



*Rapport :*  
**Projet Data Science**



*Groupe 1*  
**Thomas Fournier, Loris Bouchez, Océan Sicard-Baux, Emma Mathieu**

# Sommaire

<b>Sommaire</b>	<b>2</b>
<b>1 - Introduction</b>	<b>4</b>
<b>2 - Préparation</b>	<b>7</b>
<b>3 - Production</b>	<b>11</b>
3.1 - Data mining description	12
3.2 - Choix des méthodes de prédiction	13
3.3 - Mise en œuvre des méthodes pour prédire la production d'électricité	13
3.3.1 - <i>Prédictions pour la bioénergie</i>	14
3.3.2 - <i>Prédictions pour le charbon</i>	16
3.3.3 - <i>Prédictions pour l'éolien</i>	18
3.3.4 - <i>Prédictions pour le gaz</i>	19
3.3.5 - <i>Prédictions pour l'hydraulique</i>	21
3.3.6 - <i>Prédictions pour le nucléaire</i>	23
3.3.7 - <i>Prédictions pour le solaire</i>	25
3.4 - Efficacité des modèles de prédiction utilisés	27
3.4.1 - <i>Efficacité du modèle pour la bioénergie</i>	27
3.4.2 - <i>Efficacité du modèle pour le charbon</i>	28
3.4.3 - <i>Efficacité du modèle pour l'éolien</i>	28
3.4.4 - <i>Efficacité du modèle pour le gaz</i>	29
3.4.5 - <i>Efficacité du modèle pour l'hydraulique</i>	29
3.4.6 - <i>Efficacité du modèle pour le nucléaire</i>	30
3.4.7 - <i>Efficacité du modèle pour le solaire</i>	30
3.5 - Synthèse des résultats sur la production	30
<b>4 - Consommation</b>	<b>35</b>
4.1 - Data Mining descriptif	36
4.2 - Observation des données	36
4.3 - Mise en œuvre des méthodes pour prédire la consommation d'électricité	37
4.3.1 - <i>Prédictions pour la consommation d'électricité (de manière globale)</i>	37
4.3.2 - <i>Prédictions pour la consommation d'électricité (secteur agriculture)</i>	40
4.3.3 - <i>Prédictions pour la consommation d'électricité (secteur industrie)</i>	41
4.3.4 - <i>Prédictions pour la consommation d'électricité (secteur tertiaire)</i>	43
4.3.5 - <i>Prédictions pour la consommation d'électricité (secteur résidentiel)</i>	44
4.3.6 - <i>Prédictions pour la consommation d'électricité (secteur inconnu)</i>	45
4.3.7 - <i>Critique</i>	47

<b>5 - Coût</b>	<b>48</b>
5.1 - Comparaison entre les différentes filières dont le nucléaire	49
5.2 - Méthodologie de calcul adaptée aux économies d'échelles	52
5.3 - Comparaison entre les différentes filières sans le nucléaire	53
5.4 - Comparaison des coûts de production de l'électricité avec et sans le nucléaire	54
<b>6 - Pollution</b>	<b>58</b>
6.1 - Comparaison entre les différentes filières dont le nucléaire	59
6.2 - Comparaison entre les différentes filières sans le nucléaire	62
<b>7 - Conclusion</b>	<b>66</b>
7.1 - Résumé	67
7.1.1 - <i>Production</i>	67
7.1.2 - <i>Consommation</i>	67
7.1.3 - <i>Coût</i>	68
7.1.4 - <i>Pollution</i>	68
7.2 - Corrélation	69
7.2.1 - <i>Corrélation production et consommation</i>	69
7.2.2 - <i>Corrélation production et coût</i>	70
7.2.3 - <i>Corrélation production et pollution</i>	71
7.2.4 - <i>Corrélation coût et pollution</i>	72
7.3 - Comparaison	72
7.3.1 - <i>Comparaison entre les attentes et les véritables résultats concernant la production et la consommation</i>	72
7.3.2 - <i>Comparaison entre les attentes et les véritables résultats concernant le prix et la pollution</i>	73

**GitHub:** <https://github.com/thmsfnr/DataScience>

**Dashboard:** <https://thmsfnr.shinyapps.io/EnergieEnFrance/>

# ***1 - Introduction***

Ce projet a initialement pour ambition de permettre à la fois d'approfondir certaines connaissances acquises au cours de notre formation, d'en découvrir d'autres, d'identifier des méthodes appropriées pour répondre à un objectif donné, de combiner des méthodes complémentaires, de comparer des méthodes concurrentes pour répondre à des objectifs donnés, d'interpréter et éventuellement de comparer les résultats des méthodes mises en œuvre, ainsi que d'avoir l'esprit critique face à la collecte de données, aux méthodes mises en œuvre et à leur paramétrage.

Concernant le sujet, nous avions le choix entre différents thèmes et nous avons choisi celui relatif à l'étude de données ouvertes, plus précisément relatif au projet [Open Data University](#) porté par l'association Latitudes en lien avec le site institutionnel national [data.gouv](#). Les jeux de données de ce thème sont alors multiples, sachant que nous avons fait le choix d'étudier celui associé à l'[énergie en France](#).

Nous disposons donc de nombreuses données ouvertes mises à disposition par l'équipe de l'Open Data University pour répondre au sujet proposé. Ces données proviennent de diverses sources.

La première d'entre-elle est ORE (Observatoire des Ressources et des Énergies). Cette agence est chargée de suivre et d'analyser les données sur les ressources et les énergies en France. Elle produit des données et des analyses sur les énergies renouvelables, la production d'électricité, la consommation d'énergie, les prix de l'énergie, etc.

La seconde est RTE (Réseau de Transport d'Électricité) . Cette entreprise est responsable du transport d'électricité en France. Elle gère le réseau de transport d'électricité à haute tension et assure l'équilibre entre la production et la consommation d'électricité. Elle fournit des données sur la production d'électricité, la consommation d'électricité, les prix de l'électricité, etc.

La suivante est EDF (Electricité de France). Cette entreprise est un des principaux producteurs et fournisseurs d'électricité en France. Elle produit de l'électricité à partir de différentes sources d'énergie, notamment le nucléaire, les énergies renouvelables et le charbon. Elle fournit des données sur la production d'électricité, les prix de l'électricité, etc.

Celle d'après est GRDF (Gaz Réseau Distribution France). Cette entreprise est responsable de la distribution de gaz naturel en France. Elle gère le réseau de distribution de gaz et assure l'acheminement du gaz naturel jusqu'aux consommateurs. Elle fournit des données sur la consommation de gaz, les prix du gaz, etc.

La suivante est l'ADEME (Agence de l'Environnement et de la Maîtrise de l'Énergie). Cette agence est chargée de promouvoir une transition énergétique durable en France. Elle produit des données et des analyses sur les émissions de gaz à effet de serre, l'efficacité énergétique, les énergies renouvelables et les politiques environnementales. Elle propose également des solutions et des outils pour aider les entreprises et les particuliers à réduire leur impact environnemental.

Celle d'après est l'INSEE (Institut National de la Statistique et des Études Économiques). Cet institut est chargé de produire des données statistiques sur l'économie et la société française. Il produit des données sur la consommation d'énergie, les prix de l'énergie, les dépenses énergétiques des ménages, etc.

La dernière est l'OCDE (Organisation de Coopération et de Développement Économiques). Cette organisation regroupe les pays développés et a pour objectif de promouvoir la croissance économique et le bien-être social. Elle produit des données et des analyses sur les politiques énergétiques, les tendances énergétiques, les prix de l'énergie, les émissions de gaz à effet de serre et les investissements dans les énergies renouvelables dans les

différents pays membres. Les données de l'OCDE peuvent être utilisées pour des comparaisons internationales et pour comprendre les tendances énergétiques à l'échelle mondiale.

Au sujet de la problématique, nous en avons choisi une qui diffère de celle proposée sur le projet. En effet, au lieu d'orienter notre étude sur les conséquences énergétiques de la guerre en Ukraine, nous allons déterminer les conséquences de l'arrêt du nucléaire dans la production d'électricité en France.

Cela va alors passer par une multitude d'objectifs, à savoir déterminer la production d'électricité en fonction des filières de production, déterminer la consommation d'électricité, prévoir de la production d'électricité dans le futur sans le nucléaire, prévoir la consommation d'électricité dans le futur, analyser l'influence de l'arrêt du nucléaire sur l'environnement, analyser l'influence de l'arrêt du nucléaire sur le prix, mettre en lien les prévisions de production et de consommation sans le nucléaire avant de terminer par la comparaison avec d'autres études.

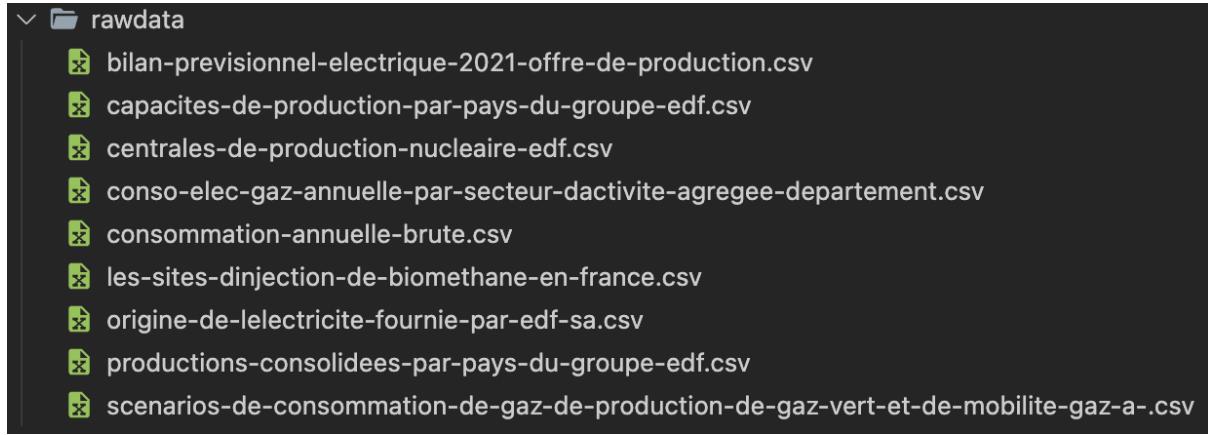
## ***2 - Préparation***

À la suite de l'écriture de la problématique et des différents objectifs, la première étape a consisté en l'inventaire des données mises à disposition par l'équipe de l'Open Data University. Pour cela, nous avons listé l'ensemble des fichiers, sous la forme d'une source, ORE par exemple, d'une description, et du lien vers les données. De plus, après avoir effectué cela pour toutes les données, nous avons sélectionné minutieusement celles qui avaient une utilité dans le cadre de notre problématique. Ainsi, nous obtenons une liste de la forme suivante :

Source	Description	Lien	Utilité
ORE (Opé	Consommation annuelle d'électricité et gaz par IRIS et par code NAF	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricite-et-gaz-par-iris-et-par-code-naf/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricite-et-gaz-par-iris-et-par-code-naf/</a>	
ORE	Voitures particulières immatriculées par commune et par type de recharge	<a href="https://www.data.gouv.fr/fr/datasets/voitures-particulières-immatriculées-par-commune-et-par-type-de-recharge/">https://www.data.gouv.fr/fr/datasets/voitures-particulières-immatriculées-par-commune-et-par-type-de-recharge/</a>	
ORE	Installations de production de la filière "Eolien" par commune	<a href="https://www.data.gouv.fr/fr/datasets/installations-de-production-de-la-filière-eolien-par-commune/">https://www.data.gouv.fr/fr/datasets/installations-de-production-de-la-filière-eolien-par-commune/</a>	
ORE	Installations de production de la filière "Solaire" par commune	<a href="https://www.data.gouv.fr/fr/datasets/installations-de-production-de-la-filière-solaire-par-commune/">https://www.data.gouv.fr/fr/datasets/installations-de-production-de-la-filière-solaire-par-commune/</a>	
ORE	Consommation annuelle d'électricité et gaz par région et par code NAF	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricite-et-gaz-par-region-et-par-code-naf/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricite-et-gaz-par-region-et-par-code-naf/</a>	
ORE	Productions d'électricité d'origine renouvelable aux enchères des Garanties D'Origine	<a href="https://www.data.gouv.fr/fr/datasets/productions-delectricité-d'origine-renouvelable-aux-enchères-des-garanties-dorigine/">https://www.data.gouv.fr/fr/datasets/productions-delectricité-d'origine-renouvelable-aux-enchères-des-garanties-dorigine/</a>	
ORE	Consommation agrégée demi-horaire des points de soutirage inférieurs à 36kVA par région	<a href="https://www.data.gouv.fr/fr/datasets/consommation-agrégée-demi-horaire-des-points-de-soutirage-inferieurs-a-36kva-par-region/">https://www.data.gouv.fr/fr/datasets/consommation-agrégée-demi-horaire-des-points-de-soutirage-inferieurs-a-36kva-par-region/</a>	
ORE	Production demi-horaire agrégée par région	<a href="https://www.data.gouv.fr/fr/datasets/production-demi-horaire-agrégée-par-region/">https://www.data.gouv.fr/fr/datasets/production-demi-horaire-agrégée-par-region/</a>	
ORE	Consommation agrégée demi-horaire des points de soutirage supérieurs à 36kVA par région	<a href="https://www.data.gouv.fr/fr/datasets/consommation-agrégée-demi-horaire-des-points-de-soutirage-supérieurs-a-36kva-par-region/">https://www.data.gouv.fr/fr/datasets/consommation-agrégée-demi-horaire-des-points-de-soutirage-supérieurs-a-36kva-par-region/</a>	
ORE	Installations de production de la filière "Hydraulique" par commune	<a href="https://www.data.gouv.fr/fr/datasets/installations-de-production-de-la-filière-hydraulique-par-commune/">https://www.data.gouv.fr/fr/datasets/installations-de-production-de-la-filière-hydraulique-par-commune/</a>	
ORE	Consommation annuelle d'électricité et gaz par EPCI et par secteur d'activité	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-epci-et-par-secteur-d'activité/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-epci-et-par-secteur-d'activité/</a>	
ORE	Consommation annuelle d'électricité et gaz par commune et par secteur d'activité	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-commune-et-par-secteur-d'activité/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-commune-et-par-secteur-d'activité/</a>	
ORE	Installations de production d'électricité et de gaz	<a href="https://www.data.gouv.fr/fr/datasets/installations-de-production-delectricité-et-de-gaz/">https://www.data.gouv.fr/fr/datasets/installations-de-production-delectricité-et-de-gaz/</a>	
ORE	Consommation annuelle d'électricité et gaz par IRIS et par secteur d'activité	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-iris-et-par-secteur-d'activité/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-iris-et-par-secteur-d'activité/</a>	
ORE	Consommation annuelle d'électricité et gaz par commune et par code NAF	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-commune-et-par-code-naf/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-commune-et-par-code-naf/</a>	
ORE	Consommation annuelle d'électricité et gaz par département et par secteur d'activité	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-département-et-par-secteur-d'activité/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-département-et-par-secteur-d'activité/</a>	
ORE	Consommation annuelle d'électricité et gaz par département et par code NAF	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-département-et-par-code-naf/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-département-et-par-code-naf/</a>	
ORE	Nombre de points de livraison par distributeur	<a href="https://www.data.gouv.fr/fr/datasets/nombre-de-points-de-livraison-par-distributeur/">https://www.data.gouv.fr/fr/datasets/nombre-de-points-de-livraison-par-distributeur/</a>	
ORE	Consommation annuelle d'électricité et gaz par EPCI et par code NAF	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-epci-et-par-code-naf/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-epci-et-par-code-naf/</a>	
ORE	Lignes électriques aériennes Basse Tension (BT)	<a href="https://www.data.gouv.fr/fr/datasets/lignes-electriques-aériennes-basse-tension-bt/">https://www.data.gouv.fr/fr/datasets/lignes-electriques-aériennes-basse-tension-bt/</a>	
ORE	Lignes électriques aériennes moyenne tension (HTA)	<a href="https://www.data.gouv.fr/fr/datasets/lignes-electriques-aériennes-moyenne-tension-hta/">https://www.data.gouv.fr/fr/datasets/lignes-electriques-aériennes-moyenne-tension-hta/</a>	
ORE	Lignes électriques souterraines moyenne tension (HTA)	<a href="https://www.data.gouv.fr/fr/datasets/lignes-electriques-souterraines-moyenne-tension-hta/">https://www.data.gouv.fr/fr/datasets/lignes-electriques-souterraines-moyenne-tension-hta/</a>	
ORE	Postes Source et Postes de Répartition HTA/HTB	<a href="https://www.data.gouv.fr/fr/datasets/postes-source-et-postes-de-répartition-hta-htb-1/">https://www.data.gouv.fr/fr/datasets/postes-source-et-postes-de-répartition-hta-htb-1/</a>	
ORE	Référentiel des distributeurs d'énergie	<a href="https://www.data.gouv.fr/fr/datasets/referentiel-des-distributeurs-d'énergie/">https://www.data.gouv.fr/fr/datasets/referentiel-des-distributeurs-d'énergie/</a>	
ORE	Distributeurs de Gaz et d'électricité par commune	<a href="https://www.data.gouv.fr/fr/datasets/distributeurs-de-gaz-et-delectricité-par-commune/">https://www.data.gouv.fr/fr/datasets/distributeurs-de-gaz-et-delectricité-par-commune/</a>	
ORE	Lignes électriques souterraines Basse Tension (BT)	<a href="https://www.data.gouv.fr/fr/datasets/lignes-electriques-souterraines-basse-tension-bt/">https://www.data.gouv.fr/fr/datasets/lignes-electriques-souterraines-basse-tension-bt/</a>	
ORE	Infrastructures de réseau de gaz	<a href="https://www.data.gouv.fr/fr/datasets/infrastructures-de-réseau-de-gaz/">https://www.data.gouv.fr/fr/datasets/infrastructures-de-réseau-de-gaz/</a>	
ORE	Consommation annuelle d'électricité et gaz par région et par secteur d'activité	<a href="https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-region-et-par-secteur-d'activité/">https://www.data.gouv.fr/fr/datasets/consommation-annuelle-delectricité-et-gaz-par-region-et-par-secteur-d'activité/</a>	
ORE	Postes de distribution publique (postes HTA/BT)	<a href="https://www.data.gouv.fr/fr/datasets/postes-de-distribution-publique-postes-hta-bt-1/">https://www.data.gouv.fr/fr/datasets/postes-de-distribution-publique-postes-hta-bt-1/</a>	
RTE	Consumption	<a href="https://www.data.gouv.fr/fr/datasets/consumption/">https://www.data.gouv.fr/fr/datasets/consumption/</a>	
RTE	Électricité: consommation, production, CO2 et échanges	<a href="https://www.data.gouv.fr/fr/datasets/electricité-consommation-production-co2-et-échanges/">https://www.data.gouv.fr/fr/datasets/electricité-consommation-production-co2-et-échanges/</a>	
RTE	Tempo Like Supply Contract	<a href="https://www.data.gouv.fr/fr/datasets/tempo-like-supply-contract/">https://www.data.gouv.fr/fr/datasets/tempo-like-supply-contract/</a>	
RTE	Ntc	<a href="https://www.data.gouv.fr/fr/datasets/ntc/">https://www.data.gouv.fr/fr/datasets/ntc/</a>	
RTE	Actual generation	<a href="https://www.data.gouv.fr/fr/datasets/actual-generation/">https://www.data.gouv.fr/fr/datasets/actual-generation/</a>	
RTE	Congestion	<a href="https://www.data.gouv.fr/fr/datasets/congestion/">https://www.data.gouv.fr/fr/datasets/congestion/</a>	
RTE	Wholesale Market	<a href="https://www.data.gouv.fr/fr/datasets/wholesale-market/">https://www.data.gouv.fr/fr/datasets/wholesale-market/</a>	
RTE	Balancing Capacity	<a href="https://www.data.gouv.fr/fr/datasets/balancing-capacity/">https://www.data.gouv.fr/fr/datasets/balancing-capacity/</a>	
RTE	Bre Imbalance Reconstitution	<a href="https://www.data.gouv.fr/fr/datasets/bre-imbalance-reconciliation/">https://www.data.gouv.fr/fr/datasets/bre-imbalance-reconciliation/</a>	
RTE	Losses Public Transmission System	<a href="https://www.data.gouv.fr/fr/datasets/losses-public-transmission-system/">https://www.data.gouv.fr/fr/datasets/losses-public-transmission-system/</a>	
RTE	Certification Obligation Parameter	<a href="https://www.data.gouv.fr/fr/datasets/certification-obligation-parameter/">https://www.data.gouv.fr/fr/datasets/certification-obligation-parameter/</a>	
RTE	Bre Temporal Reconciliation	<a href="https://www.data.gouv.fr/fr/datasets/bre-temporal-reconciliation/">https://www.data.gouv.fr/fr/datasets/bre-temporal-reconciliation/</a>	
RTE	Tempo Like Supply Contract	<a href="https://www.data.gouv.fr/fr/datasets/tempo-like-supply-contract/">https://www.data.gouv.fr/fr/datasets/tempo-like-supply-contract/</a>	
RTE	Demand Response	<a href="https://www.data.gouv.fr/fr/datasets/demand-response/">https://www.data.gouv.fr/fr/datasets/demand-response/</a>	
RTE	Consolidated Consumption	<a href="https://www.data.gouv.fr/fr/datasets/consolidated-consumption/">https://www.data.gouv.fr/fr/datasets/consolidated-consumption/</a>	
RTE	Signal	<a href="https://www.data.gouv.fr/fr/datasets/signal/">https://www.data.gouv.fr/fr/datasets/signal/</a>	
RTE	Cross Zonal	<a href="https://www.data.gouv.fr/fr/datasets/cross-zonal-capacity/">https://www.data.gouv.fr/fr/datasets/cross-zonal-capacity/</a>	
RTE	Speed Physical Showcase	<a href="https://www.data.gouv.fr/fr/datasets/speed-physical-showcase/">https://www.data.gouv.fr/fr/datasets/speed-physical-showcase/</a>	
RTE	Certified Capacities Registry	<a href="https://www.data.gouv.fr/fr/datasets/certified-capacities-registry/">https://www.data.gouv.fr/fr/datasets/certified-capacities-registry/</a>	
RTE	Speed Metering Showcase	<a href="https://www.data.gouv.fr/fr/datasets/speed-metering-showcase/">https://www.data.gouv.fr/fr/datasets/speed-metering-showcase/</a>	
RTE	Unavailability Additional Information	<a href="https://www.data.gouv.fr/fr/datasets/unavailability-additional-information/">https://www.data.gouv.fr/fr/datasets/unavailability-additional-information/</a>	
RTE	Generation Forecast	<a href="https://www.data.gouv.fr/fr/datasets/generation-forecast/">https://www.data.gouv.fr/fr/datasets/generation-forecast/</a>	
RTE	Exchange Schedule	<a href="https://www.data.gouv.fr/fr/datasets/exchange-schedule/">https://www.data.gouv.fr/fr/datasets/exchange-schedule/</a>	
RTE	Generation Storage Park	<a href="https://www.data.gouv.fr/fr/datasets/generation-storage-park/">https://www.data.gouv.fr/fr/datasets/generation-storage-park/</a>	
RTE	Référentiels Consommation Contrôle Mesure	<a href="https://www.data.gouv.fr/fr/datasets/référentiels-consommation-contrôle-mesure/">https://www.data.gouv.fr/fr/datasets/référentiels-consommation-contrôle-mesure/</a>	

Une fois cette première sélection effectuée, nous avons étudié en détail chacune des ces données. En effet, pour chaque donnée, nous avons à la fois écrit une description très détaillée de la composition de cette donnée, notamment les différentes colonnes présentes, et les possibles applications, c'est-à-dire les colonnes et lignes à sélectionner. Cela c'est donc traduit par une nouvelle liste de la forme suivante :

Après avoir terminé cette liste, nous avons effectué une seconde sélection, notamment de par le fait que beaucoup de données illustrent les mêmes éléments. Cette sélection finale a ainsi été le point de départ de la phase de nettoyage des données. Effectivement, pour chaque donnée, nous avons mis en place un ETL sur le fichier csv associé. Il est d'ailleurs important de préciser que la liste des fichiers sélectionnés est celle ci-dessous :



Chaque ETL a donc consisté en un fichier Python utilisant la bibliothèque Pandas, qui fournit justement des structures de données et des outils de manipulation de données de haut niveau. Cela se traduit sous la forme suivante :

```

import pandas as pd

# Open file with pandas
df = pd.read_csv("rawdata/productions-consolidées-par-pays-du-groupe-edf.csv",
sep=";")

# Keep only the columns we need
df = df[["annee", "filiere", "production","perimetrespatial"]]

# Rename columns
df = df.rename(columns={"production": "production_MWh"})

# Keep only the rows we need
df = df[df["filiere"] != "Chaleur"]
df = df[df["perimetrespatial"].str.contains("France", na=False)]

# Keep only the columns we need
df = df[["annee", "filiere", "production_MWh"]]

# Save the file
df.to_csv("cleaneddata/productions-consolidées-par-filiere-en-france.csv",
index=False)

```

La fin de la transformation de chaque fichier a alors signée la fin de cette phase et le début de la mise en place de diverses méthodes d'analyses relatives aux objectifs définis précédemment.

## ***3 - Production***

Cette partie se concentre sur l'étude de la production d'électricité en France en fonction des différentes filières (nucléaire, gaz, éolien...). On va notamment chercher à prédire des productions jusqu'en 2030 pour avoir un aperçu de celle-ci et notamment les comparer entre elles pour établir leur part de production par rapport à la production totale pour mesurer l'impact du nucléaire sur la production.

### 3.1 - Data mining description

Afin de réaliser des prédictions de production, en tout premier lieu, il faut avoir un grand jeu de données. De ce fait, les premières données récoltées sur le site RTE n'étaient pas suffisantes, il a alors fallu réaliser de nombreuses recherches supplémentaires pour compléter celles-ci. Les données étaient parfois non accessibles ou non fiables. Les données sur la production sont représentées par une série temporelle. Il s'agit de valeur en fonction des années. Ce sont des données quantitatives.

Nous avons appliqué un ETL pour les nettoyer. Nous avons par exemple converti les productions en MWh, changé le format des données ou encore renommé les colonnes d'un CSV.

Voici un exemple en ligne de code pour les données sur la production d'électricité du nucléaire :

```
##-----Nettoyage des
données-----##
df = pd.read_csv('cleaneddata/production_electricite_nucleaire.csv')
# Formater les dates
#conversion TWh en MWh
df['production électricité TWh'] = df['production électricité TWh'] *
1000000
print(df.head())

df['annee'] = pd.to_datetime(df['annee'],format="%Y")
print(df)
#Formater les noms des colonnes pour lancer Prophet
df.columns = ['ds', 'y']
```

Pour réaliser des prédictions sur des données quantitatives et temporelles, nous avons cherché des librairies compatibles avec Python pour réaliser des prédictions qui prennent en compte ces critères.

Nous avons trouvé plusieurs librairies telle que ARIMA ou Prophet de Facebook et nous avons retenu Prophet pour faire ces prédictions. En effet, ARIMA était devenu obsolète et les autres ne permettant pas de réaliser l'objectif voulu en fonction de nos données.

## 3.2 - Choix des méthodes de prédiction

Prophet est donc une librairie permettant de réaliser des prédictions sur des séries temporelles. C'est un modèle additif qui permet de gérer des données saisonnières et tendancielles ce qui est le cas avec la production d'électricité puisqu'elle dépend par exemple de la saison (hiver, automne...) mais aussi des tendances (grèves, crise économique...)

Il décompose une série temporelle en utilisant la formule suivant :

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

Avec

$g(t)$  : tendance

$s(t)$  : saisonnalité

$h(t)$  : effet de vacances

$e(t)$  : erreur

Il va donc réaliser une régression linéaire.

## 3.3 - Mise en œuvre des méthodes pour prédire la production d'électricité

Après avoir nettoyé les données, nous avons commencé par modéliser l'évolution de la production d'électricité en fonction des filières.

Voici un exemple du code utilisé pour réaliser cela :

```
plt.figure(figsize=(17, 8))
plt.plot(df['ds'], df['y']) (1)
plt.xlabel('Année')
plt.ylabel("Production d'électricité en MWh")
plt.title("Production d'électricité en MWh pour le bioénergie en
fonction des années")
plt.grid(False)
```

(1) Construction de la figure

Ensuite, nous avons réalisé les prédictions avec Prophet.

Le but du modèle de Prophet est de s'entraîner sur des données présentes pour prédire des données sur lesquelles il ne se serait pas entraîné. Il va ensuite utiliser des données historiques pour réaliser au mieux les prédictions. Voici un exemple du code :

```

#Creation objet prophet
m = Prophet() (1)
#Entrainement du modèle
m.fit(df) (2)
#prédition des 11 prochaines années par défaut ça compare avec les
données historiques
future = m.make_future_dataframe(periods=11, freq='Y') (3)
forecast = m.predict(future) (4)

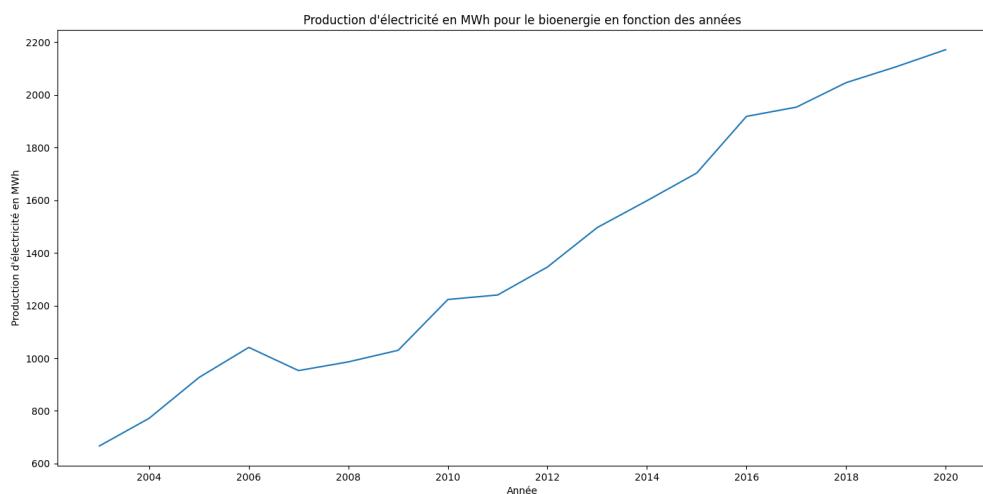
```

- (1) Création de l'objet prophet
- (2) Entrainement du modèle sur les données
- (3) Configuration du modèle pour réaliser la prédition des 11 prochaines années en ajoutant à chaque itération, une année
- (4) Réalisation de la prédition

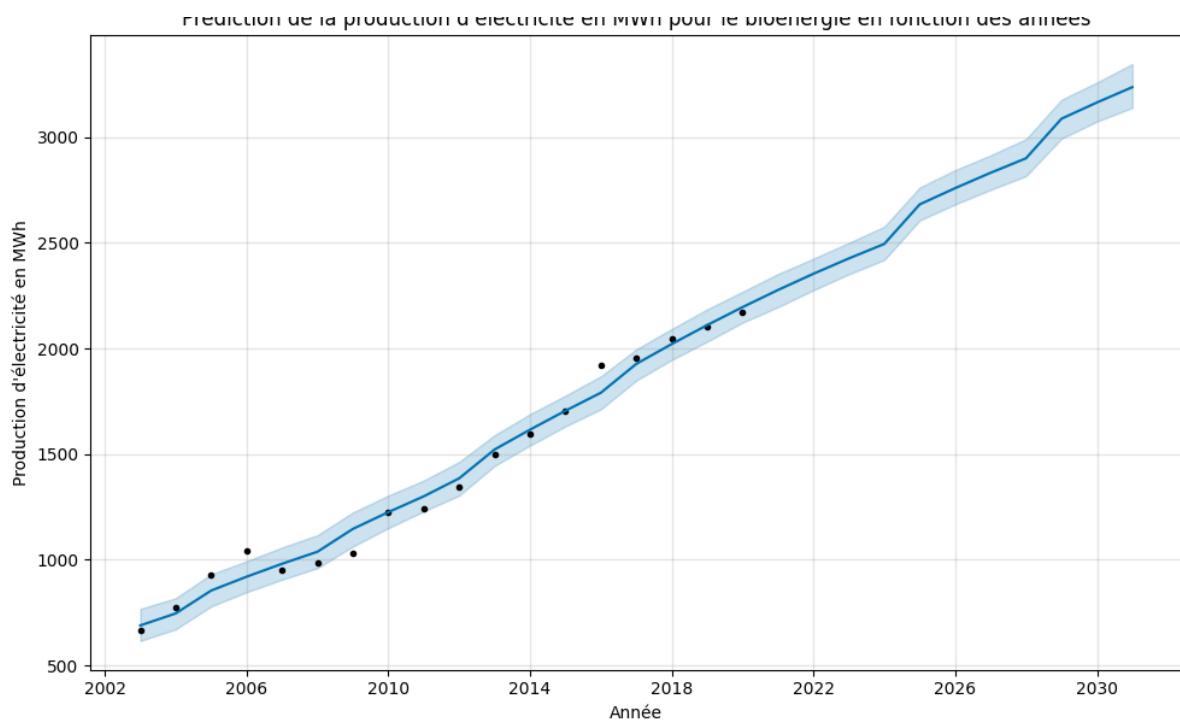
Enfin, nous avons réalisé un graphique représentant l'analyse de la prédition qui présente la différence entre les données réelles et les données prédites. Cela indique notamment si les prédictions sont correctes ou non.

### *3.3.1 - Prédictions pour la bioénergie*

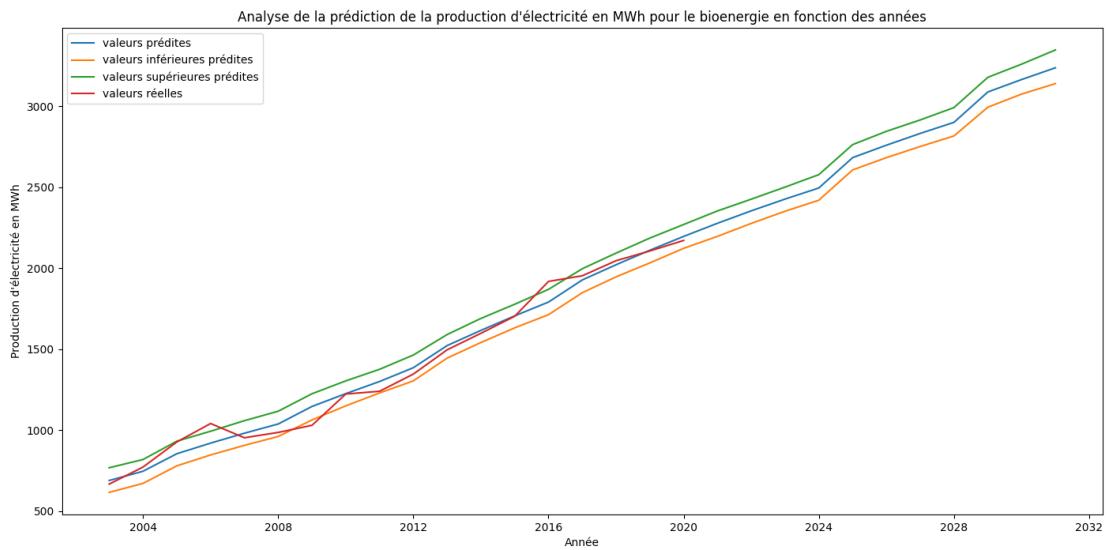
Au niveau de la bioénergie, voici le graphique montrant l'évolution de la production d'électricité pour la bioénergie de 2003 à 2020 :



On voit sur ce graphique que la production d'électricité pour la bioénergie ne cesse d'augmenter d'année en année passant de 600 MWh à 2200 MWh. Voici le graphique montrant la courbe de régression pour la production de l'électricité en fonction des années :



On voit que cette production est censée augmenter pour les prochaines années passant de 600 MWh en 2003 à plus de 3000 MWh en 2030. Enfin, voici un graphique montrant l'analyse de la prédition réalisée pour la production d'électricité pour le bioénergie.



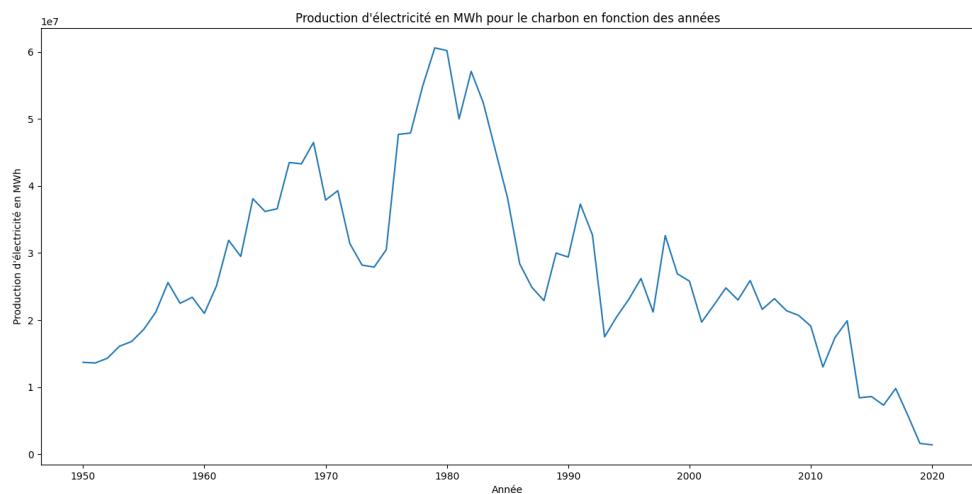
La courbe rouge représente les données réelles, la bleu, les valeurs prédites de production et les courbes vertes et orange représentent un intervalle dans lequel la production évolue (le maximum et le minimum).

On voit avec ce graphique que les valeurs prédites sont bien représentatives des valeurs actuelles ce qui montre que le modèle a réalisé des prédictions correctes.

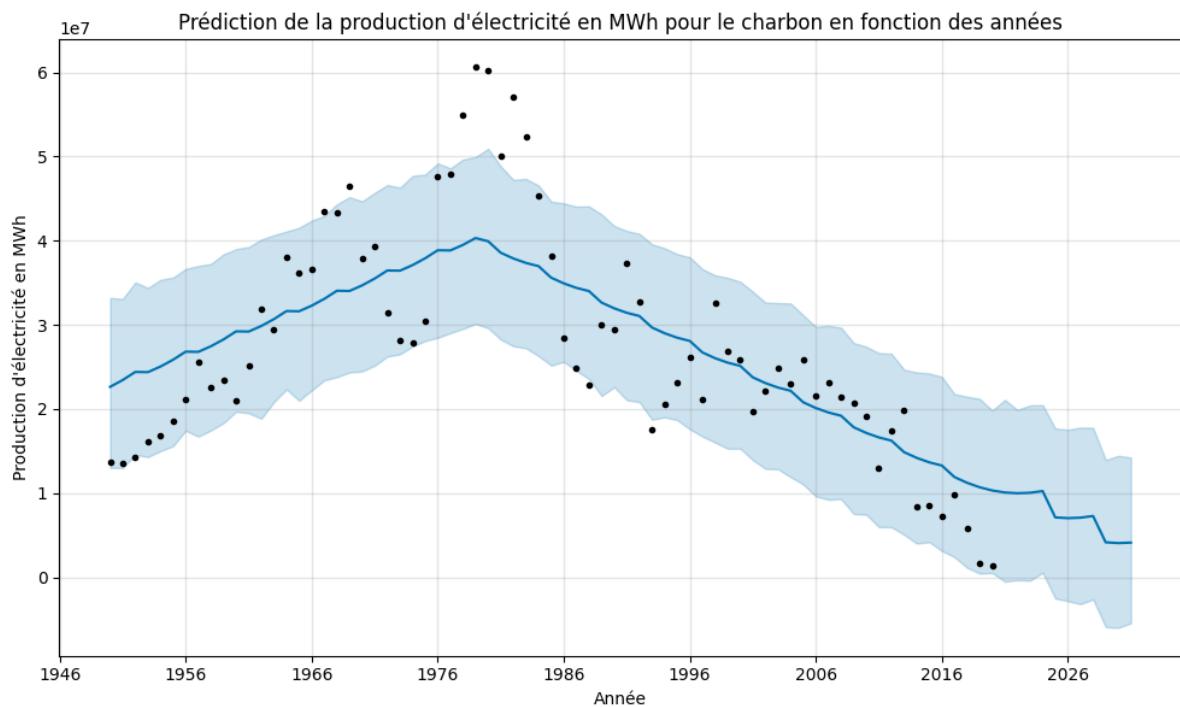
Néanmoins, on peut s'interroger sur l'efficacité du modèle étant donné que nous disposons de peu de données. L'efficacité sera testé dans la partie suivante.

### 3.3.2 - *Prédictions pour le charbon*

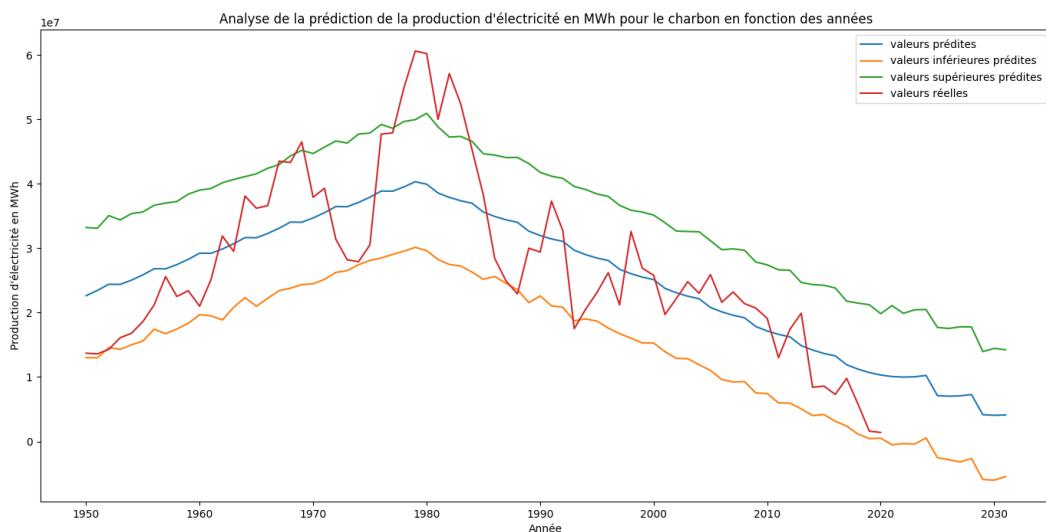
Au niveau du charbon voici le graphique montrant l'évolution de la production d'électricité pour le charbon de 2003 à 2020 :



D'après le graphique, la production d'électricité par le charbon est plutôt instable. Elle n'a cessé d'augmenter de 1950 à 1980 et de baisser à partir de 1980 ce qui correspond au fait que le nucléaire a été en expansion avec les nouvelles politiques sur l'environnement. Voici le graphique montrant la courbe de régression pour la production de l'électricité en fonction des années :



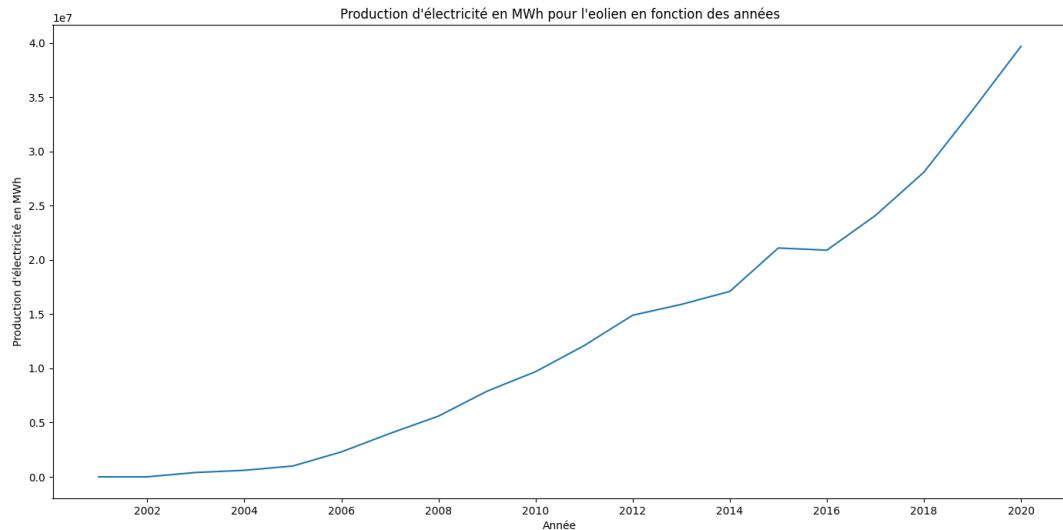
Les résultats de prédiction correspondent avec l'évolution de celle-ci présentée dans le graphique précédent. En effet, celle-ci est en baisse dans les années qui suivent passant en dessous des 10 000 000 MWh en 2030 alors qu'en 1976, celle-ci était de 40 000 000 MWh. Enfin, voici un graphique montrant l'analyse de la prédiction réalisée pour la production d'électricité pour le charbon :



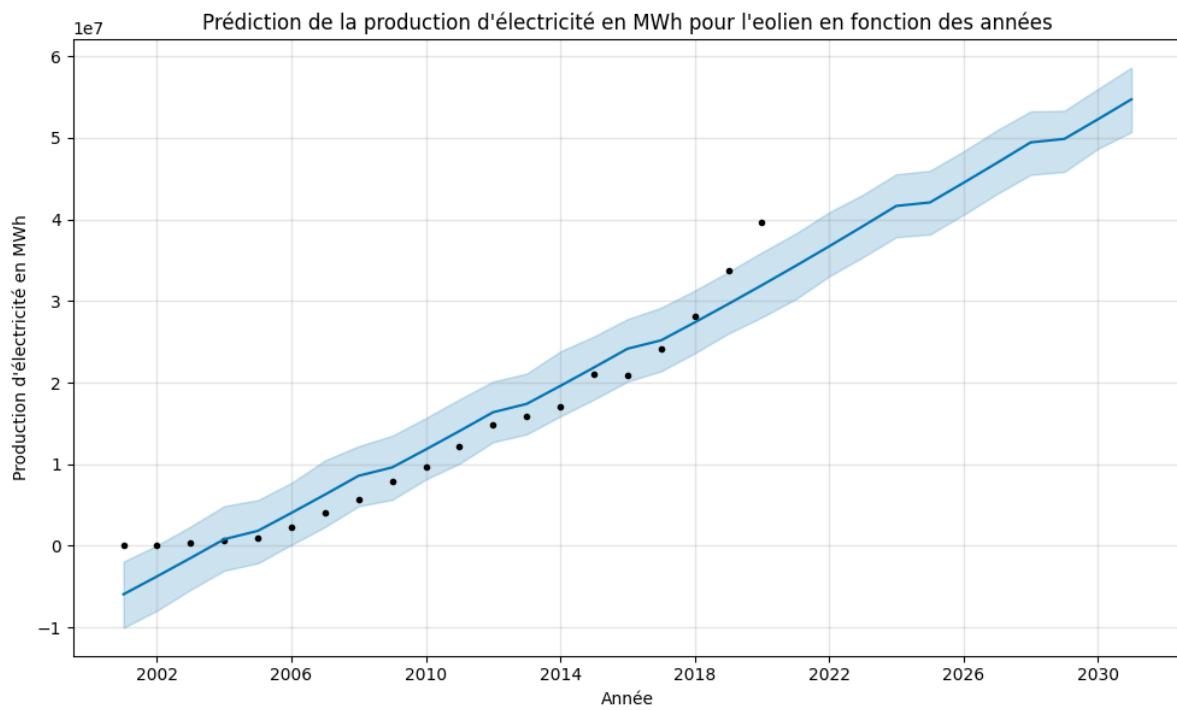
La courbe rouge représente les données réelles, la bleu, les valeurs prédictes de production et les courbes vertes et orange représentent un intervalle dans lequel la production évolue (le maximum et le minimum). On peut voir que les valeurs prédictes représentent en général plutôt bien les données réelles ce qui montrent que le modèle est plutôt pertinent sur ce jeu de données. Reste à savoir si cela est vrai dans la partie suivante de ce rapport.

### 3.3.3 - Prédictions pour l'éolien

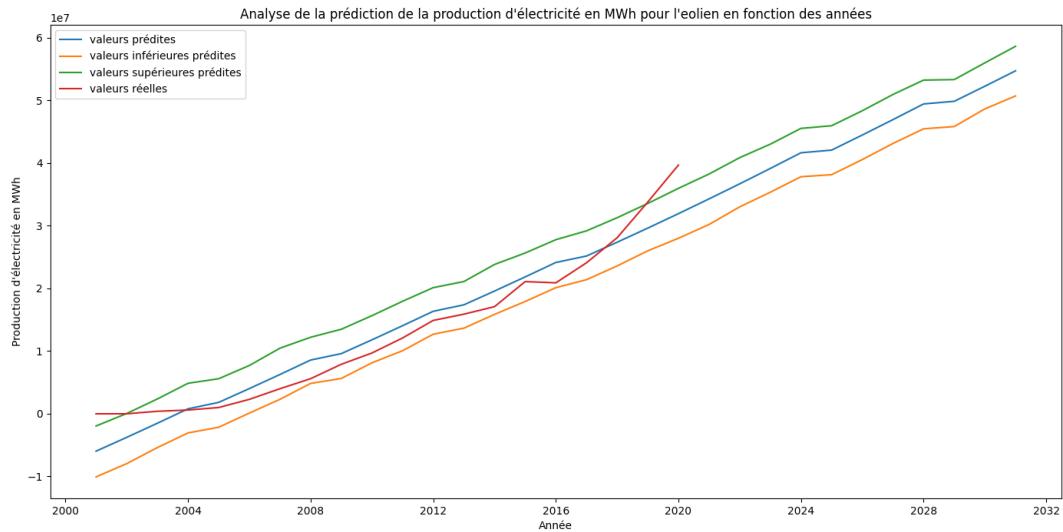
Au niveau de l'éolien, voici le graphique montrant l'évolution de la production d'électricité pour l'éolien de 2001 à 2020 :



À travers ce graphique, on voit que cette production est en constante augmentation depuis 2001 ce qui correspond avec les nouvelles politiques sur le développement des énergies renouvelables en France. Voici le graphique montrant la courbe de régression pour la production de l'électricité en fonction des années :



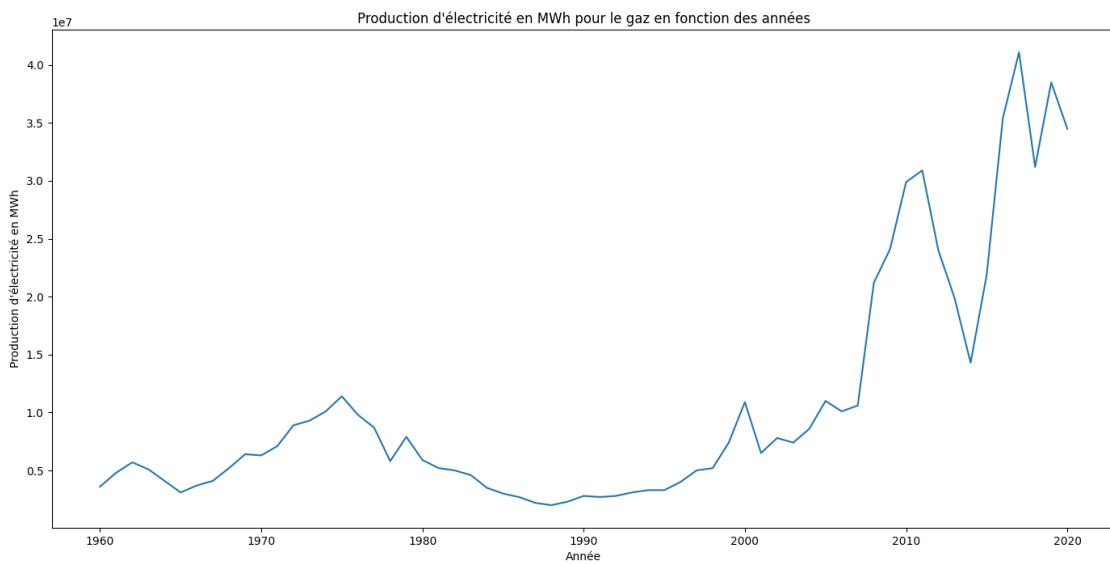
Comme pour le graphique précédent, on voit une nette progression de la production de 2001 à 2030 passant de 0 MWh à 60 000 000 MWh. Enfin, voici un graphique montrant l'analyse de la prédition réalisée pour la production d'électricité pour l'éolien :



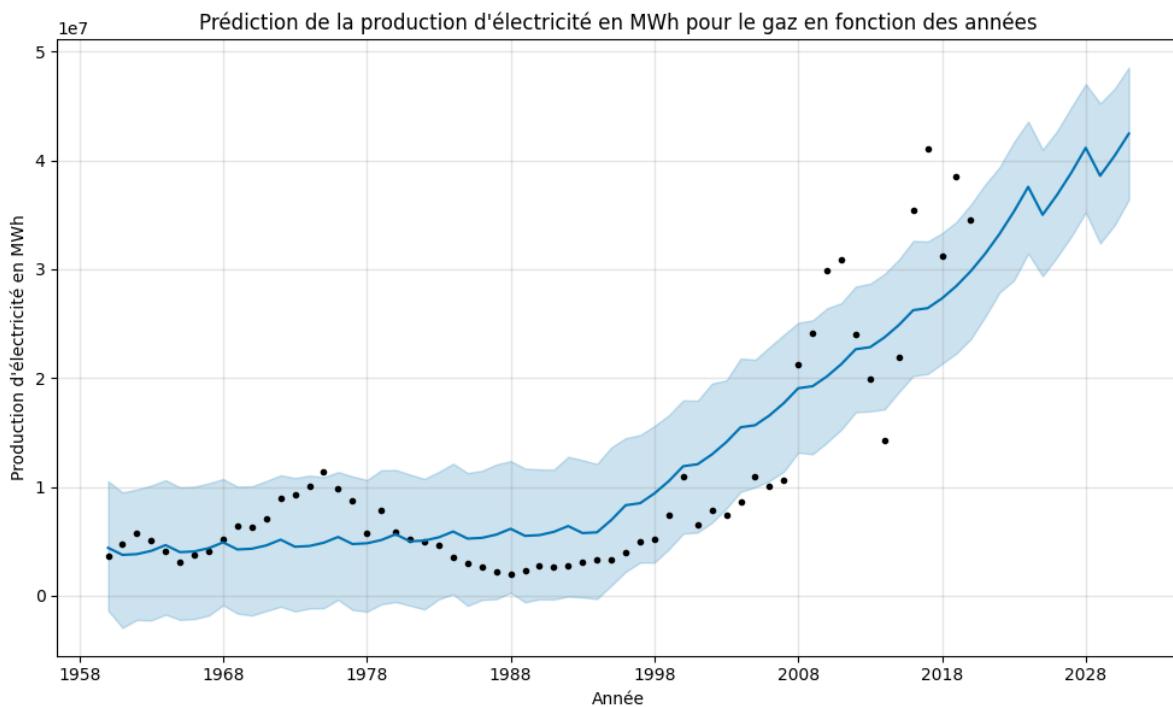
La courbe rouge représente les données réelles, la bleu, les valeurs prédictes de production et les courbes vertes et orange représentent un intervalle dans lequel la production évolue (le maximum et le minimum). On peut voir que les valeurs prédictes représentent en général bien les vraies valeurs donc à priori, sur ce jeu de données, la prédition semble correcte.

### 3.3.4 - Prédictions pour le gaz

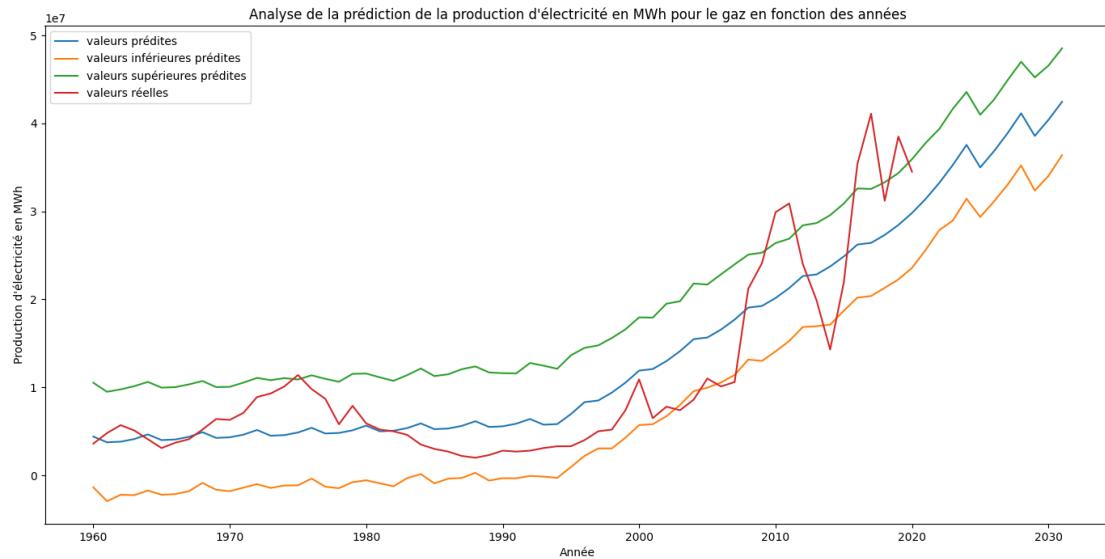
Pour le gaz, voici le graphique montrant l'évolution de la production d'électricité de 1960 à 2020 :



On peut voir sur graphique que la production d'électricité par le gaz est plutôt restée stable de 1960 à 2009 et qu'elle a connu une augmentation à partir de 2009 avec des périodes de chute notamment en 2015. Elle est donc plutôt instable à partir des années 2010. Voici le graphique montrant la courbe de régression pour la production de l'électricité en fonction des années :



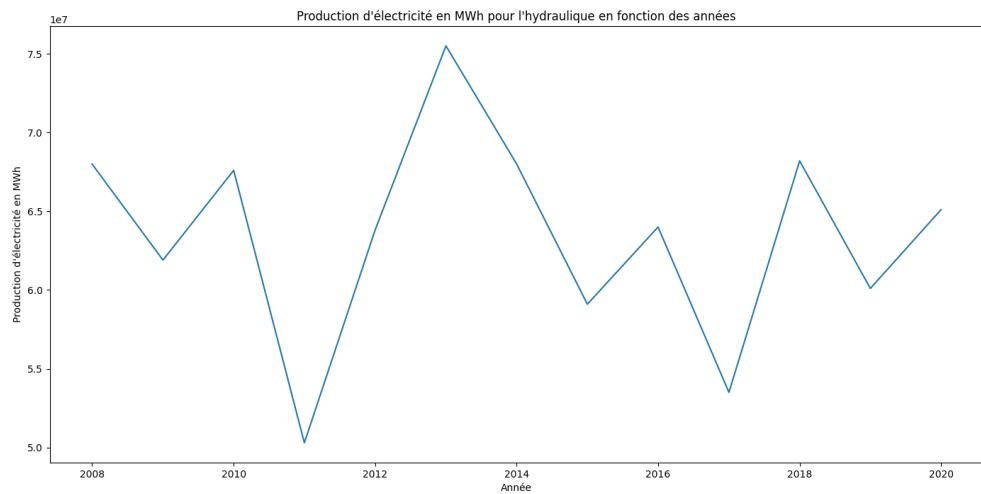
Ce graphique est bien représentatif de l'évolution de la production par le gaz et on peut voir notamment la tendance de croissance de celle-ci pour les années 2030. Pour finir, voici un graphique montrant l'analyse de la prédition réalisée pour la production d'électricité pour le gaz :



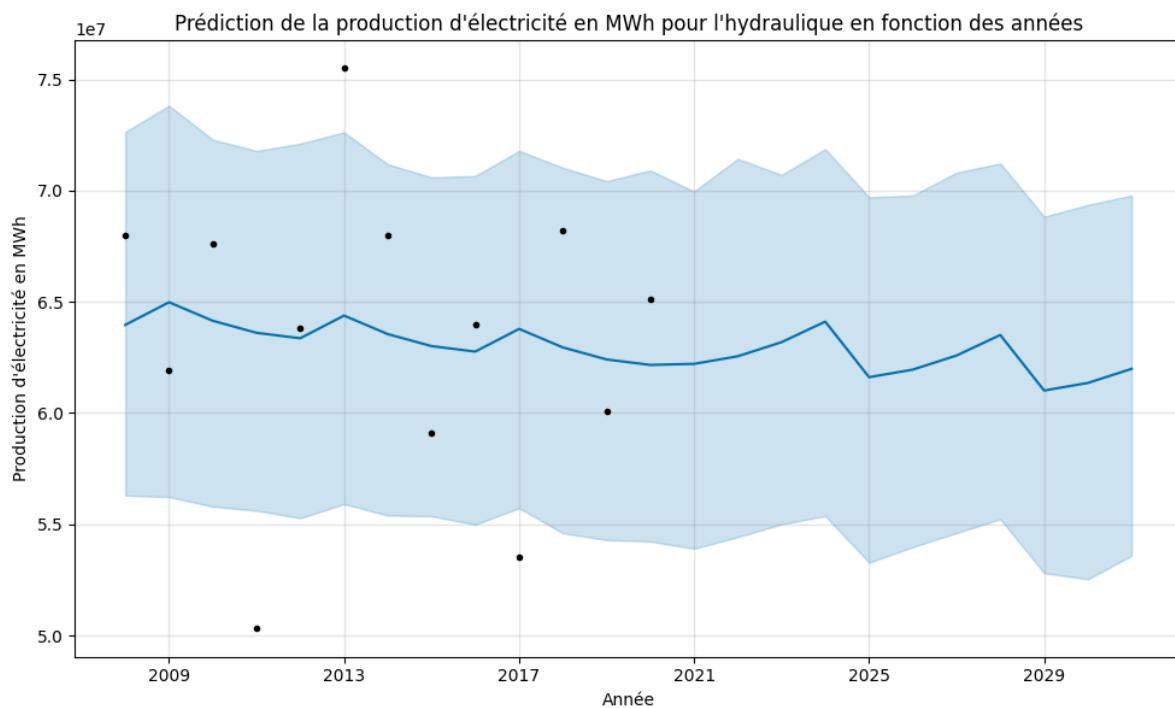
On voit que les données réelles (courbe rouge) sont bien comprises entre l'intervalle des valeurs prédictives (courbes verte et orange) et qu'elles sont bien représentées par la courbe bleue (les valeurs prédictives). Cela montre encore une fois la pertinence du modèle mais il faudra le vérifier après.

### 3.3.5 - Prédictions pour l'hydraulique

Maintenant au tour de la production de l'électricité par l'énergie hydraulique, voici son graphique d'évolution de 2008 à 2020 :

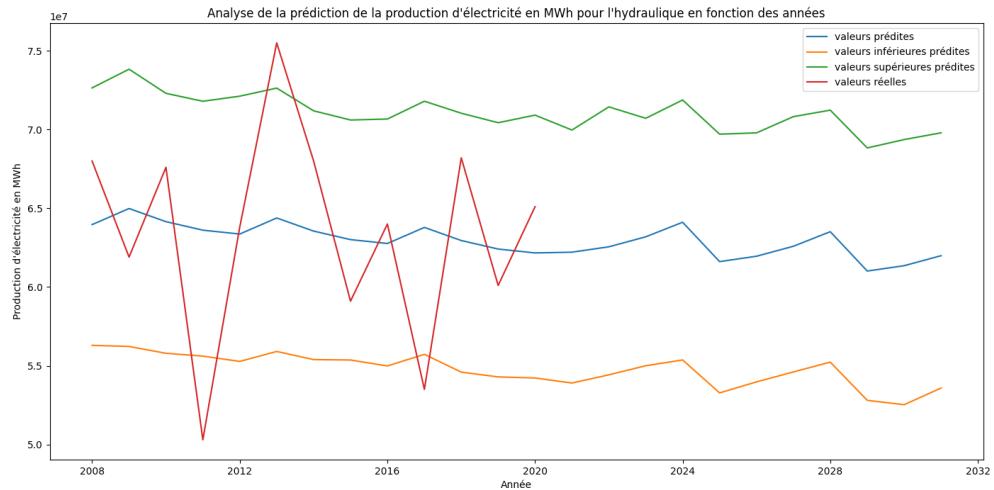


D'après le graphique, la production est très difficile à modélisé, elle ne fait qu'augmenter et baisser en fonction des années. Par exemple, on a commencé à presque 70 000 000 MWh en 2008 pour descendre à 50 000 000 MWh en 2011 et remonter à 75 000 000 en 2014. Voici le graphique montrant la courbe de régression pour la production de l'électricité en fonction des années :



Sur ce graphique, on peut voir que la zone de prédiction de production est très large, le modèle a eu du mal à bien représenter les données et les prédictions risquent donc d'être incorrectes. Il a été difficile d'approcher la courbe précédente avec une courbe linéaire sur des données irrégulières et peu nombreuses.

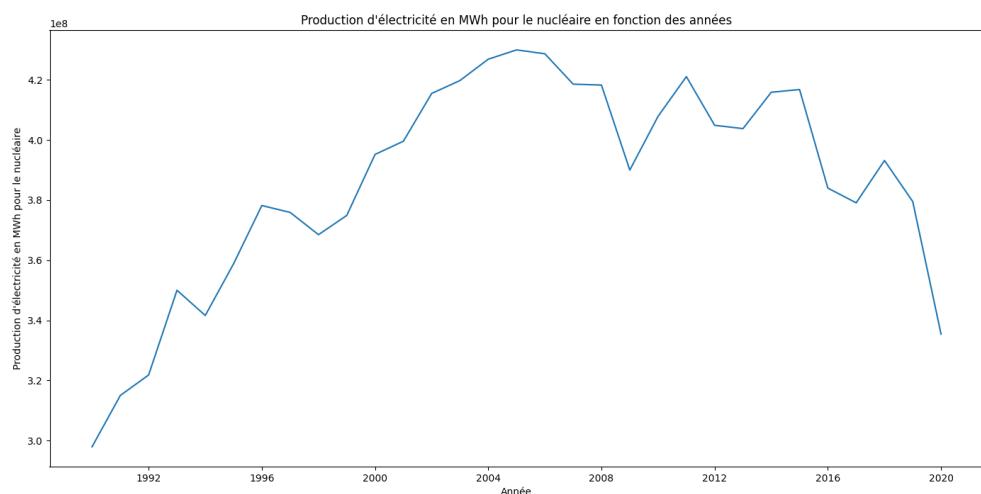
Le graphique montrant l'analyse de la prédition pour la production d'électricité par l'hydraulique devrait présenter les mêmes résultats :



Le graphique ci-dessus montre bien la difficulté à réaliser les prévisions étant donné que la ligne bleue (valeur prédictée) ne correspond pas vraiment à la courbe rouge (valeur réelle) donc les prédictions futures sont incertaines. Le taux de variation est trop élevé, on a un minimum à 55 000 000 MWh et un maximum à 75 000 000 MWh.

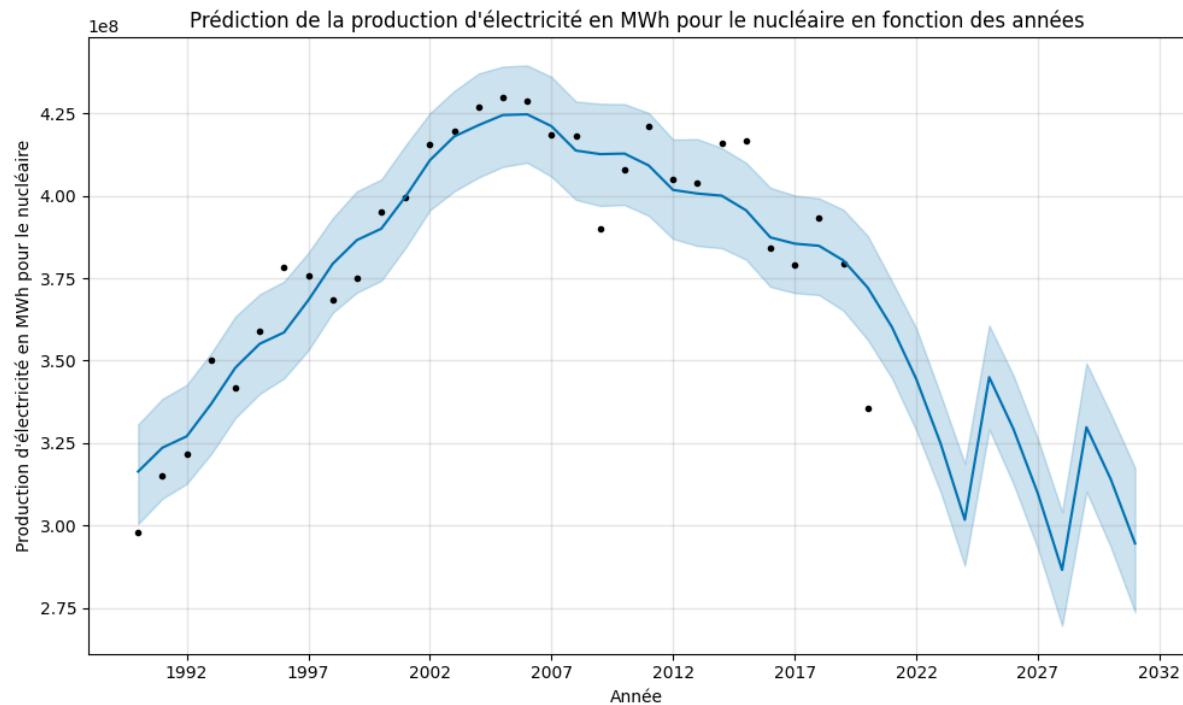
### 3.3.6 - Prédictions pour le nucléaire

Parlons maintenant du nucléaire, voici donc le graphique représentant l'évolution de la production d'électricité pour le nucléaire de 1990 à 2020 :



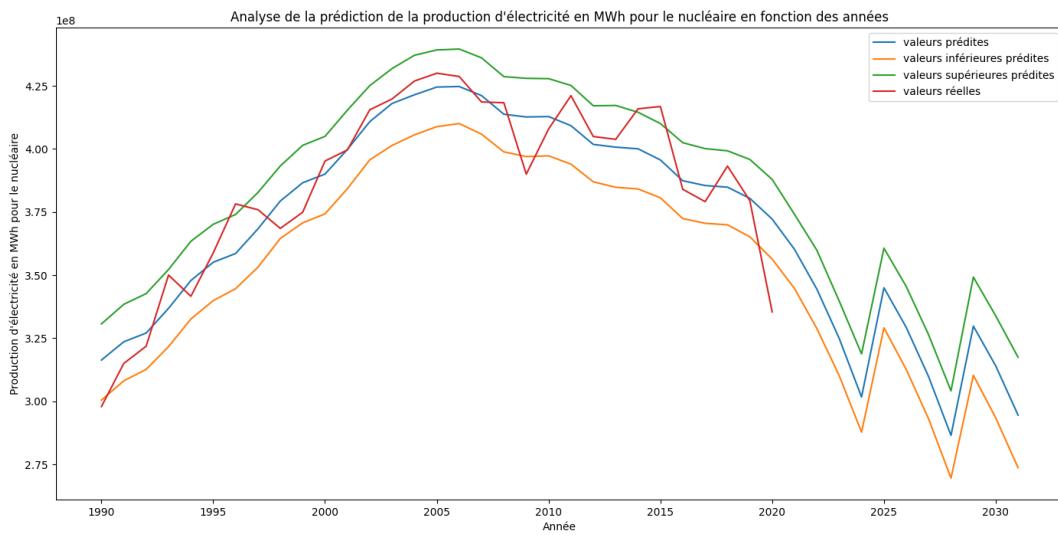
On voit que depuis 1990, la production ne cesse d'augmenter passant de 300 000 000 MWh à 420 000 000 MWh en 2007 avec quelques périodes de chute à partir de 2007.

Néanmoins, à partir de 2017, celle-ci n'a fait que baisser. Cela peut être en rapport avec la fermeture des centrales ou des grèves. Voici le graphique montrant la courbe de régression pour la production de l'électricité pour le nucléaire en fonction des années :



Cette courbe est bien représentative de la production du nucléaire, on voit bien la tendance à l'augmentation de la production jusqu'en 2007 avec le pic le plus haut. De plus, il y a la diminution de la production à partir de cette même année comme dans le graphe précédent. La prédition semble être bien représentative. Néanmoins, même si on prévoit des baisses et des augmentations pour les années à venir, cela semble un peu trop irrégulier. Il faudra bien vérifier la cohérence du modèle dans la partie d'après.

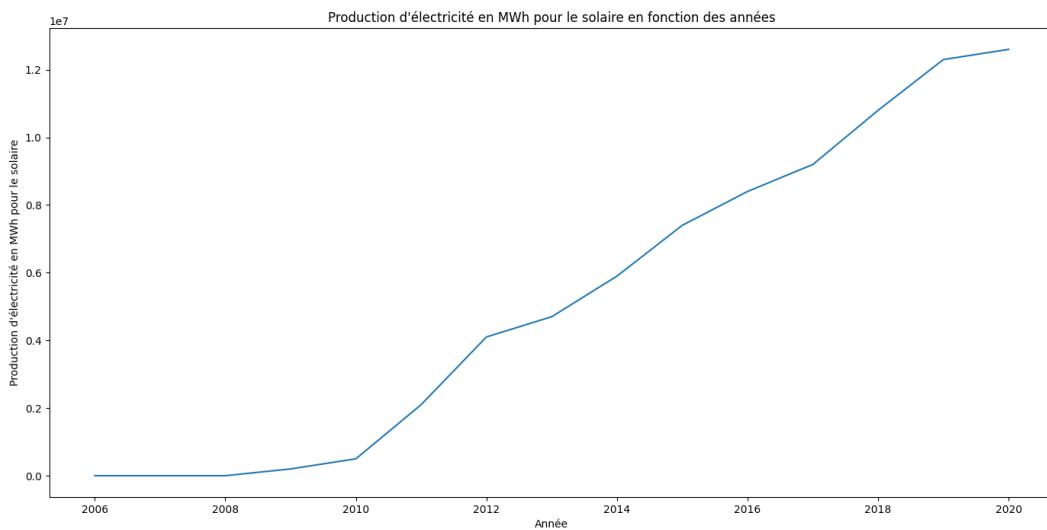
Pour finir, voici un graphique montrant l'analyse de la prédition réalisée pour la production d'électricité :



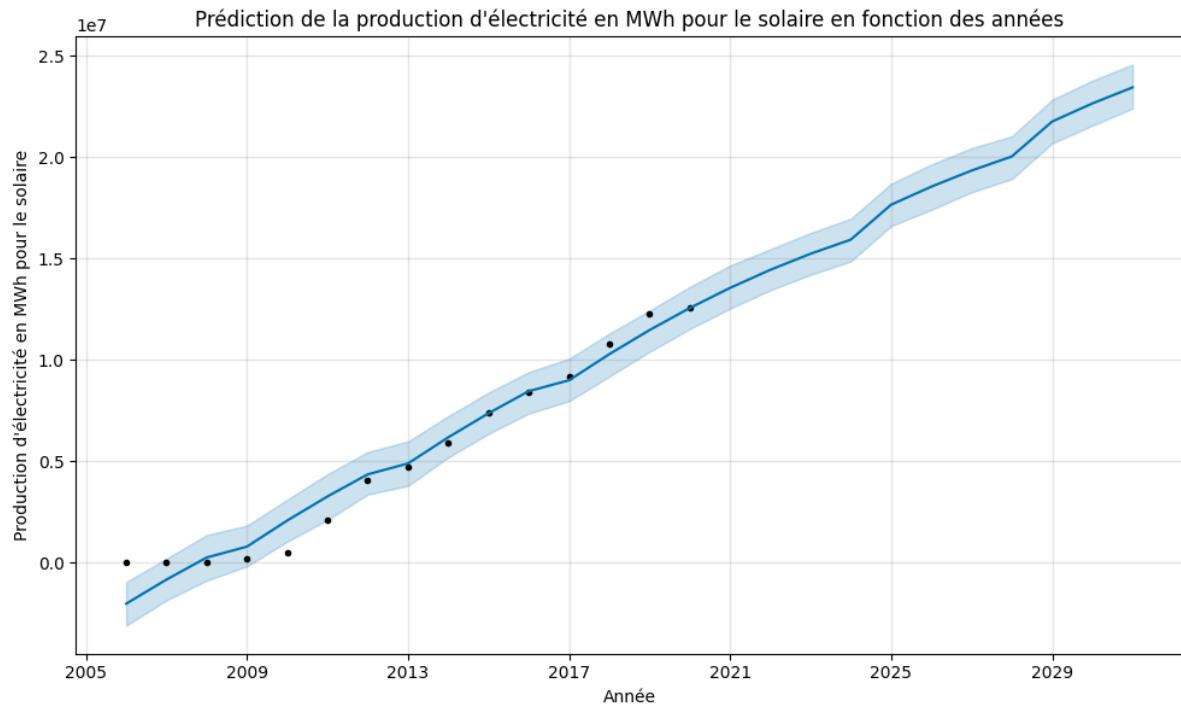
Sur ce graphique, on retrouve les mêmes problèmes cités précédemment, il s'agit clairement d'une bonne représentation mais les valeurs prédites maximales et minimales pour les dernières années sont un peu trop chaotiques.

### 3.3.7 - *Prédictions pour le solaire*

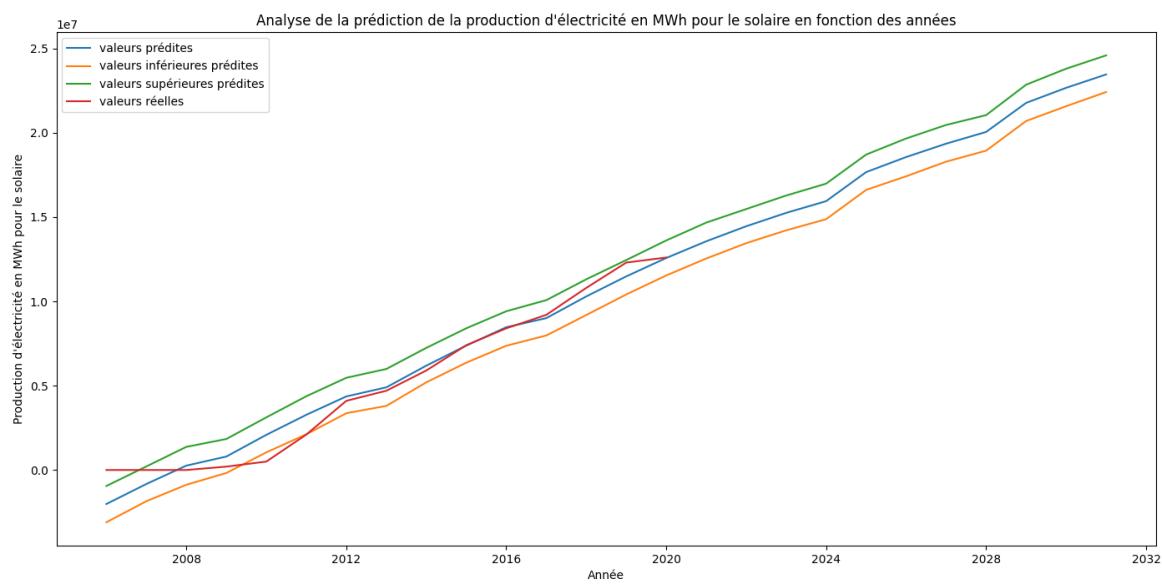
Pour finir, parlons de la production d'électricité par l'énergie solaire, voici donc le graphique représentant l'évolution de la production d'électricité pour le solaire de 2006 à 2020 :



D'après le graphique précédent, la production est en constante augmentation passant de 0 MWh en 2006 à 12 000 000 MWh en 2020. Voyons maintenant le graphique de prédiction de la production d'électricité pour le solaire :



Sur ce graphique, on peut voir que la courbe de prédiction suit bien les données du précédent graphe et montre qu'en 2030, la production sera encore plus haute, passant de 0 en 2006 à 25 000 000 MWh en 2030. Le dernier graphique montrant l'analyse de la prédiction réalisée pour la production d'électricité illustre bien ce fait :



## 3.4 - Efficacité des modèles de prédition utilisés

Afin de tester l'efficacité du modèle Prophet pour réaliser les prévisions de production, nous avons réalisé une cross validation qui est intégrée à la librairie Prophet. Voici le code utilisé pour réaliser la cross validation :

```
# 90 % entrainement et 10 % test ( 365 * 4 ans =  jours) (de 2006 à  
2010) donc je veux tester sur les 10 ans restant soit 365*10= 3650 days  
df_cv = cross_validation(m, initial = '1460 days', period='365 days',  
horizon = '3650 days') (1)
```

- (1) Le paramètre initial correspond à l'ensemble des données sur lequel le modèle s'entraîne, le paramètre period indique le nombres de données à calculer à chaque itération (nous prédisons année par année) et le paramètre horizon indique le nombre de données à prédire

Le but de cette cross-validation est d'entraîner le modèle sur un maximum de données possibles et de lui demander de prédire les valeurs des données sur lesquelles il ne s'est pas entraîné. Dans notre cas, pour toutes les filières, nous avons entraîné notre modèle sur toutes les données en enlevant 10 ans (En général, le modèle s'entraîne sur les données jusqu'en 2010 et il prévoit les 10 dernières années (jusqu'en 2020) puisque les données récoltés s'arrêtent en 2020)

### 3.4.1 - Efficacité du modèle pour la bioénergie

Pour mesurer l'efficacité du modèle avec la cross-validation pour les prédictions de la bioénergie, nous allons nous intéresser aux paramètres suivants :

INDICATEUR PERFORMANCE								
	horizon	mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	1102.269878	33.200450	33.200450	0.026775	0.026775	0.027138	1.0
1	728 days	13730.638773	117.177808	117.177808	0.087056	0.087056	0.091018	0.0
2	1094 days	16429.685827	128.178336	128.178336	0.085681	0.085681	0.089516	0.0
3	1459 days	35237.990994	187.717849	187.717849	0.117470	0.117470	0.124801	0.0
4	1824 days	67857.171189	260.494091	260.494091	0.152962	0.152962	0.165629	0.0
5	2189 days	205636.354965	453.471449	453.471449	0.236429	0.236429	0.268126	0.0
6	2555 days	122130.662403	349.471977	349.471977	0.178941	0.178941	0.196524	0.0
7	2920 days	160009.192101	400.011490	400.011490	0.195509	0.195509	0.216692	0.0
8	3285 days	183002.342828	427.787731	427.787731	0.203128	0.203128	0.226091	0.0
9	3650 days	221619.769427	470.765089	470.765089	0.216843	0.216843	0.243212	0.0

Le paramètre le plus important dans le tableau ci-dessus est l'Erreur Absolue Moyenne (MAE) qui représente en moyenne ici 282,827 MWh. Etant donné que le minimum de production pour la bioénergie de 2003 à 2020 est de 667 MWh et que le maximum est de

2171 MWh, les prédictions sont bonnes à 282 MWh près ce qui est plutôt satisfaisant. On voit notamment que plus le nombre d'années passent, plus le MAE augmentent ce qui est logique étant donné le peu de données dont nous disposons pour l'entraînement.

### 3.4.2 - Efficacité du modèle pour le charbon

Passons maintenant à la mesure de l'efficacité du modèle avec la cross-validation pour le charbon :

	horizon	mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	3.007050e+14	1.734085e+07	1.734085e+07	1.333911	1.333911	0.800208	0.0
1	728 days	1.859813e+14	1.363750e+07	1.363750e+07	0.783764	0.783764	0.563097	1.0
2	1094 days	8.144403e+13	9.024634e+06	9.024634e+06	0.453499	0.453499	0.369675	1.0
3	1459 days	4.485905e+14	2.117995e+07	2.117995e+07	2.521423	2.521423	1.115323	0.0
4	1824 days	4.689984e+14	2.165637e+07	2.165637e+07	2.518183	2.518183	1.114688	0.0
5	2189 days	5.594654e+14	2.365302e+07	2.365302e+07	3.240140	3.240140	1.236662	0.0
6	2555 days	3.625276e+14	1.904016e+07	1.904016e+07	1.942873	1.942873	0.985511	0.0
7	2920 days	5.614757e+14	2.369548e+07	2.369548e+07	4.085427	4.085427	1.342692	0.0
8	3285 days	8.163532e+14	2.857189e+07	2.857189e+07	17.857434	17.857434	1.798564	0.0
9	3650 days	8.683951e+14	2.946854e+07	2.946854e+07	21.048961	21.048961	1.826456	0.0

Au niveau de la MAE mesurée ici, elle est de 20 726 839.96602 en moyenne.

La production minimum pour le charbon est de 1 400 000 MWh et la maximum est de 60 600 000 MWh de 1950 à 2020. Celle-ci est quand même importante mais cela peut sembler satisfaisant étant donné que l'écart entre la production minimale et maximale est énorme.

### 3.4.3 - Efficacité du modèle pour l'éolien

Au niveau de l'efficacité du modèle pour l'éolien, voici le tableau des paramètres de la cross-validation :

	horizon	mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	9.836186e+12	3.136270e+06	3.136270e+06	0.259196	0.259196	0.297789	0.0
1	728 days	2.700418e+13	5.196554e+06	5.196554e+06	0.348762	0.348762	0.422425	0.0
2	1094 days	1.622787e+13	4.028383e+06	4.028383e+06	0.253357	0.253357	0.290108	0.0
3	1459 days	2.003030e+13	4.475523e+06	4.475523e+06	0.261726	0.261726	0.301134	0.0
4	1824 days	5.974442e+13	7.729452e+06	7.729452e+06	0.366325	0.366325	0.448467	0.0
5	2189 days	4.610052e+13	6.789736e+06	6.789736e+06	0.324868	0.324868	0.387871	0.0
6	2555 days	6.117688e+13	7.821565e+06	7.821565e+06	0.324546	0.324546	0.387413	0.0
7	2920 days	1.225162e+14	1.106870e+07	1.106870e+07	0.393904	0.393904	0.490511	0.0
8	3285 days	2.567248e+14	1.602263e+07	1.602263e+07	0.474042	0.474042	0.621305	0.0
9	3650 days	4.487160e+14	2.118292e+07	0.533575	0.533575	0.727722	0.0	

La MAE est de 8 745 173.99403034 MWh en moyenne. La production minimum pour l'éolien est de 0.0 MWh et la maximum est de 39 700 000 MWh de 2001 à 2020.

Avec cette MAE, on peut en déduire que le modèle est assez bon mais la valeur du MAE reste assez élevée ce qui est inquiétant.

### 3.4.4 - Efficacité du modèle pour le gaz

Pour l'efficacité du modèle pour la production faites par le gaz, voici son tableau de paramètres :

	horizon	mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	4.003343e+14	2.000836e+07	2.000836e+07	0.647520	0.647520	0.957529	0.0
1	728 days	1.697160e+14	1.302751e+07	1.302751e+07	0.542813	0.542813	0.745015	0.0
2	1094 days	7.074248e+13	8.410855e+06	8.410855e+06	0.422656	0.422656	0.535909	0.0
3	1459 days	7.969821e+12	2.823087e+06	2.823087e+06	0.197419	0.197419	0.219040	1.0
4	1824 days	1.079284e+14	1.038886e+07	1.038886e+07	0.474377	0.474377	0.621880	0.0
5	2189 days	5.668215e+14	2.380801e+07	2.380801e+07	0.672543	0.672543	1.013280	0.0
6	2555 days	8.404990e+14	2.899136e+07	2.899136e+07	0.705386	0.705386	1.089724	0.0
7	2920 days	3.649473e+14	1.910359e+07	1.910359e+07	0.612295	0.612295	0.882456	0.0
8	3285 days	6.953436e+14	2.636937e+07	2.636937e+07	0.684919	0.684919	1.041637	0.0
9	3650 days	4.967782e+14	2.228852e+07	2.228852e+07	0.646044	0.646044	0.954306	0.0

La MAE est de 17 521 952.47582602 MWh en moyenne. La production minimum pour le gaz est de 2 000 000 MWh et la maximum est de 41 100 000 MWh de 1960 à 2020.

Ici, la MAE est énorme comparé à l'intervalle de production donc le modèle n'est pas si pertinent pour ce jeu de données pourtant il contient plus de données que certains autres jeux. Cela peut s'expliquer par l'irrégularité dans la production d'électricité observé durant ces années là comme on peut le voir dans le graphique montrant cette évolution dans la partie précédente.

### 3.4.5 - Efficacité du modèle pour l'hydraulique

Au niveau de l'efficacité du modèle pour l'hydraulique, voici les paramètres de la cross-validation :

	horizon	mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	5.313719e+14	2.305150e+07	2.305150e+07	0.458280	0.458280	0.372846	0.0
1	728 days	2.357976e+14	1.535570e+07	1.535570e+07	0.240685	0.240685	0.214832	0.0
2	1094 days	5.974479e+12	2.444275e+06	2.444275e+06	0.032375	0.032375	0.032907	0.0
3	1459 days	1.156854e+14	1.075572e+07	1.075572e+07	0.158172	0.158172	0.146580	0.0
4	1824 days	6.455388e+14	2.540746e+07	2.540746e+07	0.429906	0.429906	0.353846	0.0
5	2189 days	6.923032e+14	2.631165e+07	2.631165e+07	0.411120	0.411120	0.341020	0.0
6	2555 days	9.432071e+14	3.071168e+07	3.071168e+07	0.574050	0.574050	0.446029	0.0
7	2920 days	4.713965e+14	2.171167e+07	2.171167e+07	0.318353	0.318353	0.274637	0.0
8	3285 days	1.264756e+15	3.556341e+07	3.556341e+07	0.591737	0.591737	0.456634	0.0
9	3650 days	1.322603e+15	3.636761e+07	3.636761e+07	0.558642	0.558642	0.436671	0.0

La MAE est de 22 768 066.70 MWh. La production minimum pour l'hydraulique est de 50 300 000 MWh et la maximum est de 75 500 000 MWh de 2008 à 2020. La MAE est très élevée donc le modèle pour l'hydraulique n'est pas très adapté. Cela s'explique par le fait que c'est l'un des jeux de données avec le moins de données collectées.

### 3.4.6 - Efficacité du modèle pour le nucléaire

Passons maintenant au nucléaire, voici les paramètres de la cross-validation :

horizon		mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	1.104968e+14	1.051175e+07	1.051175e+07	0.024963	0.024963	0.025278	1.0
1	728 days	9.092664e+13	9.535546e+06	9.535546e+06	0.023550	0.023550	0.023276	1.0
2	1094 days	3.623905e+13	6.019888e+06	6.019888e+06	0.014908	0.014908	0.014798	1.0
3	1459 days	9.442991e+13	9.717505e+06	9.717505e+06	0.023365	0.023365	0.023641	1.0
4	1824 days	1.104994e+14	1.051187e+07	1.051187e+07	0.025220	0.025220	0.025543	1.0
5	2189 days	6.830603e+14	2.613542e+07	2.613542e+07	0.068061	0.068061	0.065821	0.0
6	2555 days	6.980040e+14	2.641976e+07	2.641976e+07	0.069691	0.069691	0.067344	0.0
7	2920 days	7.538358e+13	8.682372e+06	8.682372e+06	0.022081	0.022081	0.021840	1.0
8	3285 days	5.057105e+14	2.248801e+07	2.248801e+07	0.059257	0.059257	0.057552	0.0
9	3650 days	4.961131e+15	7.043530e+07	7.043530e+07	0.210004	0.210004	0.190048	0.0

On voit que la MAE moyenne est de 20 045 742.0915 MWh. La production minium par le nucléaire est de 297 900 000 MWh, la maximum est de 430 000 000 MWh de 1960 à 2020 . Étant donné l'ordre de grandeur de la production minimale et maximale, la MAE est tout à fait acceptable donc on peut dire que le modèle est plutôt bien adapté pour les données portant sur le nucléaire.

### 3.4.7 - Efficacité du modèle pour le solaire

Pour finir, nous avons obtenu le tableau suivant pour la mesure de l'efficacité du modèle sur le solaire :

horizon		mse	rmse	mae	mape	mdape	smape	coverage
0	363 days	2.750488e+12	1.658460e+06	1.658460e+06	0.789743	0.789743	1.305082	0.0
1	728 days	1.296149e+13	3.600207e+06	3.600207e+06	0.878099	0.878099	1.565378	0.0
2	1094 days	1.553668e+13	3.941660e+06	3.941660e+06	0.838651	0.838651	1.444271	0.0
3	1459 days	2.579297e+13	5.078678e+06	5.078678e+06	0.860793	0.860793	1.511214	0.0
4	1824 days	4.248551e+13	6.518091e+06	6.518091e+06	0.880823	0.880823	1.574055	0.0
5	2189 days	5.564920e+13	7.459839e+06	7.459839e+06	0.888076	0.888076	1.597368	0.0
6	2555 days	6.402067e+13	8.001292e+06	8.001292e+06	0.869706	0.869706	1.538901	0.0
7	2920 days	9.097935e+13	9.538309e+06	9.538309e+06	0.883177	0.883177	1.581588	0.0
8	3285 days	1.205104e+14	1.097772e+07	1.097772e+07	0.892498	0.892498	1.611731	0.0
9	3650 days	1.258765e+14	1.121947e+07	1.121947e+07	0.890434	0.890434	1.605014	0.0

On voit que la MAE est de 6 799 372.9004 MWh en moyenne. La production minimum est de 0 MWh et la maximum de 12 600 000 MWh de 2006 à 2020 .

La MAE correspond à plus de la moitié de la production maximum ce qui est énorme donc le modèle n'est pas adapté pour ce jeu de données.

## 3.5 - Synthèse des résultats sur la production

Nous avons vu précédemment, la production d'électricité pour chaque filière et la prévision de production d'électricité jusqu'en 2030.

Nous avons donc réalisé des calculs pour calculer la production moyenne par filière, la valeur de la production minimum et maximum ainsi que l'écart-type de production pour les données actuelles et prédictes afin de mieux interpréter tout cela.

Les prédictions ont été réalisées jusqu'en 2030 pour toutes les filières.

Voici un tableau récapitulatif de ces résultats :

	Bioénergie	Charbon	Eolien	Gaz	Hydraulique	Nucléaire	Solaire
Production Moyenne	1398.7	2762253 5.2	12960 000	10604 918.0 32	63469230.76 9	38602580 6.451	521333 3.33
Production minimum	667	1400000	0	20000 00	50300000	29790000 0	0.0
Production maximum	2171	6060000 0	39700 000	41100 000	75500000	43000000 0	126000 00
Ecart-type production	489.7	1361911 6.78	12000 105.2 6	10207 814.4 28	6670880.07	35727934. 614	468567 2.183
Prévision production moyenne	1909.92	2490555 0.02	24221 676.8 7	14688 525.9 407	62965275.50 9	36920810 5.815	108043 11.566
Prévision production minimum	689.31	4065391. 84	-5962 281.7 4	37541 74.43 7	61013260.38 6	28647986 9.311	-20186 45.99
Prévision production maximum	3236.7	4031573 8.693	54738 076.1 77	42448 935.7 598	64982809.46 3	42472932 6.367	234549 80.143
Ecart-type prévision production	790.4	1037918 6.5	18378 620.5 17	12194 497.3 33	1017466.147 7	40708582. 0559	779589 4.454

Afin de réaliser ce tableau, nous avons utilisé le code python suivant :

```

print("Minimum production : " + str(df["y"].min())) (1) #minimum de la production
print("Maximum production : " + str(df["y"].max())) (2) #maximum de la production
print("Moyenne production : " + str(df["y"].mean())) (3) #moyenne de la production
print("Ecart type production : " + str(df["y"].std())) (4) #écart type de la production
moyennePrevision = forecast['yhat'].mean() (5)

```

```

minProductionPrevision = forecast['yhat'].min() (6)
maxProductionPrevision = forecast['yhat'].max() (7)
ecartTypeProductionPrevision = forecast['yhat'].std()
forecast.to_csv('prevision-production/bioenergie/prevision-bioenergie.csv',
index=False)

```

- (1) Calcul de la valeur minimum du jeu de données
- (2) Calcul de la valeur maximum du jeu de données
- (3) Calcul de la moyenne de production
- (4) Calcul de l'écart type
- (5) Calcul de la moyenne de production prévue
- (6) Calcul de la valeur minimum de la production prévue
- (7) Calcul de la valeur du maximum de la production prévue

A travers ce tableau, on peut voir que la somme des productions d'électricité de toutes les filières en moyenne sans le nucléaire vaut : 119 871 416 MWh alors que la production moyenne du nucléaire vaut à elle seule : 386 025 806.451 MWh sur la période avant 2020. Actuellement, les filières ne peuvent pas compenser l'électricité produite par le nucléaire qui est 2 fois plus élevée.

Si on regarde la somme des capacités maximales de production pour chaque filière sans le nucléaire, on voit que celle-ci vaut : 229 502 171 MWh. Encore une fois, même en prenant les capacités maximales, nous n'arrivons pas à couvrir la production faite par le nucléaire en moyenne.

Par ailleurs, en faisant les mêmes calculs pour la période après 2020 avec les valeurs de productions prédites, on retrouve que la somme des prévisions de productions d'électricité de toutes les filières sans le nucléaire est égale à : 137 587 250 MWh en moyenne et celle du nucléaire vaut : 369 208 105.815 MWh en moyenne. Même après les années 2020, nous n'avons toujours pas rattrapé la production du nucléaire avec les autres filières. En prenant la capacité maximum de production prédites pour toutes les filières sauf le nucléaire, on obtient 225 943 777 MWh ce qui est toujours insuffisant pour rattraper la production d'électricité moyenne prédite du nucléaire.

D'autre part, grâce aux données récoltées, nous avons calculé la part de production en pourcentage de chaque filière de 2020 à 2030 comme suit :

	Bioénergie	Charbon	Eolien	Gaz	Hydraulique	Nucléaire	Solaire
Part de production 2020	0.000444	0.28647	8.123 55	7.059 51	13.32099	68.63075 7	2.5782 5
Part de production 2021	0.000469	1.99256	7.320 41	6.627 90	12.47676	68.69921	2.8826 7

Part de production 2022	0.00049	2.05933	8.02653	7.23200	12.95245	66.60400	3.12516
Part de production 2023	0.00052	2.17457	8.84115	7.970392	13.60580	64.02452	3.38301
Part de production 2024	0.00052	1.39968	8.27731	6.88167	12.11749	67.84898	3.474313
Part de production 2025	0.00055	1.41097	8.93295	7.39129	12.438776	66.09962	3.72581
Part de production 2026	0.00058	1.46196	9.68886	8.02079	12.91663	63.91766	3.99348
Part de production 2027	0.00061	1.55678	10.5682	8.79241	13.57330	61.22458	4.28402
Part de production 2028	0.00061	0.82241	9.87340	7.6359	12.07769	65.28042	4.30944
Part de production 2029	0.00063	0.82165	10.56692	8.16483	12.40028	63.46550	4.58016
Part de production 2030	0.00067	0.85643	11.37390	8.82037	12.88032	61.19463	4.87365

Afin de réaliser ce tableau, nous avons utilisé le code python suivant :

```
#sélection production totale de 2021 à 2030 et production par année des filières &
prévision production
for année in range(2021, 2031):
    #bioénergie
    df_prevision =
pd.concat([df_prevision_bioénergie[['ds', 'yhat']] [df_prevision_bioénergie['ds'] ==
str(année) + '-12-31'], df_prevision_eolien[['ds', 'yhat']] [df_prevision_eolien['ds'] ==
str(année) +
'-12-31'], df_prevision_charbon[['ds', 'yhat']] [df_prevision_charbon['ds'] ==
str(année) + '-12-31'], df_prevision_gaz[['ds', 'yhat']] [df_prevision_gaz['ds'] ==
str(année) +
'-12-31'], df_prevision_hydraulique[['ds', 'yhat']] [df_prevision_hydraulique['ds'] ==
str(année) +
'-12-31'], df_prevision_nucleaire[['ds', 'yhat']] [df_prevision_nucleaire['ds'] ==
str(année) +
```

```

'-12-31'],df_prevision_solaire[['ds','yhat']][df_prevision_solaire['ds'] ==
str(annee) + '-12-31'])] (1)
    df_prevision = df_prevision.rename(columns={'yhat': 'production_'+str(annee)}) (2)
    df_prevision = df_prevision['production_'+str(annee)].sum() (3)
    df_stock_bioenergie =
df_prevision_bioenergie['yhat'][df_prevision_bioenergie['ds'] == str(annee) +
'-12-31'] (4) #df_stock est une série

    for production_actuelle in df_stock_bioenergie.values:
        df_part_bioenergie.loc[len(df_part_bioenergie)] = [str(annee),
(production_actuelle / df_prevision)*100]

```

- (1) Formation d'un DataFrame pour ajouter les prévisions des ventes pour toutes les filières
- (2) Renomme les colonnes pour la production
- (3) Réalise la somme de toutes les productions des filières pour l'année actuelle
- (4) Stockage des prévisions du bioénergie pour l'année

Ce tableau nous montre encore une fois la supériorité de la production d'électricité par le nucléaire par rapport aux autres filières. En effet, on voit qu'en 2020, la part de production d'électricité par le nucléaire est de 68.6 %, elle reste la plus élevée comparés aux autres. La plus faible étant celle du bioénergie avec 0.0004 %.

En 2030, la part de production du nucléaire a baissé, on est passé de 68.6 % à 61.2 % mais elle est toujours majoritaire par rapport aux autres.

## ***4 - Consommation***

Cette partie se concentre sur l'étude de la consommation d'électricité en France. On va essayer de prédire la consommation jusqu'en 2030 pour estimer les besoins, et voir aussi la consommation en fonction des différents secteurs (agriculture, industrie, tertiaire, résidentiel...).

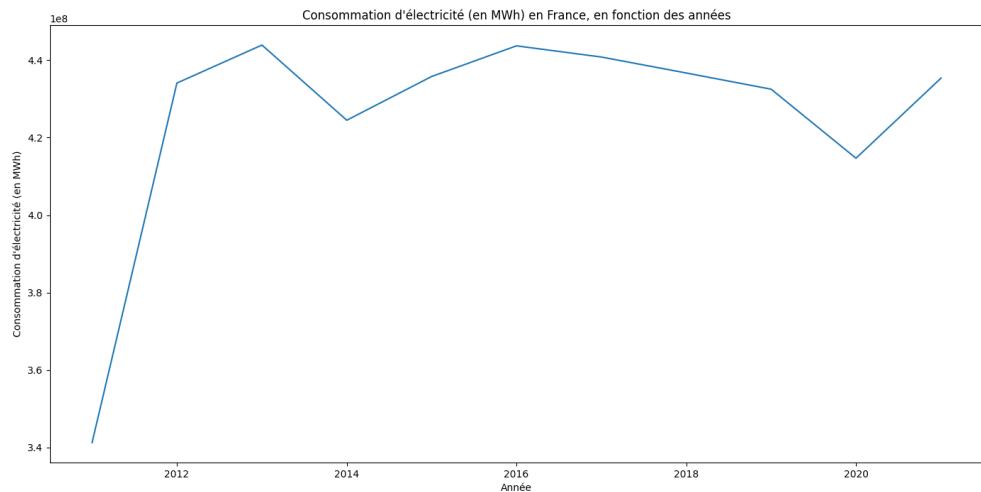
## 4.1 - Data Mining descriptif

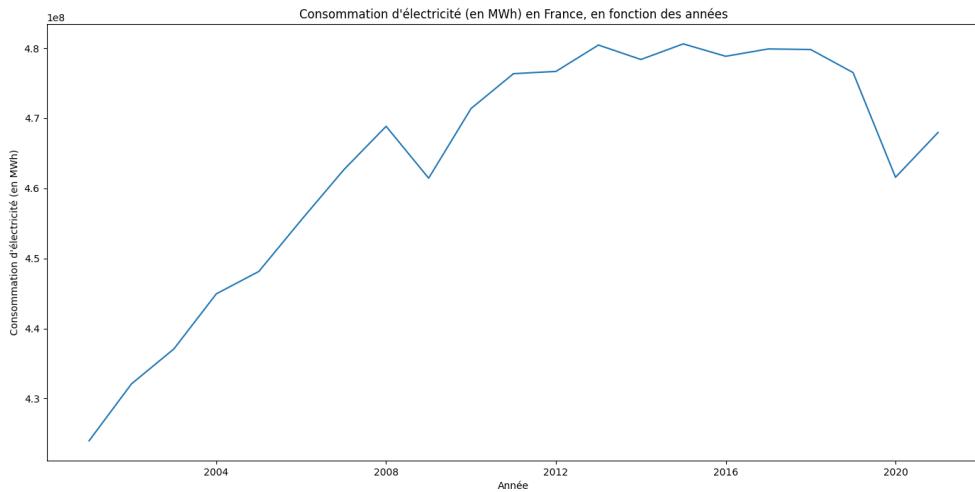
Nous avons utilisé deux jeux de données, les données de l'ORE (Opérateur de réseaux d'Énergie), [Consommation annuelle d'électricité et gaz par département et par secteur d'activité - data.gouv.fr](#) qui indiquent la consommation par secteurs, mais ne prend en compte que les années 2011-2021, et les données de RTE (Réseau de transport d'électricité) disponibles sur ORDÉ (Open Data Réseaux Énergies), [Consommation annuelle brute aux réseaux de transport](#) qui ne détaillent pas la consommation par secteur, mais possède les données des années 2001-2021.

Puisqu'on voulait observer et essayer de prédire la consommation en électricité, nous avons gardé uniquement les données concernant l'électricité, et supprimé les détails.

```
df = df[df["filiere"] == "Électricité"]
df = df.groupby("année").sum(numeric_only=True).reset_index()
# conversion de l'année
df["année"] = pandas.to_datetime(df["année"], format="%Y")
```

## 4.2 - Observation des données





On peut voir que quelque soit la source des données (ORE pour le premier graphe, RTE pour le second), la consommation d'électricité augmente.

Elle a diminué en 2018, 2019 et 2020 mais est repartie à la hausse ensuite, sans revenir au niveau précédemment atteint. Cela est dû à des hiver doux en 2018 et 2019, et à la crise sanitaire en 2020.

En 2021 elle atteint 467 TWh, alors qu'elle était à 483 TWh en 2017.

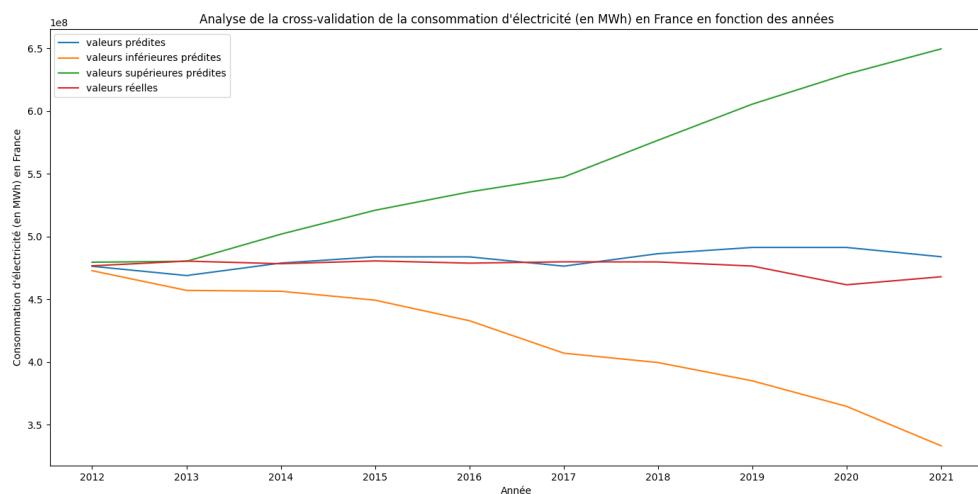
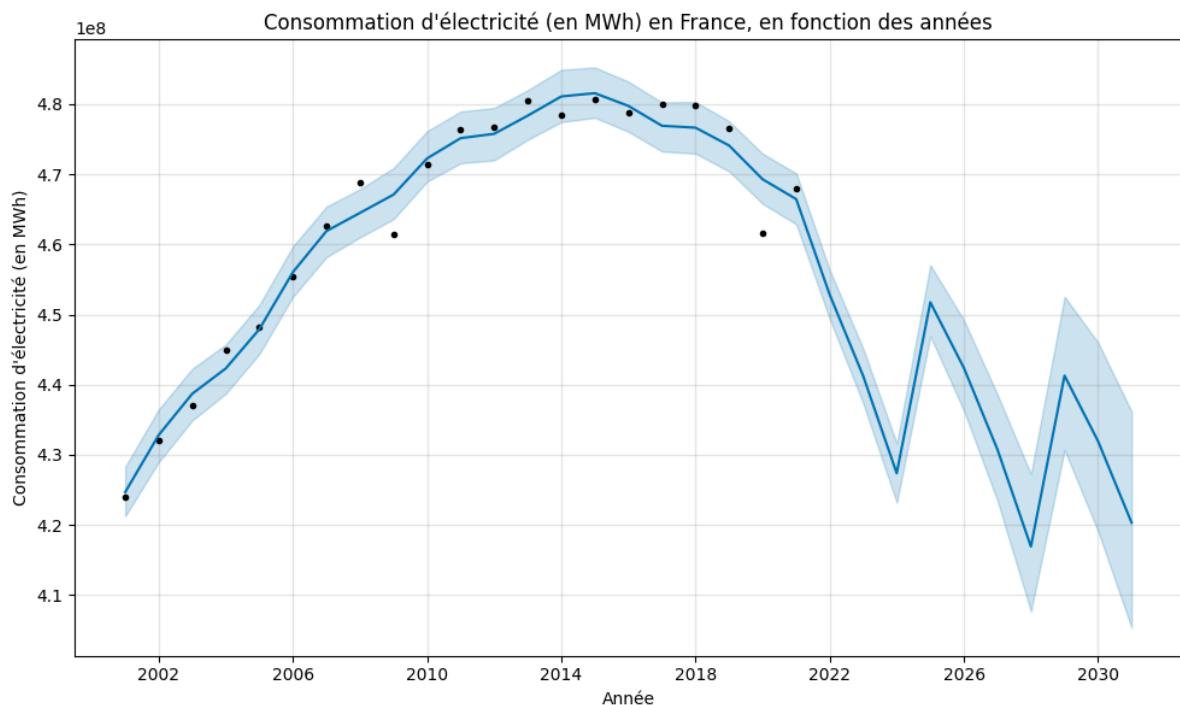
## 4.3 - Mise en œuvre des méthodes pour prédire la consommation d'électricité

Puisque les données étaient du même format que pour la production : année - consommation, nous avons utilisé les mêmes méthodes.

Pour chaque secteur, nous avons préparé les données :

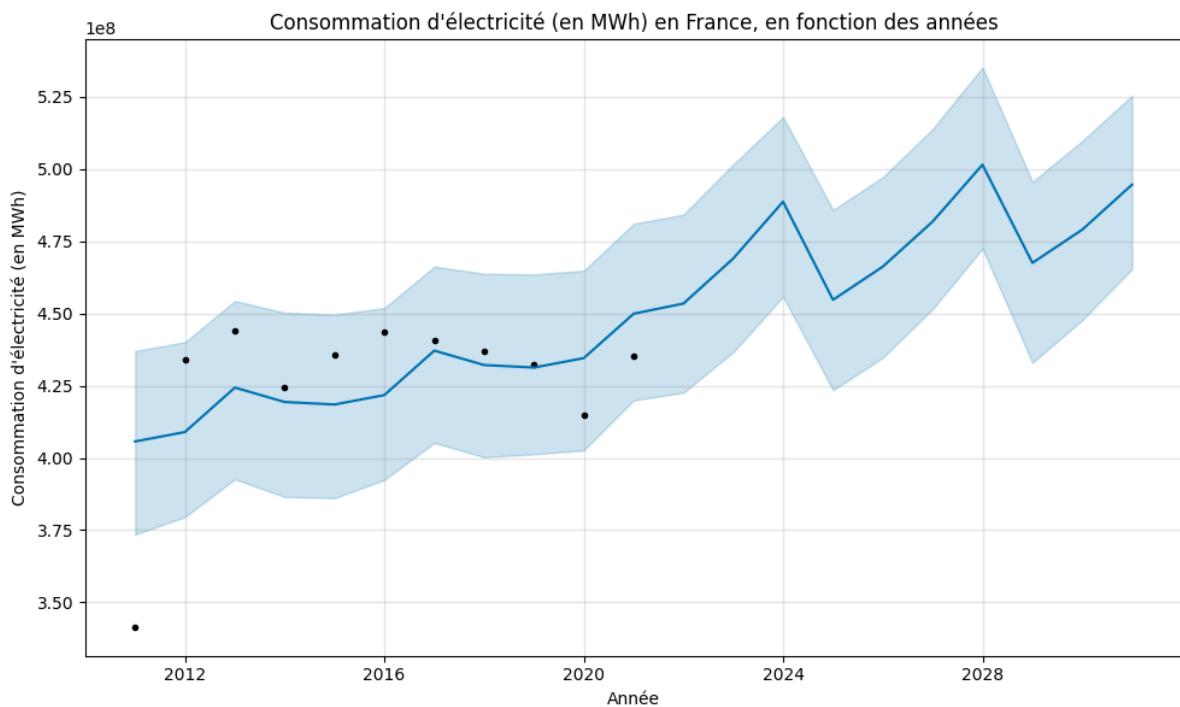
```
conso_agriculture = df[["annee", "conso_agriculture"]]
```

### 4.3.1 - Prédictions pour la consommation d'électricité (de manière globale)



Avec les données de RTE (2001-2021), la prédiction a été fortement influencée par les années 2018-2020 et la prédiction est à la baisse.

La cross validation indique que Prophet a prédit des données (courbe bleue) très similaires aux données réelles (courbe rouge) pour les années dont on possède les données donc le modèle est très fiable.

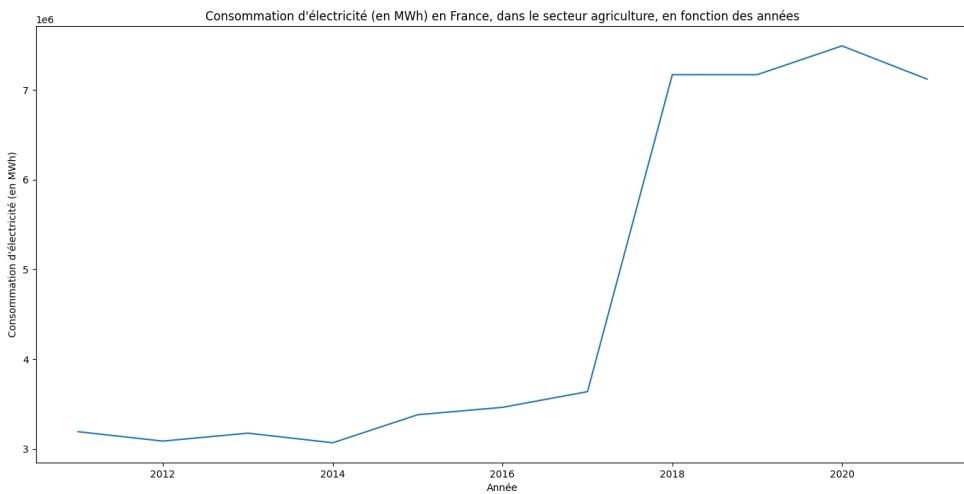


Avec les données de l'ORE (2011-2021), la prédiction est à la hausse, mais elle est incertaine à cause de la variabilité dans les données. On atteindrait environ 500 TWh de consommation vers 2030.

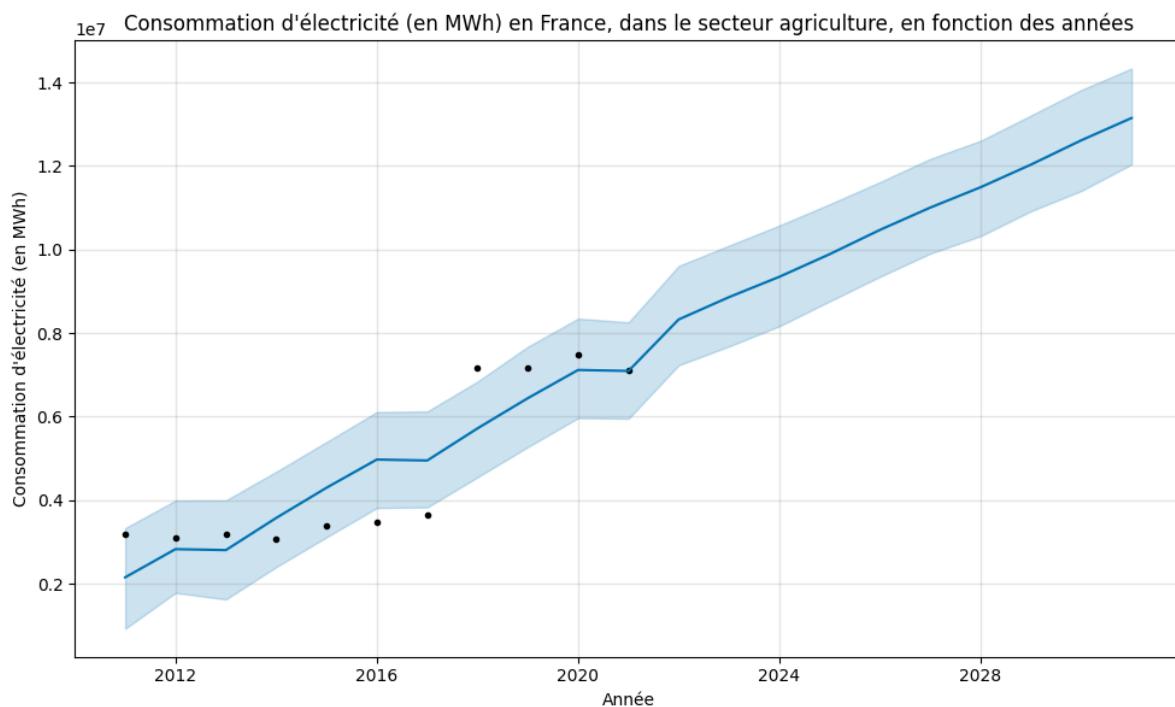
Le MAE est de 72 702 621.66429742 MWh. La consommation minimale est de 341 243 009.81700003 MWh et le maximum est de 443 911 428.697 MWh de 2011 à 2021. Le modèle est donc moyennement voire peu fiable sur ces données.

Par la suite, nous allons analyser la consommation par secteur. Pour cela nous utilisons les données de l'ORE (2011-2021).

### 4.3.2 - Prédictions pour la consommation d'électricité (secteur agriculture)



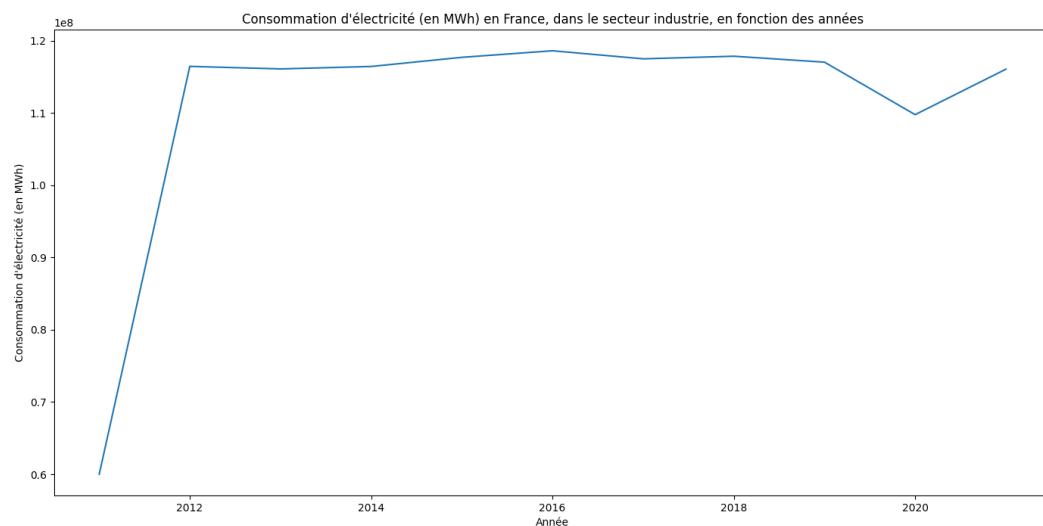
On remarque que la consommation dans le secteur de l'agriculture augmente grandement en 2018, passant d'environ 4 TWh à 8 TWh. Nous nous sommes demandé ce qu'il s'est passé cette année, et nous avons cherché d'autres données. [Le bilan énergétique de la France pour 2018](#) montre page 82 que la consommation d'électricité dans le secteur de l'agriculture est stable, à 8-9 TWh. Cela est confirmé dans la partie 5.7 du rapport, "Agriculture-pêche : une consommation stable depuis le début de la décennie" (page 103). On suppose donc une erreur dans les données.



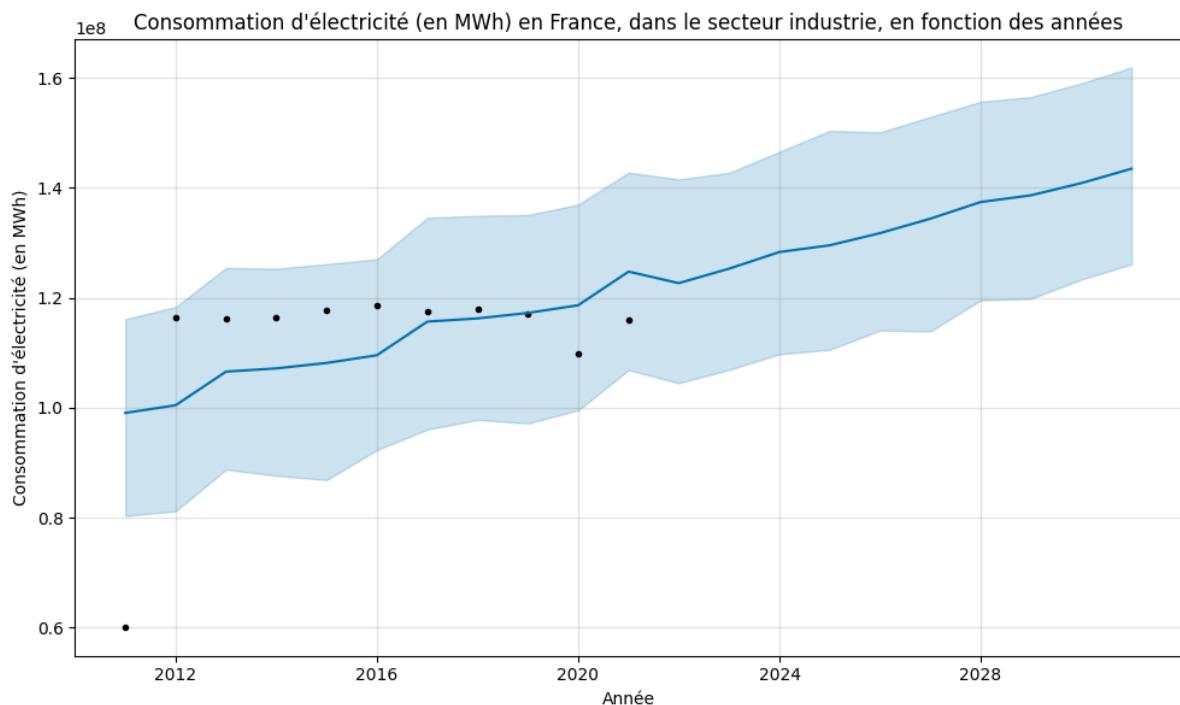
Cela nous donne une prédiction à la hausse, mais comme dit précédemment les données ne concordent pas avec d'autres, et la consommation dans l'agriculture devrait rester stable.

Le MAE est de 3 000 713.682962583 MWh. La consommation minimale pour le secteur agricole est de 3 067 822.83 MWh et le maximum est de 7 492 312.638364623 MWh de 2011 à 2021. Le modèle est donc moyennement adapté pour ce jeu de données.

### *4.3.3 - Prédictions pour la consommation d'électricité (secteur industrie)*



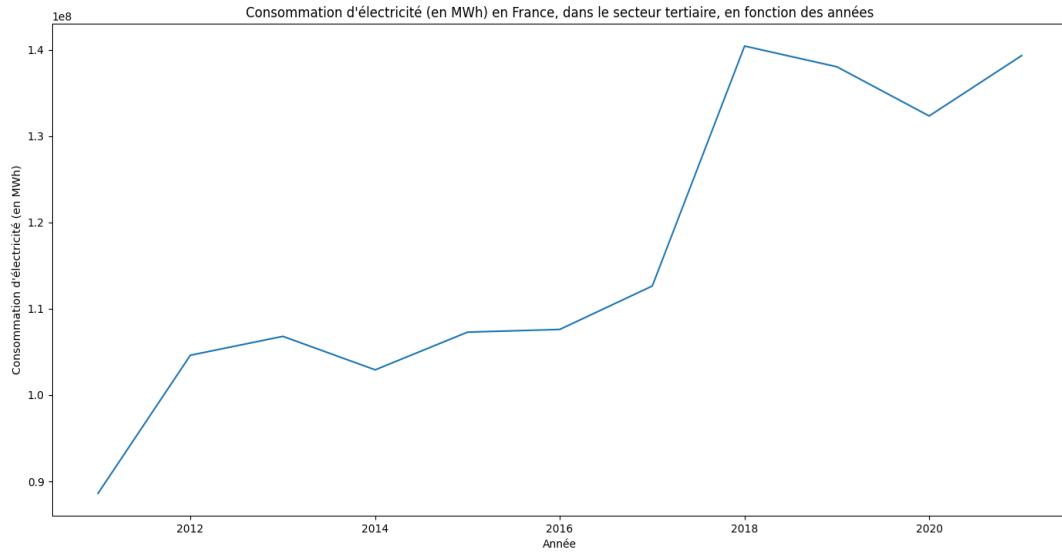
On remarque que la consommation dans le secteur de l'industrie dans les années 2011-2021 est très stable, environ 120 TWh, en légère hausse sauf les années 2020-2021 où on observe une chute de la consommation à cause du COVID.



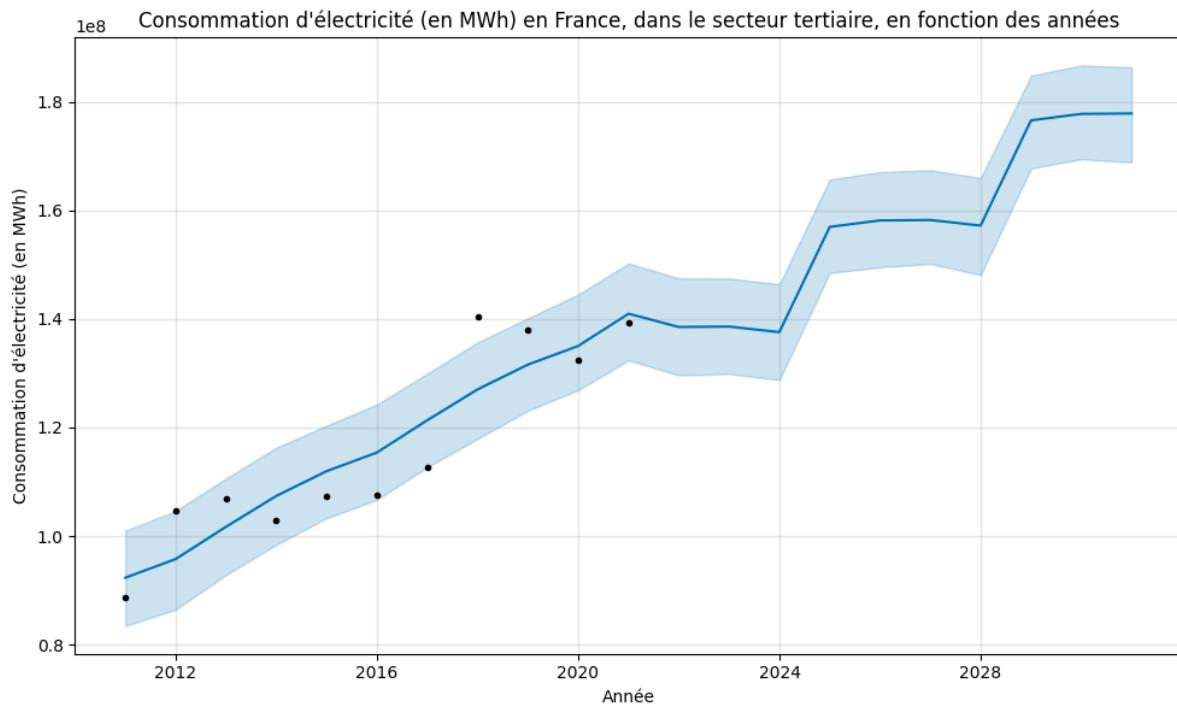
Les prédictions confirment nos observations précédentes. On devrait atteindre plus de 140 TWh de consommation dans le secteur de l'industrie en 2030.

Le MAE est de 41 701 319.80592166 MWh. La consommation minimale pour le secteur industriel est de 60 005 862.608 MWh et le maximum est de 118 603 403.26 MWh de 2011 à 2021. Le modèle est donc moyennement fiable sur ces données puisqu'on les prédit à 40 millions de MWh près ce qui est quand même important.

#### 4.3.4 - Prédictions pour la consommation d'électricité (secteur tertiaire)



La consommation d'électricité dans le secteur tertiaire avoisine les 110 TWh dans le début des années 2010, augmente pour atteindre les 140 TWh dans les années 2018 et 2021.

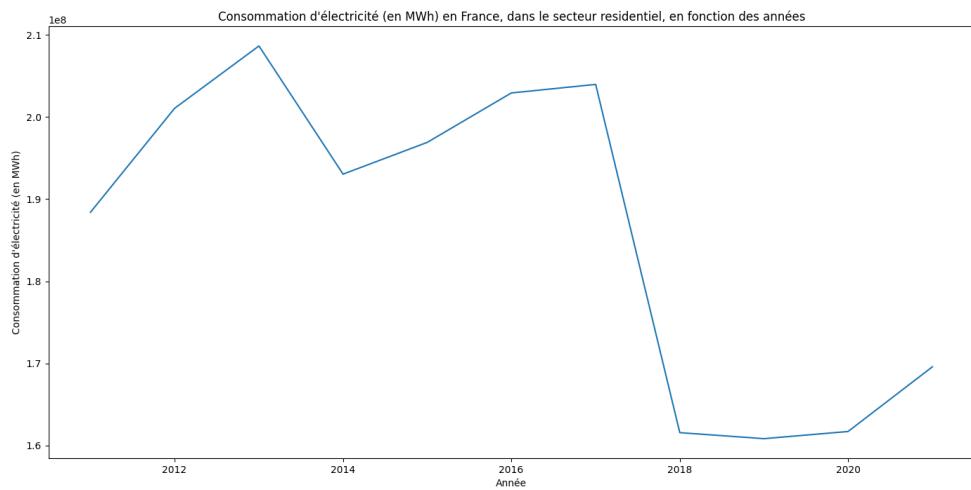


Notre modèle prédit aussi une hausse malgré une incertitude visible, et la consommation dans le secteur tertiaire devrait atteindre 180 TWh en 2030.

Le MAE est de 14 248 552.715957036 MWh. La consommation minimale pour le secteur tertiaire est de 88 612 405.391 MWh et le maximum est de 140 419 689.43022436

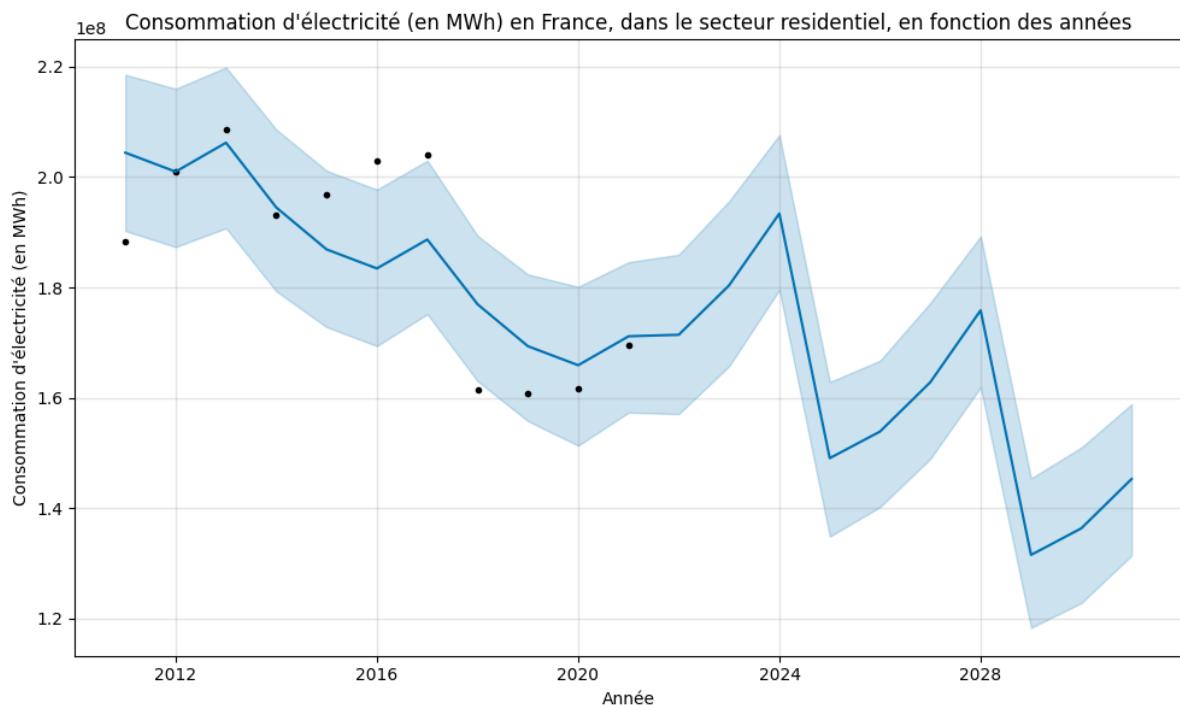
MWh de 2011 à 2021. Avec ce MAE, on peut conclure que le modèle est bon, les prédictions sont réalisées à 14 Millions d'erreur près sachant que les MWh varient de 88 millions à 140 millions environ.

#### *4.3.5 - Prédictions pour la consommation d'électricité (secteur résidentiel)*



Dans le secteur résidentiel, la consommation n'est pas stable, environ 200 TWh dans les années 2011-2017, avec une chute à 160 TWh en 2018.

Pourtant le rapport [Bilan énergétique de la France pour 2019](#) indique (page 81) une consommation stable : 157 TWh en 2015, 160 en 2019. Nous doutons de la qualité des données, et il est difficile de prédire avec des données comme cela.

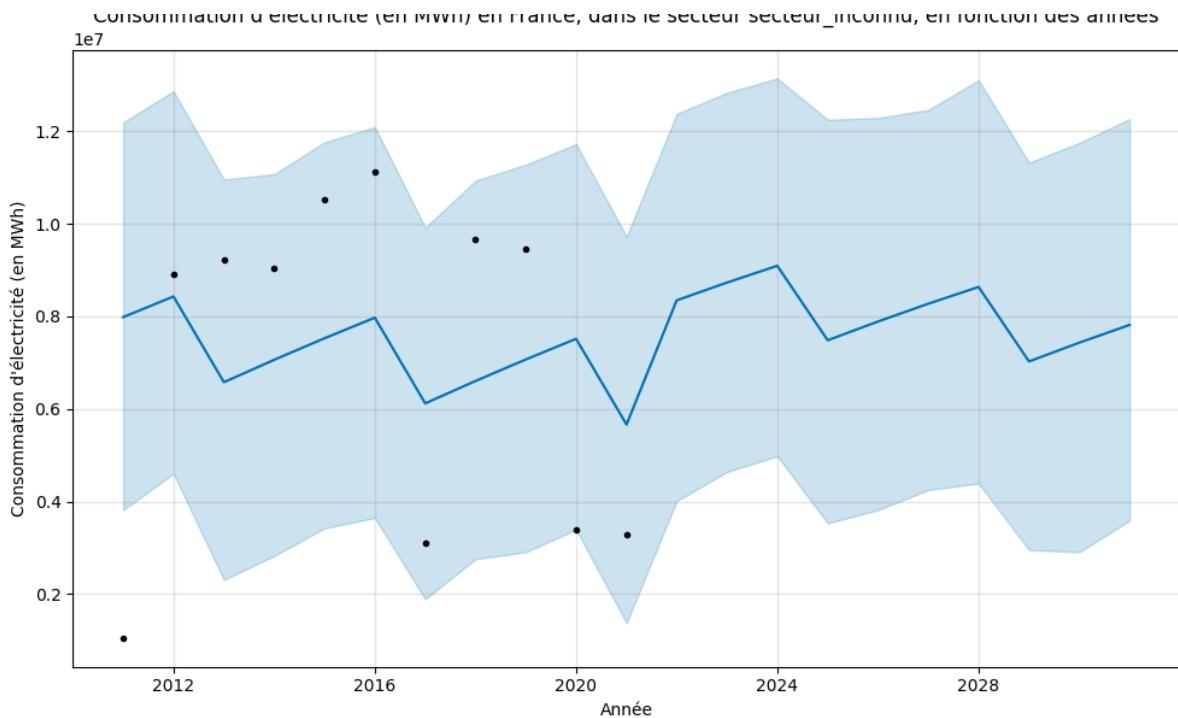
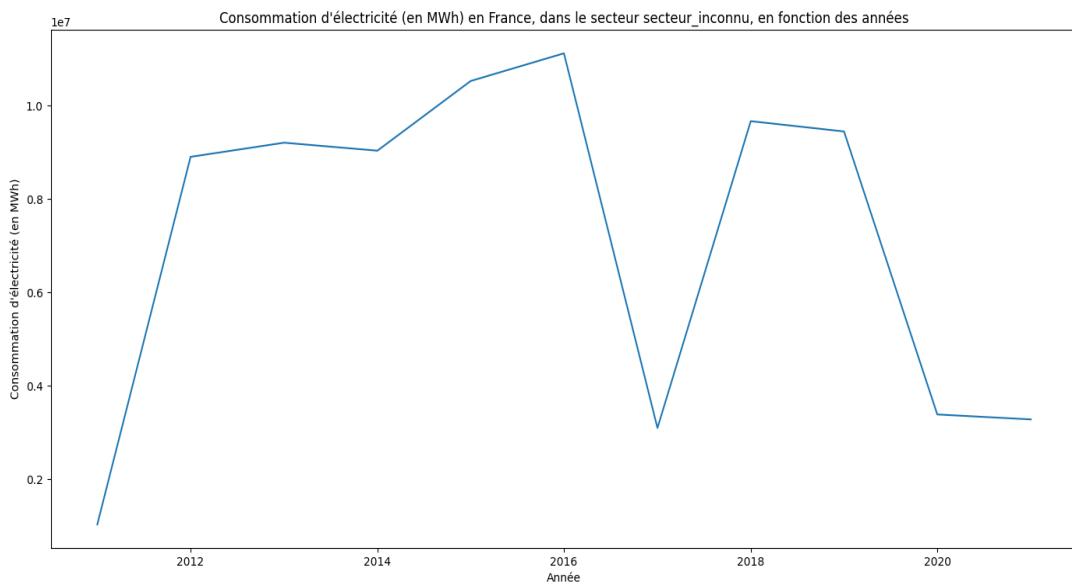


Avec ces données, cela nous prédit une baisse de la consommation dans le secteur résidentiel, avec environ 140 TWh en 2030.

Le MAE est de 35 607 572.08461699 MWh. La consommation minimale pour le secteur résidentiel est de 160 845 271.5029954 MWh et le maximum est de 208 638 268.112 MWh de 2011 à 2021. Avec ce MAE, on peut en déduire que le modèle est assez bon pour prédire les données, mais c'est à nuancer avec la remarque précédente sur la qualité des données.

#### 4.3.6 - *Prédictions pour la consommation d'électricité (secteur inconnu)*

ORE a classifié la consommation non affectée dans l'un des quatre autres secteurs dans le secteur inconnu. Les données sont très variables, entre 4 et 10 TWh selon les années.



La prédition est globalement stable, mais le MAE est de 11 422 443.226798484 MWh. La consommation minimale pour le secteur inconnu est de 1 034 921.75 MWh et le maximum est de 11 123 063.21 MWh de 2012 à 2021.

Avec ce MAE, on peut en déduire que le modèle de prédition n'est pas du tout fiable pour ces données, puisque les prévisions sont à 11 millions près alors que le maximum est de 11 millions de MWh, c'est énorme.

#### ***4.3.7 - Critique***

Les données par secteurs sont assez mauvaises. On voit des changements brusques dans les données notamment en 2018, alors que d'autres sources qui détaillent la consommation par secteur ne montrent pas de changements brusques.

De plus, il est difficile de prédire à cause des facteurs externes (comme l'épidémie de COVID-19) mais aussi d'autres facteurs comme l'électrification.

## **5 - Coût**

## 5.1 - Comparaison entre les différentes filières dont le nucléaire

Dans cette première partie, l'objectif était de prouver que le prix de production de l'électricité dépend de la filière. Cependant, nous ne disposions pas des informations suffisantes pour cela. Ainsi, nous avons recherchés ces informations sur internet et nous sommes tombés sur le document suivant [ADEME\\_étude\\_mix-electrique.pdf](#) ainsi que sur le site [suivant](#).

Nous avons alors croisé ces deux sources pour obtenir la liste suivante (pour 2021) :

- Nucléaire - 59.8 €/MWh à 109 €/MWh
- Éolien - 90 €/MWh à 200 €/MWh
- Hydraulique - 15 €/MWh à 20 €/MWh
- Photovoltaïque - 142 €/MWh
- Gaz - 70 €/MWh à 100 €/MWh
- Charbon - 40 €/MWh
- Fioul - 70 €/MWh à 100 €/MWh (par supposition grâce au gaz)

Comme nous pouvons le voir, ce sont évidemment des données quantitatives qui indiquent une fourchette du prix de production de l'électricité pour les principales filières.

Ainsi, avec ces données nous avons construit le fichier csv suivant par l'intermédiaire d'un programme en python :

```
range,filiere,prix_MWh
low,Nucléaire,59.8
high,Nucléaire,109.0
low,Éolien,90.0
high,Éolien,200.0
low,Hydraulique,15.0
high,Hydraulique,20.0
low,Photovoltaïque,142.0
high,Photovoltaïque,142.0
low,Gaz,70.0
high,Gaz,100.0
low,Charbon,40.0
high,Charbon,40.0
low,Fioul,70.0
high,Fioul,100.0
```

Deux études distinctes ont alors fait suite à cela.

La première d'entre-elles a été très manuelle puisque que nous avons tirés des conclusions de la simple observation suivante :

	count	mean	std	min	25%	50%	75%	max
<b>filiere</b>								
Charbon	2.0	40.0	0.000000	40.0	40.00	40.0	40.00	40.0
Fioul	2.0	85.0	21.213203	70.0	77.50	85.0	92.50	100.0
Gaz	2.0	85.0	21.213203	70.0	77.50	85.0	92.50	100.0
Hydraulique	2.0	17.5	3.535534	15.0	16.25	17.5	18.75	20.0
Nucléaire	2.0	84.4	34.789654	59.8	72.10	84.4	96.70	109.0
Photovoltaïque	2.0	142.0	0.000000	142.0	142.00	142.0	142.00	142.0
Éolien	2.0	145.0	77.781746	90.0	117.50	145.0	172.50	200.0

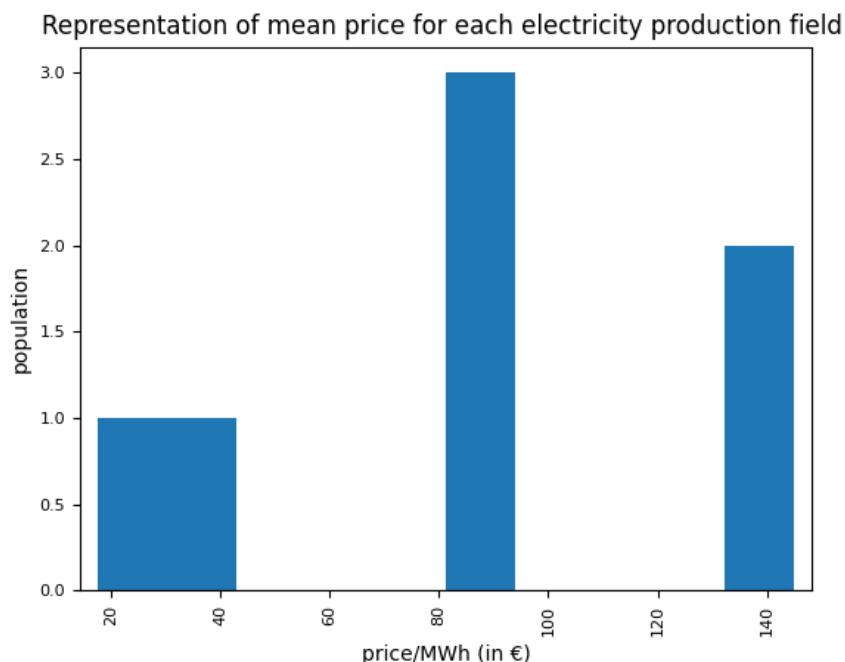
Nous avons alors à partir de ça construit le tableau ci-dessous :

	count	mean	std	min	25%	50%	75%	max
<b>filiere_group</b>								
Cout faible – Hydraulique, Charbon	4.0	28.75	13.149778	15.0	18.75	30.0	40.0	40.0
Cout moyen – Nucléaire, Gaz, Fioul	6.0	84.80	20.546533	59.8	70.00	85.0	100.0	109.0
Cout élevé – Eolien, Photovoltaïque	4.0	143.50	44.940702	90.0	129.00	142.0	156.5	200.0

Nous avons donc conclut de cela que le coût de la production d'électricité dépend de la filière et on peut rassembler :

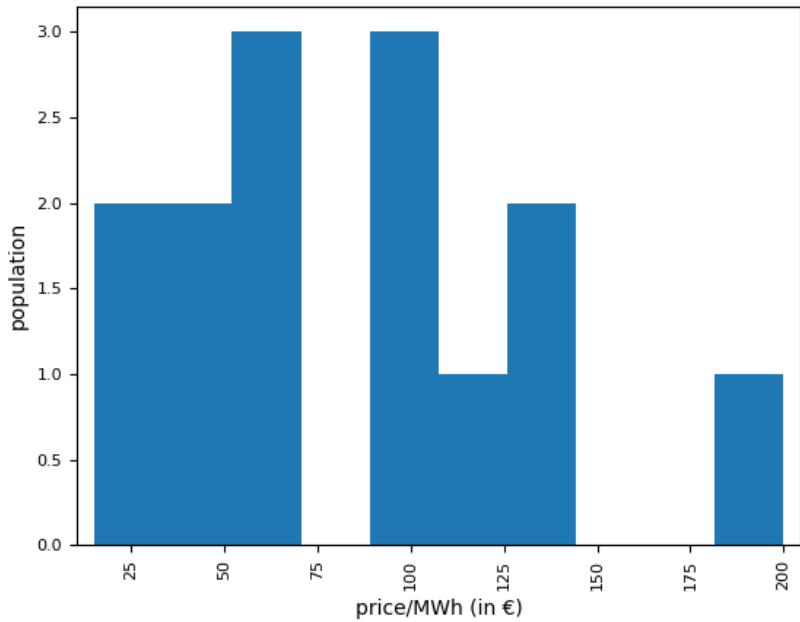
- L'hydraulique et le charbon avec un coût moyen de 28.75 €/MWh.
- Le nucléaire, le gaz et le fioul avec un coût moyen de 84.80 €/MWh.
- L'éolien et le photovoltaïque avec un coût moyen de 143.50 €/MWh.

Évidemment en raison de la méthode utilisée, cela n'était pas très pertinent. Ainsi, nous avons mis en place une méthode de clustering pour obtenir un résultat similaire. Mais avant toute chose, pour avoir une vision plus claire de nos données nous avons représenté sur le graphique suivant le prix moyen de chaque filière de production d'électricité :

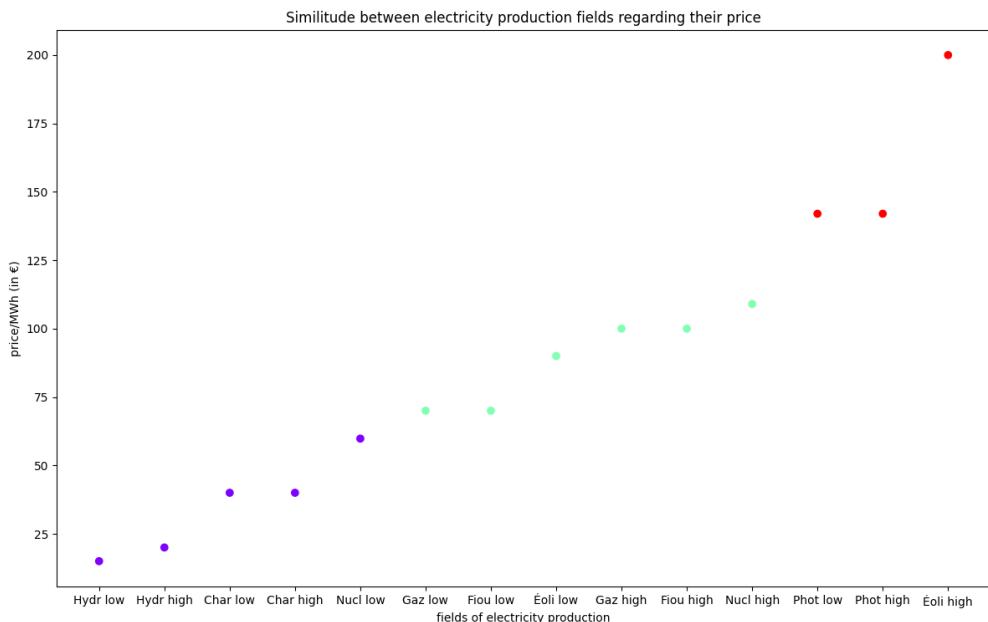


Cependant, cela ne représentant que la moyenne du prix de chaque filière, nous avons également souhaité représenter les extrêmes de ces mêmes filières sur le graphique ci-dessous :

Representation of extremum prices for each electricity production field



Ces deux représentations nous permettent alors de constater qu'il y a trois groupes de filières. Cela nous a donc renseigné sur le nombre de clusters souhaités avec donc la méthodes des K-means puisque cela se prête particulièrement à notre cas. Le résultat est alors le suivant :



Nous avons donc grâce à cela rassemblé les filières de chaque cluster pour construire le tableau de résultat suivant :

Nucl high, Éoli low, Gaz low, Gaz high, Fiou lo...	89.833333	2.0
Nucl low, Hydr low, Hydr high, Char low, Char high	34.960000	1.0
Éoli high, Phot low, Phot high	161.333333	0.0

À présent nous sommes certains de nos conclusions et nous pouvons donc affirmer que le prix de production de l'électricité dépend de la filière.

## 5.2 - Méthodologie de calcul adaptée aux économies d'échelles

Dans cette seconde partie, l'objectif était de mettre en place une méthode calcul des prix de production d'électricité de chaque filière réaliste avec des parts de marché différentes de celles officielles. En effet, cela servira de base dans une section de notre dashboard final. Pour cela, nous avons, en plus du jeu de données créé précédemment, utilisé ce [document](#) en le croisant avec [celui-ci](#).

Ainsi, nous avons commencé par croiser ces deux documents, le premier ne donnant pas assez de précision sur les énergies renouvelables. Cela a donné le résultat suivant :

```
filiere,part_de_production
Nucléaire,76.88
Fioul,0.8
Hydraulique,8.56
Charbon,0.81
Gaz,7.0
Autres Renouvelables,5.106433105848653
Eolien,0.7647861595767812
Solaire,0.07878073457456555
```

Évidemment, comme vous pouvez le constater, ce sont une nouvelle fois des données quantitatives.

À partir de cela, nous avons mis en place des coefficients pour calculer le nouveau coût de production en fonction de la part de chaque filière. En effet, avec l'évolution des parts, le prix de production est impacté en raison de nombreuses économies d'échelle.

Cela donne donc pour un exemple dans lequel on augmente la part de charbon les résultats suivant :

	filiere	prix_MWh	part_de_production	factor
0	Charbon	40.0	0.810000	0.907500
1	Fioul	85.0	0.800000	1.000000
2	Gaz	85.0	7.000000	1.000000
3	Hydraulique	17.5	8.560000	0.999535
4	Nucléaire	84.4	76.880000	1.015090
5	Solaire	142.0	0.078781	1.001001
6	Eolien	145.0	0.764786	1.009255

Une fois réalisé, cela permet de calculer les coûts ci-dessous avec la nouvelle répartition :

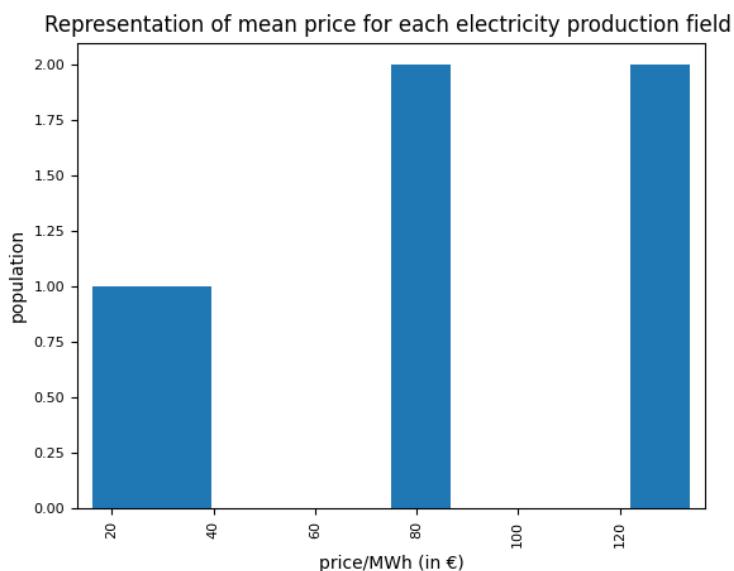
```
Charbon : 36.3 €/MWh
Fioul : 85.0 €/MWh
Gaz : 85.0 €/MWh
Hydraulique : 17.49186046511628 €/MWh
Nucléaire : 85.67358083832336 €/MWh
Solaire : 142.1421337302414 €/MWh
Eolien : 146.34199901980475 €/MWh
The global electricity price is 75.51836617556104 €/MWh
```

Nous avons donc à présent une méthode de calcul réaliste basée sur les données de 2021.

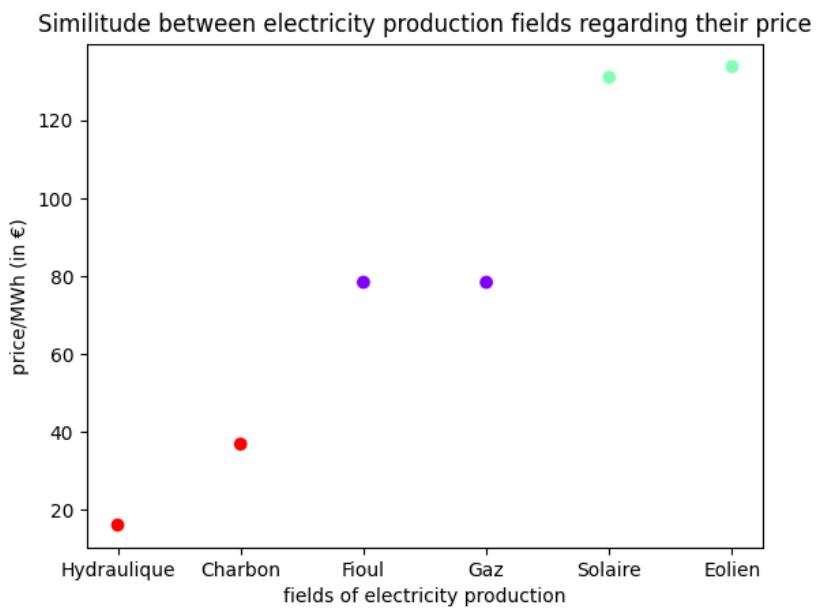
## 5.3 - Comparaison entre les différentes filières sans le nucléaire

Dans cette troisième partie, l'objectif était d'étudier de nouveau la dépendance entre le prix de production de l'électricité et la filière de production, mais sans le nucléaire. Pour cela, nous sommes donc parti des données créées au début de la première partie et nous les avons croisées avec les prix suivant que nous avons calculés avec les éléments de la partie précédente.

Ainsi, nous avons mis en place une méthode de clustering. Mais avant toute chose, pour avoir une vision plus claire de nos données nous avons représenté sur le graphique suivant le prix moyen de chaque filière de production d'électricité :



Le nombre de groupes étant de trois sur cette représentation, nous allons choisir le même nombre de clusters. Le résultat est alors le suivant avec la méthode des K-means :



Nous avons donc grâce à cela rassemblé les filières de chaque cluster pour construire le tableau de résultat suivant :

Charbon, Hydraulique	26.53970	2.0
Fioul, Gaz	78.46520	0.0
Solaire, Eolien	132.46772	1.0

À présent nous avons une représentation de la dépendance du prix des différentes filières de production d'électricité sans le nucléaire.

## 5.4 - Comparaison des coûts de production de l'électricité avec et sans le nucléaire

Dans cette dernière partie, l'objectif était de comparer le coût de la production d'électricité avec et sans le nucléaire.

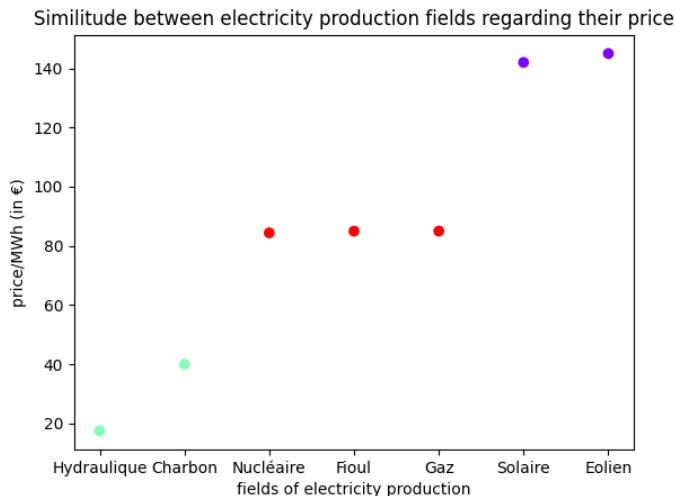
Pour cela, nous avons tout d'abord simuler la répartition de la production des différentes filières ainsi que les coûts associés en utilisant des éléments de la seconde partie. Il est d'ailleurs important de préciser que pour calculer la nouvelle répartition, nous nous sommes exclusivement basés sur celles actuelles en augmentant proportionnellement à cela la part de chaque filière. En effet, cela constitue un scénario de base que nous trouvons intéressant d'utiliser ici. cela nous a donc permis de créer le tableau suivant :

	filiere	prix_MWh	part_de_production	factor
0	Charbon	36.759317	4.266994	0.918983
1	Fioul	78.113548	4.214315	0.918983
2	Gaz	78.113548	36.875260	0.918983
3	Hydraulique	16.082201	45.093175	0.918983
5	Solaire	131.259708	0.323347	0.924364
6	Eolien	133.252523	4.028813	0.918983

A partir de ces données et des données de production réelles, nous avons alors calculé les coûts suivants :

Price with nuclear : 74.55952857448221  
 Price without nuclear : 46.709937572266796

Comme nous pouvons le constater, en prenant en compte les parts de chaque filière dans le calcul du coût, celui-ci est inférieur au coût actuel avec l'utilisation du nucléaire. Il est alors nécessaire de s'intéresser à la raison de cela en effectuant une comparaison plus en détail. Ainsi, pour permettre cela nous allons effectuer un clustering des données réelles étant donné que nous ne disposons que d'un clustering correspondant aux extrêmes des coûts des filières et pas directement de celui de la moyenne des coûts de chaque filière comme pour les données simulées. Évidemment, nous allons une nouvelle fois utiliser la méthode des K-means avec trois groupes pour que les analyses soient comparables. Nous obtenons alors le résultat ci-dessous :



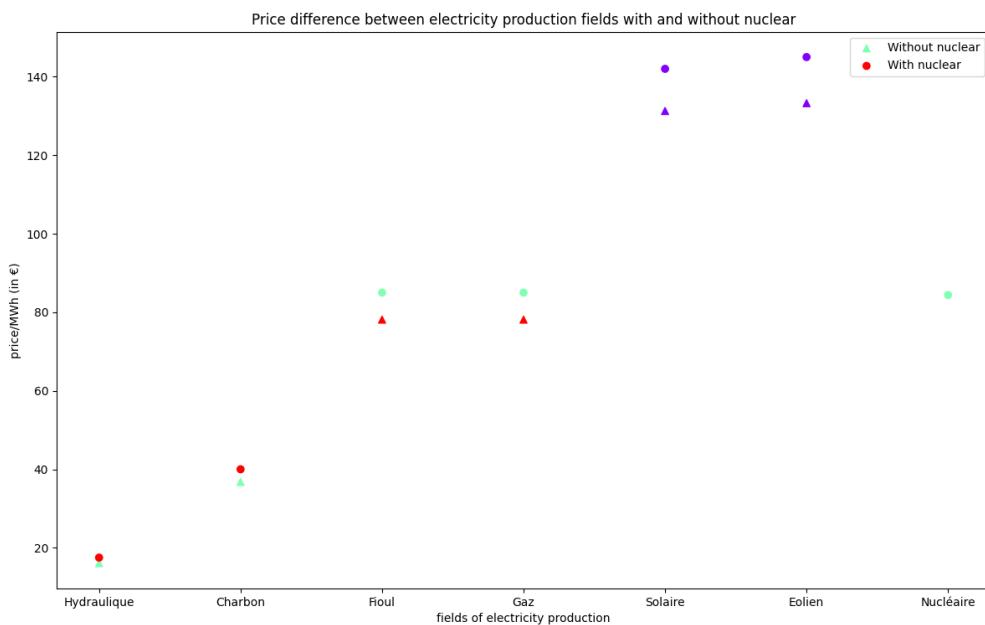
De plus, les données que nous allons utiliser pour la comparaison sont les suivantes :

	filiere	prix_MWh	cluster	filiere_group
4	Hydraulique	17.5	2	Charbon, Hydraulique
0	Charbon	40.0	2	Charbon, Hydraulique
5	Nucléaire	84.4	1	Fioul, Gaz, Nucléaire
2	Fioul	85.0	1	Fioul, Gaz, Nucléaire
3	Gaz	85.0	1	Fioul, Gaz, Nucléaire
6	Solaire	142.0	0	Eolien, Solaire
1	Eolien	145.0	0	Eolien, Solaire

Précisons d'ailleurs que les données issus de la simulation sont les suivantes :

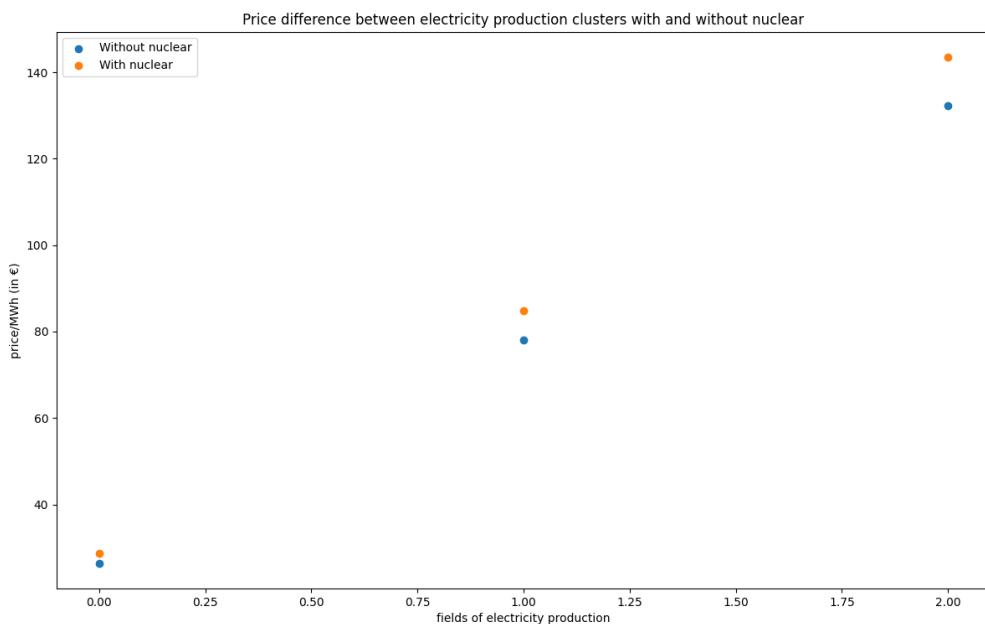
	prix_MWh	filiere	cluster	filiere_group
3	16.082201	Hydraulique	1	Charbon, Hydraulique
0	36.759317	Charbon	1	Charbon, Hydraulique
1	78.113548	Fioul	2	Fioul, Gaz
2	78.113548	Gaz	2	Fioul, Gaz
4	131.259708	Solaire	0	Solaire, Eolien
5	133.252523	Eolien	0	Solaire, Eolien

Ainsi, en comparant ces deux jeu de données, nous avons d'abord obtenu la différence entre les filières de production d'électricité avec et sans le nucléaire ci-dessous :

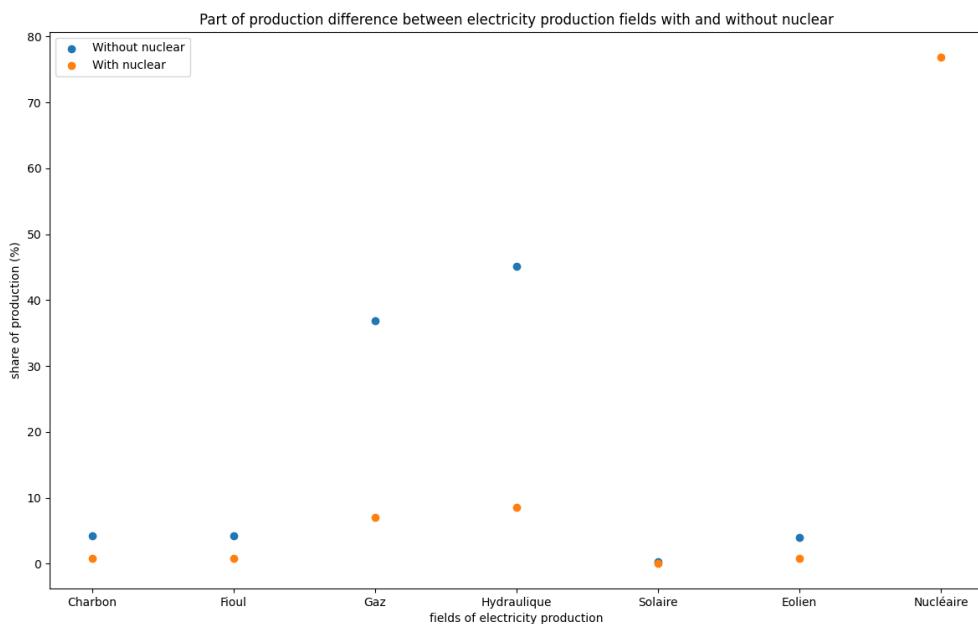


Nous remarquons alors qu'en l'absence du nucléaire, les coûts de production de toutes les filières ont diminué. En effet, cela est le résultat du calcul du facteur d'échelle que nous avons mis en place pour augmenter la pertinence de nos prévisions puisque lorsque la part d'une filière augmente, de nombreuses économies peuvent être réalisées.

De plus, en regroupant cette affichage par cluster, nous retrouvons les mêmes conclusions que précédemment sur le graphique suivant :



Maintenant que avons étudié précisément la différence de prix, il est temps de conclure en croissant cela avec les parts de chaque filière avec et sans le nucléaire du graphique ci-dessous :



Comme nous pouvons le constater, la production est globalement répartie entre l'hydraulique et le gaz, le premier ayant un coût bien moins élevé que le nucléaire et le second ayant un coût comparable. C'est alors cela qui explique le coût global inférieur sans le nucléaire.

## ***6 - Pollution***

## 6.1 - Comparaison entre les différentes filières dont le nucléaire

Dans cette première partie, l'objectif sera de prouver que le taux de pollution (émission de CO<sub>2</sub>) dépend de la filière.

On débute par identifier la production par filière qui sont les suivants (en 2021) :

- Nucléaire – 76.88 %
- Gaz – 7.0 %
- Hydraulique – 8.56 %
- Autres Renouvelables – 5.95 %
- Charbon – 0.81 %
- Fioul – 0.8 %
- Émission de CO<sub>2</sub> – 45.15 g/kWh fourni
- Déchets radioactifs – 2.07 mg/kWh fourni

Ensuite on construit le dataset avec python, ce qui donne le résultat suivant :

annee	filiere	valeur	unite
2021	Nucléaire	76.88	%
2021	CO2	45.15	g/kWh fourni
2021	Hydraulique	8.56	%
2021	Gaz	7.00	%
2021	Autres Renouvelables	5.95	%
2021	Déchets radioactifs	2.07	mg/kWh fourni
2021	Charbon	0.81	%
2021	Fioul	0.80	%

Comme on peut le voir pour le moment, le nucléaire est l'énergie la plus utilisée, elle est 8,98 fois plus grosse que la deuxième (Hydraulique) et 10,98 fois plus grosse que la troisième (Gaz).

Nous voyons alors le pourcentage de participation des sources d'électricité en France en 2021. Ce qui va nous permettre après de calculer leur participation dans la production de pollution.

Maintenant, il s'agit de voir si les déchets lâchés par ces derniers sont proportionnels à la taille de leur production.

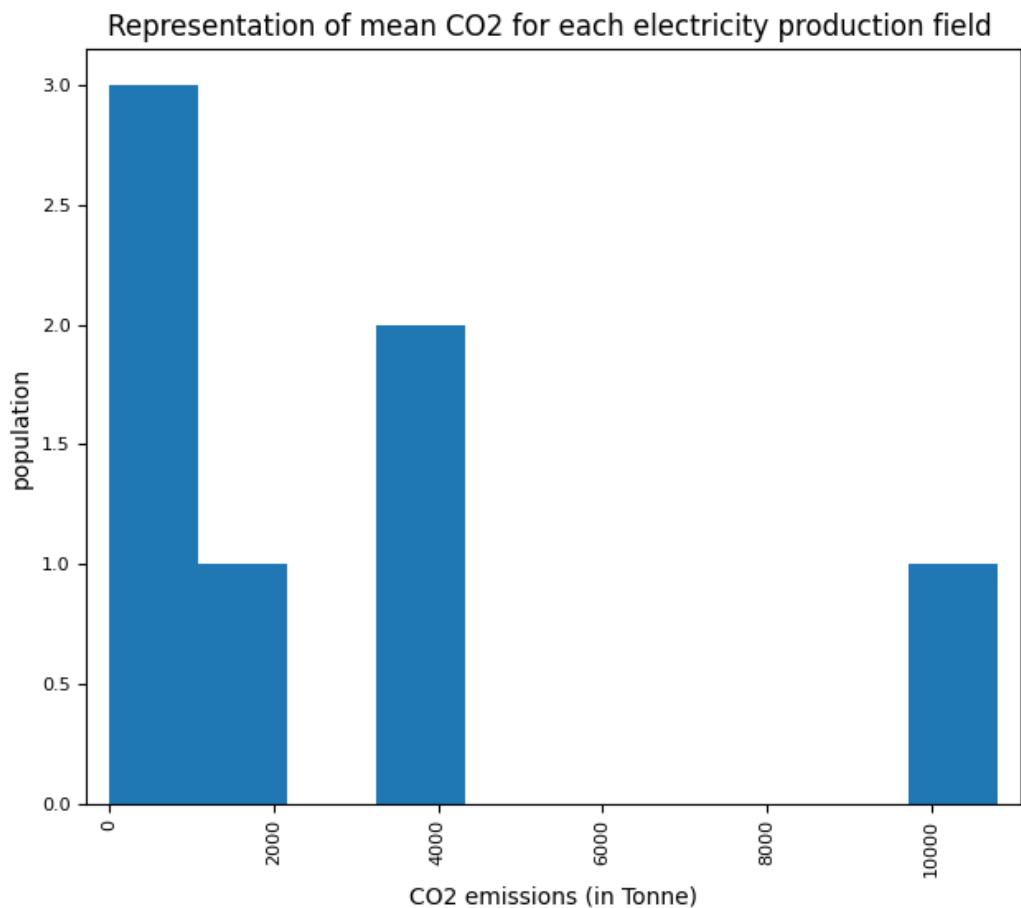
Nous avons récupéré les informations sur l'émission de CO<sub>2</sub> par filière sur [ce site](#).

Puis nous avons par une simple multiplication obtenu la production de CO<sub>2</sub> en tonnes par filière, à cela, on a ajouté leur part en pourcent de la production d'électricité et de CO<sub>2</sub> globale.

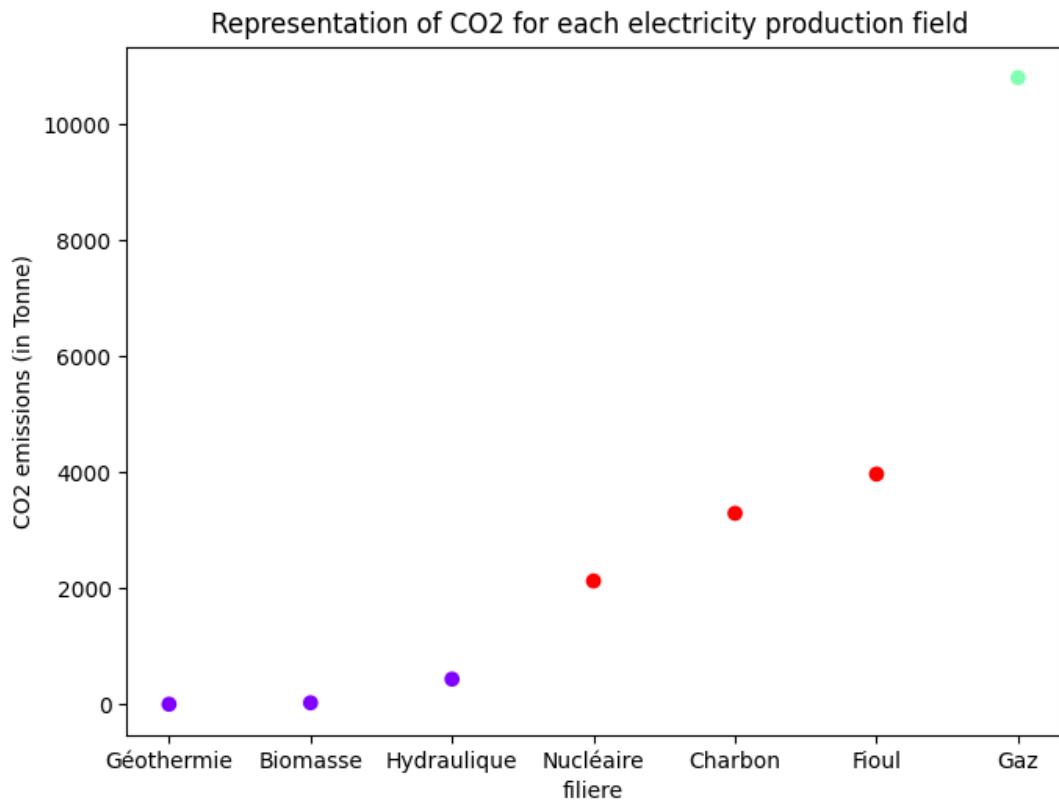
filiere	CO2 emissions [gCO2/kWh]	production_MWh	CO2 emissions [tCO2]	percentage_energy_production	percentage_CO2_emissions
Biomasse	32.0	0.767765	24.568484	0.177951	0.118930
Charbon	1058.0	3.113277	3293.847144	0.721588	15.944710
Fioul	730.0	5.439878	3971.110935	1.260841	19.223178
Gaz	443.0	24.398621	10808.589271	5.655051	52.321741
Géothermie	38.0	0.003746	0.142345	0.000868	0.000689
Hydraulique	10.0	43.330790	433.307899	10.043102	2.097538
Nucléaire	6.0	354.394205	2126.365230	82.140599	10.293215

Les chiffres (pourcentages) peuvent varier un peu de ce qu'on a eu en début de partie car certaines énergies (Autres Renouvelables qui comprennent l'éolien) ne sont plus prises en compte (par manque de données). Ces données représentent seulement la production et non pas la création d'infrastructures.

Nous avons ensuite utilisé ces informations pour tracer le graphe afin de visualiser leurs émissions de CO2 (en tonnes) par rapport à leur filière.



Ce qui nous permet de déjà remarquer qu'il y a trois groupes de filières qui se forment. Cela nous a donc renseigné sur le nombre de clusters souhaités avec donc la méthodes des K-means puisque cela se prête particulièrement à notre cas. Le résultat est alors le suivant :



Nous avons donc grâce à cela rassemblé les filières de chaque cluster pour construire le tableau de résultat suivant :

filiere_group	CO2 emissions (in Tonne)	cluster
Biomasse, Géothermie, Hydraulique	152.672910	0.0
Charbon, Fioul, Nucléaire	3130.441103	2.0
Gaz	10808.589271	1.0

À présent nous sommes certains de nos conclusions et nous pouvons donc affirmer que le prix de production de CO2 dépend de la filière.

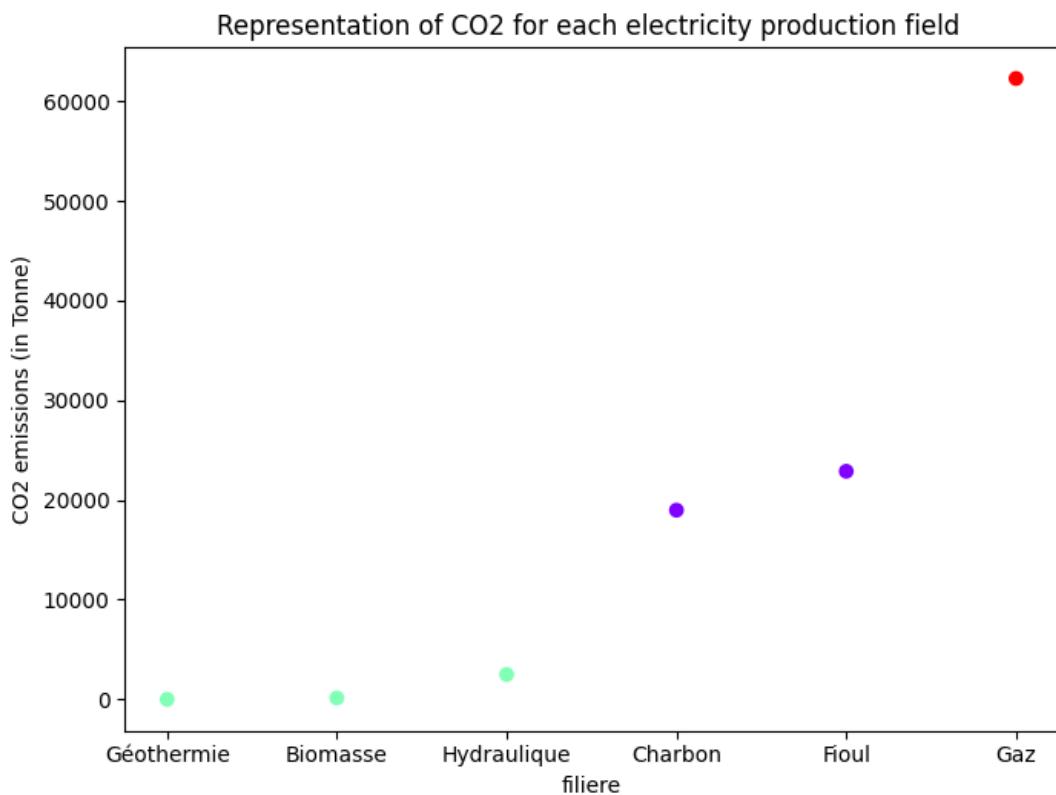
## 6.2 - Comparaison entre les différentes filières sans le nucléaire

Dans cette partie, l'objectif était d'étudier de nouveau la dépendance entre la production de CO<sub>2</sub> et la filière de production, mais sans le nucléaire. Pour ce faire, nous avons repris les données de la première partie, que nous avons transformées afin de trouver leur production d'énergie sans le nucléaire ainsi que leur nouvelle production de CO<sub>2</sub>.

	filiere	part_production	CO2 emissions [gCO2/kWh]	production_kWh	CO2 emissions [tCO2]
0	Biomasse	0.996398	32.0	4.424796e+06	141.593481
1	Charbon	4.040379	1058.0	1.794249e+07	18983.152353
2	Fioul	7.059818	730.0	3.135119e+07	22886.369829
3	Gaz	31.664283	443.0	1.406145e+08	62292.233940
4	Géothermie	0.004861	38.0	2.158858e+04	0.820366
5	Hydraulique	56.234260	10.0	2.497247e+08	2497.246990

Ainsi, nous avons mis en place une nouvelle fois une méthode de clustering. Mais avant toute chose, pour avoir une vision plus claire de nos données nous avons représenté sur le graphique suivant le prix moyen de chaque filière de production d'électricité :

Le nombre de groupes étant de trois sur cette représentation, nous allons choisir le même nombre de clusters. Le résultat est alors le suivant avec la méthode des K-means :



Nous avons donc grâce à cela rassemblé les filières de chaque cluster pour construire le tableau de résultat suivant :

	CO2 emissions (in Tonne)	cluster
filiere_group		
Biomasse, Géothermie, Hydraulique	152.672910	2.0
Charbon, Fioul, Nucléaire	3130.441103	0.0
Gaz	10808.589271	1.0

À présent nous avons une représentation de la dépendance de la production de CO2 des différentes filières de production d'électricité sans le nucléaire.

## 6.3 - Comparaison de la pollution de production de l'électricité avec et sans le nucléaire

Dans cette partie finale, l'objectif est de comparer la pollution de la production d'électricité sans le nucléaire.

Pour cela, nous avons tout d'abord simuler la répartition de la production des différentes filières ainsi que la pollution associée en utilisant des éléments de la seconde partie. Il est d'ailleurs important de préciser que pour calculer la nouvelle répartition, nous nous sommes exclusivement basés sur celles actuelles (2021) en augmentant proportionnellement à cela la part de chaque filière. En effet, cela constitue un scénario de base que nous trouvons intéressant d'utiliser ici. cela nous a donc permis de créer le tableau suivant :

	filière	part_production	CO2 emissions [gCO2/kWh]	production_kWh	CO2 emissions [tCO2]
0	Biomasse	0.996398	32.0	4.424796e+06	141.593481
1	Charbon	4.040379	1058.0	1.794249e+07	18983.152353
2	Fioul	7.059818	730.0	3.135119e+07	22886.369829
3	Gaz	31.664283	443.0	1.406145e+08	62292.233940
4	Géothermie	0.004861	38.0	2.158858e+04	0.820366
5	Hydraulique	56.234260	10.0	2.497247e+08	2497.246990

À partir de ces données et des données de production réelles, nous avons alors calculé les productions suivante :

- Avec le nucléaire :

	filière	part_production	CO2 emissions (in Tonne)
1	Biomasse	0.177951	24.568484
4	Charbon	0.721588	3293.847144
5	Fioul	1.260841	3971.110935
6	Gaz	5.655051	10808.589271
0	Géothermie	0.000868	0.142345
2	Hydraulique	10.043102	433.307899
3	Nucléaire	82.140599	2126.365230
20657.931309687337 Tonne de CO2 produite par an			

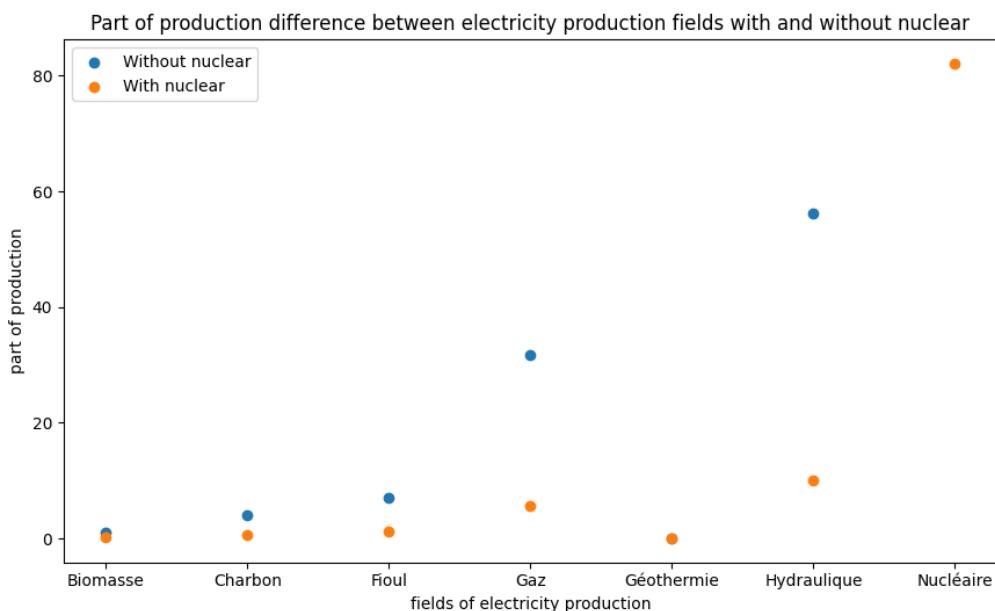
- Sans le nucléaire :

	filiere	part_production	C02 emissions (in Tonne)
0	Biomasse	0.996398	141.593481
1	Charbon	4.040379	18983.152353
2	Fioul	7.059818	22886.369829
3	Gaz	31.664283	62292.233940
4	Géothermie	0.004861	0.820366
5	Hydraulique	56.234260	2497.246990
		106801.4169598105	Tonne de C02 produite par an

On obtient alors :

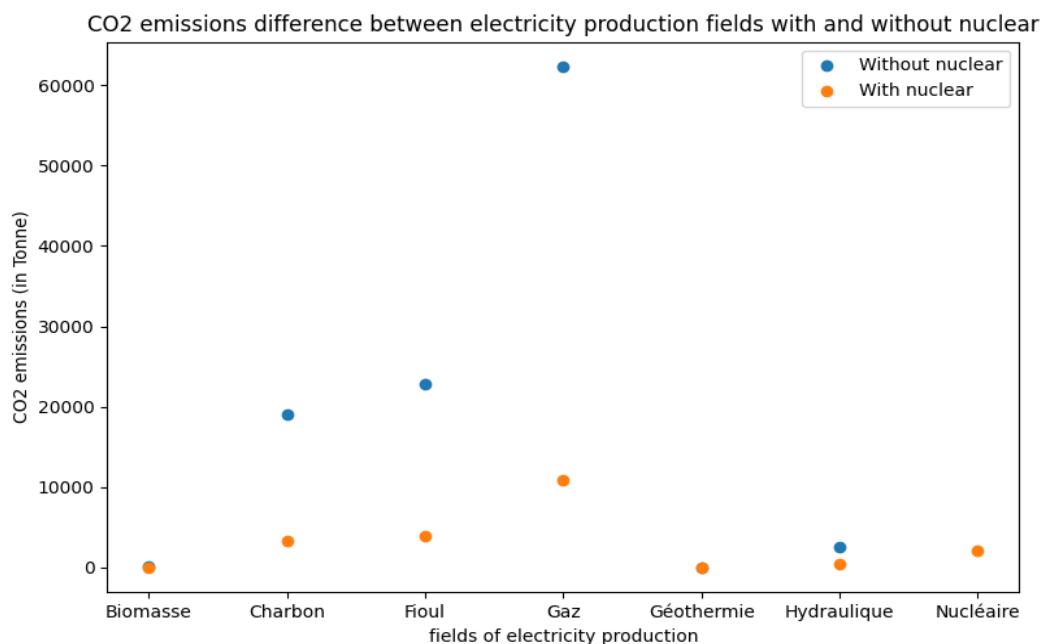
```
Différence entre la production de C02 (Avec - Sans) : -86143.48565012317 Tonne de C02
```

Maintenant si on compare les deux jeux de données, nous obtenons une différence de part tel que :



Comme nous pouvons le constater, la production est globalement répartie entre l'hydraulique et le gaz, le premier ayant une production de CO2 à peine plus élevée que le nucléaire et le second ayant une production 73,8 fois plus grande. C'est alors cela qui explique la différence de 86 143 Tonnes de CO2 en plus.

Nous pouvons aussi voir la différence d'émissions de CO2 plus en détails :



Ainsi, nous voyons la différence d'émission entre les filières avec et sans nucléaire. Nous voyons bien les parts Hydraulique et Gaz qui ont augmenté comme vu plus tôt. Nous voyons aussi l'émission du Gaz et du Fioul qui ont plus que doublé aussi.

Il faut cependant ajouter que cette étude consiste juste de l'émission du CO<sub>2</sub>, or le nucléaire produit aussi des déchets radioactifs et comme c'est le seul. Il faudrait voir lequel peut être considéré comme le pire.

La différence de taille avec des déchets industriels classiques, les déchets issus des centrales sont radioactifs, donc non recyclables. Leur radioactivité peut s'estomper relativement rapidement (plusieurs semaines) ou durer plusieurs millions d'années. La majorité des déchets, dits à « vie courte », voient leur radioactivité disparaître au bout d'une centaine de jours. Les déchets, dits à « vie longue », environ 10 % des déchets existants, demeurent radioactifs pendant au moins trois décennies... jusqu'à plusieurs millions d'années. Cependant avec le temps, si on continue à utiliser le nucléaire, l'humain pourrait trouver une solution pour ces déchets à vie longue et ainsi réduire la pollution engendrée.

## **7 - Conclusion**

## 7.1 - Résumé

### 7.1.1 - Production

Au niveau de la production d'électricité, nous avons comparé les différentes capacités de production des différentes filières en fonction des années.

Il nous est apparu qu'en France, la filière produisant le plus d'électricité est le nucléaire. En effet, celle-ci produit en moyenne 386 025 806.451 MWh sachant que la somme des productions moyennes de toutes les filières sans le nucléaire vaut 119 871 416 MWh.

Ensuite, nous avons réalisé des prédictions de production en utilisant la librairie Prophet pour produire des régressions linéaires en fonction des filières.

À travers ces prédictions, nous avons obtenu les mêmes résultats que précédemment à savoir que le nucléaire domine toujours le marché de la production d'électricité avec une moyenne de 369 208 105.815 MWh comparé aux autres qui en étant réunis produisent 137 587 250 MWh en moyenne.

De plus, nous avons aussi calculé les parts de production de chaque filière. En 2020, le nucléaire possédait 68.6 % de part de production et en 2030, le nucléaire devrait posséder 61.2 % de part de production.

Pour finir, nous avons réalisé une cross-validation sur le modèle de Prophet pour vérifier la fiabilité des prédictions pour chaque filière. Nous avons pu déterminer que les prédictions de production pour la bioénergie, le charbon et le nucléaire étaient fiables alors que les prévisions de production pour l'éolien, le gaz, l'hydraulique et le solaire ne l'étaient pas. En effet, certaines filières comme l'énergie éolienne, hydraulique ou solaire ont été développées assez tardivement (2003, 2006) comparé à d'autres qui sont présentes depuis l'âge industrielle comme le gaz ou le charbon ce qui fait que nous n'avons pas assez de données pour déterminer une bonne prédition. Cependant, il arrive que des filières notamment le gaz présentent des productions très irrégulières au fil des années à cause d'événements particuliers (développement des énergies renouvelables, hausse des prix, ...) ce qui rend les prédictions difficiles.

### 7.1.2 - Consommation

La prédition est difficile du fait du faible nombre de données disponibles (10 ou 20 ans) et de la variabilité des facteurs externes : climat, catastrophes naturelles, mais aussi du comportement humain : nouvelles lois, électrification pour réduire la consommation d'énergies fossiles...

Il est donc difficile de prédire les données de consommation. Malgré tout, on prévoit une hausse de la consommation.

Nous avons trouvé d'autres sources qui confirment ce fait. RTE a notamment publié un rapport [Futurs énergétiques 2050](#) indiquant la consommation possible en 2050 selon différentes hypothèses (sobriété, réindustrialisation...).

### ***7.1.3 - Coût***

Concernant, la partie consacrée au coût de production de l'électricité, nous pouvons relever divers éléments. Tout d'abord, il est important de noter que ce coût dépend de la filière, que ce soit avec la présence de la filière du nucléaire ou non. En effet, les filières charbon et hydraulique peuvent être vu comme celles ayant le coût de production le plus faible. Au sujet de celles qui ont le coût le plus élevé, nous pouvons noter les filières éolien et solaire. À cela s'ajoutent les filières fioul et gaz, complétées de par la filière nucléaire si elle est prise en compte, qui ont un coût de production entre les deux précédents. Évidemment, les résultats relatifs à l'absence du nucléaire dépendent d'un choix de calcul des coûts selon la part de production de cette filière afin de prendre en compte les économies d'échelles induites sur le long terme. Enfin, la filière nucléaire ayant un coût de production moyen, une répartition sans cette dernière n'implique alors pas nécessairement un coût de production global plus élevé, notamment si elle implique une part importante des filières gaz et fioul. De plus, ce coût peut même être bien inférieur avec la forte présence des filières hydraulique et charbon.

### ***7.1.4 - Pollution***

Concernant, la partie consacrée à la pollution dû à la production d'électricité, nous pouvons relever divers éléments. Tout d'abord, il est important de noter que la pollution dépend de la filière, que ce soit avec la présence de la filière du nucléaire ou non. En effet, les filières nucléaire et les énergies renouvelables (hydraulique, biomasse et géothermie) peuvent être vu comme celles ayant la production de CO<sub>2</sub> la plus faible. Au sujet de celles qui ont la pollution la plus élevée, nous pouvons noter les filières charbon, fioul et gaz (énergies fossiles). Évidemment, les résultats relatifs à l'absence du nucléaire dépendent d'un choix de calcul de pollution selon la part de production de cette filière afin de prendre en compte les émissions induites sur le long terme. Enfin, la filière nucléaire ayant un taux de pollution très bas (CO<sub>2</sub>), une répartition sans cette dernière implique alors une pollution globale plus élevée, notamment si elle implique une part importante des filières d'énergie fossiles. Cependant les déchets radioactifs ne seraient plus un problème et il n'y aurait pas à devoir les enterrer pendant plusieurs centaines d'années.

## 7.2 - Corrélation

### 7.2.1 - Corrélation production et consommation

Dans la partie 3 traitant de la production d'électricité en fonction des filières, nous avons pu calculer la production par filière en 2020, voici un résumé pour les filières :

- **nucléaire : 335 400 000 MWh**
- bioénergie : 2171 MWh
- charbon : 1 400 000 MWh
- éolien : 39 700 000 MWh
- gaz : 34 500 000 MWh
- hydraulique : 65 100 000 MWh
- solaire : 12 600 000 MWh

Nous avons ensuite déterminer la prévision de production d'électricité pour 2030 :

- **nucléaire : 294 505 420.8691064 MWh**
- bioénergie : 3236.707748375197 MWh
- charbon : 4 121 698.7312554345 MWh
- eolien : 54 738 076.177761815 MWh
- gaz : 42 448 935.759819515 MWh
- hydraulique : 61 987 869.46652835 MWh
- solaire : 23 454 980.143268142 MWh

Dans cette partie, nous aimerais savoir par rapport à la consommation d'électricité en fonction de secteurs en 2020 et 2030, si l'on peut la satisfaire avec la production du nucléaire et sans la production du nucléaire.

De ce fait, si nous regardons la consommation d'électricité générale en 2020, d'après les graphiques de RTE présentés dans la partie 4, elle était de 440 000 000 MWh. Si l'on décide de prendre en compte la production du nucléaire, rien qu'avec celle-ci, on couvre déjà 335 400 000 MWh sur les 440 000 000 MWh de consommation. Il nous reste 104 600 000 MWh à couvrir avec les filières restantes.

D'après les valeurs de production ci-dessous, nous pouvons faire plusieurs combinaisons pour la couvrir (sachant qu'en utilisant certaines des combinaisons, il restera un surplus d'électricité) tels que :

- hydraulique + éolien
- gaz + hydraulique + charbon + bioénergie + solaire
- bioénergie + charbon + éolien + gaz + hydraulique + solaire
- ....

Maintenant voyons s'il on peut couvrir la consommation de 440 000 000 MWh en 2020 sans utiliser la production nucléaire :

$$2171+1\ 400\ 000+39\ 700\ 000+34\ 500\ 000+65\ 100\ 000+12\ 600\ 000 = 153\ 302\ 171\ \text{MWh}$$

On voit qu'il est impossible avec la somme des productions de toutes les filières de satisfaire la consommation générale sans utiliser le nucléaire avec le parc électrique existant.

En réalisant les mêmes calculs en 2030, on obtient une consommation égale à 450 000 000 MWh en générale et la consommation totale que l'on peut atteindre avec toutes les autres filières sans le nucléaire est de 186 754 794 MWh ce qui est toujours insuffisant alors qu'avec le nucléaire, on se rend compte qu'on arrive bien à satisfaire la consommation de 2030. Si on prend la prévision à partir des données ORE, la consommation est plus élevée que celle pour RTE donc c'est encore plus insuffisant.

### *7.2.2 - Corrélation production et coût*

Avec la méthode des K-means réalisé précédemment pour classer les filières en fonction de leur coût, nous avons obtenu les groupes suivants :

- Coût élevé: Éolien et photovoltaïque
- Coût moyen : Nucléaire, gaz, fioul
- Coût faible: hydraulique, charbon

Nous pourrions alors nous demander si les coûts influencent la production des filières. Par exemple, est-ce que le groupe qui a les coûts de production les plus élevés produit le moins ?

D'après les résultats sur la production, voici le classement au niveau des parts de production en 2021 :

- Le charbon a une part de production de 1.99256 %
- L'éolien a une part de production de 7.32041 %
- Le gaz a une part de production de 6.62790 %
- L'hydraulique a une part de production de 12.47676 %
- Le nucléaire a une part de production de 68.69921 %
- Le solaire a une part de production de 2.88267 %

On voit que l'éolien et le solaire sont classés dans le groupe avec les coûts les plus élevés et leurs productions n'est pas similaire. L'éolien est le 3ème plus gros producteur d'électricité et le solaire l'avant dernier donc il n'y a pas de lien entre le fait d'avoir un coût élevé et la production des filières de ces groupes.

Ensuite, pour les filières qui ont un coût moyen, on retrouve la même chose, le nucléaire est celui qui produit le plus comparé au gaz qui a deux fois moins de production que le nucléaire.

Enfin, l'hydraulique et le charbon sont les deux filières les moins chères et au niveau production, elles sont opposées l'une de l'autre, l'hydraulique produit énormément plus que le charbon.

Si l'on compare entre les groupes, on voit que les filières qui produisent le plus à savoir l'hydraulique et le nucléaire ont des prix différents. On peut en conclure que les énergies qui sont les plus récentes (éolien et le solaire) sont les plus chères mais ne sont pas celles qui produisent le moins de même pour les filières avec un coût faible.

En calculant les coût de production de chaque filière en fonction de leurs parts en 2030, on obtient les résultats suivant :

Charbon : 39.811764705882354 €/MWh
Gaz : 83.24603174603175 €/MWh
Hydraulique : 16.91304347826087 €/MWh
Nucléaire : 86.56413793103448 €/MWh
Solaire : 128.0297097394166 €/MWh
Eolien : 131.47532095988245 €/MWh

Nous pouvons alors nous poser la même question que précédemment. Pour cela, il est important de rappeler les parts des différentes filières en 2030 ci-dessous :

- Le charbon a une part de production de 0.85 %
- L'éolien a une part de production de 11.37 %
- Le gaz a une part de production de 8.82 %
- L'hydraulique a une part de production de 12.88 %
- Le nucléaire a une part de production de 61.19 %
- Le solaire a une part de production de 4.87 %

Nous constatons ainsi qu'il n'y a toujours pas de lien entre les parts et le coût de production. En effet, le nucléaire, qui est la filière largement majoritaire a un coût de production bien supérieur à d'autres filières moins importantes comme l'hydraulique. Cela est également renforcé par la différence de parts entre les filières éolienne et solaire qui ont pourtant un coût de production presque identique. Il y a donc bien des facteurs extérieurs qui empêchent cette corrélation d'exister.

### 7.2.3 - Corrélation production et pollution

Au sujet de la corrélation entre la production d'électricité et la pollution sont plutôt bien corrélées. Les filières qui ont le moins bon rendement en énergies sont les filières "vertes" avec l'éolien, le solaire, la biomasse et la géothermie, qui à cause de dépendre principalement de la nature, leur production électrique est moins efficace. Cependant le charbon possède aussi une production d'énergie très faible alors que c'est la filière la plus polluante de très loin.

Pour ce qui est du fioul et du gaz (les énergies fossiles), ces filières sont capables de fournir relativement beaucoup d'électricité malgré la relativement grande émission de CO<sub>2</sub>, surtout le gaz. Cependant, elles sont aussi accompagnées de l'hydraulique qui produit beaucoup mais pollue peu. Et enfin, le nucléaire est la filière qui produit le plus d'électricité en plus

d'émettre la plus petite dose de CO<sub>2</sub> pour produire un MW. C'est la raison pour laquelle cette filière est beaucoup utilisée en France. Cependant, c'est également la seule filière à produire des déchets radioactifs, difficiles à traiter en totalité.

### **7.2.4 - Corrélation coût et pollution**

Au sujet de la corrélation entre le coût et la pollution, nous allons comparer les production MWh de chacun sans prendre en compte les parts de production de la filière en question. Ainsi, les corrélations entre les coûts de production et la pollution engendrée sont assez faibles. D'abord, l'hydraulique a un coût de production très faible tout comme ses émissions de CO<sub>2</sub>. Ensuite, les filières éolienne, photovoltaïque, biomasse et géothermie ont un coût de production très élevé mais des émissions de CO<sub>2</sub> relativement faibles. Dans le même temps, les filières gaz et fioul ont un coût de production moyen tout en ayant des émissions de CO<sub>2</sub> assez élevées. De plus, le charbon, qui a un coût de production relativement bas, est la filière la plus émettrice de CO<sub>2</sub> et de loin. Enfin, le nucléaire est la filière de production la moins émettrice de CO<sub>2</sub> tout en ayant un coût de production moyen. Cependant, c'est également la seule filière à produire des déchets radioactifs, difficiles à traiter en totalité.

## **7.3 - Comparaison**

### **7.3.1 - Comparaison entre les attentes et les véritables résultats concernant la production et la consommation**

Afin d'avoir un aperçu critique sur la consommation et la production de nos prévisions, nous avons cherché des bilans prévisionnels sur RTE en 2030 disponibles [sur ce rapport](#).

Au niveau de la consommation en 2030 (par secteur) d'après RTE (page 880) :

- Consommation intérieure : 508,2 TWh
- résidentiel : 151,4 TWh
- tertiaire : 125,6 TWh
- industrie : 113,3 TWh
- transport : 32,5 TWh
- énergie (dont pertes réseau) : 52,5 TWh
- production d'hydrogène : 25 TWh

Nous avons calculé :

- Consommation intérieure : 500 TWh
- résidentiel : 140 TWh

- tertiaire : 180 TWh
- industrie : 140 TWh
- agriculture : 13 TWh
- secteur inconnu : 8 TWh
- somme des précédents secteurs : 481 TWh

On remarque qu'on a des valeurs similaires pour la consommation globale, mais le détail par secteur diffère grandement notamment pour les secteurs tertiaires et industriels. Il y a aussi des nouvelles catégories qui apparaissent dans le rapport de RTE comme la production d'hydrogène.

Au niveau de la production en 2030, nous avions calculé les chiffres suivants :

- **nucléaire : 294, 505421 TWh**
- bioénergie : 0,00332367 TWh
- charbon : 4,12169 TWh
- eolien : 54,73807 TWh
- gaz : 42,4489 TWh
- hydraulique : 61,987 TWh
- solaire : 23,4549 TWh

Les données obtenus sur le bilan prévisionnel sont les suivantes (page 899) :

- **nucléaire : 343,2 Twh**
- bioénergie = 9,6 Twh
- charbon = 0,0 TWH
- eolien : 86,7 TWh
- hydraulique : 65,6 TWh
- solaire : 42,3 TWH

On voit qu'en général, les valeurs de production des filières prédites sont plus hautes que celles annoncées par RTE sauf pour l'hydraulique ou elles sont similaires.

De ce fait, il y a une énorme différence au niveau du charbon puisque l'objectif prévu de la France pour 2030 est d'être 0 carbone ce qui signifie que le charbon devrait être à 0 TWh ce qui n'est pas du tout pris en compte dans nos prédictions et cela entraîne une augmentation plus importante du nucléaire et des autres énergies renouvelables.

### *7.3.2 - Comparaison entre les attentes et les véritables résultats concernant le prix et la pollution*

Comme nous l'avions envisagé, les énergies fossiles sont effectivement celles qui ont le coût de production le plus faible tout en ayant les émissions de CO<sub>2</sub> les plus élevées. Dans le même temps, les énergies renouvelables ont bien le coût de production le plus élevé et les émissions de CO<sub>2</sub> les plus faibles. De plus, nous avons confirmé que le nucléaire est évidemment à part en raison de ses rejets de déchets radioactifs malgré sa basse émission de CO<sub>2</sub>. D'ailleurs, lors de cette étude, nous avons constaté que le coût de production du nucléaire n'est pas si bon marché en comparaison à ce que nous pensions.

Évidemment, ce coût reste relativement plus faible que pour d'autres filières, mais c'est important de le souligner. Enfin, l'hydraulique est la filière qui nous a le plus surpris de par son coût de production, qui est notamment inférieur à celui du charbon.