs. Il peut cependant arriver lors des crises et de la hausse des prix à terme plus rapide que les prix de vente. De ce fait, le fournisseur se retrouve à vendre de l'énergie au client moins cher que ce qu'elle lui a coûté créant une perte. Les effets de la crise énergétique 2022 sont encore présents mais s'estompent. On se retrouve dans un cadre d'après-crise où les prix de vente sont plus élevés que les prix à terme qui ont diminué avec la baisse des prix spot.

Dans un contexte après-crise avec une baisse des prix spot, le fournisseur a plus de chance d'être en perte en revendant le surplus d'énergie sur le marché à un prix inférieur à son coût d'achat qu'en achetant le manquant d'énergie à prix avantageux sur le marché.

En pratique, on commence par la première étape en calculant la prime de risque de Flex sur les écarts négatifs. On a la distribution des données sur les écarts négatifs sur le graphique suivant :

Distribution de la prime journalière Flex (2024)



Au pas journalier, on a une prime moyenne unitaire de 3,82€/MWh avec des primes comprises dans l'intervalle de confiance à 90% suivant : [1,42 ; 5,94] €/MWh. Au pas horaire, on a une prime moyenne de 7,18€/MWh. Une prime quasiment deux fois plus élevée au pas horaire témoignant de la forte volatilité des écarts négatifs mais également de l'important effet de compensation entre les primes.

On a dans le tableau ci-dessous les primes par segment et portefeuille permettant de constater que les primes sont bien plus élevées pour le segment C2 que le segment C4. Les primes sont positives c'est-à-dire que chaque MWh vendu a coûté économiquement 3,82€ par MWh en moyenne.

	Prime journalière
C2	6,32
C4	1,08
Portefeuille	3,82

Tableau 3 : Primes moyennes journalières en €/MWh sur les écarts négatifs

En l'état, nous ne pouvons pas conclure que ces primes constituent réellement une perte financière. Donc on effectue la seconde étape, en calculant la prime de risque marché sur les écarts négatifs. On a les résultats des primes moyennes de marché sur écarts négatifs suivants :

	Prime journalière
C2	3,65
C4	0,60
Portefeuille	3,29

Tableau 4 : Primes de marché moyennes journalières sur écarts négatifs en €/MWh

On peut déjà supposer qu'on a un coût effectif présent sur les écarts négatifs avec des primes de marché positives. Le but maintenant est de quantifier cette potentielle perte par MWh et de différencier donc le coût d'opportunité du coût effectif comme on l'a présenté dans la dernière étape de cette démarche.

On a un coût effectif (une perte réelle pour le fournisseur) totale de 1 074 909,15€. Si on rapporte à la consommation totale des segments C2 et C4, on a une prime moyenne de 3,90€/MWh. En d'autres termes, chaque MWh en surplus qui devait être revendu au prix spot a coûté 3,90€ au fournisseur. Le surplus d'énergie a été revendu à un prix spot non seulement inférieur au prix de vente proposé au client mais également au prix à terme, prix d'achat de l'énergie par le fournisseur. Les prix d'achat étant encore élevés en 2024 par rapport à un prix spot en baisse, ont l'avantage en cas de sous-consommation de payer moins cher le manque à acheter mais d'enregistrer une perte effective en cas d'énergie à revendre au spot. Le cœur de la gestion des risques dans ce contexte où, encore en 2025, nous avons des prix à terme élevés et un prix spot en baisse, s'axe sur la vente du surplus d'énergie et donc sur les modèles de prévision de la demande qui surestime la consommation réelle ce qui après crise a un coût non négligeable pour le fournisseur.

On utilise les indicateurs de risque VaR et CVaR qui sont pertinents dans ce cas :

- Il y a une probabilité de 5% que la perte unitaire Flex soit pire que 15,47 €/MWh. En d'autres termes, la prime de risque Flex par MWh ne dépassera pas 15,47€ dans 95% des cas.
- Dans les 5% des pires des cas, la perte moyenne attendue est de 22,66 €/MWh.

Il est cependant intéressant de calculer la prime du risque de Flex sur la globalité des écarts pour voir si on a une compensation des écarts négatifs et positifs sur l'année. On a une prime moyenne sur écarts positifs au pas horaire de -3,14€/MWh et sur les écarts négatifs (coût effectif seulement) de 3,90€/MWh.

On a une prime de 0€/MWh en moyenne lorsqu'on effectue les calculs sur l'ensemble des écarts. Ce résultat reste à nuancer.

Le manque d'énergie achetée (écarts positifs) ne constitue pas un coût pour le fournisseur à l'échelle du portefeuille mais un gain de marge supplémentaire sur la marge du prix de vente de l'énergie où le consommateur va payer en fonction de sa consommation en €/MWh. Le surplus d'énergie sera cependant impacté directement comme un coût pour le fournisseur par rapport au prix d'achat de l'énergie.

On a vu que la prime de risque que ce soit de marché ou de Flex n'a pas la même valeur en fonction de la temporalité choisie de calcul et de l'échelle des individus (portefeuille ou par segments). On va avoir des primes plus faibles à l'échelle du portefeuille ou annuellement dû à une compensation des risques entre eux à une échelle large grâce à la diversité des profils de consommation horaire. Ce mécanisme s'appelle l'effet de foisonnement qui sera développé dans la section suivante.

3.1.3. Effet de foisonnement

L'une des stratégies de gestion classique consiste à diversifier les clients dans un portefeuille afin d'avoir tout type de clients et de compenser les risques entre les profils divers de consommation. Cet effet de compensation aussi appelé effet de foisonnement correspond donc à la réduction de la volatilité et du coût des écarts de consommation de volume observés dans un portefeuille de clients agrégés. En effet, les profils de consommation individuels présentent des variations aléatoires, mais l'agrégation permet souvent de compenser partiellement ces fluctuations grâce à la diversification des clients au sein d'un portefeuille. Plus la maille d'agrégation est large et plus la corrélation entre les profils est faible, plus l'effet de foisonnement est marqué, diminuant ainsi le risque global du portefeuille.

On observe cet effet dans les tableaux des primes trimestrielles et annuelles de chaque risque où en fonction des trimestres, on a des primes plus élevées ou plus faibles, positives ou négatives, puis une agrégation du coût des risques à l'échelle annuelle. On a également cette disparité des valeurs de prime entre segments avec des segments qui vont contribuer négativement ou positivement au risque, c'est-à-dire qu'ils vont accroître le coût du risque ou le diminuer respectivement.

Afin de formaliser cet effet, on commence par définir les composants entraînant cet effet.

Soit un portefeuille composé de N clients, chacun présentant un écart de consommation horaire $\Delta V_{i,t} = C_{i,t}^{réelle} - C_{i,t}^{prévision}$. L'écart agrégé à l'instant t s'écrit :

$$\Delta V_t^{ptf} = \sum_{i=1}^N \Delta V_{i,t} = \sum_{i=1}^N (C_{i,t}^{réelle} - C_{i,t}^{prévision})$$

La variance agrégée du portefeuille est donnée par :

$$Var(\Delta V_t^{ptf}) = \sum_{i=1}^N Var(\Delta V_{i,t}) + 2 \sum_{i \neq j} Cov(\Delta V_{i,t}, \Delta V_{j,t})$$

Où $Cov(\Delta V_{i,t}, \Delta V_{j,t})$ est la covariance entre les écarts des clients i et j permettant de déterminer la corrélation entre le client i et le client j .

On a deux cas extrêmes qui s'illustrent pour l'effet de foisonnement sur la base de la covariance :

- Les clients sont parfaitement corrélés : $Cov(\Delta V_{i,t}, \Delta V_{j,t}) = Var(\Delta V_{i,t})$. Il y a donc peu d'effet de foisonnement, la variance agrégée est égale à la somme des variances individuelles.
- Les clients sont indépendants ou faiblement corrélés : $Cov(\Delta V_{i,t}, \Delta V_{j,t}) \approx 0$. La variance agrégée croît moins vite que la somme des variances individuelles, l'effet de foisonnement est maximal ce qui réduit le risque total.

La réduction relative du risque grâce au foisonnement peut être exprimée par le pourcentage de compensation mutuelle :

$$Foisonnement = \left(1 - \frac{\sigma_{ptf}}{\sum_{i=1}^N \sigma_i}\right) * 100$$

Où $\sigma_{ptf} = \sqrt{Var(\Delta V_t^{ptf})}$ et $\sigma_i = \sqrt{Var(\Delta V_{i,t})}$

On a un effet de foisonnement de 7,01% sur l'année et sur le portefeuille signifiant que le risque global est réduit de 7% par rapport à la somme des risques individuels des segments. En d'autres termes, les variations de consommation des différents segments ne sont pas parfaitement corrélées. Elles se compensent partiellement entre elles grâce à la diversification des profils de consommation.

Cependant, cet effet est limité car, lors de l'analyse par clustering, des gros groupes très similaires (c'est-à-dire, très fortement corrélés et/ou présentant des profils de consommation similaires) sont ressortis, ce qui limite les effets possibles de la diversification. Même si certains profils sont différents, le poids de ces groupes va dominer le calcul des primes de risque. De ce fait, le risque global ne va pas beaucoup diminuer par rapport à la somme des risques individuels ce qui est le cas ici mais l'effet n'est pas inexistant et a donc un impact sur la valeur globale des primes de risque.

Au pas trimestriel, on a un effet bien plus faible de 0,62% ce qui indique que le foisonnement est plus efficace sur des périodes longues qui bénéficient d'un effet de lissage temporel sur l'année et entre les saisons où la consommation va fortement varier.

En regardant plus en détail les valeurs des σ (sigmas) de chaque segment, la dispersion des écarts de consommation présente des valeurs trimestrielles bien plus élevées que celle annuelles. En effet, annuellement, on a des valeurs de l'ordre de 0,15 à 1,77 MWh, traduisant une faible dispersion des écarts et une variabilité agrégée très faible alors que trimestriellement, les valeurs sont nettement plus élevées allant de 2 390 à 5 353 MWh, présentant donc une variabilité plus forte qui signifie que le foisonnement ne compense pas les pics ou creux de consommation sur des périodes courtes. Néanmoins, les profils se compensent bien sur une longue durée (annuelle) et l'effet de foisonnement est plus efficace.

	C2	C4	C5-PRO	C5-RES	Portefeuille
Sigma trimestriel	4 565,48	5 353,58	5 112,66	2 390,63	17314,35
Sigma annuel	1,77	1,22	0,61	0,15	3,49

Tableau 5 : Valeurs des sigmas trimestriels et annuels par segment en MWh

L'effet de foisonnement va dépendre fortement :

- Du nombre de clients agrégés : plus le portefeuille est large, plus la moyenne statistique réduit les écarts extrêmes.
- De la corrélation entre profils : une forte corrélation (par exemple des clients résidentiels soumis aux mêmes contraintes climatiques) réduit l'efficacité du foisonnement.
- Du pas temporel d'agrégation : des agrégations horaires fines montrent un risque résiduel plus élevé que des agrégations journalières ou hebdomadaires, car les fluctuations intra-journalières peuvent s'annuler partiellement.

Cette analyse permet de calibrer la taille optimale des portefeuilles et de définir les segments clients ou les sous-groupes qui maximisent la réduction de risque.

Connaitre l'effet de foisonnement du portefeuille de clients est essentiel, mais il est d'autant plus important de pouvoir le quantifier financièrement. Il faut ainsi comparer la valorisation des écarts en intégrant la variance agrégée réduite par le foisonnement en reprenant la formule du calcul de prime (ici avec le risque de marché mais c'est le même raisonnement avec le risque de Flex où on modifie le prix de vente par le prix d'achat de l'énergie à terme) :

$$L^{ptf} = \sum_{t=1}^T (\Delta V_t^{ptf}) * (P_t^{spot} - P_{vente})$$

Avec la somme des pertes potentielles individuelles (indépendamment de l'effet de foisonnement). Cela permet de quantifier la réduction effective du risque apportée par le foisonnement.

Pour ce faire, on va donc comparer la volatilité de la perte portefeuille avec la somme des volatilités individuelles :

$$\sigma_{ptf}^{financier} = std(\Delta V_t^{ptf} * (P_t^{spot} - P_{vente}))$$

$$\sigma_{segment}^{financier} = \sum_s std(\Delta V_t^s * (P_t^{spot} - P_{vente}))$$

Puis, on mesure le gain relatif de l'effet de foisonnement :

$$Gain\ relatif = \left(1 - \frac{\sigma_{ptf}^{financier}}{\sigma_{segment}^{financier}} \right) * 100$$

On a pu quantifier physiquement l'effet de foisonnement en MWh, mais il est intéressant d'également le quantifier financièrement. Le gain financier de l'effet de foisonnement pour le risque de Flex est de 12,28% et de 12,94% pour le risque de marché. Cela signifie, qu'à l'échelle du portefeuille, le foisonnement permet de réduire de plus de 12% le coût financier des risques avec d'abord un effet physique de compensation des écarts qui réduisent les écarts globaux mais aussi un effet marché où si les écarts résiduels tombent dans des plages horaires moins couteuses en prix spot, le gain financier est alors amplifié d'où un pourcentage de foisonnement financier plus élevé que le physique.

Ces pourcentages correspondent à la réduction de la volatilité monétaire horaire liée aux écarts de consommation qu'on retrouve dans le tableau suivant et où on constate bien une variation plus faible à l'échelle du portefeuille :

	Risque marché	Risque Flex
Sigma portefeuille (avec foisonnement)	222,87	551,82
Sigma individuel (sans foisonnement)	255,99	629,06

Tableau 6 : Valeurs des sigmas avec et sans foisonnement en €

Les primes de risque et l'effet de foisonnement ont été calculés sur des données historiques et ne présentent les résultats que d'une année. Le but est de généraliser le calcul et la valeur des primes dans le temps en prenant en compte plus d'historique de données afin de déterminer une prime plus long terme avec d'autres méthodes plus poussées telles que les simulations présentées dans la section suivante.

3.2. Simulation des primes

Les simulations vont permettre de modéliser différents scénarios futurs en intégrant la variabilité des prix spot ou des écarts contrairement aux calculs sur données historiques qui ne vont se baser que sur ce qui a eu lieu dans le passé. Elles vont également permettre de générer des situations, sur la base d'un historique de données plus important qu'une seule année, qui ne se sont jamais produites mais qui restent possibles ce qui aide à mieux évaluer les risques et à prendre en compte les scénarios existants passés et les potentiels futurs.

Chacune des simulations ont été réalisé sur les deux types de risque étudiés dans ce mémoire.

3.2.1. Simulation Monte Carlo : méthode directe

Les simulations ont été réalisées en fonction de deux méthodes. La première méthode est directe, c'est-à-dire qu'on applique la simulation Monte Carlo sur les prix spot pour le risque de marché et sur les écarts pour le risque de Flex. La deuxième méthode se base sur la simulation des variations des variables cibles, une méthode plus robuste pour approcher la volatilité de ces variables et ne pas entièrement dépendre des données historiques.

Pour le risque de marché, on va modéliser le prix spot horaire $P_{spot,t}$ comme une série temporelle en générant un certain nombre de scénarios (10 000 dans le cas de cette section) comme suit :

$$P_{spot,t}^{(i)} = P_{spot,t} + \varepsilon_t^{(i)}, \quad \varepsilon_t^{(i)} \sim \mathcal{L}$$

Où \mathcal{L} est la loi de simulation choisie et $i = 1, \dots, N$ correspond aux simulations.

La prime de risque de marché simulée va alors s'écrire :

$$Prime_{marché,t}^{(i)} = (P_{spot,t}^{(i)} - P_{terme}) * \Delta V_t^+$$

Où ΔV_t^+ correspond aux écarts positifs entre consommation réelle et prévisionnelle.

Pour le risque de Flex, on simule l'ensemble des écarts (positifs et négatifs) de consommation ΔV_t :

$$\Delta V_t^{(i)} = \Delta V_t + \eta_t^{(i)}, \quad \eta_t^{(i)} \sim \mathcal{L}$$

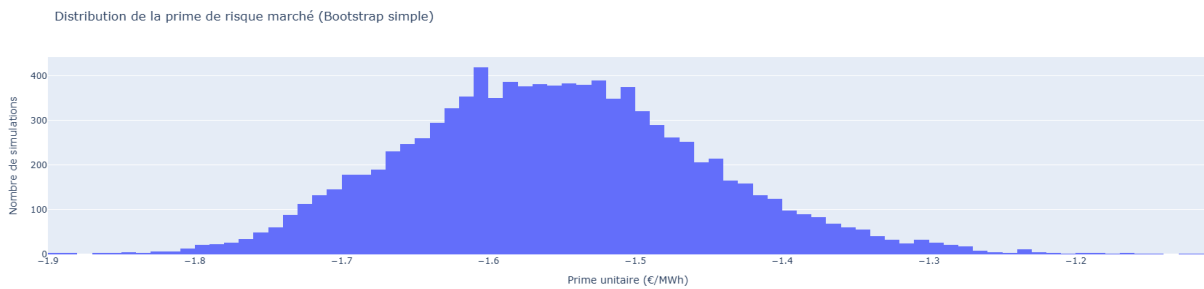
Et la prime Flex simulée s'écrit alors :

$$Prime_{flex,t}^{(i)} = (P_{spot,t} - P_{vente}) * \Delta V_t^{(i)}$$

Pour réaliser ces simulations, il faut donc maintenant choisir la loi qui suit au mieux la distribution des données en fonction des risques.

Le test des lois dans la partie 2 a permis de déterminer que la loi Student était la plus appropriée pour représenter la distribution des données de prix car elle permet de tenir compte des queues épaisses des distributions historiques, capturant mieux les événements extrêmes. Cette loi reste une approximation et pour conserver la structure empirique des données sans imposer de paramétrisation, on peut opter pour une méthode non-paramétrique, via notamment le Bootstrap²⁸ qui permet un rééchantillonnage des résidus historiques. L'approche avec le Bootstrap sera développée ici, celle avec la loi Student sera présentée en Annexe 11 (pour les deux méthodes de 3.2.1 et 3.2.2).

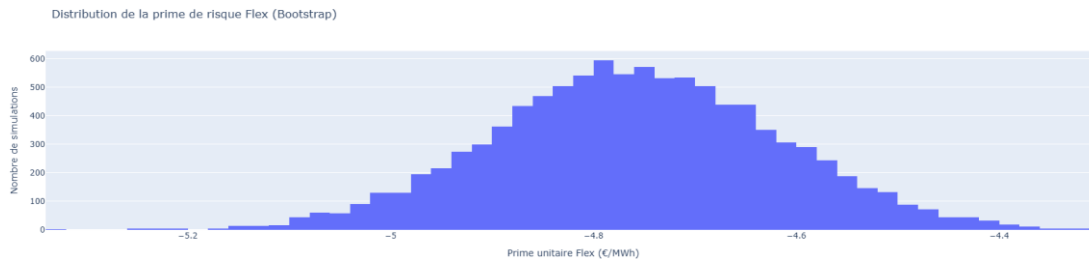
Pour le risque de marché, on a des primes qui présentent une distribution plus dispersées mais sur un intervalle plus resserré et des queues moins longues comme on peut le voir sur le graphique suivant :



On a en effet un intervalle de confiance à 90% qui comprend les primes dans $[-1,72 ; -1,38]$ €/MWh avec une prime moyenne de $-1,55$ €/MWh. L'écart-type témoigne d'ailleurs d'une variabilité bien moins importante ($0,1$ €/MWh) qu'avec la loi Student ($1,86$ €/MWh).

Pour la prime du risque de Flex, on calcule la prime de risque Flex sur les écarts positifs. On a une distribution plus centrée et avec des queues moins longues en utilisant l'approche Bootstrap représentée sur le schéma ci-dessous :

²⁸ Cette méthode permet de simuler de nombreux échantillons avec remise à partir des données historiques disponibles. Pour chaque échantillon, on calcule une statistique d'intérêt (ici la prime de risque unitaire moyenne). La distribution empirique de ces statistiques fournit une approximation de la distribution de la statistique réelle (la prime de risque réelle non-approximée). Ces statistiques vont nous permettre d'établir un intervalle de confiance à $1 - \alpha$ obtenu à partir des percentiles de la distribution Bootstrap.



On a une prime unitaire moyenne à peu près similaire de $-4,76\text{€/MWh}$ mais une variabilité plus stable, par rapport à la simulation avec la loi Student, avec un écart-type de $13,63\text{€/MWh}$ et des primes comprises dans un intervalle de confiance à 90% suivant : $[-29,10 ; 0] \text{€/MWh}$.

Pour les risques sur écarts négatifs, on a la distribution suivante :



On a des primes positives caractéristiques de perte (de perte d'opportunité au sens large, mais globalement les pertes sont effectives quand on compare par rapport au risque de marché sur écarts négatifs). On a une prime de risque moyenne de $8,08\text{€/MWh}$ mais avec un écart-type (dispersion) assez marquée de $14,67\text{€/MWh}$ et un intervalle de confiance à 90% suivant $[0, 34,59] \text{€/MWh}$ témoignant de prime pouvant être bien plus élevée sur certaines périodes de l'année.

Cette méthode est simple d'application mais elle suppose l'indépendance d'une heure à l'autre alors que les heures de prix spot sont fortement auto-corrélés entre elles. Cette méthode ne va pas non plus prendre en compte la variabilité des prix et leur continuité dans le temps.

Une approche différente, toujours dans le cadre d'une marche aléatoire simple type Monte Carlo, a été testé. Au lieu d'effectuer la simulation sur les prix, elle est faite sur la variation absolue des prix d'une heure à une autre.

3.2.2. Simulation Monte Carlo : méthode des variations

Cette méthode permet d'intégrer plus facilement les changements de volatilité ou les chocs de marché en simulant les variations de prix puis en reconstruisant les prix spot à partir de ces variations.

On constate néanmoins, que dans la littérature financière, il est plus commun d'effectuer des simulations sur les rendements de prix en utilisant notamment les rendements logarithmiques. Or, dans le cas du marché spot où les prix sont fortement volatils et peuvent être négatifs, le calcul avec des rendements logarithmiques devient instable en présence de valeurs nulles ou négatives ne permettant pas, dans notre cas d'utiliser cette méthode mais plutôt d'opter pour les variations absolues.

On reprend la même logique de calcul des simulations présentée dans la précédente section à laquelle on va se baser sur la simulation des variations absolues $\Delta P_{spot,t} = P_{spot,t} - P_{spot,t-1}$ pour le prix spot et $\Delta(\Delta V_t) = \Delta V_t - \Delta V_{t-1}$ pour les écarts de consommation.

Pour le risque de marché, on aura donc la simulation suivante :

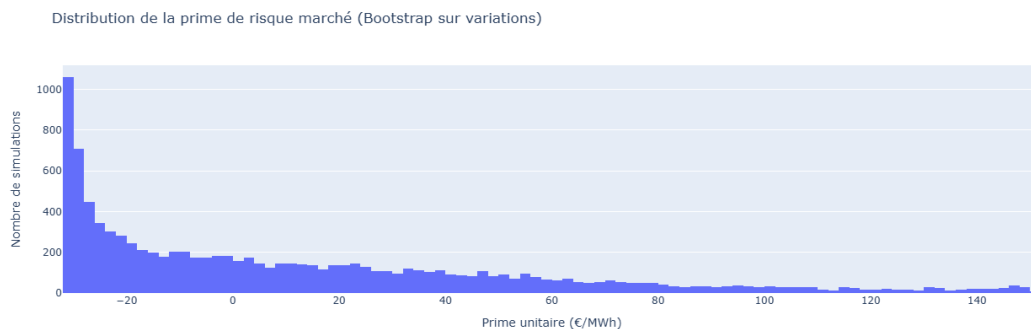
$$P_{spot,t}^{(i)} = P_{spot,t-1} + \Delta P_{spot,t}^{(i)}, \quad \Delta P_{spot,t}^{(i)} \sim \mathcal{L}$$

Et pour le risque de Flex, on aura :

$$\Delta V_t^{(i)} = \Delta V_{t-1} + \Delta(\Delta V_t)^{(i)}, \quad \Delta(\Delta V_t)^{(i)} \sim \mathcal{L}$$

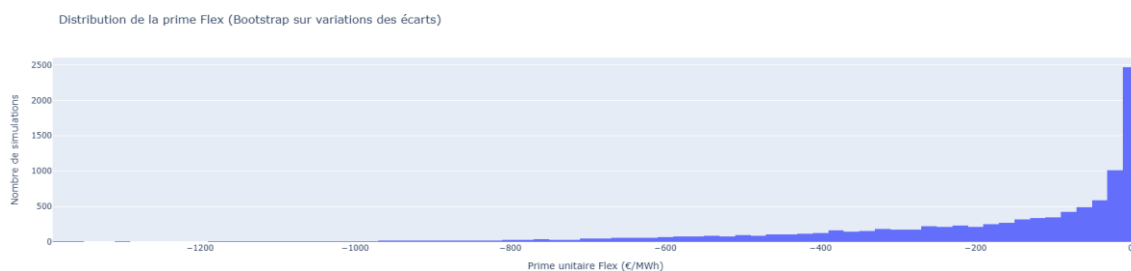
En pratique, la simulation sur variations pour le risque de Flex conserve la dynamique positive et négative c'est-à-dire qu'on effectue la simulation sur l'ensemble des écarts puis on filtre pour ne garder que les écarts positifs ou négatifs qu'au moment du calcul de la prime afin de garder la logique économique de la mesure de la perte financière pour le fournisseur.

On effectue donc la simulation avec l'approche Bootstrap en prenant en compte les variations du prix spot pour le risque de marché. On a une distribution asymétrique dérivant à l'infini à gauche rendant la simulation des primes instables comme le présente le graphique ci-dessous :



La prime moyenne est de 12,30€/MWh avec un écart-type de 43,73€/MWh élevé et un intervalle de confiance à 90% très dispersés et peu représentatif de la réalité : [-30,61 ; 104,17] €/MWh, cela malgré la pondération de chaque échantillon Bootstrap effectué pour éviter les dérives infinies.

On retrouve des résultats très volatils avec des primes extrêmes et irréalistes dans la simulation de la prime de risque de Flex sur les écarts positifs. La distribution Bootstrap présente une volatilité moins importante que la loi Student (écart-type de 365,54€/MWh contre 678,98€/MWh avec la loi Student) et on a des valeurs moins impressionnantes mais elles restent irréalistes comme le présente le graphique suivant :



On a d'ailleurs une prime moyenne très fortement négative s'établissant à -181,31€/MWh mais d'autant plus avec la simulation Student où la prime moyenne est de -2 263,82€/MWh.

On a les mêmes constats sur les écarts négatifs pour la prime de risque Flex comme on peut le voir sur la distribution ci-dessous :



Avec une prime moyenne de 166,32€/MWh et un écart-type de 355,72€/MWh, le problème de dérive infinie ne permet pas d'obtenir une simulation viable des primes avec cette seconde méthode sur les écarts négatifs.

Les méthodes Monte Carlo précédentes, qu'il s'agisse de tirages à partir d'une loi de Student ou du Bootstrap sur les variations historiques, reposent en réalité sur une marche aléatoire simple. Dans ces approches, les trajectoires de prix simulées sont obtenues par une somme cumulative des variations, ce qui signifie que le processus n'a ni mémoire ni force de rappel vers un prix moyen. Par conséquent, certaines simulations dérivent vers des prix très bas, générant des primes très positives/négatives ou une forte pondération de prime nulle comme ce qui a pu être observé dans les graphiques précédents, créant une asymétrie marquée dans la distribution des primes. Le Bootstrap, bien que non paramétrique, ne corrige pas ce problème fondamental de dérive infinie, qui n'est d'ailleurs pas réaliste pour les prix de l'énergie. Pour obtenir des simulations de prix plus réalistes et éviter ces dérives, il est nécessaire d'utiliser un modèle intégrant un retour à la moyenne, comme le processus d'Ornstein-Uhlenbeck. Ce processus va contraindre les prix à fluctuer autour d'un niveau d'équilibre ou d'une moyenne, ce qui est beaucoup plus représentatif de la réalité des marchés de l'énergie.

3.2.3. Simulation avec le processus Ornstein-Uhlenbeck

Ce processus va supposer qu'il existe un prix d'équilibre à long terme μ et que si le prix actuel s'en écarte, une « force de rappel » θ le ramène progressivement vers cette moyenne en suivant la logique suivante :

- Si les prix sont très élevés, la demande baisse et/ou une production plus coûteuse mais disponible est activée, ce qui fait baisser les prix.
- Si les prix sont très bas, la demande augmente et/ou les producteurs les moins rentables s'arrêtent, ce qui fait remonter les prix.

Il va donc intégrer le concept de l'offre et de la demande qui tend à stabiliser les prix autour d'un équilibre permettant ainsi d'empêcher les trajectoires de prix de diverger à l'infini et de produire des distributions de résultats plus stables et réalistes.

Le processus Ornstein-Uhlenbeck permet également de tester différents scénarios en faisant varier les paramètres du modèle (μ, θ, σ) initialement estimés à partir des données historiques.

On va alors appliquer aux deux types de risque ce processus à la simulation. Pour le risque de marché, on va chercher à simuler le prix spot de l'électricité en utilisant la formule discrète du processus (qu'on peut directement appliquer dans les calculs) :

$$P_{spot,t} = P_{spot,t-1} + \theta(\mu - P_{spot,t-1})\Delta t + \sigma\varepsilon_t$$

Où :

- $P_{spot,t}$ est le prix spot à l'instant t
- θ est la vitesse de retour à la moyenne (aussi appelée mean reversion)
- μ est le prix moyen à long terme (qui sert de niveau de référence)
- σ est la volatilité du processus
- ε_t est le choc aléatoire simulé via une loi Student ou un Bootstrap sur les variations historiques

La prime de risque de marché s'obtient alors comme :

$$Prime_{marché,t} = (P_{spot,t} - P_{terme}) * \max(C_t^{réelle} - C_t^{prévue}, 0)$$

On a la même logique pour le risque de Flex mais pour lequel on applique la simulation sur les écarts entre consommation réelle et prévisionnelle. Le processus va se formaliser de la façon suivante :

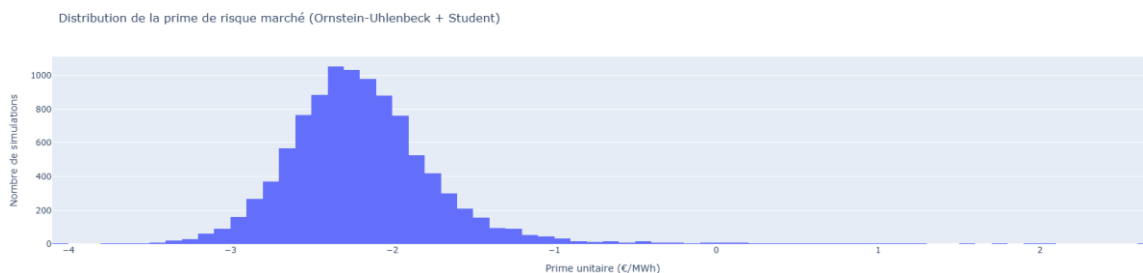
$$\Delta V_t = \Delta V_{t-1} + \theta(\mu - \Delta V_{t-1})\Delta t + \sigma\varepsilon_t$$

Avec $\Delta V_{t-1} = \max(C_t^{réelle} - C_t^{prévue}, 0)$

La prime de risque de Flex va s'écrire alors :

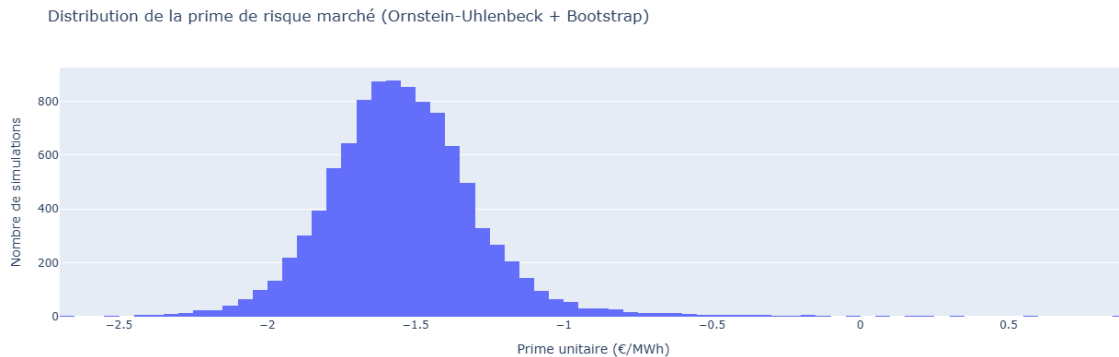
$$Prime_{marché,t} = (P_{spot,t} - P_{terme}) * \Delta V_t^+$$

Pour la prime de risque de marché, les distributions (de chaque approche) des primes sont plus stables, centrées et réalistes mais la loi Student présente encore une queue de distribution assez longue à droite :

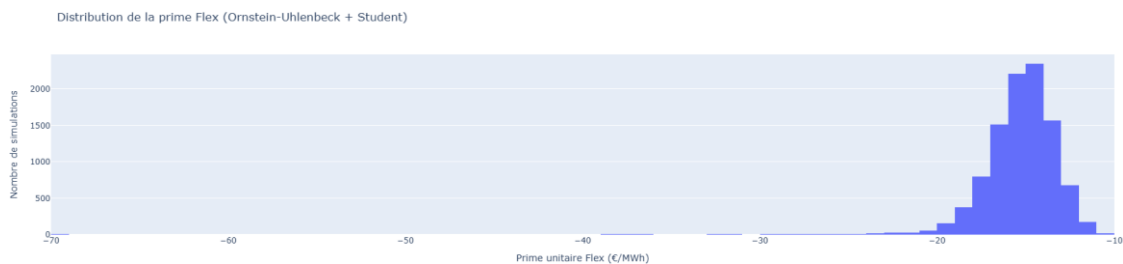


On a une prime moyenne de -2,19€/MWh avec un écart-type (ou dispersion) de 0,48€/MWh ce qui traduit une volatilité contrôlée. On a 90% des primes comprises entre -2,85€/MWh et -1,42€/MWh.

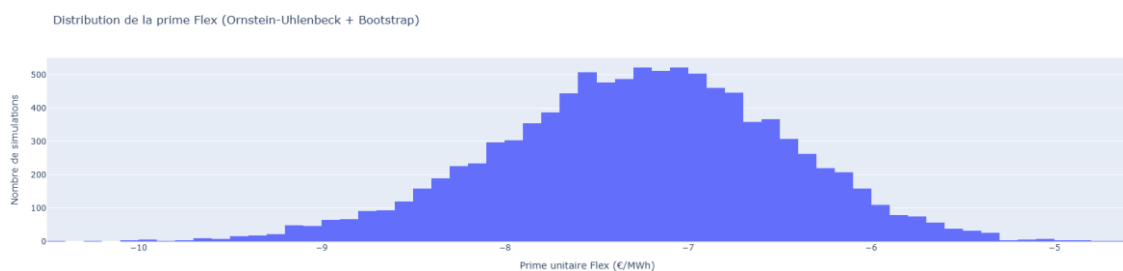
On a une volatilité plus faible avec la distribution Bootstrap où l'écart-type est de 0,26€/MWh. La prime moyenne est un peu moins faible et s'établit à -1,24€/MWh. 90% des primes sont comprises dans un intervalle plus réduit [-1,92 ; -1,12] €/MWh. On voit d'ailleurs ce resserrement des primes sur la distribution suivante :



Pour la prime de Flex sur les écarts positifs, on a une distribution encore assez asymétrique avec la loi de Student comme le montre le graphique ci-dessous mais qui présente des primes unitaires moyennes plus réalistes avec une prime moyenne de -15,29€/MWh et un écart-type de 2,01€/MWh.



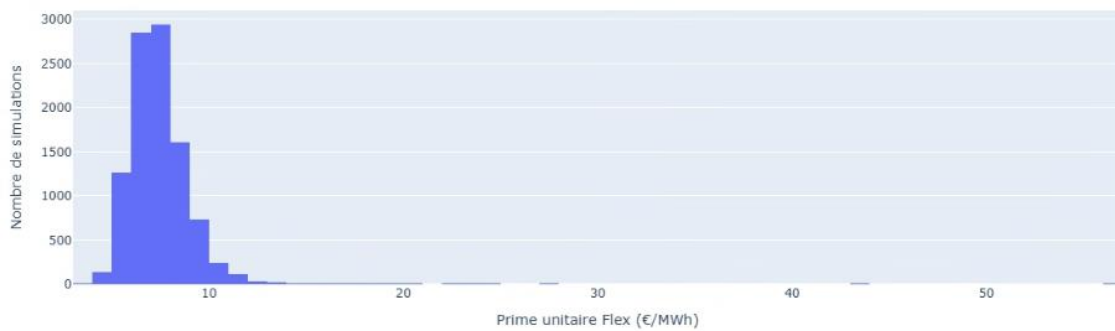
On va donc plus se concentrer sur la distribution avec le Bootstrap où on a une prime moyenne de -7,28€/MWh, des primes unitaires moyennes mieux réparties et avec des queues bien moins extrêmes comme le présente le graphique suivant :



L'écart-type est plus faible et s'établit à 0,79€/MWh avec des primes moyennes sont comprises à 90% dans l'intervalle [-8,62 ; -6,05] €/MWh.

Sur les écarts négatifs, avec les mêmes paramètres, on a des primes positives du même ordre de grandeur que les primes sur écarts positifs avec pour la simulation avec la loi de Student, la distribution suivante :

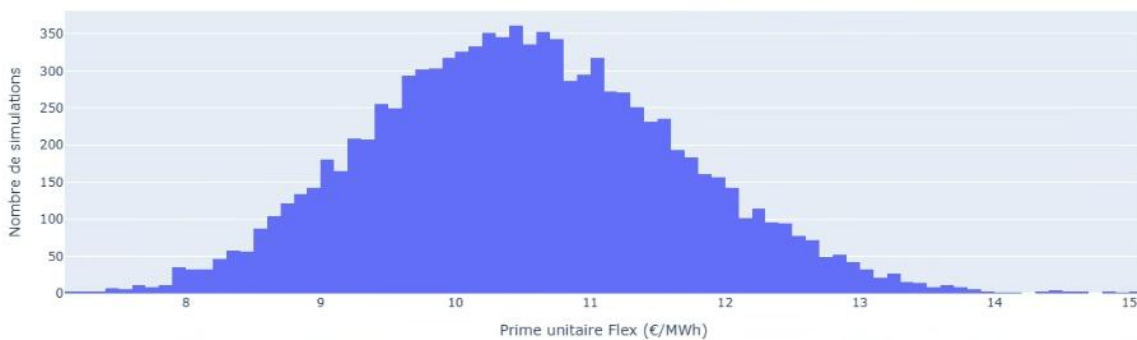
Distribution de la prime Flex (Ornstein-Uhlenbeck + Student)



Avec une prime moyenne de 7,42€/MWh et un écart-type plus faible que les simulations à marche aléatoire simple qui est de 1,66€/MWh. Les primes sont comprises à 90% dans l'intervalle de [5,44 ; 9,89] €/MWh.

Pour la distribution des primes Flex sur écarts négatifs avec Bootstrap, on a une forme de la courbe qui semble suivre le tracé d'une loi normale comme on le voit sur le graphique suivant :

Distribution de la prime Flex (Ornstein-Uhlenbeck + Bootstrap)



Avec une prime moyenne plus élevée qu'avec la loi de Student qui s'établit à 10,51€/MWh mais un écart-type plus faible de 1,14€/MWh. Les primes sont comprises dans un intervalle de confiance à 90% suivant : [8,70 ; 12,45] €/MWh.

Cette dernière méthode a permis de simuler des trajectoires d'écarts en corrigeant la dérive à l'infini avec un retour à la moyenne tout en prenant en compte l'évolution des écarts ou des prix et leur volatilité. Mais, ces simulations ne prennent pas en compte les interactions complexes entre différentes variables explicatives telles que la météo, la saisonnalité, les comportements de consommation, ... Des facteurs non-linéaires qui exercent une influence non-négligeable sur les prix spot et les volumes de consommation. Ces derniers présentent également des corrélations temporelles complexes qui ne peuvent être simplement et justement modéliser que par une dynamique linéaire.

De ce fait, pour améliorer la précision des prévisions et affiner les calculs des primes de risque, il est nécessaire de considérer des modèles plus avancés pouvant capturer ces relations complexes.

3.3. Modélisations avancées

3.3.1. Variables explicatives

Le choix des variables explicatives (aussi appelées Features) doit être cohérent avec les influences possibles sur les prix et/ou la consommation. L'objectif de la construction de ces Features est de fournir aux différents modèles testés un ensemble de variables riches et diversifiées, capables de capturer les dynamiques temporelles, les effets météorologiques et les particularités calendaires qui influencent la consommation électrique et les prix sur le marché court terme. Pour ce faire, on crée plusieurs variables qui pourraient exercer une influence sur l'estimation des prix ou des écarts de consommation qu'on utilise ensuite dans les modèles et pour lesquelles on vérifiera leur importance.

La consommation d'électricité et les prix spot présentent une forte saisonnalité intra-journalière, hebdomadaire et annuelle. Pour capturer ces variations, des indicateurs horaires, journaliers, mensuels et annuels ont été créés ainsi que des marqueurs temporels plus spécifiques comme le jour de la semaine et le week-end. En effet, la consommation diffère d'un jour à l'autre, d'une heure à l'autre et d'un mois à l'autre due à l'activité économique principalement. Pour au mieux prendre en compte les heures de forte demande de celles à faible demande, des variables binaires `is_peak` et `off_peak` ont été créés. Elles permettent de distinguer ces deux périodes d'heure de pointe et d'heure creuse en semaine.

D'autres périodes temporelles sont intéressantes à prendre en compte car elles sont synonymes d'une consommation plus faible même en heure de pointe. Elles correspondent aux jours fériés et aux périodes de vacances scolaires où les comportements de consommation diffèrent de la normale avec une demande industrielle plus faible ce qui modifie les profils horaires. Des variables binaires ont été créées pour représenter ces périodes : `holiday` comprend les jours fériés, `vacances` comprend les vacances scolaires, `vacances_noel` une variable spécifique pour cette période qui, en plein cœur de l'hiver avec des températures faibles, présente une demande bien plus faible par rapport au reste du mois de décembre, et `vacances_ete` où de nombreuses entreprises cessent leur activité temporairement notamment au mois d'août. Ces différenciations pour les vacances d'été et d'hiver permettent de faire comprendre au modèle, en théorie, qu'en été, dans le cas d'une demande faible, elle peut être encore plus faible combinée à la période des vacances, et qu'en hiver, avec une demande thermosensible et donc plus élevée, elle peut être plus faible en période de vacances.

Toujours dans le volet temporalité, des lags temporels ont été constitués. Ils permettent aux modèles de comprendre la structure auto-corrélée des écarts et la persistance des prix sur plusieurs heures consécutives en intégrant les valeurs passées à différents pas de temps. En effet, un lag va correspondre à 1 heure, 2 heures, 12 heures, 24 heures ou 168 heures (soit une semaine) qui correspondent respectivement au prix à l'heure précédente, deux heures avant, de la même heure la veille et de la même heure la semaine précédente. Ces lags permettent de capturer la dépendance temporelle courte (lags 1 et 2), la saisonnalité journalière (lag 24) et la saisonnalité hebdomadaire (lag 168). On va également intégrer les moyennes glissantes (ou `rolling mean` en anglais) à différents horizons (24, 168 ou 365 heures) qui vont permettre de lisser les fluctuations à court terme et de capturer les tendances sous-jacentes. Elles vont faciliter dans ce sens la détection des anomalies par rapport au comportement attendu.

En plus de ces effets calendaires, la météo et notamment la température dans le cadre de la consommation, va influencer directement la consommation en énergie à travers les besoins de chauffage ou de climatisation. La variable directe `temperature` a été créée mais deux autres indicateurs classiques

retrouvés en littérature ont également été créés comme dérivées de la température : les Heating Degree Days (HDD) et les Cooling Degree Days (CDD). Ils représentent respectivement les jours avec un besoin de chauffage et les jours avec un besoin de refroidissement ou climatisation, et se calculent de la manière suivante :

$$HDD = \max(18 - \text{temperature}_t, 0)$$

$$CDD = \max(\text{temperature}_t - 22, 0)$$

Ces variables ont été enrichies par des mesures d'anomalies climatiques, qu'on définit comme la différence entre la température actuelle et la moyenne glissante des 15 derniers jours. Cela permet de modéliser les épisodes météorologiques inhabituels qui peuvent provoquer des variations de consommation significatives et donc impacter les primes de chaque risque.

Des relations peuvent également exister entre les effets calendaires et les effets météorologiques qui ont été traduites en plusieurs Features d'interaction :

- `hour_x_HDD` : certaines heures de peak couplées à des températures froides ou très froides peuvent augmenter anormalement la demande et donc pousser les prix vers le haut à très court terme
- `is_peak_x_anomalie_temp`, `holiday_x_anomalie_temp` et `weekend_x_anomalie_temp` : les anomalies où les températures sont plus faibles ou plus élevées que la moyenne peut avoir un impact sur la demande mais couplée à une heure de pointe, aux jours fériés ou aux jours du week-end, peut modifier le profil horaire de la demande
- `vacances_x_HDD`, `holiday_x_HDD` et `weekend_x_HDD` : l'impact des heures à température froide et poussant naturellement la demande en chauffage vers le haut peut être moins importante en période de vacances scolaires, pendant les jours fériés et les week-ends
- `is_2h_x_vacances`, `is_2h_x_weekend` et `is_2h_x_anomalie_temp` : on a remarqué dans les données que 2 heures du matin était spécifique et présentait l'horaire avec le plus grand écart de prévision pour la consommation. De ce fait, en la couplant avec plusieurs Features, l'idée était d'essayer de capter les spécificités qui pourraient y avoir lors de cette heure en fonction des périodes ou températures.

Il y a enfin des épisodes exceptionnels ou spécifiques qui doivent être modéliser pour un meilleur apprentissage. On a :

- `evenement_extreme` : qui représente un événement marqué par des vacances, jours fériés, week-end ou anomalie de température $> 5^{\circ}\text{C}$
- `low_consumption_day` : qui correspond aux jours avec structurellement une faible consommation : week-end, vacances, jours fériés
- `is_2h` : une heure spécifique comme détaillée précédemment
- `weekday_vacation` et `weekend_vacation` : pour potentiellement capter les différences de consommation entre jours de la semaine et week-end pendant les vacances et ne pas juste partir du principe que la période de vacances présente globalement et peu importe les jours, un niveau de consommation plus faible que la normale

Le détail de la formulation des différentes variables explicatives dans la programmation de ces variables se trouve en Annexe 12.

D'autres Features pourrait être explorer mais les principaux facteurs pouvant expliquer les variations des prix et de la consommation ont été présentés, avec des liens d'interaction pour essayer de capter au mieux les relations complexes même entre variables explicatives.

Ces variables seront donc enfin utilisées et testées dans les différents modèles utiliser pour le calcul des primes via l'estimation des prix spot ou des écarts de consommation. Elles seront ensuite triées en fonction de leur importance et de ce qu'elles apportent en termes de précision aux modèles afin de simplifier l'équation et le nombre de variable mais aussi de limiter le risque de surapprentissage sur des variables peu pertinentes. Concrètement, les modèles utilisés permettent d'estimer l'importance relative de chaque variable en mesurant la réduction de l'impureté moyenne ou erreur moyenne ou encore en mesurant la contribution au gain lors des splits. Dans notre étude, un seuil de tolérance de 0,5% a été appliqué donc seules les Features avec un impact relatif dépassant 0,5% seront retenues.

3.3.2. Modèles testés

Les modèles testés doivent prendre en compte la forme des données et leur nature. Les modèles doivent être adaptés pour les séries temporelles et prendre en compte les corrélations entre les heures des variables cibles et des variables explicatives.

On commence par implémenter un modèle classique des séries temporelles mais qui prend en compte la saisonnalité des données : le modèle SARIMA (Seasonal AutoRegressive Integrated Moving Average). Le modèle va capturer les tendances temporelles, les cycles saisonniers et les dépendances temporelles du prix spot ou des écarts de consommation sans nécessiter de variables explicatives externes comme celles qui ont été construites auparavant.

L'objectif est de modéliser les prix spot pour le risque de marché et les écarts de consommation pour le risque de Flex en tenant compte de la saisonnalité horaire et quotidienne.

Le modèle se formalise de la façon suivante :

$$\Phi_p(L)(1-L)^d(1-L^s)^D y_t = \Theta_q(L)\Theta_Q(L^s)\varepsilon_t$$

Avec :

- y_t la série temporelle de la variable cible (prix spot ou écarts de consommation)
- L l'opérateur de retard (aussi appelé lag)
- p, d, q les paramètres d'ordres AR (p) et MA (q) ainsi que de différenciation (d)
- P, D, Q, s les paramètres d'ordres saisonniers et période s
- ε_t le résidu ou bruit blanc

Les paramètres du modèle testé et les plus significatifs pour les prix spot sont $(p, d, q) = (1, 1, 1)$ et $(P, D, Q, s) = (1, 1, 1, 24)$. Les mêmes valeurs ont été utilisés pour les paramètres de simulation des écarts de consommation.

On effectue ensuite l'entraînement en divisant l'échantillon de données en deux : l'échantillon de test et l'échantillon d'apprentissage (ou train en anglais) et on évalue les performances du modèle afin de pouvoir déterminer le niveau d'explicabilité du modèle et de comparer ses métriques aux autres modèles testés.

Pour pouvoir intégrer des éléments externes potentiellement explicatifs d'une partie de la variance du prix spot ou des écarts de consommation en fonction des risques étudiés, on utilise des modèles de Machine Learning qui permettent de prendre en compte de plus grande dimension de problème et d'intégrer des variables explicatives exogènes.

Dans cette étude, deux modèles basés sur des arbres ont été testés : XGBoost (XGB) et Random Forest (RF). Ces modèles permettent de capturer les relations non linéaires, sont robustes aux interactions complexes et aux données bruitées. En effet, les prix spot et les écarts de consommation dépendent de nombreux facteurs exogènes tels que la température, les heures pleines ou creuses, les événements exceptionnels (canicule intense et longue, hiver très froid, ...) qui ne sont pas nécessairement liés de manière linéaire à la variable cible. Les arbres permettent également de gérer les interactions entre variables explicatives sans nécessiter des transformations de variable au préalable.

Le Random Forest, ou Forêt aléatoire en français, est un modèle d'arbres de décision T_1, T_2, \dots, T_N construits sur des sous-échantillons aléatoires de données (de la variable cible, donc soit le prix spot, soit les écarts de consommation) et en fonction des variables explicatives. La prédiction finale est la moyenne des prédictions de chaque arbre qui constituent la forêt. Le modèle se formalise de la façon suivante :

$$\hat{y}_{RF(X)} = \frac{1}{N} \sum_{n=1}^N T_n(X)$$

Avec :

- N le nombre total d'arbres
- $T_n(X)$ la prédiction de l'arbre n pour le vecteur de variables explicatives X
- X correspond à l'ensemble des variables explicatives sélectionnées. Dans le cas de la sélection des variables les plus importantes, on a comme Features :
 - ⇒ Pour le risque de marché : [*temperature, HDD, lag₁, lag₂₄, lag₁₆₈*]
 - ⇒ Pour le risque de Flex : [*is_peak, lag₁, lag₂₄, lag₁₆₈*]

Chaque arbre T_n est construit en choisissant aléatoirement un sous-ensemble des observations et des variables explicatives, puis en effectuant des sous-échantillons successifs pour minimiser la variance au sein des feuilles terminales.

On choisit différents paramètres dans ce modèle et dans notre cas :

- $n_estimators$: Le nombre d'arbre de décision que va construire le modèle. Plus il y a d'arbres, plus la forêt est robuste mais il y a un risque de surajustement du modèle (qui peut être compensé par la régularisation dans le modèle XGBoost)

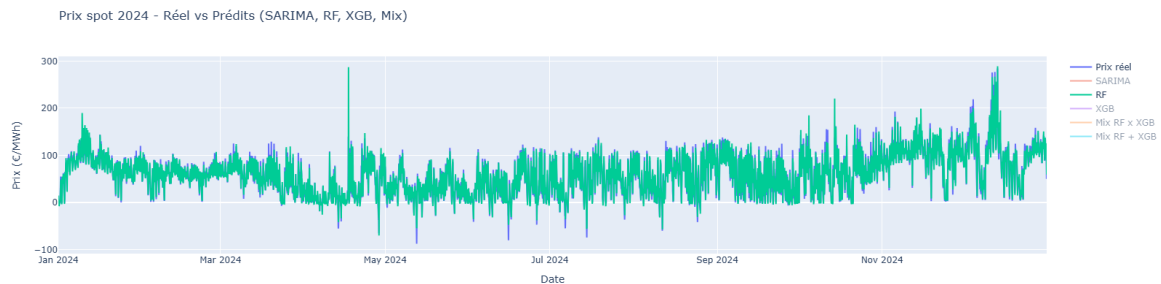
- `max_depth` : La profondeur de chaque arbre où plus la profondeur est grande, plus l'arbre va capturer les relations complexes mais il y a également un risque d'overfitting. Si elle est trop faible cependant, le modèle va manquer de flexibilité et rater des schémas complexes.
- `random_state` : La graine aléatoire qui est utilisée pour contrôler le caractère aléatoire du modèle c'est-à-dire qu'on va intégrer du hasard dans le tirage des échantillons et des variables utilisées à chaque découpage. Elle est fixée à 0^{29} et permet d'avoir toujours les mêmes résultats à chaque exécution ce qui assure la répliquabilité des modèles.

```
rf_model = RandomForestRegressor(n_estimators=100, max_depth=10, random_state=0)
rf_model.fit(X_train, y_train)
spot_pred_rf = rf_model.predict(X_test)
```

Code 1 : Modélisation Random Forest pour les prix spot en Python

Après avoir défini ces paramètres pour chaque risque, on entraîne le modèle sur les échantillons train et on le teste sur un autre échantillon (test).

On visualise enfin le résultat en comparant avec les données réelles et les données estimées par le modèle :



XGBoost est un modèle qui se base aussi sur les arbres mais auquel on intègre un volet Boosting, c'est-à-dire que les arbres sont construits séquentiellement pour corriger les erreurs des arbres précédents via la minimisation d'une fonction de perte. La prédiction finale est la somme pondérée des prédictions de chaque arbre :

$$\hat{y}_{XGB(X)} = \frac{1}{N} \sum_{k=1}^K f_k(X), \quad f_k \in \mathcal{F}$$

Où :

- K est le nombre d'arbres construits séquentiellement
- f_k est un arbre de décision optimisé pour réduire la fonction de perte $L(y, \hat{y})$ qui est l'erreur quadratique dans le cas de cette étude et qui se formalise de la façon suivante :

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{k=1}^K \Omega(f_k)$$

²⁹ La valeur du nombre (fixé à 0 ou couramment à 42) n'a pas de signification particulière. L'objectif c'est de soit mettre un nombre qui permet d'avoir toujours les mêmes résultats soit ne pas en mettre (`random_state = None`) où chaque entraînement peut donc donner des résultats légèrement différents ce qui peut être utile pour tester la robustesse du modèle.

Avec $\Omega(f_k)$ le terme de régularisation pénalisant la complexité de l'arbre afin de limiter le sur-apprentissage.

On définit les mêmes paramètres avec une lecture légèrement différente :

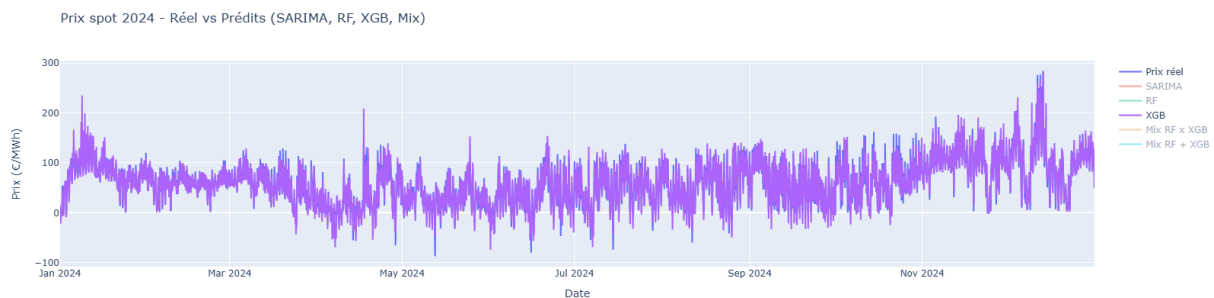
- `n_estimators` correspond au nombre d'itérations de boosting, c'est-à-dire le nombre d'arbres construits séquentiellement
- `max_depth` : même logique que le modèle de forêt aléatoire
- `random_state` : certaines étapes dans le cadre du XGB va inclure du hasard lors du sous-échantillonnage des observations ou des variables explicatives par exemple

On entraîne ensuite le modèle et on le teste :

```
xgb_model = XGBRegressor(n_estimators=100, max_depth=6, random_state=0)
xgb_model.fit(X_train, y_train)
spot_pred_xgb = xgb_model.predict(X_test)
```

Code 2 : Modélisation de XGBoost pour les prix spot en Python

On visualise les données réelles et les données simulées avec ce modèle pour comparer les deux :



Pour bénéficier des points forts de chaque modèle et réduire la variance des prédictions, on a combiné les prédictions RF et XGB avec deux méthodes.

La première méthode consiste en la moyenne arithmétique des deux modèles comme suit :

$$\hat{y}_{mix_{avg}}(X) = \frac{\hat{y}_{XGB(X)} + \hat{y}_{RF(X)}}{2}$$

Cette approche donne le même poids aux deux modèles et permet de lisser les prédictions, réduisant l'effet des erreurs extrêmes. Elle permet également de prendre en compte les prix négatifs dans le cadre de la modélisation des prix spot.

La deuxième méthode prend la racine géométrique des deux modèles multipliés entre eux comme suit :

$$\hat{y}_{mix_{geo}}(X) = \sqrt{\hat{y}_{XGB(X)} * \hat{y}_{RF(X)}}$$

On suppose avec ce modèle que l'information apportée par chacun des modèles est complémentaire et multiplicative (et non additive comme avec la moyenne arithmétique). On va également émettre

l'hypothèse que les erreurs se compensent de manière proportionnelle (ce qui peut être le cas pour le risque de Flex).

Cependant, la racine carrée implique qu'on doit travailler avec des valeurs strictement positives. Donc, il ne permet pas de prendre en compte les prix spot nuls ou négatifs pour le risque de marché.

Ces deux modèles mélangés vont se programmer de la manière suivante :

```
ecart_pred_mix = (ecart_pred_rf.clip(min=0) * ecart_pred_xgb.clip(min=0)) ** 0.5
ecart_pred_avg = (ecart_pred_rf + ecart_pred_xgb) / 2
```

*Code 3: Modélisation des modèles XGB + RF et XGB * RF pour le risque de Flex en Python*

Une fois les modélisations formalisées et programmées, on passe à l'évaluation et la comparaison de chacun des modèles puis aux calculs des primes de risque.

3.3.3. Comparaison des performances et résultats pour chaque risque

Ces modèles ainsi testés présentent différentes métriques d'évaluation qui vont permettre de les comparer entre eux et de juger de leur qualité d'estimation du prix spot pour le risque de marché et des écarts de consommation pour le risque Flex.

Pour le risque de marché, avec uniquement les Features les plus importantes, plusieurs métriques de performance ont été calculées : l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R^2 et R^2 ajusté), le biais moyen des prédictions, ainsi que la variance des valeurs prédites. Ces indicateurs permettent d'évaluer la précision, la stabilité et la qualité explicative de chaque approche et sont calculés pour chaque modèle testé ci-dessous :

```
--- SARIMA ---
RMSE : 1342.27 €/MWh
R2 : -1087.3450
R2 ajusté : -1087.4689
Biais : 1134.11 €/MWh
Variance des prédictions : 499027.85
```

Pour le modèle SARIMA³⁰, on a des métriques très mauvaises avec un R^2 fortement négatif qui signifie que le modèle explique moins bien la variance des prix que la moyenne historique. On a d'ailleurs un RMSE et un biais très élevé confirmant le fait que ce modèle n'est pas adapté et ne permet pas d'avoir des prédictions utilisables.

```
--- Random Forest (features importantes)
RMSE : 13.37 €/MWh
R2 : 0.8920
R2 ajusté : 0.8919
Biais : -0.47 €/MWh
Variance des prédictions : 1549.28
```

Le modèle Random Forest présente le meilleur R^2 qui indique que le modèle explique près de 90% de la variance observée. On a une erreur bien plus faible que SARIMA (RMSE = 13,37€/MWh) et un biais très faible, proche de 0 traduisant une absence de tendance du modèle à sur ou sous-estimer la variable cible. On a d'ailleurs une variance plus modérée des prédictions montrant une bonne stabilité.

```
--- XGBoost (features importantes) ---
RMSE : 16.63 €/MWh
R2 : 0.8330
R2 ajusté : 0.8329
Biais : 0.49 €/MWh
Variance des prédictions : 1745.90
```

³⁰ Le choix des Features n'a pas d'impact sur le modèle SARIMA qui n'intègre pas de variables externes et inclut dans sa formulation mathématique des retards ou lags temporels.

--- Mix RF x XGB (features importantes) On a le même constat pour le modèle XGBoost avec des métriques un peu moins performantes que le Random Forest avec 83% d'explication par le modèle de la variance expliquée, une erreur et un biais contrôlés mais plus élevés.
RMSE : 13.96 €/MWh
R² : 0.8823
R² ajusté : 0.8822
Biais : -0.94 €/MWh
Variance des prédictions : 1466.43

--- Mix RF + XGB (features importantes) Enfin, le mix des modèles avec les deux méthodes présente une méthode géométrique légèrement plus performante que la méthode de la moyenne et un biais et une erreur moins élevés également.
RMSE : 14.21 €/MWh
R² : 0.8781
R² ajusté : 0.8780
Biais : 0.01 €/MWh
Variance des prédictions : 1621.97

Ces comparaisons sont différentes si on prend en compte toutes les variables explicatives créées. Mais prendre en compte l'ensemble de ces Features complexifient les modèles qui présentent des dimensions très grande de résolution. La présentation de ces métriques est néanmoins faite en annexe 13.

Ces modèles ont estimé une prime de risque marché pour 2024 :

- Avec le modèle SARIMA : -86,09€/MWh
- Avec le modèle Random Forest : -0,94€/MWh
- Avec le modèle XGBoost : -0,99€/MWh
- Avec le modèle mix RF x XGB : -0,94€/MWh
- Avec le modèle mix RF + XGB : -0,96€/MWh

Ces primes sont négatives et signifie que le recours au marché pour acheter les blocs d'énergie manquants afin de répondre aux besoins de la demande non couvert par les prévisions a été moins couteux de presque 1€/MWh que le prix d'achat à terme de l'énergie.

On effectue la même démarche de comparaison des modèles et calculs de primes pour le risque de Flex avec des variables explicatives les plus importantes légèrement différentes et n'incluant pas, pour la simulation des écarts de consommation, le volet météorologique. Dans les variables explicatives sélectionnées par importance, on retrouve les lags temporels de 1 heure, 24 heures et 168 heures ainsi que la variable is_peak.

Pour ces variables, les modèles entraînés ont de très bonnes métriques qu'on va détailler ci-dessous :

--- SARIMA (écart) --- Pour le modèle SARIMA, comme avec le risque de marché, les métriques sont très mauvaises avec un R² fortement négatif et une erreur quadratique élevée par rapport aux autres modèles. Il n'est donc pas adapté pour obtenir des prédictions fiables.
RMSE : 60.78 MWh
R² : -26.4474
R² ajusté : -26.4531
Biais : -52.84 MWh
Variance des prédictions : 927.69

--- Random Forest (Flex) --- Le modèle Random Forest est celui qui présente le meilleur R² (0,95) avec plus de 95% de la variance observée expliquée par le modèle. L'erreur de prédiction est d'ailleurs très faible avec un RMSE de 2,55€/MWh et un biais proche de 0. On a, de plus, une variance des prédictions modérée démontrant une bonne stabilité du modèle.
RMSE : 2.55 MWh
R² : 0.9515
R² ajusté : 0.9515
Biais : 0.04 MWh
Variance des prédictions : 128.34

--- XGBoost (Flex) ---
RMSE : 2.74 MWh
R² : 0.9443
R² ajusté : 0.9442
Biais : 0.03 MWh
Variance des prédictions : 127.16

Le modèle XGBoost a de très bonnes performances aussi, légèrement plus faible que le modèle Random Forest avec un R² de 0,94, une erreur faible et un biais quasi nul aussi. La variance des prédictions est comparable à celle du précédent modèle ce qui confirme la robustesse de ce modèle à arbre pour capturer les dynamiques de l'écart de consommation.

--- Mix RF x XGB (géométrique) ---
RMSE : 6.89 MWh
R² : 0.6470
R² ajusté : 0.6467
Biais : -2.90 MWh
Variance des prédictions : 58.55

On a ensuite fait, comme pour le risque de marché, un mix géométrique et de moyenne des deux modèles avec des métriques moins bonnes pour le mix géométrique (R² de 0,65) mais de très bonnes métriques pour le modèle de moyenne avec une erreur quasi-similaire aux modèles de Random Forest et XGBoost et un biais presque nul.

--- Mix RF + XGB (moyenne) ---
RMSE : 2.60 MWh
R² : 0.9500
R² ajusté : 0.9499
Biais : 0.03 MWh
Variance des prédictions : 127.47

L'analyse des performances des modèles qui utilisent l'ensemble des Features et le calcul des primes ont été également faits en Annexe 14.

Ces modèles ont estimé une prime de risque marché pour 2024 :

- Avec le modèle SARIMA : -133,39€/MWh
- Avec le modèle Random Forest : -1,79€/MWh
- Avec le modèle XGBoost : -1,84€/MWh
- Avec le modèle mix RF x XGB : -6,60€/MWh
- Avec le modèle mix RF + XGB : -1,81€/MWh

Hors primes SARIMA et mix RF x XGB, les primes sont quasi-similaires mais plus élevées négativement que le risque de marché.

Ces primes négatives signifient que le coût d'achat d'énergie supplémentaire sur le marché est en moyenne inférieur d'environ 1,80€/MWh du prix de vente au client permettant de préserver la marge du fournisseur sur l'année 2024 à l'échelle du portefeuille. Cela est à nuancer car les modèles rencontrent certaines limites notamment la minimisation des écarts négatifs ou des écarts qui tendent vers 0 quand ils sont négatifs ce qui biaise les résultats malgré d'excellentes performances des modèles.

3.4. Prédiction des primes pour l'année 2026

Les modélisations effectuées précédemment vont servir de base pour pouvoir effectuer une projection de la valeur des primes pour 2026 en prenant en compte les relations complexes entre les variables explicatives et les variables cibles dans un environnement volatil et sensible à la temporalité.

3.4.1. Features utilisées

On reprend les variables explicatives créées et testées dans la précédente section qui ont une importance significative (avec une tolérance à 0,01).

Pour certains Features, notamment celle dépendante de la temporalité, on doit les adapter pour l'année 2026 en fonction des risques simulés.

Pour le risque de marché où on cherche à modéliser et estimer le prix spot, on doit avoir les prix à terme de 2026 et les consommations réelles et prévisionnelles de 2026.

Si on détient l'information pour 2026 des prix à terme, ce n'est pas le cas pour les consommations réelles qui ne sont effectives qu'au moment de la consommation en 2026. Pour ce faire, on extrapole les valeurs des écarts de l'année 2024 à l'année 2026. On garde donc les écarts de consommation de 2024 pour la projection 2026 dans la formule de calcul de prime.

Il faut également adapter les variables explicatives utilisées notamment les Features de température. En effet, nous ne connaissons pas à l'avance de manière précise la température qu'il fera à chaque heure ou jour de 2026, donc on doit constituer la température pour cette année.

Pour ce faire, on construit la variable de température en 2026 en projetant sur l'année 2026 un profil horaire moyen historique de température obtenu par regroupement selon les mois et heures de l'historique de météo à disposition. Pour chaque couple (mois, heure), on va calculer la température moyenne historique afin de constituer une matrice de référence des profils température selon la saison et l'heure. On projette ensuite sur l'année 2026 en attribuant la température moyenne correspondante à chaque heure de chaque mois ce qui nous permet finalement d'obtenir une série temporelle complète de température horaire pour 2026. Cette méthode ne prend pas cependant en compte la variabilité interannuelle (c'est-à-dire les hivers doux ou les canicules exceptionnelles) et reste une vision moyenne de la météo. Mais cette méthode est simple et permet de capturer les principales structures temporelles de la température.

Une fois la variable température récupérée, on peut construire la variable HDD pour 2026 qui fait partie des variables, avec la température, les plus significatives

En plus des variables explicatives de températures et HDD, les lags temporels font également parti des Features les plus significatifs notamment les lags de 1 heure, 24 heures et 168 heures. Mais pour l'année 2026, nous n'avons pas encore de valeurs réelles de prix spot ce qui est nécessaire pour calculer les lags. On va donc approcher les valeurs passées par des estimations basées sur l'historique en commençant par calculer la moyenne historique pour chaque couple (mois, heure) ce qui va permettre d'avoir un profil horaire saisonnier moyen pour chaque mois et chaque heure de la journée. Puis, pour chaque pas horaire de 2026, on récupère la valeur moyenne historique correspondant au mois et à l'heure. On aura donc pour chaque lag, une approximation des prix passés décalée d'une, vingt-quatre ou cent soixante-huit heures en fonction du lag.

Pour le risque de Flex où on cherche à modéliser les écarts de consommation future, on a à disposition les prix de vente 2026 et les prix spot prévisionnels 2026 construit à partir de la courbe HPFC³¹. Cette courbe est une méthode standard pour construire une trajectoire des prix futurs, ne les connaissant pas à l'avance, à partir des prix Forward cotés mensuellement, trimestriellement ou annuellement. Utiliser cette courbe comme approximation des prix spot 2026 est cohérente dans le cadre du calcul de la prime de risque car elle prend en compte les anticipations du marché. En effet, les prix Forward donnent une indication sur le niveau moyen attendu des prix en fonction des périodes dans l'année. Se baser sur cette

³¹ Hourly Price Forward Curves ou courbe de prix Forward au pas horaire en français

courbe permet de prendre comme hypothèse que le marché spot suivra en 2026 une dynamique conforme aux courbes de prix Forward qui sont ajustées à des profils historiques et prennent en compte la saisonnalité et la variabilité historiques des prix. Cependant, la HPFC ne capture pas les événements extrêmes futurs qu'il pourrait y avoir en 2026 tels que les événements climatiques, les stress de marché, les périodes de prix négatif par exemple. Mais pour le calcul de la prime de risque, cette méthode reste justifiée notamment pour utiliser des prix spot fixe.

En plus des lags temporels, la variable explicative `is_peak` est significative pour l'estimation des écarts de consommation pour le risque de Flex. Elle permet de distinguer les heures de pics de consommation de celle de creux de façon binaire : 1 si l'heure est comprise sur la plage horaire de 8h à 20h, du lundi au vendredi (correspondant généralement aux heures de forte demande) et 0 sinon.

Une fois ces ajustements effectués, on peut appliquer les meilleurs modèles en fonction des risques à chaque calcul de prime afin d'obtenir une estimation du coût de la prime en 2026.

3.4.2. Risque de marché

Pour le risque de marché, le modèle d'estimation avec les meilleurs métriques et le meilleur couple biais/variance est le mix de modèle additif XGBoost et Random Forest qu'on va appliquer pour estimer la prime de risque de marché 2026 avec les Features 2026 qui ont été préparé en amont.

L'objectif est de déterminer une prime unitaire pondérée pour l'année 2026 à l'échelle du portefeuille en prenant bien en compte que le calcul de la prime se base sur une approximation de plusieurs variables explicatives et un entraînement de modèle qui a été influencé notablement par des prix spot encore élevé issus de la crise énergétique.

On se base également sur des écarts de consommation historique de l'année 2024 qui ne seront surement pas les mêmes, à la même heure, journée, mois qu'en 2026. La consommation étant variable d'une heure à l'heure et sensible à la météo, élément non certain à l'avance également, les écarts peuvent et seront surement fluctuants et différents pour l'année 2026.

La prime de risque estimée pour 2026 est de -0,07 €/MWh donc un gain par MWh pour le fournisseur de 7 centimes et non une perte. On a une prime encore négative mais très proche de 0 due à l'effet de foisonnement où les écarts se compensent entre eux mais aussi due à une baisse des prix à terme pour l'année 2026 ainsi qu'une tendance à la baisse des prix spot. En effet, pour 2024, le prix à terme moyen était de 95,21€/MWh contre 71,29€/MWh pour les prix négociés à terme de 2026. Cette prime qui ne correspond pas à une perte est à l'échelle du portefeuille et ne permet pas de distinguer les différents segments.

De ce fait, on effectue la projection 2026 sur chaque segment et on calcule les primes de risque de marché qu'on retrouve dans le tableau suivant :

	C2	C4	C5-PRO	C5-RES	Portefeuille
Prime de risque marché	0,04	-0,32	-0,20	-0,11	-0,07

Tableau 7 : Prime de risque marché moyenne par segment en €/MWh

La majorité des primes sont négatives sauf pour le segment C2 où la prime est légèrement positive ce qui implique une légère perte et donc un coût sur ce segment pour le fournisseur à prix à terme moyen fixe et écarts fixes historiques non effectifs.

Ces primes par segment comprennent toujours l'effet de foisonnement entre différents profils de consommation, d'où le fait qu'elles tendent vers 0.

On peut également calculer la prime de risque de marché en modifiant le prix à terme et le réduisant de 3, 5, 10 ou 20€/MWh pour voir l'effet d'une baisse moyenne des prix Forward comme c'est le cas entre 2024 et 2026 sur la prime de risque marché moyenne globale sur l'ensemble du portefeuille puis par segment :

	Recalcul de la prime de risque en soustrayant un montant X au prix à terme			
	-3€/MWh	-5€/MWh	-10€/MWh	-20€/MWh
Portefeuille	0,08	0,19	0,45	0,98
C2	0,19	0,29	0,54	1,05
C4	-0,03	0,17	0,65	1,63
C5-PRO	0,02	0,16	0,51	1,22
C5-RES	-0,06	-0,02	0,08	0,27

Tableau 8 : Primes de risque marché moyenne par segment et portefeuille selon différents scénarios de baisse des prix à terme en €/MWh

La baisse du prix à terme exerce la même influence mais à différents niveaux en fonction du segment où la prime sera plus rapidement couteuse pour certains segments que d'autres. On constate également des primes plus élevées pour le dernier scénario sur certains segments pour lesquels la prime était très faible voir négative pour le premier scénario. Par exemple, la prime de risque pour le segment C4 commence à -0,03€/MWh mais atteint les 1,63€/MWh. Si on compare avec le segment C2 qui a une prime légèrement positive dès le premier scénario (0,19€/MWh), elle n'augmente pas aussi rapidement pour le dernier scénario et s'établit à 1,05€/MWh. Cela traduit une sensibilité à la baisse des prix à terme plus accrue pour le segment C4 que le segment C2 qui peut avoir plusieurs raisons : les écarts à la consommation du segment C4 ont lieu plus souvent sur des heures ou périodes tendues, les écarts à la consommation pour le segment C2 ont lieu plus souvent sur des périodes creuses moins couteuses par exemple.

La réalisation de ces scénarios nous permet d'avoir une vision plus large de la sensibilité au risque de chaque segment et de la quantifier afin de guider le fournisseur sur une tarification justifiée et transparente. On effectue la même démarche pour le risque de Flex dans la section suivante.

3.4.3. Risque de Flex

On pourrait reprendre un modèle entraîné avec les meilleures métriques comme on l'a fait pour le risque de marché mais les tests effectués n'ont pas été concluants car ils présentaient beaucoup trop d'incertitudes sur les données et dans la méthode d'application.

En effet, les écarts de consommation doivent être calculer en faisant la différence entre la consommation réelle et la consommation prévisionnelle mais pour l'année 2026, on ne dispose pas des valeurs réelles de consommation. Le modèle s'appuie seulement sur l'historique des écarts pour les prédire et doit retenir uniquement les écarts positifs. Cependant, ces écarts au pas horaire sont très faibles ou pas assez nombreux sur une année pour permettre au modèle d'estimer d'autres écarts ce qui résulte en pratique à

des écarts simulés anormalement faibles et une prime moyenne de risque Flex fortement négative non réaliste de -32,36€/MWh.

Lorsqu'on applique le calcul de la prime de risque sur les écarts négatifs, on constate en détail que le modèle XGBoost n'a généré quasiment que des écarts positifs. Cela s'explique d'une part avec les données historiques utilisées pour l'entraînement du modèle qui présentent une asymétrie avec des écarts positifs moins nombreux en quantité sur l'année mais qui ont une valeur plus importante en amplitude. D'autre part, l'algorithme XGBoost, en raison de sa structure basée sur des arbres de décision optimisés, a tendance à privilégier la minimisation de l'erreur globale plutôt qu'une répartition équilibrée des résidus. Ainsi, les événements rares mais de forte intensité influencent davantage l'apprentissage et biaisent le modèle vers des prédictions qui reproduisent cette asymétrie. Enfin, cette limitation reflète une faiblesse générale des modèles d'arbres en présence de distributions déséquilibrées : ils captent bien les tendances dominantes, mais reproduisent difficilement la variabilité des queues de distribution, ce qui se traduit par une sous-estimation systématique des écarts négatifs.

Partant de ce constat, il est préférable de calculer la prime Flex pour 2026 en utilisant des méthodes plus simples mais plus robustes telles que les simulations Monte-Carlo ou avec processus Ornstein-Uhlenbeck. Ces méthodes capturent la variabilité réaliste des écarts, ce qui est recherché dans cette étude pour quantifier la prime de risque Flex.

On va donc reprendre les calculs et résultats de la simulation qu'on va adapter pour l'année 2026 en appliquant certains ajustements :

- On simule les écarts de consommation avec une distribution Bootstrap, méthode non-paramétrique la plus appropriée pour coller au mieux à la distribution empirique des données.
- On effectue 10 000 simulations.
- On calcule la prime de risque de Flex en récupérant les données de prix de vente 2026 et les prix spot de 2024 qui sont deux variables fixes.

On obtient une prime simulée moyenne pour 2026 de 3,00€/MWh sur les segments C2 et C4 regroupés avec la prise en compte de l'ensemble des écarts. La prime est cohérente pour ces deux segments regroupant de gros consommateurs et dans le contexte encore de prix à terme élevé par rapport à des prix spot faible où les modèles de prévision de consommation utilisée en interne ont tendance à surestimer la consommation ce qui est coûteux dans la structure des prix actuels et à venir pour 2026.

Une prime positive indique un coût pour le fournisseur qui résulte de deux cas :

- On a une consommation réelle supérieure à la consommation prévisionnelle (surconsommation, manque à acheter sur le marché) et un prix spot supérieur au prix de vente ;
- On a une consommation réelle inférieure à la consommation prévisionnelle (sous-consommation, surplus à vendre sur le marché) et un prix spot inférieur au prix de vente.

On peut également calculer la prime de risque de Flex en modifiant le prix de vente et le réduisant de 5, 10, 20 ou 40€/MWh pour voir l'effet d'une baisse moyenne des prix dans le cadre de renégociation de contrat ou de renouvellement de contrat dans un contexte de marché plus stable et avec des prix à terme moins élevés. Cela entraînera des conséquences sur la prime de risque Flex moyenne globale de l'ensemble du portefeuille puis par segment :

	Recalcul de la prime de risque en soustrayant un montant X au prix de vente			
	-5€/MWh	-10€/MWh	-20€/MWh	-40€/MWh
Portefeuille	2,68	2,36	1,73	0,45
C2	4,05	3,57	2,62	0,70
C4	8,06	7,09	5,16	1,28

Tableau 9 : Primes de risque Flex par segment et portefeuille selon différents scénarios de baisse des prix de vente en €/MWh

On a toujours des primes positives en faisant fluctuer le prix de vente même si elles sont un peu plus faibles. Les périodes où il y a un cas de surconsommation ou sous-consommation avec respectivement des prix spot élevés et plus faibles, sont des périodes à risque qui doivent être surveiller et quantifier. Le prix de vente n'est pas la variable la plus déterminante dans le calcul mais reste une variable informative et intéressante à faire varier notamment pour le pôle commercial.

Cette méthode rencontre une certaine limite notamment l'usage des prix spot 2024 dans la formule de prime pour l'année 2026 sans modification. On n'applique pas non plus de poids sur les années de crise qui ont des prix spot bien plus élevés que la normal pouvant influencer à la hausse la valeur des primes.

Pour essayer de contrer l'effet de cette limite, on va simuler les prix spot en prenant l'ensemble de l'historique des données spot de 2005 à 2025 avec le prix de vente 2026. On a une prime négative (du fait encore des prix de vente élevés par rapport à la normale. En effet, les prix en 2019 été de 70€/MWh en moyenne et même autour de 53€/MWh en 2015) sur le portefeuille (segments C2 et C4) de -2,26€/MWh. Les primes moyennes sont comprises dans un intervalle à 90% suivant [-3,02 ; -1,33] €/MWh. Cette simulation des prix spot prend en compte l'ensemble des années sans différenciation des années de crise qui influence la valeur des primes. Pour réduire l'effet des années de crise, on ajoute des poids aux années 2008, 2009, 2020, 2021, 2022, 2023 et 2024 où 2008, 2020, 2021 et 2022 ont un poids réduit de moitié (on donne moins d'importance à ces prix lors de la simulation) et les autres années ont un poids de 0,7 (années après-crise qui sont des années dites de « transition »).

A cette simulation avec poids, on peut modifier le prix de vente moyen à titre informatif avec les prix avant crise. Par exemple, pour un prix de vente moyen à 70€/MWh, on a une prime moyenne sur le portefeuille de 1,45€/MWh. Et pour un prix de vente moyen à 53€/MWh, on a une prime moyenne de 0,28€/MWh. Donc un coût potentiel si les prix reviennent à ces niveaux de 0,46€ par MWh. Cette différence de valeur de prime où à prix de vente plus élevé, on a une prime plus élevée s'explique par le fait que le calcul de la prime de risque se base sur la volatilité du prix spot notamment. En effet, lorsque le prix spot baisse fortement, l'écart entre le prix de vente et le prix spot s'accroît et la revente peut constituer une perte plus élevée par exemple. Alors que quand le prix de vente est plus faible, l'écart se réduit résultant sur un risque plus faible et donc une prime moins élevée.

Les prix des clients C2 et C4 sont assez élevés par rapport aux clients C5 pour qui le risque de Flex est effectivement un sujet à détailler mais la complexité de la construction des données ne permet pas pour l'instant dans le cadre de ce mémoire de calculer la prime de risque de Flex. On suppose cependant, de par la nature des volumes globaux qu'il y a des primes positives pour les clients professionnels et résidentiels de ce portefeuille mais qui ont un impact plus faible en termes de volume de consommation totale dans le portefeuille que les clients C2 et C4.

Une autre approche de comparaison entre les prix spot 2024 utilisés initialement et les prix spots des autres années a été testé afin de comparer, avec un prix de vente fixe correspondant à la moyenne des prix proposé par le fournisseur au client pour 2026, la valeur de la prime moyenne en fonction des

années. En pratique, pour chaque année, on applique le prix spot de l'année considérée dans la formule de calcul de la prime de risque ce qui donne les résultats suivants :

Prime Flex moyenne 2026 (prix spot 2005) :	-3.29 €/MWh
Prime Flex moyenne 2026 (prix spot 2006) :	-3.22 €/MWh
Prime Flex moyenne 2026 (prix spot 2007) :	-3.78 €/MWh
Prime Flex moyenne 2026 (prix spot 2008) :	-2.04 €/MWh
Prime Flex moyenne 2026 (prix spot 2009) :	-3.48 €/MWh
Prime Flex moyenne 2026 (prix spot 2010) :	-3.20 €/MWh
Prime Flex moyenne 2026 (prix spot 2011) :	-3.09 €/MWh
Prime Flex moyenne 2026 (prix spot 2012) :	-3.28 €/MWh
Prime Flex moyenne 2026 (prix spot 2013) :	-3.53 €/MWh
Prime Flex moyenne 2026 (prix spot 2014) :	-3.94 €/MWh
Prime Flex moyenne 2026 (prix spot 2015) :	-3.70 €/MWh
Prime Flex moyenne 2026 (prix spot 2016) :	-3.83 €/MWh
Prime Flex moyenne 2026 (prix spot 2017) :	-3.42 €/MWh
Prime Flex moyenne 2026 (prix spot 2018) :	-3.11 €/MWh
Prime Flex moyenne 2026 (prix spot 2019) :	-3.70 €/MWh
Prime Flex moyenne 2026 (prix spot 2020) :	-4.09 €/MWh
Prime Flex moyenne 2026 (prix spot 2021) :	0.31 €/MWh
Prime Flex moyenne 2026 (prix spot 2022) :	10.36 €/MWh
Prime Flex moyenne 2026 (prix spot 2023) :	-0.36 €/MWh
Prime Flex moyenne 2026 (prix spot 2024) :	-2.77 €/MWh
Prime Flex moyenne 2026 (prix spot 2025) :	-2.55 €/MWh

A l'échelle du portefeuille (clients C2 et C4), on a des primes estimées pour 2026 en majorité négatives (qui ne constitue pas une perte pour le fournisseur mais une amélioration de sa marge à prix fixe moyen 2026 encore au-dessus des normales de prix de vente avant crise). Mais on constate cependant, une prime moyenne élevée en plein cœur de la crise énergétique en 2022 à hauteur de 10,36€/MWh (de perte estimée par MWh achetée ou vendu en fonction respectivement des cas de sous-consommation ou surconsommation).

La projection du calcul des primes permet de quantifier en avance les potentiels pertes sur la vente ou l'achat d'énergie afin de faire face à la demande. Elle sert d'outil de tarification transparent et de guide pour une meilleure gestion des risques mais chaque méthode comprend des limites qui nécessitent des choix stratégiques et réalistes en termes de quantification.

4. Recommandations stratégiques et limites de l'étude

Cette étude a permis de prendre du recul sur la gestion des risques de consommation au sein d'Énergie d'Ici et de pouvoir émettre des recommandations tout en présentant les limites et les perspectives d'évolution de la démarche.

4.1. Recommandations concrètes

Le calcul des primes de risque sur différentes périodicités (annuelle, trimestrielle, mensuelle) permet de mieux appréhender les effets de foisonnement tant au niveau global du portefeuille qu'à l'échelle des segments. Cela permet de mesurer l'impact relatif de ce segment sur le risque total, et par conséquent identifier les marges de manœuvre possibles lors de négociation de contrats par exemple. Ces calculs permettent également de détecter les variations liées à des facteurs exogènes tels que les périodes hivernales dépendantes de la température, les périodes de canicule, les crises ponctuelles ou encore les

évolutions réglementaires. Cette approche met en évidence la diversité des profils de consommation et l'importance de préserver cet équilibre : plus la composition du portefeuille reste hétérogène, plus l'effet de foisonnement contribue à réduire le risque global.

L'identification des groupes à risque au sein des segments permettra d'isoler ces clients et de calculer des primes de risque spécifiques notamment pour les clients qui présentent une consommation imprévisible ou fortement volatile qui augmentent le risque global du portefeuille. Il sera alors pertinent d'effectuer le calcul des primes par PDL. A l'inverse, les clients à risques faibles c'est-à-dire plus stables, mieux prévisibles ou faiblement volatils peuvent avoir un effet bénéfique en réduisant le risque global ce qui ouvre la voie à des ajustements commerciaux sous forme de réductions (en proposant une tarification plus avantageuse sur la base de la prime calculée permettant de diminuer le risque global) ou de bonus tarifaires (primes réduites ou négatives) dans le cadre de négociations / renégociations contractuelles ou fidélisation clients.

Dans un souci de représentativité plus dynamique des conditions économiques, il sera pertinent de revaloriser et mettre à jour régulièrement les primes (sur une base trimestrielle ou semestrielle) afin d'intégrer notamment les nouveaux entrants et sortants du portefeuille, les évolutions conjoncturelles (crise, changement réglementaire, variations de prix de marché) ainsi que les particularités saisonnières (pics hivernaux, canicules).

Le calcul de ces primes permet d'avoir un élément de la facturation plus transparent pour les clients qu'une prime standard ne reflétant pas le profil du client ni son impact dans le risque global. Cela rend la tarification plus lisible notamment pour la compréhension de la constitution des risques.

D'un point de vue financier, la quantification précise des primes favorise une meilleure gestion des provisions, tant au niveau du segment que du client individuel. Cela ouvre la voie à une gestion plus fine des risques et à une sécurisation accrue des marges.

Enfin, la constitution de bases de données robustes et actualisées permettra de garantir la fiabilité des analyses et d'améliorer la qualité des modélisations. La rigueur de la collecte et la régularité des mises à jour conditionneront donc la pertinence des résultats et leur capacité à éclairer la stratégie de gestion des risques.

4.2. Limites générales

Toute démarche de calculs et de modélisations comporte nécessairement des limites tant liées au contexte conjoncturel qu'aux hypothèses de travail admises.

L'étude s'est basée sur l'année 2024 en sortie de crise énergétique où le marché a gardé encore des traces de cette crise avec une décroissance des prix importante mais une volatilité toujours aussi marquée. De ce fait, les primes de risque, et notamment de risque de marché, tendent vers des primes très faibles voire positives (c'est-à-dire qu'il n'y a pas de risque financier au vu du comportement du marché spot qui rend son usage plus avantageux que les prix à terme qui était également encore fortement impacté par la crise). Cette situation atypique reflète un contexte transitoire qui ne saurait être extrapolé à long terme : l'achat de volume sur le marché spot, a été temporairement plus avantageux que l'achat via des contrats de gré à gré ou les marchés à terme ce qui fait peser encore un coût de l'énergie élevé sur les clients par rapport à la réalité actuelle. De ce fait, ce contexte exceptionnel ne reflète pas une situation structurelle et limite la portée des résultats dans le temps.

De plus, la qualité et la disponibilité des données ont limité l'analyse et la modélisation sur tous les segments des risques. En effet, le manque d'informations détaillées sur le segment C5 où seuls les volumes globaux étaient disponibles mais non répartis par client, ont réduit la finesse des analyses à l'échelle du portefeuille. Quelques contraintes de données ont également été rencontrées durant cette analyse notamment sur l'absence temporaire de certaines données ou lignes de données dans une base, due à des maintenances ou mise à jour des sites hébergeurs.

Certaines méthodes de calcul peuvent également poser question ou remettre en cause la robustesse des analyses notamment le choix des formules de calculs, des modèles d'estimation et de prévision ou des Features utilisées et construits. D'autres variables explicatives pourraient également être prises en compte pour l'entraînement des modèles et l'amélioration des métriques d'évaluation. La méthode d'imputation des volumes prévisionnelles pour les segments C2 et C4 lors de la partie d'analyse du risque peut également manquer de justesse. En effet, cette méthode se base sur le calcul du volume prévisionnel de chaque PDL en fonction de leur part de consommation réelle dans la consommation globale réelle du segment au pas horaire. Cette méthode a permis de reconstruire les volumes prédits par heure et par PDL mais peut sous-estimer ou surestimer la consommation réellement prédite d'un PDL créant potentiellement des écarts plus ou moins importants sans pouvoir vérifier s'ils sont effectivement corrects ce qui est le cas avec une majoration anormale d'écarts positifs par PDL.

Ces calculs et modèles se basent sur certaines hypothèses de travail. En effet, dans la troisième partie, on a considéré la volatilité journalière comme stable et constante sur l'année ce qui n'est pas le cas en réalité. On a également utilisé un prix de vente (prix de contrats des clients) moyen par segment pour simplifier les calculs mais il ne reflète pas la diversité réelle des prix de chaque contrat pour chaque client et ne prend pas en compte les clients en segment C5 (en l'absence de ces données). Nous avons utilisé la même logique pour les prix à terme (prix d'achat sur le marché à terme ou négocié avec les producteurs) avec un prix moyen par année (moyenne des prix sur 2024 pour l'estimation et sur 2026 pour la modélisation).

On ajoute à cela, la difficulté à anticiper les variables exogènes. En effet, les effets de la météo, du changement climatique, des évolutions réglementaires, ou encore des comportements de consommation et des dynamiques du portefeuille (entrées/sorties de clients) sont difficilement anticipables. Ces incertitudes limitent la validité des projections à long terme (2026 et au-delà) et augmentent le risque de biais dans l'évaluation des primes.

Ces limites rappellent que les résultats doivent être interprétés avec prudence et replacés dans le contexte particulier de l'année étudiée.

4.3. Perspectives

Ce travail peut faire l'objet d'amélioration en prenant en compte les limites présentées au-dessus et également d'autres pistes d'amélioration possibles.

Le calcul des primes pourrait être raffiné à échelle de temps trimestrielle ou mensuelle plutôt qu'annuelle notamment pour faciliter l'intégration des variations saisonnières, des entrées de nouveaux clients en cours d'année et des évolutions conjoncturelles. Reste à arbitrer entre un lissage des primes (favorisant la stabilité et la répartition des coûts sur une année) ou une transparence de la tarification (révélant des primes très élevées à certaines périodes et nulles, voire négatives, à d'autres).

Dans la continuité des calculs plus précis des primes de risque, il est pertinent de les calculer en fonction des clusters déterminés pour distinguer les clients très risqués de ceux moins risqués afin d'avoir des primes plus adaptés au profil risque de chacun. Pour les clients très risqués, on pourrait encore segmenter à l'intérieur du groupe afin de déterminer une prime personnalisée pour ces PDL impactant négativement la structure de risque du portefeuille. On peut avoir la même logique pour les clients à risques faibles, contribuant positivement à l'effet de foisonnement pour calculer une réduction (prime négative sous forme de « bonus ») sur le coût de l'énergie.

Dans le cadre des modèles de prévision de consommation, les calculs des volumes prédits pourraient être effectués par PDL et non juste globalement par segment (un calcul qui est très utile cependant pour l'agrégation des blocs de volume d'énergie à acheter) afin de déterminer les écarts par PDL et leur contribution effective au risque global du portefeuille.

Ces prévisions d'ailleurs peuvent comporter des biais et des incertitudes qui ne permettent pas de déterminer avec exactitude le montant des primes de risque à appliquer, le montant se vérifiant que ex-post c'est-à-dire après la consommation réelle effective et la détermination des écarts par la suite. De ce fait, il est pertinent d'appliquer une marge de sécurité au montant global de la prime de risque soit arbitrairement (+10% par exemple) soit sur la base de l'erreur moyenne de prévision ou la volatilité moyenne de la consommation du portefeuille par exemple.

Dans la même logique d'incertitude, pour les clients profilés du segment C5, on pourrait utiliser des méthodes de reconstitution de courbe de charge avec des coefficients dynamiques (et non statiques comme c'est le cas actuellement) ce qui permet de prendre en compte les données d'historique récentes de consommation, les variations saisonnières, le changement de comportement de consommation et les événements (tels que les crises ou changement réglementaire) par exemple. Cela permettra donc d'avoir une définition de la courbe de charge par profil plus dynamique par heure et mis régulièrement à jour (Enedis fournit ces coefficients et les met régulièrement à jour sur une périodicité de publication hebdomadaire). On pourrait également, dans le cadre du calcul de la prime de risque de Flex, creuser le traitement et la reconstruction des données des clients du segment afin de déterminer une prime par segment puis par PDL.

L'analyse gagnerait à être élargie sur les autres risques non-traités mais mentionnés en première partie permettant de mieux capturer l'ensemble des potentiels coûts pour l'entreprise mais aussi d'anticiper certains changements comme les changements réglementaires avec la fin du dispositif de l'ARENH au 1^{er} janvier 2026.

La spécificité d'Énergie d'Ici pourrait également être prise en compte notamment dans le cadre de la diversification des sources d'approvisionnement permettant de réduire le recours à la bourse d'électricité. En effet, l'entreprise est à la fois fournisseur et producteur lui permettant de sécuriser des prix fixes d'achat d'énergie et de limiter son exposition aux fluctuations de long terme (encore marquées par la crise énergétique récente) via la négociation directe avec les producteurs.

Enfin, des approches plus globales peuvent être envisagées :

- Réaliser des stress tests pour simuler des scénarios extrêmes (hausse brutale des prix, crise énergétique, arrivée massive de clients atypiques, modification structurelle du mix énergétique).
- Développer une synergie entre production et consommation, afin de tirer parti d'un foisonnement global et d'adapter la stratégie de couverture.

- Optimiser la composition du portefeuille clients, en limitant les clients à risque ou en les isolant. On pourrait mesurer en amont la prévisibilité et la volatilité des clients sur la base de leurs données historiques quand cela est possible afin d'au mieux piloter le portefeuille.
- Intégrer les évolutions technologiques (développement de technologie de stockage par batterie ou hydrogène, flexibilité de la demande) qui permettraient d'avoir une stratégie de gestion de l'équilibre moins dépendante des prix du marché spot et de limiter donc les risques.

Conclusion

À travers ce mémoire, nous avons mis en évidence l'importance cruciale de la prise en compte des divers risques auxquels l'entreprise Energie d'Ici est confrontée, ainsi que de leur quantification précise pour le portefeuille de consommation, dans un environnement particulièrement volatil qu'est le marché de l'électricité. L'analyse de ces risques révèle leur nature multiple et interconnectée. D'une part, le risque de marché, lié aux variations du prix spot, présente un niveau de volatilité élevé et des difficultés importantes de prévision. D'autre part, le risque de flexibilité (Flex), fondé sur les écarts de consommation, est fortement dépendant du profil horaire de consommation des clients, du moment où surviennent les écarts, de la saisonnalité et des conditions météorologiques. Ce dernier s'avère être le plus critique, avec des primes de risque plus élevées que celles de marché, reposant sur des variables simulées très volatiles et incertaines à l'avance, comme la consommation réelle et ses écarts potentiels.

L'étude a permis d'analyser l'impact de chaque segment de clients sur le risque global et de le quantifier financièrement, tout en proposant une approche alternative de regroupement des clients au sein d'un segment, axée sur le profil de risque. Elle a également mis en lumière les effets de foisonnement, à l'échelle annuelle et au sein du portefeuille, qui permettent de réduire le risque global grâce à la mutualisation des écarts entre clients. Ces éléments démontrent l'importance de considérer non seulement les risques individuels mais également leur interaction au sein du portefeuille.

Les différentes méthodes de modélisation appliquées ont permis de définir une prime de risque moyenne, tant historique (année 2024) que prospective (année 2026), pour chacun des risques étudiés. L'utilisation combinée de techniques statistiques classiques et de modélisations plus avancées a permis d'évaluer les écarts de consommation potentiels et les prix spot futurs, tout en soulignant le rôle central de la qualité des données et de leur structuration dans la fiabilité des calculs.

Il apparaît donc essentiel pour l'entreprise d'intégrer l'identification et la quantification de ces risques dans ses prévisions de prix et de consommation, ainsi que dans sa stratégie de tarification. Une telle intégration permettrait non seulement de se couvrir contre les fluctuations de la demande et des prix de marché, mais également d'optimiser la rentabilité et la compétitivité tout en respectant les obligations réglementaires.

Enfin, plusieurs axes d'amélioration et perspectives ont été proposés pour enrichir cette étude, notamment l'exploration de nouvelles méthodes de simulation, la prise en compte de risques complémentaires liés à la production et à la demande électrique, et le raffinement des modèles existants pour mieux capturer la volatilité et l'incertitude inhérentes au marché de l'énergie. Dans un contexte global incertain, tant sur les plans économique, politique qu'énergétique, une gestion rigoureuse et proactive de ces risques constitue un levier stratégique majeur pour l'entreprise, permettant de sécuriser ses opérations et d'assurer sa pérennité à long terme.

Bibliographie

ACCARIER Etienne, Grenoble Ecole de Management, 2020, Les énergies renouvelables : un défi pour les acteurs de marché [Thèse]

BIENVENUE Alexis, 2023, Python [Cours universitaire]

BURGER Markus, GRAEBER Bernhard et SCHINDLMAYR Gero, Wiley Finance Series, 2014, Managing Energy Risk, Chapitres 2 à 4 et 6 à 7 [Livre]

CLOT Denis, 2023, Data Mining Avancé – parallélisation des calculs et optimisation du temps de calcul [Travaux Pratiques]

Connaissances des énergies, 14 août 2024, Qu'est-ce que la logique de « merit order » et des coûts marginaux de production électrique ? <https://www.connaissancedesenergies.org/questions-et-reponses-energies/electricite-quest-ce-que-la-logique-de-merit-order>

Connaissances des énergies, 4 juillet 2024, Prix de l'électricité en France, <https://www.connaissancedesenergies.org/fiche-pedagogique/tarification-de-lelectricite>

CRE, 10 juin 2025, Rapport d'activité 2024, <https://www.cre.fr/documents/rapports-et-etudes/rapport-dactivite-2024.html>

CRE, 26 novembre 2024, Analyse de la CRE sur le phénomène de prix de l'électricité négatifs et recommandations relatives aux dispositifs de soutien aux énergies renouvelables, <https://www.cre.fr/documents/rapports-et-etudes/analyse-de-la-cre-sur-le-phenomene-de-prix-de-lelectricite-negatifs-et-recommandations-relatives-aux-dispositifs-de-soutien-aux-energies-renouvelables.html> [Note d'analyse]

DATA CAMP, Formations en ligne sur la programmation et la gestion des risques de portefeuille : Manipulating Time Series Data in Python, Intermediate Python for Finance, Introduction to SQL, Introduction to Portfolio Risk Management in Python, Time Series Analysis in Python, Quantitative Risk Management in Python, Machine Learning for Finance in Python, <https://app.datacamp.com/learn> [Formations en ligne]

DE LATOUR Arnaud (EDF), Février 2016, Electricity markets [Présentation PowerPoint]

DULDINGER Lukas, 18 avril 2022, Detriments of the Merit Order - Surging Electricity Prices, <https://courses.renewablesvaluationinstitute.com/pages/academy/surging-electricity-prices-detriments-of-the-merit-order>

EDF, Histoire, L'aventure EDF, <https://www.edf.fr/groupe-edf/edf-en-bref/histoire>

ELSAYED Shereen, THYSSENS Daniela, RASHED Ahmed, JOMAA Hadi Samer and SCHMIDT-THIENNE Lars, University of Hildesheim (Germany), Department of Computer-Science, Octobre 2021, Do We Really Need Deep Learning Models for Time Series Forecasting ? [Article]

ENEDIS, Glossaire, <https://www.enedis.fr/faq/glossaire>

ENEDIS, Le marché électrique français, <https://www.enedis.fr/nous-connaître/le-marche-de-lelectricite>

Énergie d'Ici, Blog D'Ici, 19 mai 2023, Comprendre le modèle de vente directe d'électricité, avec Antoine Garcier, Directeur Général d'Énergie d'ici, https://www.energiesdici.fr/blog-dici/272_comprendre-le-modele-de-vente-directe-delelectricite-energie-dici

Énergie d'Ici, Blog D'Ici, 20 janvier 2023, Hausse des prix de l'électricité : mieux comprendre et savoir à quoi s'attendre, [Hausse des prix de l'électricité : mieux comprendre et savoir à quoi s'attendre](#)

Énergie d'Ici, Blog D'Ici, 29 février 2024, Les énergies renouvelables : états des lieux et nouvelles formes, https://www.energiesdici.fr/blog-dici/4789_les-energies-renouvelables-etats-des-lieux-et-nouvelles-formes

Énergie d'Ici, Blog D'Ici, 4 novembre 2021, Comment expliquer la hausse des prix de l'énergie, https://www.energiesdici.fr/blog-dici/66_comment-expliquer-la-hausse-des-prix-de-lenergie

ERDF, RTE, Règles relatives à la Programmation, au Mécanisme d'Ajustement et au dispositif de Responsable d'Équilibre, Section 2, Chapitre F, Relations entre le GRD et le Responsable d'Équilibre, Dispositions applicables pour l'estimation des courbes de charge [Rapport]

ESSOMBA Ronald, Grenoble Ecole de Management, 2019, Comment un fournisseur d'électricité peut-il estimer l'exposition financière et volumétrique de sa SHAPE, courbe au pas horaire des volumes non achetés lors de la couverture de son portefeuille sur les marchés à terme ? [Thèse]

FICHEUX Françoise, BERNARD Jean-Marie, DELBECQ Jean-Michel et ZANOLIN Ludovic, Institut IED Énergie et Développement, Juillet 2017, Etude sur l'équilibre entre l'offre et la demande sur le réseau d'électricité [Etude]

FINON Dominique, DEFEUILLEY Christophe et MATRY Frédéric, Revue Économie et prévision, Firms, marchés et innovation, 2011/1, n°197-198, Pages 81 à 104, Signaux-prix et équilibre de long terme : reconsidérer les formes d'organisation sur les marchés de l'électricité, <https://shs.cairn.info/revue-economie-et-prevision-1-2011-1-page-81?lang=fr&tab=texte-integral> [Revue]

GRUET Pierre, VARIANCES, 27 novembre 2023, Spécificités des modèles de prix de l'électricité pour la gestion des risques de marché [Article]

HANSEN Jean-Pierre, PERCEBOIS Jacques et JANSSENS Alain, DEBOECK SUPERIEUR, 3^e édition, Énergie – Économie et Politique, Chapitres 1, 2, 5 et 8 [Livre]

HAYA Energy Solutions, Octobre 2023, Achat d'électricité et gestion du risque après la crise de 2022, le cas du contrat « Bloc + Spot » [Note d'analyse]

IDEA "Intelligence Data Exploiting Agency", 15 février 2024, Medium, Les algorithmes de Machine Learning pour la prévision des séries temporelles – Partie I, <https://medium.com/@ideaacademy/les-algorithmes-du-machine-learning-pour-la-pr%C3%A9vision-des-s%C3%A9ries-temporelles-partie-i-2b75abae4087> [Article]

LIU Min et WU Felix F., ScienceDirect, 7 novembre 2006, Portfolio optimization in electricity markets [Article]

LUDVIGSEN Kasper G. A., 26 octobre 2022, towards data science, Multi-step time series forecasting with XGBoost, <https://towardsdatascience.com/multi-step-time-series-forecasting-with-xgboost-65d6820bec39/> [Article]

MAUME-DESCHAMPS Véronique, 2024-2025, Agrégation par mélange de lois, Bootstrap, Forêts aléatoires et Séries temporelles [Cours universitaire]

MICROSOFT, Copilot, Outil d'aide au développement et à la correction de code

MINISTERE DE LA TRANSITION ECOLOGIQUE ET SOLIDAIRE, Programmation Pluriannuelle de l'énergie 2019-2028, Synthèse, <https://www.ecologie.gouv.fr/sites/default/files/documents/20200422%20Synthe%CC%80se%20de%20la%20PPE.pdf>

MORALES Juan M., CONEJO Antonio J., MADSEN Henrik, PINSON Pierre et ZUGNO Marco, Springer, Volume 205, 2014, Integrating Renewables in Electricity Markets [Livre]

NABATCHIAN Ehsan, 10 janvier 2024, Medium, Exploring Time Series Prediction of Energy Consumption Using XGBoost and Cross Validation, <https://medium.com/@ehsannabatchian/exploring-time-series-prediction-of-energy-consumption-using-xgboost-and-cross-validation-5d299655bec6> [Article]

OHM ENERGIE, Blog, L'ARENH : son rôle sur le marché de l'électricité et sa disparition fin 2025, <https://ohm-energie.com/blog/arenh>

OPENAI, ChatGPT, Versions 4.1 et 5, Outil d'assistance pour la reformulation, la révision et la correction lors de la rédaction du mémoire

PERCEBOIS Jacques, Revue d'économie politique, 2001/6, Volume 111, Pages 815 à 860, Énergie et théorie économique : un survol, <https://shs.cairn.info/revue-d-economie-politique-2001-6-page-815?lang=fr#s1n3> [Revue]

RTE, 15 mai 2022, Le transport d'électricité, comment ça fonctionne ?, [Le transport d'électricité, comment ça fonctionne ? | RTE](#)

RTE, Bilan électrique 2023 Principaux résultats, <https://assets.rte-france.com/analyse-et-donnees/2025-03/BE2023%20-%20Synth%C3%A8se.pdf>

RTE, Bilan électrique 2024 Rapport complet, <https://assets.rte-france.com/analyse-et-donnees/2025-05/BE2024%20-%20Rapport%20Complet.pdf> [Rapport]

RTE, Le transport d'électricité, comment ça fonctionne ? [Le transport d'électricité, comment ça fonctionne ? | RTE](#)

RTE, Reconstituer les flux de responsabilité d'équilibre, <https://www.services-rte.com/fr/decouvrez-nos-offres-de-services/le-role-des-gestionnaires-de-res/reconstituez-les-flux-re.html>

RUFFIÉ Luc, ENERGYSCAN, ENGIE Global Markets, Février 2025, Power market [Ebook]

SIA PARTNERS, Energies & Environnement, 24 mars 2014, Le profil de consommation, un outil au centre de la stratégie d'approvisionnement des fournisseurs [Article]

SIDI Sidi, Grenoble Ecole de Management, 2020, La « Flexibilité », déviation structurelle de la consommation d'un portefeuille clients, un risque majeur pour les fournisseurs d'électricité. [Thèse]

STACK OVERFLOW, Plateforme d'échange technique et résolution de bugs en programmation Python, Rubrique Questions [Site web]

VAROQUAUX William, Que sais-je ? Novembre 1996, Calcul économique et électricité [Livre]

WILMOTT Paul, MATHIEU Francis et PONCET Patrice, Edition Eyrolles, 2008, *La Finance quantitative en 50 questions* [Livre]

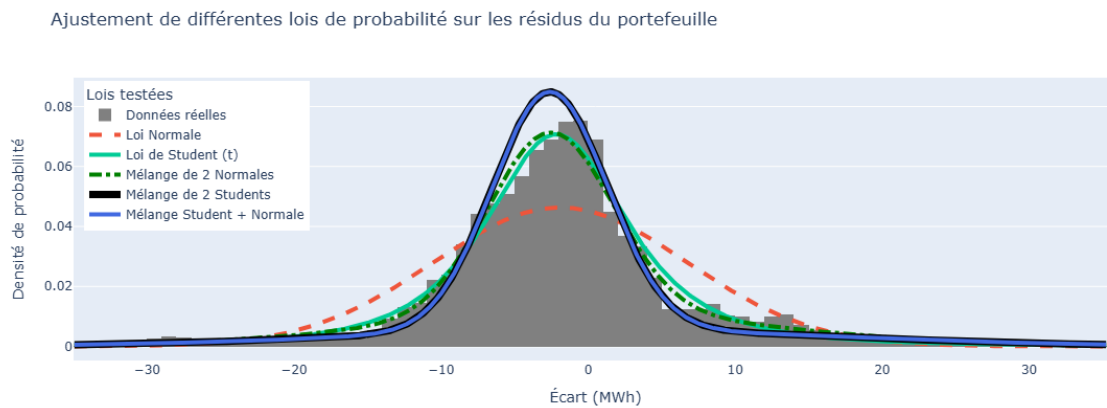
Pour les données externes :

- Prix spot et forward : EEX_8267@datasource.eex-group.com
- Météo : <https://meteonet.umr-cnrm.fr/>
- Calendrier scolaire : <https://www.data.gouv.fr/datasets/le-calendrier-scolaire/>

Accès aux codes réalisés dans le cadre de ce mémoire : <https://github.com/Meryam5/Memoire>

Annexe 1

Ont été testées la loi Student et trois mélanges de loi : deux lois normales, deux lois Student et une loi normale avec une loi Student. Elles sont présentées en comparaison d'une loi normale simple et de la distribution des résidus sur le graphique suivant :



Note de lecture : Le mélange de deux lois Student se confond avec le mélange d'une loi Student et d'une loi normale.

Le choix de ces lois se basent sur la forme des résidus qui présente des queues épaisses de chaque côté ce qui se rapproche d'une loi Student mais avec un noyau fort au centre proche d'une loi normale mais pas totalement en prenant en compte une certaine asymétrie de la distribution.

Ce graphique nous montre que la loi Student et le mélange de lois normales semblent mieux correspondre à la forme des données, capturant plus précisément les pics et les queues épaisses que la loi normale simple. Cette observation sera utile pour la suite de la modélisation en troisième partie.

Annexe 2

On a constaté que les séries temporelles des écarts par segment et à l'échelle du portefeuille sont hétéroscédastiques. De ce fait, la volatilité n'est pas constante. On a observé des périodes de forte volatilité (avec de grands écarts) suivies par d'autres périodes similaires mais non constantes. On a des sortes de clusters de volatilité sur la durée dans le temps à volatilité différente. Il est donc pertinent d'appliquer un modèle GARCH pour estimer cette volatilité.

Ce modèle va permettre de modéliser et prévoir la variance conditionnelle des résidus qui correspond à la volatilité propre des écarts de consommation sur une période.

On commence en testant un modèle GARCH simple de paramètre $(p, q) = (1, 1)$ qui signifie :

- $p = 1$: la volatilité d'aujourd'hui dépend de la taille de l'erreur d'hier
- $q = 1$: la volatilité d'aujourd'hui dépend de la prévision de la volatilité d'hier

L'étude des lois précédentes nous positionne sur l'usage d'une distribution de Student plus proche de la forme des données.

On lance le modèle et on obtient les résultats suivants :

```

Constant Mean - GARCH Model Results
=====
Dep. Variable:                Total    R-squared:                0.000
Mean Model:                   Constant Mean  Adj. R-squared:          0.000
Vol Model:                    GARCH      Log-Likelihood:         -26554.4
Distribution: Standardized Student's t  AIC:                    53118.8
Method:                       Maximum Likelihood  BIC:                    53154.2
Date:                          mer., sept. 03 2025    No. Observations:       8784
Time:                          11:00:23             Df Residuals:           8783
                               Mean Model  Df Model:               1
=====
                coef    std err          t      P>|t|   95.0% Conf. Int.
-----+-----+-----+-----+-----+-----
mu            -1.9802   9.950e-02   -19.901  4.002e-88  [-2.175, -1.785]
                               Volatility Model
=====
                coef    std err          t      P>|t|   95.0% Conf. Int.
-----+-----+-----+-----+-----+-----
omega         3.9981     0.190       21.043  2.663e-98  [ 3.626,  4.370]
alpha[1]      1.0000   1.555e-02    64.328  0.000     [ 0.970,  1.030]
beta[1]       0.0000   2.067e-03     0.000  1.000    [-4.052e-03,4.052e-03]
                               Distribution
=====
                coef    std err          t      P>|t|   95.0% Conf. Int.
-----+-----+-----+-----+-----+-----
...
nu            326.3417   25.992       12.556  3.708e-36  [2.754e+02,3.773e+02]
=====
Covariance estimator: robust

```

Le biais de la prévision (encadré « Mean Model ») est significatif (p-value inférieur à 5%) et est égal à -1,98 MWh. On a un biais systématique où le modèle surestime la consommation de 1,98 MWh en moyenne par heure.

Mais, lorsqu'on regarde en détail le second encadré (« Volatility Model »), on a deux coefficients incohérents :

- $\alpha[1] = 1.00$ et significatif à 5% : Ce coefficient mesure l'impact d'un choc passé sur la volatilité actuelle. Or, le fait qu'il soit à 1 indique que la volatilité d'aujourd'hui dépend entièrement du choc d'hier, sans aucune atténuation ce qui n'est pas réaliste. On a une volatilité explosive et non stationnaire où la variance ne revient jamais à la normale ce qui, en réalité, n'est pas possible.
- $\beta[1] = 0.00$ et non-significatif : Ici, la volatilité passée n'exerce aucune influence sur la volatilité actuelle, ce qui est totalement incohérent avec le principe du modèle GARCH.

In fine, le modèle ne capture pas correctement la structure de volatilité des écarts ni sa dynamique. On doit donc au préalable corriger l'autocorrélation des résidus (qu'on avait constaté auparavant) qui fausse l'estimation des paramètres GARCH. En effet, le modèle est conçu pour modéliser la variance conditionnelle d'une série de résidus « propres » c'est-à-dire, des résidus qui ne comportent pas de dépendance temporelle.

Pour corriger cela, on applique un modèle SARIMA en amont qui a réussi à rendre les résidus plus proches du bruit blanc et sans dépendance temporelle. On peut donc ensuite appliquer le modèle GARCH sur les résidus SARIMA et on obtient les résultats suivants :

Constant Mean - GARCH Model Results					
Dep. Variable:	None	R-squared:	0.000		
Mean Model:	Constant Mean	Adj. R-squared:	0.000		
Vol Model:	GARCH	Log-Likelihood:	-13032.3		
Distribution:	Standardized Student's t	AIC:	26074.5		
Method:	Maximum Likelihood	BIC:	26109.9		
Date:	mer., sept. 03 2025	No. Observations:	8784		
Time:	11:03:33	Df Residuals:	8783		
	Mean Model	Df Model:	1		
=====					
	coef	std err	t	P> t	95.0% Conf. Int.

mu	6.0402e-03	8.443e-03	0.715	0.474	[-1.051e-02, 2.259e-02]

Volatility Model					
	coef	std err	t	P> t	95.0% Conf. Int.

omega	0.2674	2.692e-02	9.935	2.918e-23	[0.215, 0.320]
alpha[1]	0.5656	3.963e-02	14.272	3.289e-46	[0.488, 0.643]
beta[1]	0.4344	2.965e-02	14.653	1.280e-48	[0.376, 0.493]

Distribution					
	coef	std err	t	P> t	95.0% Conf. Int.

nu	3.6337	0.152	23.836	1.407e-125	[3.335, 3.932]

Covariance estimator: robust					

Le biais caractérisé par mu est ici non-significatif ce qui indique qu'on n'a pas de biais systématique et que donc, les erreurs de prévision sont centrées autour de 0. Le modèle ne fait donc pas d'erreur et est précis (il n'y a pas de tendance à la sur ou sous-estimation des écarts de consommation).

On observe également que le modèle est bien spécifié avec des coefficients de volatilité tous significatifs. On a les paramètres suivants :

- $\omega = 0,267$: montre un niveau de volatilité de base toujours présent dans la série
- $\alpha[1] = 0,566$ signifie que la volatilité d'aujourd'hui dépend fortement du choc d'hier (erreur précédente). On a ici une valeur de coefficient inférieure à 1 donc le modèle est stationnaire (il ne présente pas de volatilité explosive)
- $\beta[1] = 0,434$ signifie que la volatilité d'hier influence aussi celle d'aujourd'hui (on a un effet de persistance dans le temps (d'une heure à l'autre) de la volatilité)

La somme des deux coefficients alpha et beta est égale à 1. Cela indique qu'on a une volatilité très persistante avec des périodes de forte volatilité qui tendent à durer sur plusieurs heures d'affilés.

Le dernier encadrant sur la distribution avec le coefficient $\nu = 3,63$ et significatif indique qu'on a des résidus avec des queues épaisses c'est-à-dire qu'on a plus d'événements extrêmes que dans une loi

normale. Donc, la distribution Student est adaptée pour modéliser les extrêmes présents dans nos données.

Annexe 3

On effectue un modèle SARIMA sur les écarts de consommation correspondant aux résidus. On automatise dans une boucle la recherche des meilleurs paramètres en bornant la recherche des paramètres AR en fixant d à 0 car la série est stationnaire et en bornant également les paramètres MA en fixant la saisonnalité à $m = 24$ car c'est à ce pas temporel qu'on a une boucle entière. On a 24 heures entre chaque cycle saisonnier.

On le formalise dans le code suivant :

```
# Plages de paramètres à tester pour (p,d,q)
p = range(0, 3)
d = range(0, 1) # d=0 car la série semble déjà stationnaire
q = range(0, 3)
pdq = list(itertools.product(p, d, q))

# Paramètres saisonniers (P,D,Q,m) avec m=24
P = range(0, 2)
D = range(1, 2) # D=1 pour la différenciation saisonnière
Q = range(0, 2)
seasonal_pdq = [(x[0], x[1], x[2], 24) for x in list(itertools.product(P, D, Q))]

print("Recherche manuelle des meilleurs paramètres SARIMA")

best_aic = float("inf")
best_pdq = None
best_seasonal_pdq = None
results_list = []

# Boucle sur toutes les combinaisons de paramètres
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = SARIMAX(residus_originaux,
                          order=param,
                          seasonal_order=param_seasonal,
                          enforce_stationarity=False,
                          enforce_invertibility=False)
            results = mod.fit(dispatch=False)

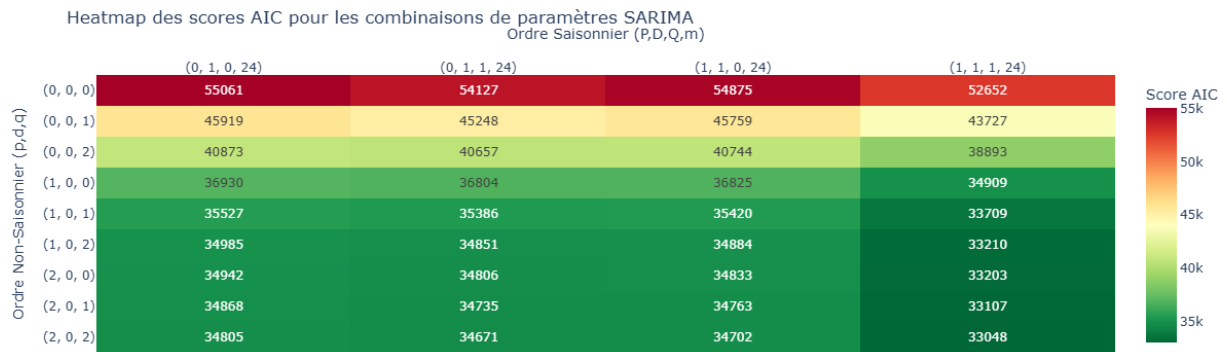
            # Stocker le résultat
            results_list.append({'param': param, 'param_seasonal': param_seasonal, 'aic': results.aic})

            if results.aic < best_aic:
                best_aic = results.aic
                best_pdq = param
                best_seasonal_pdq = param_seasonal

        except Exception as e:
            continue

print(f"\nMeilleurs paramètres SARIMA trouvés : order={best_pdq}, seasonal_order={best_seasonal_pdq}, AIC={best_aic:.2f}")
```

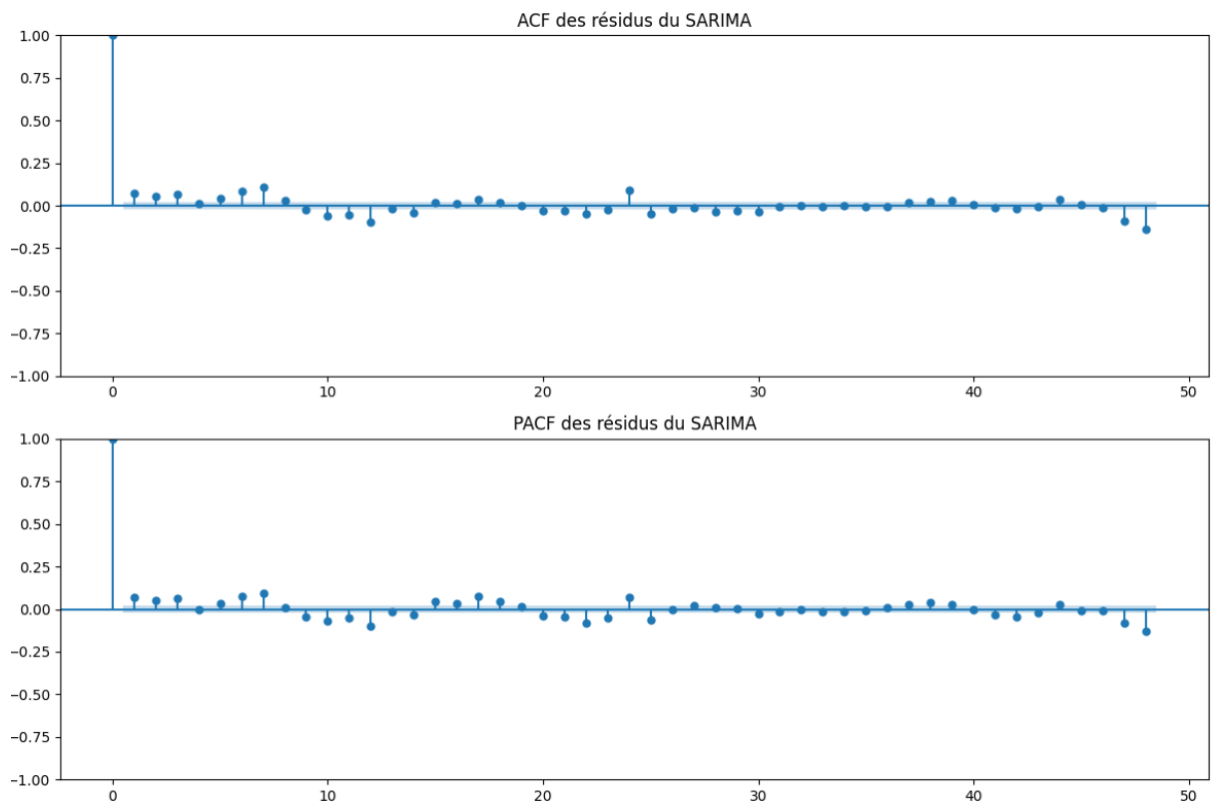
Cela nous permet d'obtenir le couple de paramètres optimal avec les critères AIC les plus faibles qu'on observe dans la Heatmap suivante :



Les meilleurs paramètres sont $(p, d, q) (P, D, Q, m) = (2, 0, 2) (1, 1, 1, 24)$ avec un AIC de 33048.

Annexe 4

Après avoir sélectionné les paramètres, on effectue une analyse des résidus pour voir s'ils sont proches du bruit blanc sans pics significatifs.



Dans le graphique ACF, les barres sont proches de 0 et ne dépassent quasiment pas les seuils de significativité indiquant que les résidus ne présentent plus de dépendance temporelle marquée. Le modèle a donc bien capturé la structure saisonnière des données.

Le second graphique (PACF) confirme qu'il n'y a aucune structure résiduelle significative. On peut donc conclure que les résidus sont assimilables à du bruit blanc.

Annexe 5

On choisit la méthode de K-means qui permet de partitionner un ensemble de données en K groupes distincts de telle sorte que la variance intra-groupe soit minimale (les points d'un même cluster doivent être proches les uns des autres) et que la variance inter-groupe soit maximisée (les clusters doivent être bien séparés).

Pour ce faire, on applique l'algorithme K-means qui suit la procédure itérative suivante :

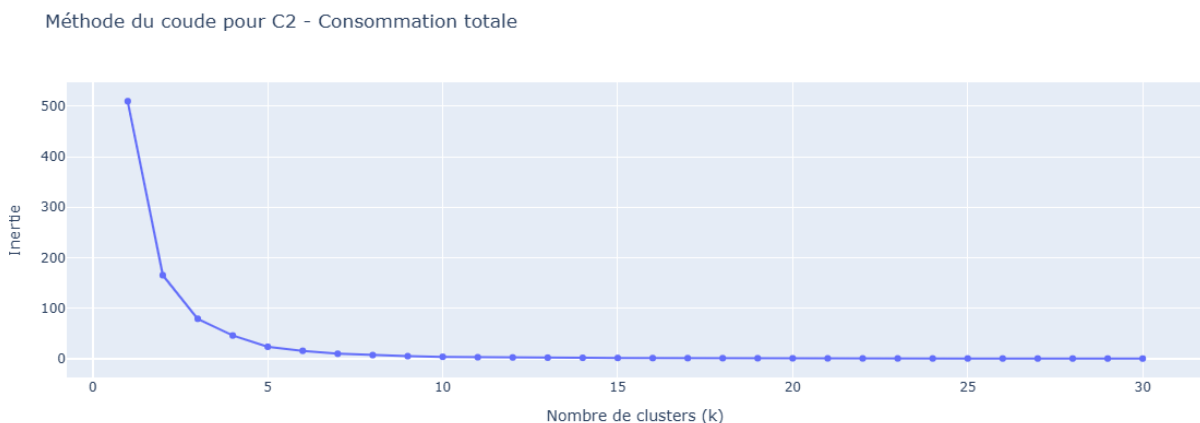
- Initialisation : Choix d'un nombre de clusters K préalable puis initialisation aléatoire de K centroïde
- Affectation : Attribution de chaque observation au cluster dont le centroïde est le plus proche (avec une distance euclidienne dans la programmation effectuée dans cette étude)
- Mise à jour : Recalcul des centroïdes comme la moyenne des observations appartenant à chaque cluster
- Itération : On répète les étapes 2 et 3 jusqu'à qu'il y ait convergence (c'est-à-dire que les centroïdes ne bougent presque plus ou que les attributions de chaque observation dans un cluster ne bougent plus)

Pour choisir le nombre de K optimal, on utilise la méthode du coude qui consiste à :

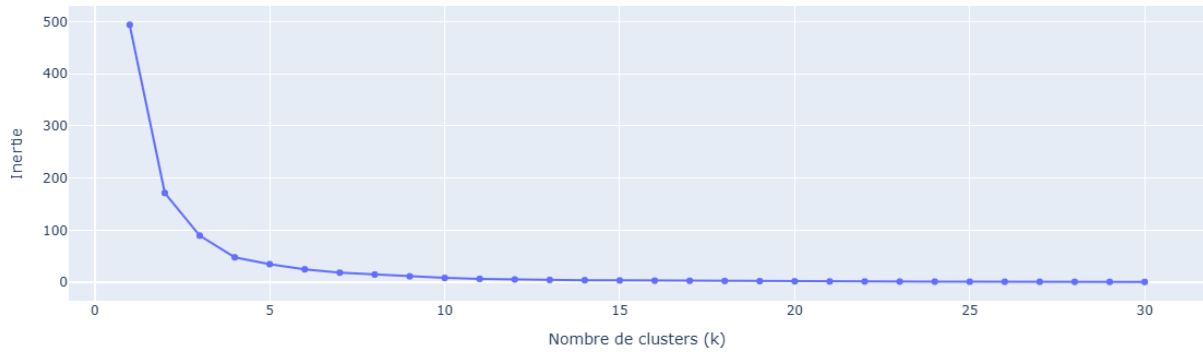
- Calculer le K-means sur plusieurs valeurs de K (de 1 à 30 dans notre cas)
- Pour chaque K, calculer l'inertie intra-groupe qui correspond à la somme des distances au carré entre chaque point et le centroïde de son cluster
- Tracer la courbe avec le numéro de cluster (K) en abscisse et l'inertie en ordonnée
- Puis, observer le point où la diminution de l'inertie ralentit fortement ce qu'on va appeler le coude. C'est ce point qui va représenter le compromis optimal entre simplicité d'étude (peu de clusters) et qualité (suffisamment de clusters homogènes).

Dans cette étude, on a plusieurs graphiques de coude en fonction du segment choisit et de la méthode de classification des PDL (consommation totale, écart absolu relatif moyen et profil de consommation).

Pour le segment C2, on a des graphiques quasiment similaires au niveau du choix du coude :



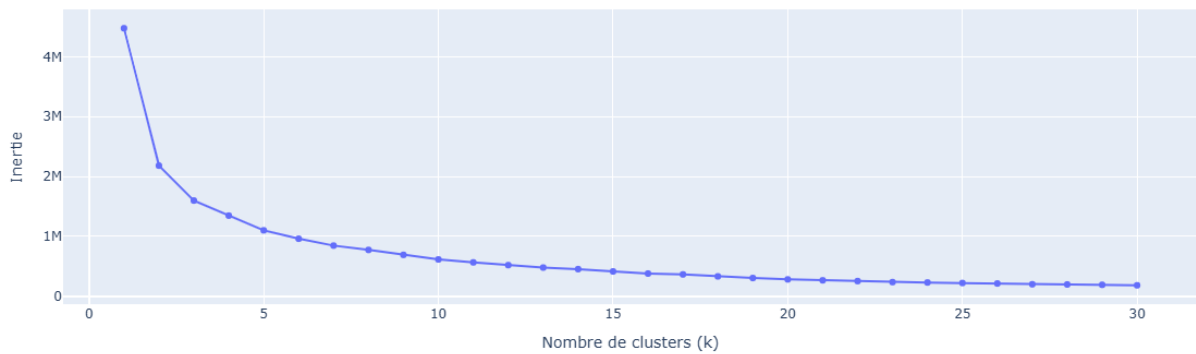
Méthode du coude pour C2 - Risque relatif absolu moyen



A partir de 10 clusters, on n'a pas de baisse d'inertie significative.

En fonction des profils, la baisse est notable au-delà de 20 clusters mais le fait d'avoir un nombre trop important de clusters va complexifier l'étude et ne pas permettre de les interpréter correctement. On l'observe sur le graphique suivant :

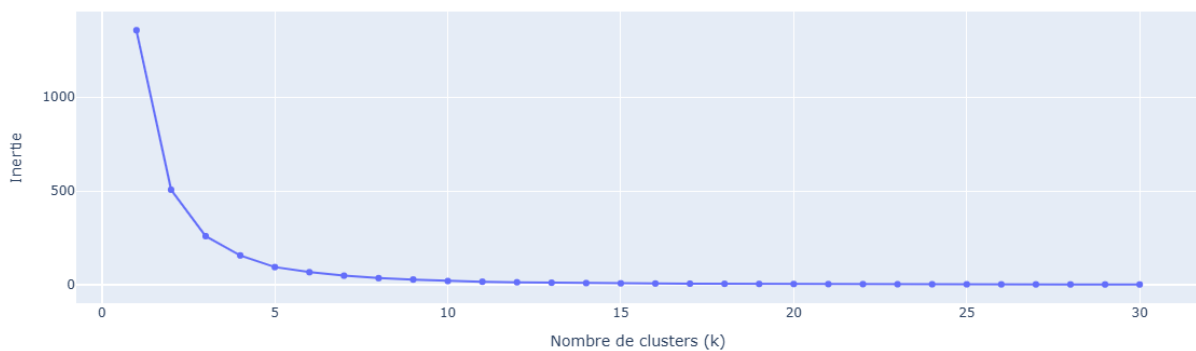
Méthode du coude pour C2 - Profil de consommation



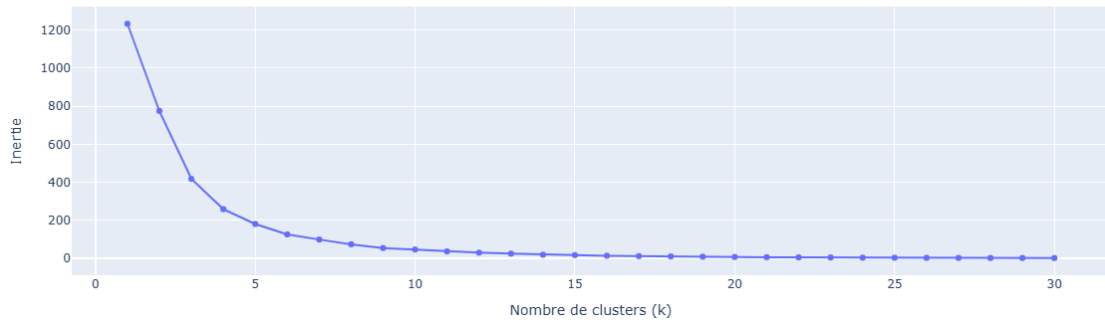
Il ne semble plus y avoir de baisse très significative à partir de 12, nombre sur lequel on s'est basé pour le clustering.

On a les mêmes constats pour le segment C4 avec pour les deux premières méthodes, une stabilité de l'inertie après 10 groupes :

Méthode du coude pour C4 - Consommation totale

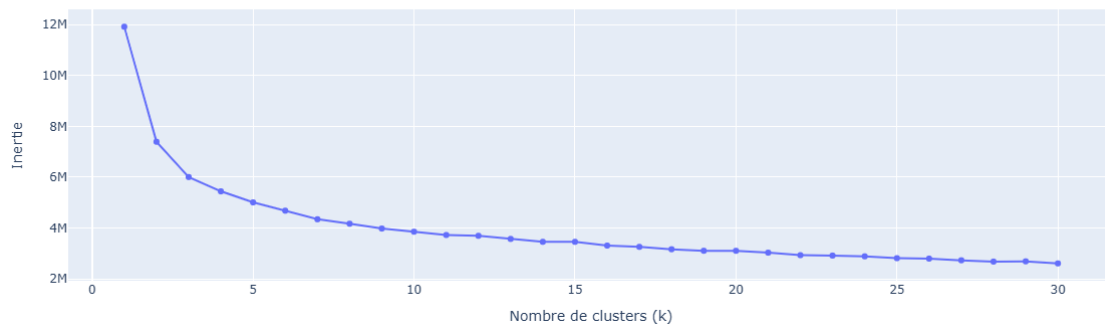


Méthode du coude pour C4 - Risque relatif absolu moyen



Et pour la méthode de groupement avec les profils de consommation, une baisse de l'inertie encore visible après 25 groupes mais moins importante :

Méthode du coude pour C4 - Profil de consommation

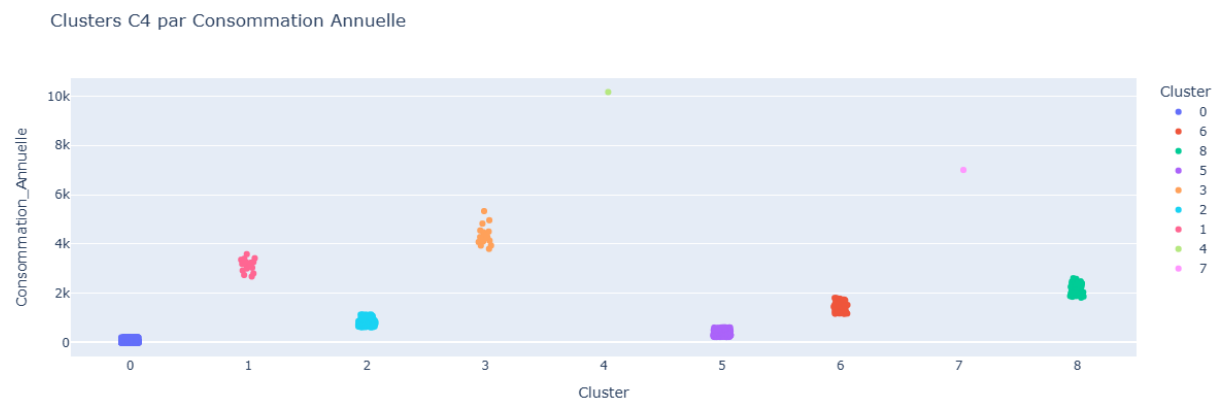


On s'arrête donc à 20 groupes pour ne pas trop partitionner les PDL.

Annexe 6

Le clustering sur le segment C4 suit la même démarche utilisée pour le segment C2.

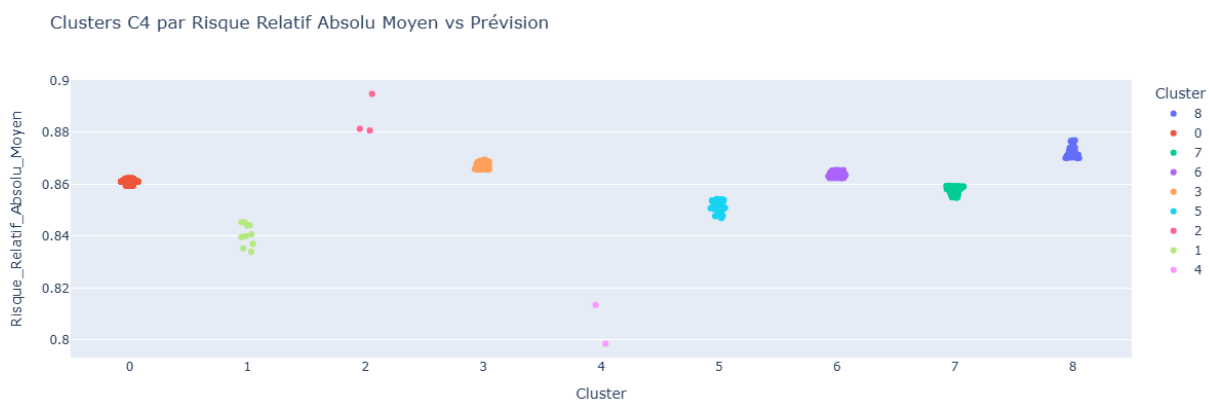
On retrouve dans la décomposition en sous-groupes dans le segment un groupe prépondérant et des plus petits en fonction des conditions de clustering. On a pour les clusters en fonction de la consommation totale :



On a 9 groupes distincts où les PDL sont répartis de la façon suivante :

Résultats du clustering C4 par consommation		
Cluster	Nombre de PDL	Pourcentage
0	741	54,61%
1	17	1,25%
2	178	13,12%
3	17	1,25%
4	1	0,07%
5	269	19,82%
6	85	6,26%
7	1	0,07%
8	48	3,54%

On a la répartition en fonction du risque relatif absolu moyen tout aussi nette et distincte dans le graphique suivant :

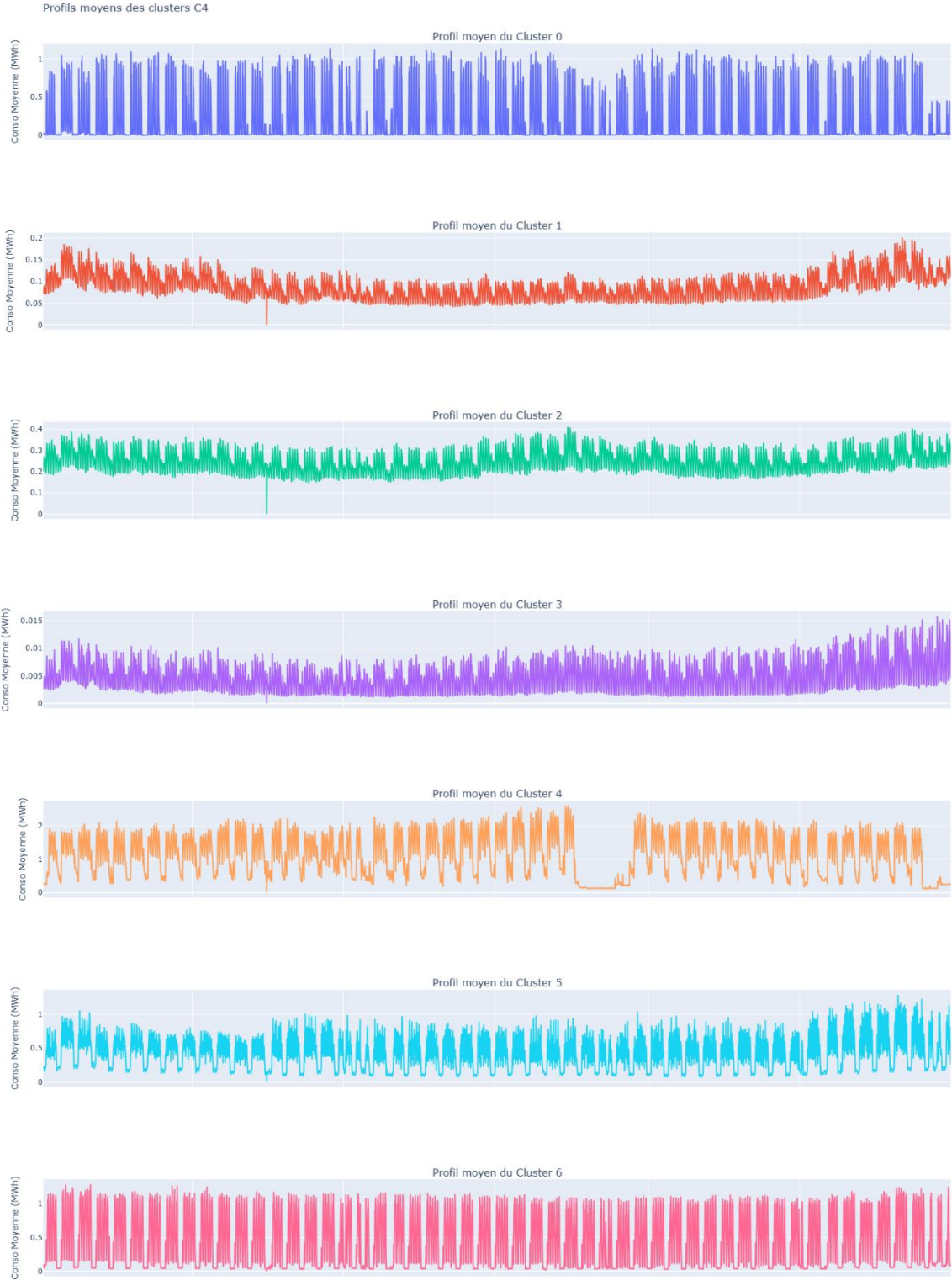


Avec 9 groupes dans lesquels sont répartis les PDL de la façon suivante :

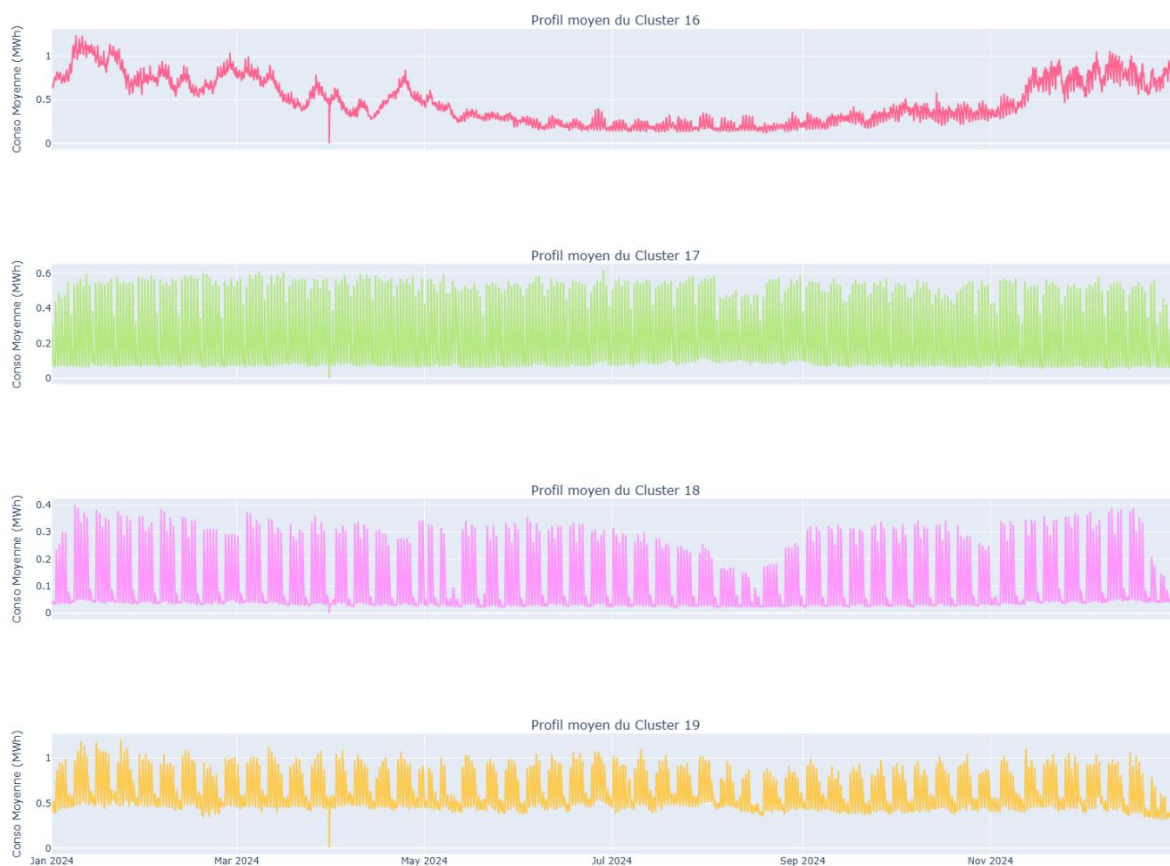
Résultats du clustering C4 par risque relatif		
Cluster	Nombre de PDL	Pourcentage
0	799	64,80%
1	10	0,81%
2	3	0,24%
3	103	8,35%
4	2	0,16%
5	35	2,84%
6	155	12,57%
7	87	7,06%
8	39	3,16%

On garde un groupe prépondérant et correspond à plus de la moitié des PDL du segment dans les deux méthodes.

Pour la dernière méthode, le nombre de sous-groupes est plus important et présente des profils de consommation très différents en fonction du cluster comme on peut le voir sur les graphiques ci-dessous :





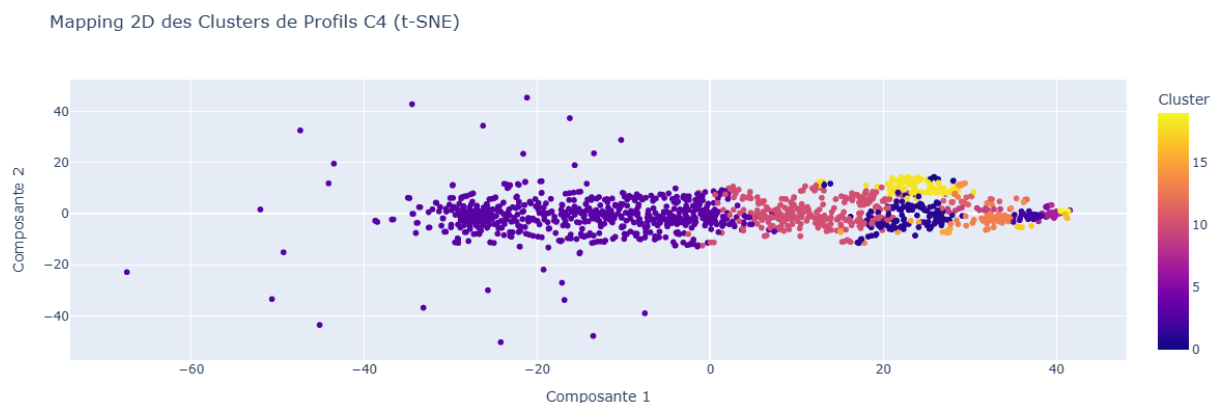


Chaque cluster comprend un nombre de PDL qu'on retrouve dans le tableau suivant (toujours avec un groupe prépondérant : le cluster 3 qui représente près de 52,47%) :

Résultats du clustering C4 par profil		
Cluster	Nombre de PDL	Pourcentage
0	5	0,37%
1	119	8,77%
2	26	1,92%
3	712	52,47%
4	1	0,07%
5	3	0,22%
6	2	0,15%
7	15	1,11%
8	1	0,07%
9	14	1,03%
10	271	19,97%
11	16	1,18%
12	5	0,37%
13	50	3,68%
14	10	0,74%
15	11	0,81%
16	4	0,29%

17	5	0,37%
18	82	6,04%
19	5	0,37%

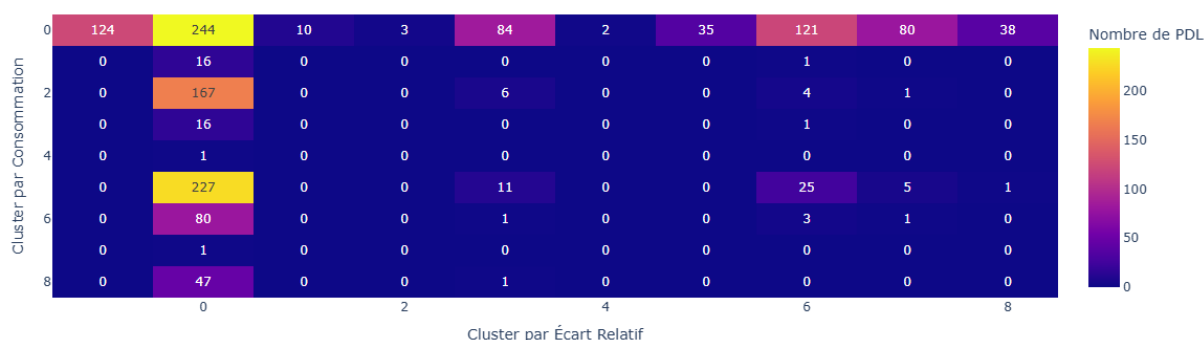
Indépendamment des conditions à choisir pour appliquer le clustering, on peut appliquer la méthode t-SNE comme fait pour le segment C2 afin de réduire la dimensionnalité des données et de répartir en fonction de deux composantes. Si deux PDL ont des profils de consommation similaires dans l'espace de 8760 heures, ils seront très proches. On voit dans le graphique ci-dessous la répartition en plusieurs groupes en fonction des couleurs :



On a des points qui se confondent dans d'autres groupes et qui pourraient être pertinent de traiter à part si on garde cette méthode de mapping pour sélectionner les sous-groupes.

On effectue également une analyse croisée des clusters avec un même nombre de groupe et une base de variable commune (consommation totale vs écart de consommation) pour observer des similitudes entre certains clusters et potentiellement subdivisé en sous-groupe pour analyser plus finement les risques :

Répartition des PDL C4 : Clusters Consommation vs. Écart



Par exemple, il y a 124 points de livraison (PDL) qui appartiennent simultanément au cluster de consommation n°0 (axe vertical) et au cluster d'écart relatif n°0 (axe horizontal). En d'autres termes, c'est le nombre de clients qui se trouvent à l'intersection de ces deux groupes spécifiques : ils ont une consommation totale et un risque relatif similaires.

On a pour le segment C4 plus de PDL se trouvant en intersection entre les deux conditions.

Annexe 7

On a la décomposition du clustering en fonction de la consommation totale puis du risque relatif où on a un groupe prépondérant et des bien plus petits groupes :

Résultats du clustering C2 par consommation		
Cluster	Nombre de PDL	Pourcentage
0	374	73,33%
1	3	0,59%
2	28	5,49%
3	2	0,39%
4	77	15,10%
5	7	1,37%
6	4	0,78%
7	15	2,94%

Résultats du clustering C2 par risque relatif		
Cluster	Nombre de PDL	Pourcentage
0	259	52,43%
1	62	12,55%
2	11	2,23%
3	70	14,17%
4	18	3,64%
5	2	0,40%
6	49	9,92%
7	23	4,66%

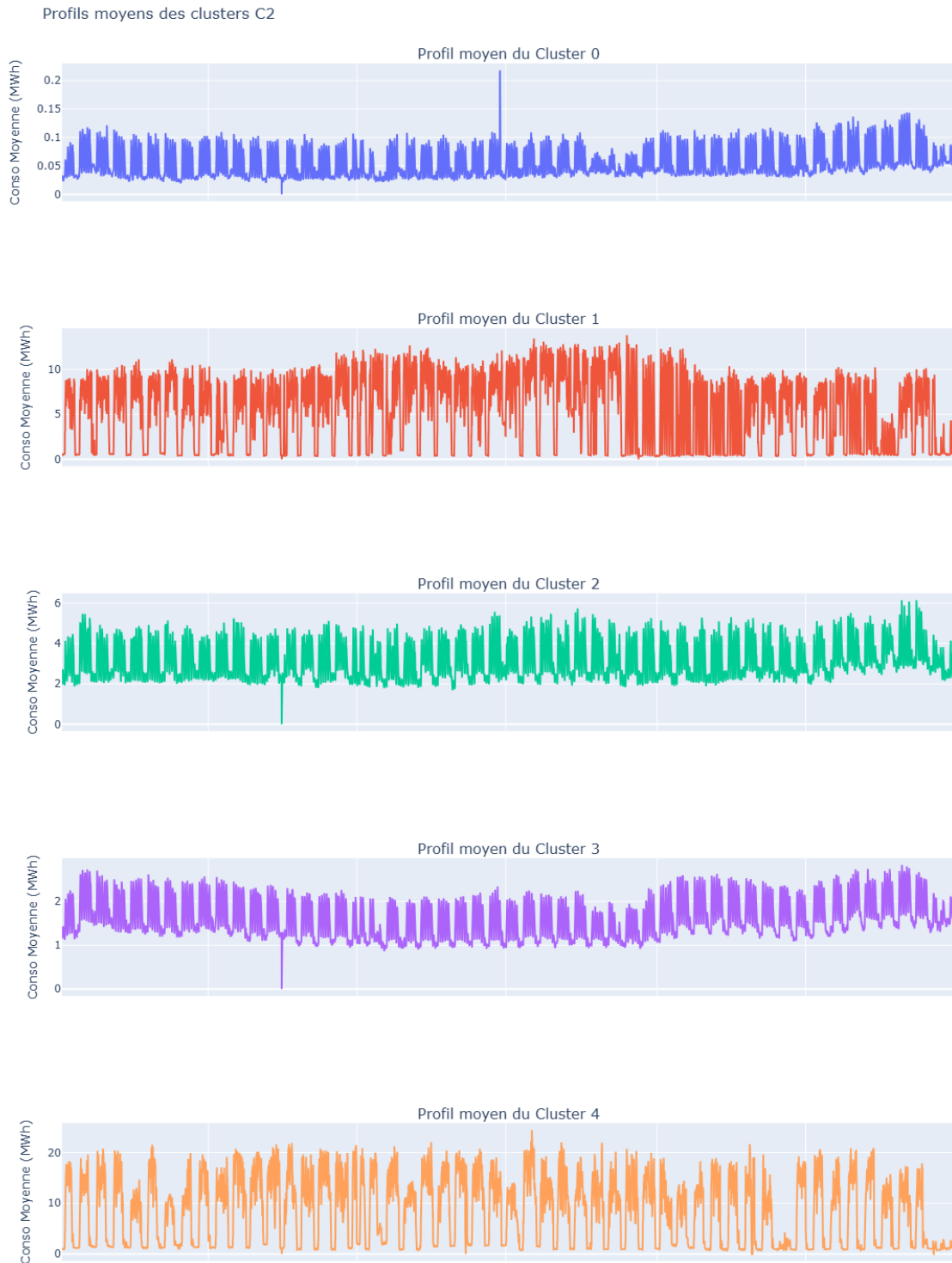
Annexe 8

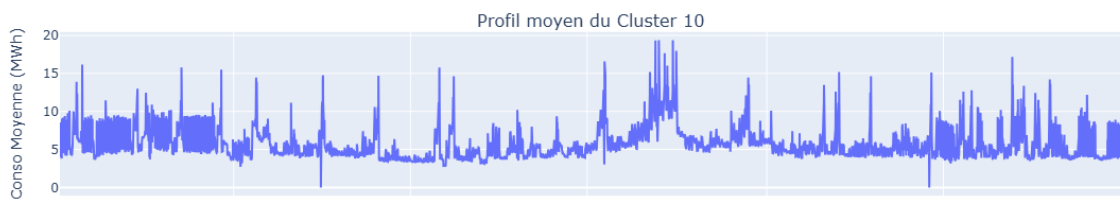
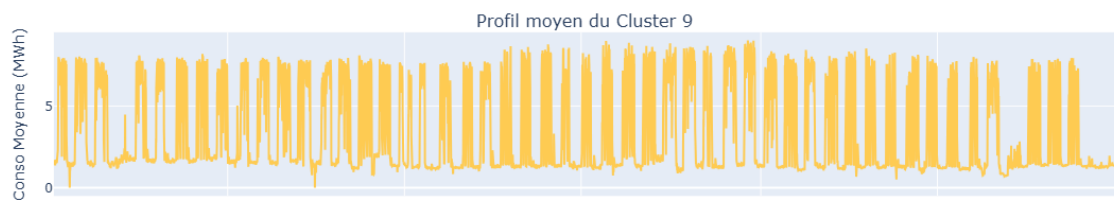
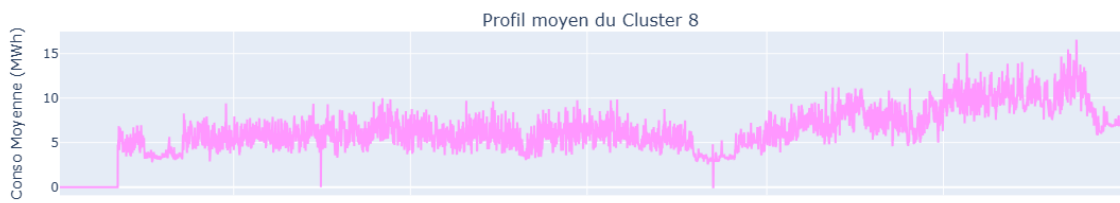
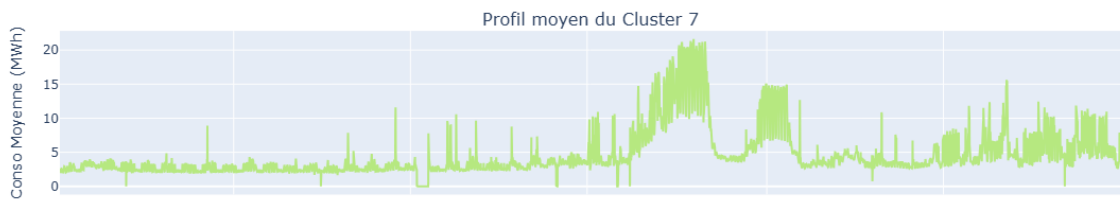
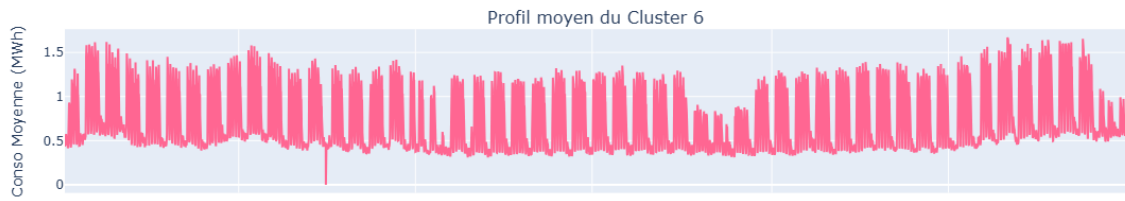
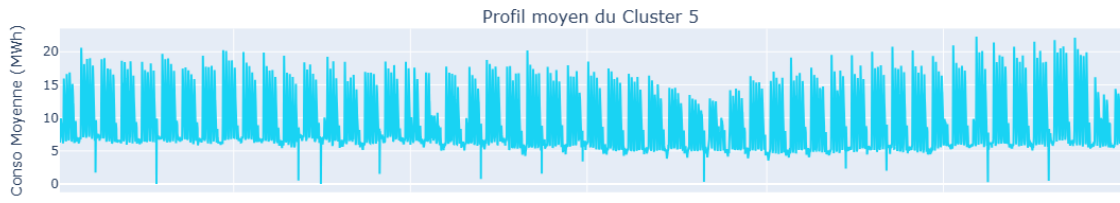
Pour le segment C2, les clusters en fonction des profils sont du nombre de 12 avec des profils chacun différents en fonction de la consommation totale. On retrouve la répartition du nombre de clients par cluster ci-dessous :

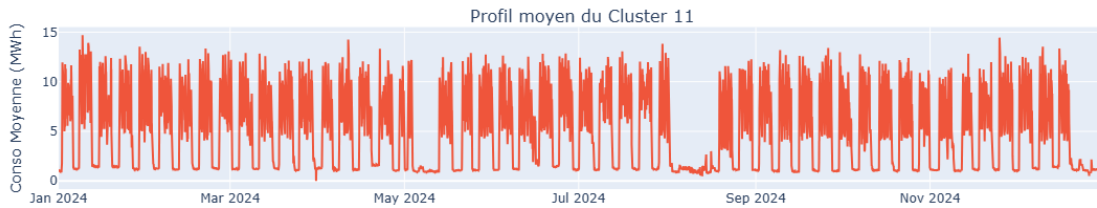
Résultats du clustering C2 par profil		
Cluster	Nombre de PDL	Pourcentage
0	393	77,06%
1	1	0,20%
2	10	1,96%
3	23	4,51%
4	1	0,20%
5	1	0,20%
6	76	14,90%
7	1	0,20%
8	1	0,20%

9	1	0,20%
10	1	0,20%
11	1	0,20%

Et la visualisation des profils de chaque cluster :







Annexe 9

En fonction de l'échelle temporelle et de la maille clients choisie pour effectuer les calculs de primes, on aura quelques différences dans la valeur des primes notamment dues à l'effet de foisonnement. On peut le constater lorsqu'on calcule les primes moyennes de risque marché par segment au pas journalier (et non plus horaire comme fait précédemment) :

	Annuel
C2	-1,58€/MWh
C4	-3,93€/MWh
C5-PRO	-2,92€/MWh
C5-RES	-0,54€/MWh
Portefeuille	-1,77€/MWh

On a des différences de quelques dizaines de centimes par primes en fonction des segments par rapport aux calculs de primes au pas trimestriel. On a globalement, pour les primes trimestrielles rapportées à l'année des primes un peu moins faibles que ce qui est présenté ici. Au pas journalier, les variations horaires de la journée et les pics ou creux de consommation sont lissés. On ne prend que la moyenne du prix spot sur la journée qui a connu des variations plus ou moins importantes positives ou négatives selon les heures. On a donc des primes « plus négatives » sauf pour le segment C5-RES et à l'échelle du portefeuille où les primes sont « moins faibles ». En effet, au pas trimestriel, pour le portefeuille, la prime est de -2,16€/MWh contre -1,77€/MWh pour la prime au pas journalier. On a donc, à l'échelle du portefeuille des prix spot plus élevés en moyenne au moment au pas journalier et moins cet effet de lissage et de foisonnement des risques qu'on retrouve au pas trimestriel où des journées vont « se compenser ».

Il est néanmoins intéressant de regarder la distribution des primes historiques, sur un intervalle de confiance à 90% :

	Quantile 5%	Quantile 95%
C2	-5,97€/MWh	0,28€/MWh
C4	-10,02€/MWh	0,31€/MWh
C5-PRO	-11,17€/MWh	0,18€/MWh
C5-RES	-2,14€/MWh	0,01€/MWh
Portefeuille	-5,82€/MWh	0,24€/MWh

On a 90% des primes qui sont comprises dans ces intervalles avec des journées plus à risque que d'autres lorsqu'il y a un écart positif (elles vont notamment correspondre aux journées hivernales en plein pic de consommation en semaine où la forte demande poussée par les besoins de chauffage va tirer les prix spot vers le haut).

Annexe 10

On a les mêmes constats que précédemment pour le risque de Flex sur écarts positifs avec les primes au pas journalier par segment suivantes :

	Annuel
C2	-0,64€/MWh
C4	-2,83€/MWh
Portefeuille	-0,96€/MWh

Ces primes pour les écarts positifs uniquement sont bien plus élevées (même si elles restent négatives en l'état et qu'elles ne caractérisent pas de perte pour le fournisseur) que les primes calculées au pas trimestriel avec un delta de plusieurs euros et un poids plus important dans la prime sur le portefeuille global du segment C2 qui est plus risqué que le segment C4 sur les écarts positifs.

Globalement, lorsqu'on combine les deux types d'écarts, au pas journalier et par segment, on a les primes suivantes :

	Annuel
C2	5,72€/MWh
C4	-1,74€/MWh
Portefeuille	3,53€/MWh

On a des primes positives qui traduit une perte pour le fournisseur de 3,53€ par MWh à l'échelle du portefeuille mais en isolant par segment, c'est le segment C2 qui porte ce risque de perte quantifié à 5,72€/MWh de prime moyenne.

Les primes pour le segment C2 sont comprises dans un intervalle de confiance à 90% entre -4,37€/MWh et 23,50€. Pour le segment C4, les primes sont comprises entre -10,41€/MWh et 6,98€/MWh. On a donc certains jours où sur le segment C4 des pertes et inversement sur le segment C2, le fournisseur ne subit pas de perte et achète ou revend à prix avantageux. C'est une nuance à prendre en compte sur l'effet de lissage des moyennes notamment lorsqu'on parle de prime moyenne propre.

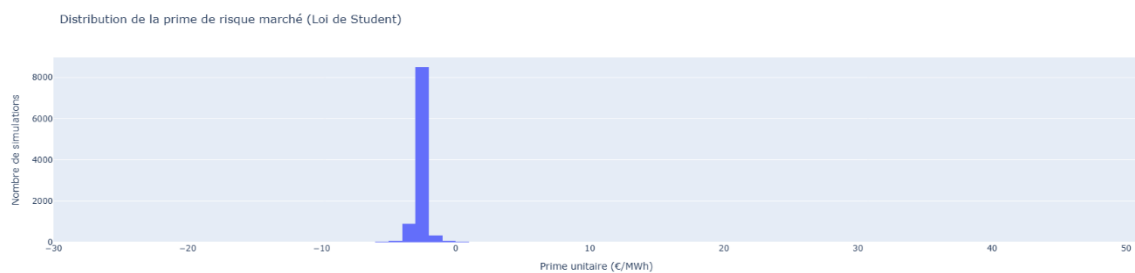
Il y a cependant une subtilité entre coût d'opportunité et coût effectif qu'on ne peut pas simplement combiner comme fait précédemment. On doit au préalable quantifier par rapport au prix à terme le coût d'opportunité afin qu'on puisse avoir un coût effectif direct et de pouvoir calculer la prime globale sur tous les écarts. Ce qui a été fait dans la sous-section.

Annexe 11

Pour chacune des deux méthodes, on a effectué des simulations avec deux approches de distribution : l'une avec la loi de Student et l'autre avec une méthode non paramétrique (le Bootstrap). La seconde approche est plus robuste car elle ne suppose pas de distribution paramétrique et s'approche de la distribution empirique des données. On développera ici en annexe la méthode avec la loi de Student pour chacune des méthodes et chacun des risques.

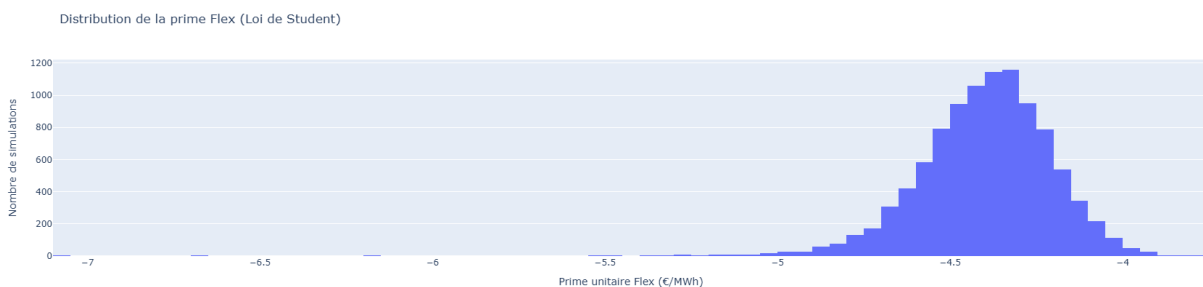
Méthode directe :

Pour le risque de marché, on a une forte proportion de primes comprises dans l'intervalle de confiance à 90% [-3,19 ; -2,08] €/MWh et qui est représentée par la barre dominante sur la distribution suivante :



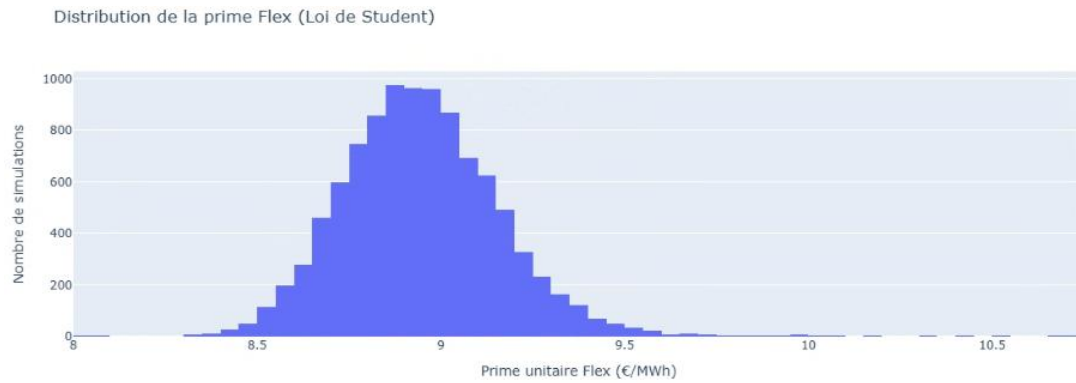
On a une prime moyenne de -2,64€ qui signifie que le recours au marché a permis de préserver 2,64€ en moyenne pour chaque MWh acheté supplémentaire de la marge.

On applique la même démarche pour le risque de Flex (d'abord sur écarts positifs) en commençant par la distribution des primes simulées avec la loi Student :



On a une prime moyenne de -4,40 €/MWh, une prime négative qui signifie qu'en moyenne, les écarts de consommation qui ont dû être compensés sur le marché spot n'ont pas entraîné de perte et à préserver la marge du fournisseur de 4,40€ en moyenne pour chaque MWh supplémentaire acheté. On a une médiane de -2,60€/MWh limitant la perte ne prenant pas en compte les effets des valeurs extrêmes. On a une longue queue à gauche créant une asymétrie conséquente témoignant de prix spot bien plus faibles aux heures des écarts positifs. Cette approche témoigne cependant d'une importante variabilité des primes calculées avec un écart-type de 574,74€/MWh et des primes comprises à 90% de confiance dans l'intervalle suivant : [-920,55 ; 913,37] €/MWh.

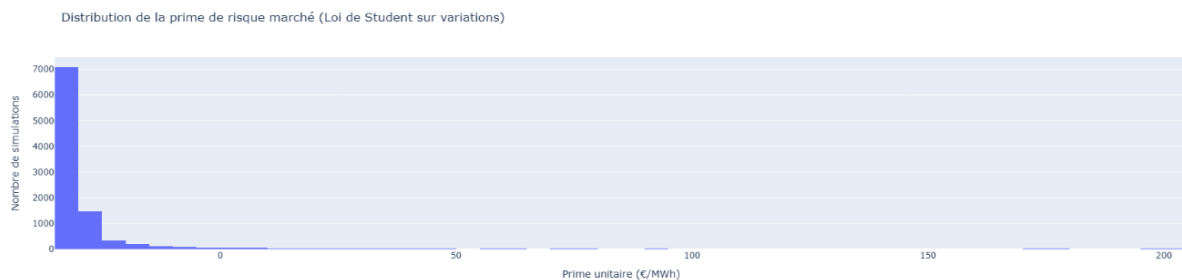
On a les mêmes constats avec le calcul de cette prime mais sur les écarts positifs avec une prime moyenne positive qui s'élève à 8,94€/MWh et une distribution des primes moyennes sur le graphique suivant :



On a cependant une repondération des extrêmes qui biaisent la moyenne qu'on peut voir avec la valeur de la médiane qui s'établit à $-0,95\text{€/MWh}$. On a d'ailleurs un écart-type tout aussi élevé que précédemment de $565,84\text{€/MWh}$.

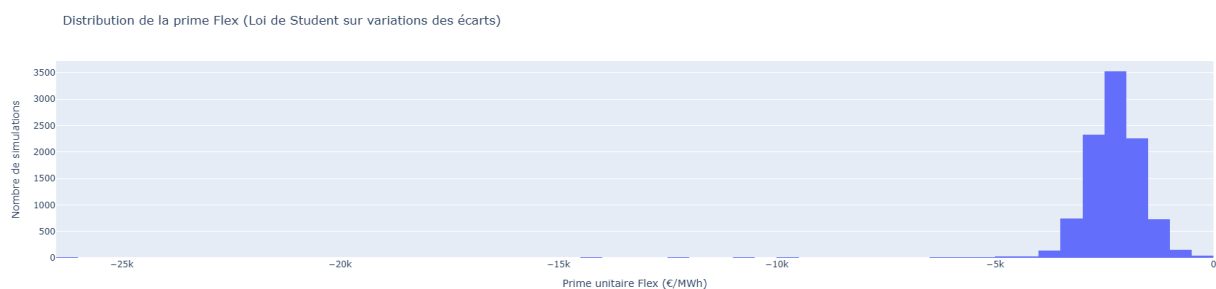
Méthode des variations :

On effectue donc les simulations en prenant en compte les variations du prix spot notamment pour le risque de marché en commençant par utiliser une loi de Student qui va présenter une distribution asymétrique marquée à droite :

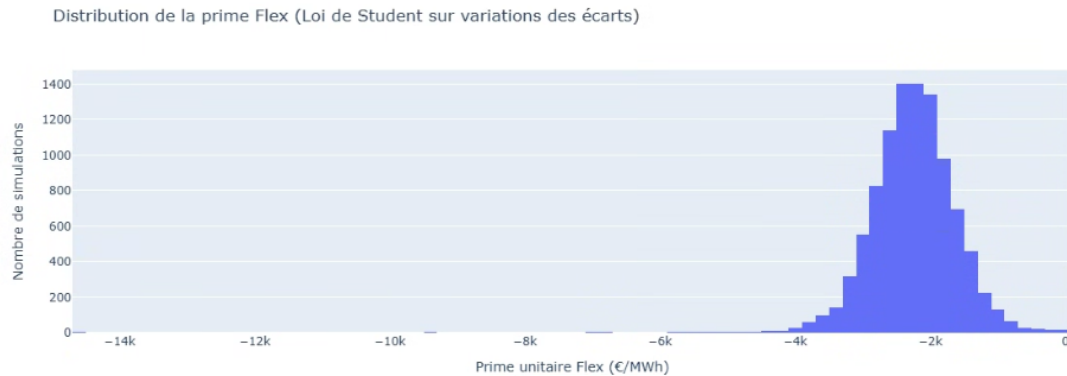


On a une prime moyenne fortement négative à $-22,36\text{€/MWh}$ avec une dispersion marquée (écart-type de $30,71\text{€/MWh}$) et des extrêmes positifs non négligeables qui, tous deux, influencent la forme de l'intervalle de confiance à 90% qui est de $[-31,11 ; 17,20] \text{€/MWh}$.

On retrouve des résultats très volatils avec des primes extrêmes et irréalistes également dans la simulation de la prime de risque de Flex sur les écarts positifs où dans la distribution suivante, on a une longue queue de distribution à gauche dépassant les $20\,000\text{€/MWh}$ de prime moyenne :



On a le même constat sur les écarts négatifs avec une distribution moins dispersée mais toujours avec des valeurs aberrantes de prime moyenne comme on peut le voir sur le graphique ci-dessous :



Il est préférable de ce fait de continuer avec une distribution Bootstrap voir d'utiliser d'autres méthodes de simulation comme il a été fait par la suite dans ce mémoire.

Annexe 12

Chacune des variables explicatives est calculée ou récupérée avec la cellule de code suivante pour le risque de marché :

```
# Features marché
# Création des features calendaires
df_spot['hour'] = df_spot.index.hour
df_spot['dayofweek'] = df_spot.index.dayofweek
df_spot['month'] = df_spot.index.month
df_spot['dayofyear'] = df_spot.index.dayofyear
df_spot['year'] = df_spot.index.year
df_spot['day'] = df_spot.index.day
df_spot['hour_of_day'] = df_spot.index.hour + df_spot.index.day * 24
df_spot['weekend'] = (df_spot.index.weekday >= 5).astype(int)
df_spot['weekday'] = (df_spot.index.weekday < 5).astype(int)
df_spot['is_peak'] = ((df_spot.index.weekday < 5) & (df_spot.index.hour >= 8) & (df_spot.index.hour <= 20)).astype(int)
df_spot['is_offpeak'] = 1 - df_spot['is_peak']

# Lags temporels
for lag in [1, 2, 12, 24, 168]:
    df_spot['lag_' + str(lag)] = df_spot['Price'].shift(lag)

# Moyennes glissantes
df_spot['rolling_mean_24'] = df_spot['Price'].rolling(window=24, min_periods=1).mean().shift(1)
df_spot['rolling_mean_168'] = df_spot['Price'].rolling(window=168, min_periods=1).mean()
df_spot['rolling_mean_365'] = df_spot['Price'].rolling(window=365, min_periods=1).mean()

# Ajout météo
df_spot = df_spot.drop(columns=['temperature'], errors='ignore')
df_spot = df_spot.merge(df_meteo_hourly[['temperature']], left_index=True, right_index=True, how='left')
df_spot['HDD'] = (18 - df_spot['temperature']).clip(lower=0)
df_spot['CDD'] = (df_spot['temperature'] - 22).clip(lower=0)
df_spot['temp_moy_15j'] = df_spot['temperature'].rolling(window=24*15, min_periods=1).mean()
df_spot['anomalie_temp'] = df_spot['Price'] - df_spot['temp_moy_15j']

# Jours fériés et vacances
df_spot = add_holidays(df_spot, holidays_csv)
df_spot['vacances'] = 0
for debut, fin in vacances:
    mask = (df_spot.index >= pd.to_datetime(debut)) & (df_spot.index <= pd.to_datetime(fin))
    df_spot.loc[mask, 'vacances'] = 1
df_spot['vacances_noel'] = 0
df_spot.loc[(df_spot.index.month == 12) & (df_spot.index.day >= 20) | (df_spot.index.month == 1) & (df_spot.index.day <= 5), 'vacances_noel'] = 1
```

```

# Features d'interaction
df_spot['hour_x_HDD'] = df_spot['hour'] * df_spot['HDD']
df_spot['is_peak_x_anomalie_temp'] = df_spot['is_peak'] * df_spot['anomalie_temp']
df_spot['holiday_x_anomalie_temp'] = df_spot['holiday'] * df_spot['anomalie_temp']
df_spot['weekend_x_anomalie_temp'] = df_spot['weekend'] * df_spot['anomalie_temp']
df_spot['vacances_x_HDD'] = df_spot['vacances'] * df_spot['HDD']
df_spot['holiday_x_HDD'] = df_spot['holiday'] * df_spot['HDD']
df_spot['weekend_x_HDD'] = df_spot['weekend'] * df_spot['HDD']

# Effets calendaires et extrêmes
df_spot['evenement_extreme'] = ((df_spot['vacances'] == 1) | (df_spot['holiday'] == 1) | (df_spot['weekend'] == 1) | (df_spot['anomalie_temp'].abs() > 5)).astype(int)
df_spot['is_2h'] = (df_spot.index.hour == 2).astype(int)
df_spot['is_2h_x_vacances'] = df_spot['is_2h'] * df_spot['vacances']
df_spot['is_2h_x_weekend'] = df_spot['is_2h'] * df_spot['weekend']
df_spot['is_2h_x_anomalie_temp'] = df_spot['is_2h'] * df_spot['anomalie_temp']
df_spot['weekday_nonvacation'] = (df_spot['weekday'] & ~df_spot['vacances']).astype(int)
df_spot['weekday_vacation'] = (df_spot['weekday'] & df_spot['vacances']).astype(int)
df_spot['weekend_vacation'] = (df_spot['weekend'] & df_spot['vacances']).astype(int)
df_spot['low_consumption_day'] = ((df_spot['weekend'] == 1) | (df_spot['vacances'] == 1) | (df_spot['holiday'] == 1)).astype(int)
df_spot['vacances_ete'] = ((df_spot['vacances'] == 1) & (df_spot['month'] >= 7) & (df_spot['month'] <= 8)).astype(int)

```

```

# Définition initiale des features
features = [
    'hour', 'dayofweek', 'month', 'dayofyear', 'lag_1', 'lag_2', 'lag_12', 'lag_24', 'lag_168',
    'is_peak', 'is_offpeak', 'holiday', 'temperature', 'HDD', 'anomalie_temp', 'hour_x_HDD',
    'is_peak_x_anomalie_temp', 'day', 'hour_of_day', 'year', 'vacances', 'vacances_noel',
    'rolling_mean_24', 'rolling_mean_168', 'rolling_mean_365', 'evenement_extreme',
    'holiday_x_anomalie_temp', 'weekend_x_anomalie_temp', 'vacances_x_HDD', 'holiday_x_HDD',
    'weekend_x_HDD', 'is_2h', 'is_2h_x_vacances', 'is_2h_x_weekend', 'is_2h_x_anomalie_temp',
    'weekday_nonvacation', 'weekday_vacation', 'weekend_vacation', 'low_consumption_day',
    'vacances_ete'
]

```

Et le code suivant pour le risque de Flex (ce sont les mêmes variables mais ajoutées à un DataFrame différent) :

```

# Features Flex
# DataFrame des écarts horaires de consommation (C2 + C4)
df_ecarts_flex = ecarts_tot.copy()

# Ajout des features calendaires
df_ecarts_flex['hour'] = df_ecarts_flex.index.hour
df_ecarts_flex['dayofweek'] = df_ecarts_flex.index.dayofweek
df_ecarts_flex['month'] = df_ecarts_flex.index.month
df_ecarts_flex['dayofyear'] = df_ecarts_flex.index.dayofyear
df_ecarts_flex['weekend'] = (df_ecarts_flex.index.weekday >= 5).astype(int)
df_ecarts_flex['weekday'] = (df_ecarts_flex.index.weekday < 5).astype(int)
df_ecarts_flex['year'] = df_ecarts_flex.index.year
df_ecarts_flex['day'] = df_ecarts_flex.index.day
df_ecarts_flex['is_peak'] = ((df_ecarts_flex.index.weekday < 5) & (df_ecarts_flex.index.hour >= 8) & (df_ecarts_flex.index.hour <= 20)).astype(int)
df_ecarts_flex['is_offpeak'] = 1 - df_ecarts_flex['is_peak']

# Ajout des features météo
if 'temperature' in df_meteo_hourly.columns:
    df_ecarts_flex = df_ecarts_flex.merge(df_meteo_hourly[['temperature']], left_index=True, right_index=True, how='left')
    df_ecarts_flex['HDD'] = (18 - df_ecarts_flex['temperature']).clip(lower=0)
    df_ecarts_flex['CDD'] = (df_ecarts_flex['temperature'] - 22).clip(lower=0)
    df_ecarts_flex['anomalie_temp'] = df_ecarts_flex['temperature'] - df_ecarts_flex['temperature'].rolling(window=24*15, min_periods=1).mean()

# Ajout des jours fériés et vacances
df_ecarts_flex = add_holidays(df_ecarts_flex, holidays_csv)
df_ecarts_flex['vacances'] = 0
for debut, fin in vacances:
    mask = (df_ecarts_flex.index >= pd.to_datetime(debut)) & (df_ecarts_flex.index <= pd.to_datetime(fin))
    df_ecarts_flex.loc[mask, 'vacances'] = 1
df_ecarts_flex['vacances_noel'] = 0
df_ecarts_flex.loc[(df_ecarts_flex.index.month == 12) & (df_ecarts_flex.index.day >= 20) | (df_ecarts_flex.index.month == 1) & (df_ecarts_flex.index.day <= 5), 'vacances_noel'] = 1

# Ajout des lags sur les écarts de consommation
for lag in [1, 2, 12, 24, 168]:
    df_ecarts_flex[f'lag_{lag}'] = df_ecarts_flex['ecart'].shift(lag)

# Liste des features
features_flex = [
    'hour', 'dayofweek', 'month', 'dayofyear', 'lag_1', 'lag_2', 'lag_12', 'lag_24', 'lag_168',
    'is_peak', 'is_offpeak', 'holiday', 'temperature', 'HDD', 'anomalie_temp', 'day', 'year',
    'vacances', 'vacances_noel'
]

```

Annexe 13

On a visuellement de meilleure performance en prenant en compte l'ensemble des variables explicatives pour chacun des risques mais il est important de regarder dans le détail chacune métrique :

--- SARIMA --- RMSE : 1342.27 €/MWh R ² : -1087.3450 R ² ajusté : -1087.4689 Biais : 1134.11 €/MWh Variance des prédictions : 499027.85	--- XGBoost --- RMSE : 13.30 €/MWh R ² : 0.8932 R ² ajusté : 0.8927 Biais : -0.14 €/MWh Variance des prédictions : 1789.38	--- Mix RF + XGB --- RMSE : 11.94 €/MWh R ² : 0.9138 R ² ajusté : 0.9134 Biais : -0.37 €/MWh Variance des prédictions : 1653.24
--- Random Forest --- RMSE : 12.45 €/MWh R ² : 0.9064 R ² ajusté : 0.9059 Biais : -0.60 €/MWh Variance des prédictions : 1563.56	--- Mix RF x XGB --- RMSE : 12.13 €/MWh R ² : 0.9112 R ² ajusté : 0.9108 Biais : -1.05 €/MWh Variance des prédictions : 1540.30	

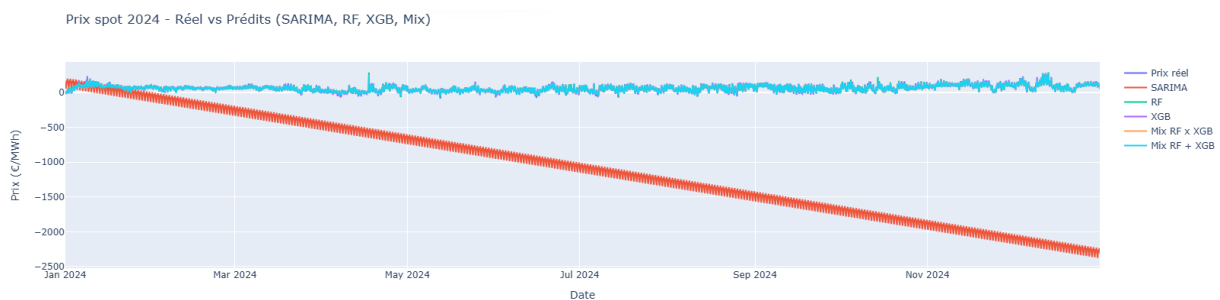
Les valeurs du modèle SARIMA restent inchangées comme ce modèle n'utilise pas de variables exogènes.

Pour le modèle Random Forest, on a un RMSE un peu plus faible mais un équilibre biais/variance légèrement dégradé. On doit avoir un surajustement de l'apprentissage du modèle sur les données avec des variables qui n'apportent finalement que très peu d'informations mais peut induire un biais un peu plus élevé.

Pour le modèle XGBoost, l'erreur RMSE est inférieure de 3,33€/MWh mais le R² est quasiment équivalent et il n'y a pas d'amélioration de l'équilibre biais/variance.

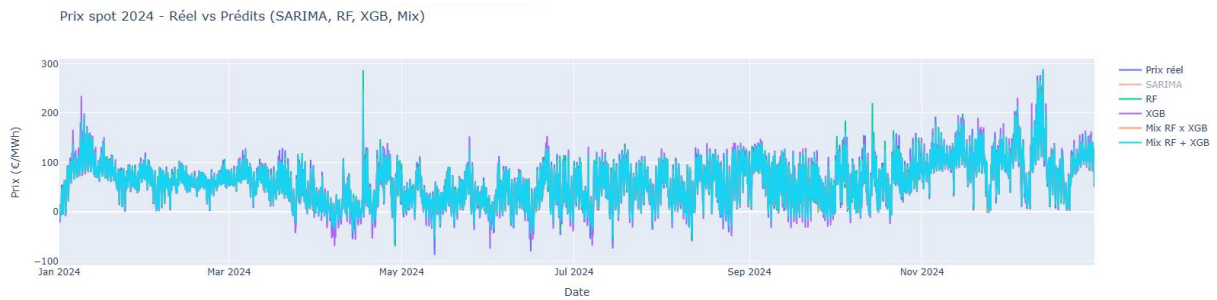
On a les mêmes constats pour les modèles mix où l'erreur est moins élevée et on gagne en explicabilité de la variance avec un R² plus proche de 1 mais l'équilibre biais/variance se détériore légèrement.

Ces différences de performance s'observe bien entre le modèle SARIMA et les autres modèles dans le graphique suivant :



Le modèle SARIMA n'est pas stable sur le long terme et dérive fortement de la courbe réelle des prix spot.

Et si on enlève le modèle SARIMA :



On observe que les autres modèles sont bien plus proches de la courbe des prix spot réelle.

Les primes de risque marché estimées pour chaque de modèle sont :

- SARIMA = -86,09€/MWh
- Random Forest = -0,97€/MWh
- XGBoost = -0,93€/MWh
- Mix RF x XGB = -0,93€/MWh
- Mix RF + XGB = -0,95€/MWh

On retrouve pour les quatre modèles (hors SARIMA) des primes similaires et de même ordre de grandeur.

Annexe 14

Comme précédemment pour le risque de marché, on compare ici les métriques d'évaluation de l'entraînement des modèles sur l'ensemble des variables explicatives :

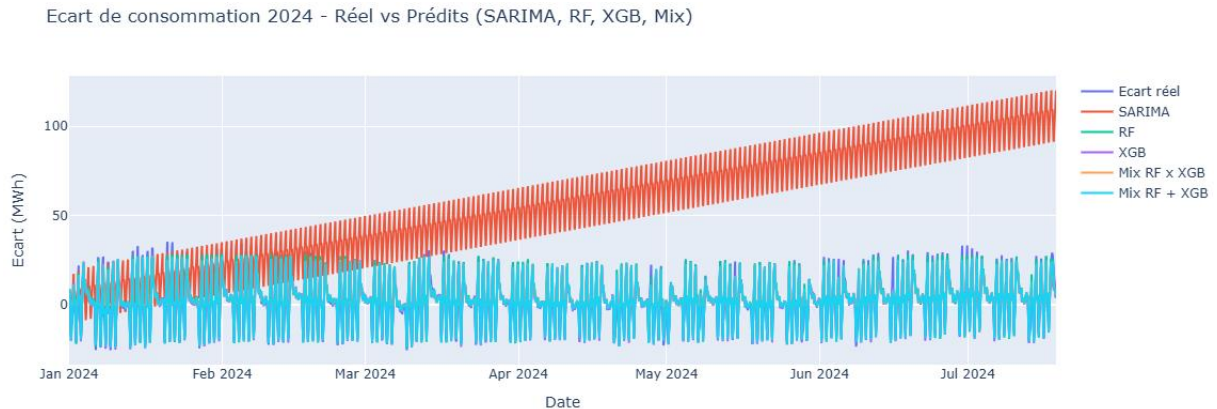
<pre> --- SARIMA (écart) --- RMSE : 60.78 MWh R² : -26.4474 R² ajusté : -26.4531 Biais : -52.84 MWh Variance des prédictions : 927.69 </pre>	<pre> --- XGBoost (écart) --- RMSE : 2.49 MWh R² : 0.9541 R² ajusté : 0.9540 Biais : 0.61 MWh Variance des prédictions : 118.56 </pre>	<pre> --- Mix RF + XGB (écart) --- RMSE : 2.29 MWh R² : 0.9610 R² ajusté : 0.9610 Biais : 0.43 MWh Variance des prédictions : 122.60 </pre>
<pre> --- Random Forest (écart) --- RMSE : 2.33 MWh R² : 0.9597 R² ajusté : 0.9596 Biais : 0.25 MWh Variance des prédictions : 127.68 </pre>	<pre> --- Mix RF x XGB (écart) --- RMSE : 6.86 MWh R² : 0.6503 R² ajusté : 0.6499 Biais : -2.56 MWh Variance des prédictions : 52.59 </pre>	

Les valeurs des métriques pour le modèle SARIMA restent inchangées. Ce modèle n'est pas du tout adéquat pour la modélisation des écarts de consommation.

Pour les autres modèles on a globalement un meilleur R^2 avec une part légèrement plus élevée de variance expliquée par les modèles dépassant les 95% sauf pour le mix RF x XGB et une erreur un peu plus faible des estimations.

On a de plus un équilibre biais/variance stable pour chacun des quatre modèles.

Comme pour le risque de marché, le modèle SARIMA ne capte pas sur la durée la courbe horaire des écarts de consommation et dérive. Mais les autres modèles arrivent à bien capter la forme des données et leur variation :



Les primes sont un peu plus dispersées entre les modèles mais restent dans le même ordre de grandeur si on ne prend pas en compte la prime SARIMA :

- Prime RF : -1,44€/MWh
- Prime XGB : -1,10€/MWh
- Prime Mix RF x XGB : -6,20€/MWh
- Prime Mix RF + XGB : -1,27€/MWh

Si on ne prend pas en compte le modèle mix RF x XGB, les primes sont proches et « moins négatives » que celles des modèles avec les Features les plus importantes.