


Projet ENERGIE

David BEAUVAIS - dbeauvais93@gmail.com 

Jean-Yves BERNARD - jybernard8@gmail.com 

Cyril BRUYERE - cyril.bruyere@gmail.com 

Mentor

Gaspard Grimm - gaspard@datascientest.com

Introduction	4
Objectif du projet	4
Black-out : consommation supérieure à la production	4
Notre motivation	4
Jeu de données	5
Données source	5
Appropriation du sujet	5
Contexte de la production d'énergie en France	5
Analyse des données	7
Production hydraulique par pompage	7
Production éolienne	7
Vérification de la réalité des données	8
Ajout des variables nécessaires à l'étude	10
Choix de l'unité de travail	10
Visualisation des datas brutes et transformées	11
Visualisation des productions/consommations	11
Visualisations régionales	13
Visualisation des types de production	15
Energies nucléaire et thermique	16
Energies renouvelables	16
Energie hydraulique	16
Energie éolienne	17
Energie solaire	17
Bioénergies	17
Table de classification des types de production	18
Black-out	19
Définition et notion de balance	19
Visualisation des Black-outs au niveau national	19
Méthodologie	20
Approches pour prédire la balance	20
Prédiction directe	20
Prédiction indirecte	20

Premiers essais de modélisation	20
Modèles utilisées	21
SARIMAX	21
Régression linéaire	21
Pas temporels testés	22
Etat des signaux donnés aux modèles	23
Signaux originaux	23
Signaux filtrés par transformation de Fourier	24
Variables explicatives ajoutées aux modèles	26
Sinusoïdales ayant des périodes correspondantes aux saisons	26
Données métiers	26
Principe de construction	27
Variables explicatives utilisées pour la prédiction directe	29
Variable explicative utilisée pour la prédiction indirecte	29
Résultats et interprétations	32
Rappel des modélisations effectuées	32
Résultat de modélisation : balance directe	33
Evaluation des modèles	33
Interprétation des résultats	34
Bilan de l'étude	35
Conclusion	36
Pilotage	37
Stratégie	37
Demain	37
Opportunités	38
Annexes	39
Bulletins météorologiques Val de Loire	39
Echanges avec RTE	40
2ème échange	40
1er échange	42
Transformée de Fourier	43

Introduction

Objectif du projet

Notre objectif est l'étude du phasage entre consommation et production d'énergie en France.

Cette production est le fruit de l'agglomérat des productions régionales ayant chacune des spécificités propres. Certaines régions produisent plus qu'elles ne consomment, d'autres se trouvent dans la situation inverse. D'autres part, certaines régions disposent d'installations nucléaires lorsque d'autres disposent d'un mix plus complet, associant productions d'énergie à partir de ressources renouvelables et non renouvelables.

Dans ce cadre de productions variées et dispersées, l'objectif initial du projet visant à étudier le phasage entre la consommation et la production énergétique devra notamment s'intéresser aux points suivants :

- Caractéristiques des séries temporelles à étudier
- Recherche de variables explicatives en corrélation avec les séries à prédire.

Black-out : consommation supérieure à la production

L'intérêt d'une telle étude trouvera son sens dans le fait d'anticiper les situations à risques où l'offre domestique ne pourrait pas satisfaire la demande. Nous pouvons donc dégager 3 axes d'approfondissement, intéressant autant d'expertises métiers.

Prévision court terme :

Une anticipation de la consommation énergétique à court terme permettra de piloter l'offre d'énergie et donc d'articuler finement les moyens de production.

L'ordre de grandeur du temps devrait alors être compris entre quelques heures et quelques jours.

Prévision moyen/long terme :

Une anticipation de la consommation à moyen/long terme permettra de piloter l'infrastructure existante, en positionnant les interventions de maintenance et facilitant l'organisation des ateliers de production.

L'ordre de grandeur du temps devrait alors être compris entre quelques semaines et quelques mois.

Dimensionnement du parc :

Une anticipation de la structuration de la consommation et de son orientation à très long terme permettra de poser les grandes stratégies et d'évaluer l'adaptation nécessaire du parc actuel aux enjeux futurs.

Notre motivation

Nous avons choisi ce projet pour les raisons suivantes :

- Sujet d'actualité
- Enjeux planétaires majeurs
- Possibilité d'être acteur du monde de demain
- Asseoir notre profil expert de la data adossée à l'énergie
- Volonté d'intégrer l'univers de la mobilité électrique

Jeu de données

Données source

Données éCO2mix régionales consolidées et définitives datant de janvier 2013 à octobre 2021 et disponible sur [Opendata](https://opendata.reseaux-energies.fr)¹.

Les données des 12 régions disponibles sont globalement consistantes :

- Le découpage régional respecte les délimitations actuelles en vigueur depuis 2015
- Le pas de 30 minutes est respecté de 2013 à 2021
- Les données sont disponibles sur la période pour l'ensemble des énergies
- Les données manquantes à corriger sont concentrées sur quelques colonnes
- Les colonnes pour les taux de couverture et taux de charge ne sont pas significatives

Appropriation du sujet

Contexte de la production d'énergie en France

L'énergie nucléaire au centre du mix Français :

1. En France, 56 réacteurs sur 18 centrales (32 de 900 MW, 20 de 1300 MW et 4 de 1450 MW)².
2. La majorité ont été mis en service entre 1977 et 1992³.
3. Les centrales ont été conçues pour fonctionner au moins 25, 30 ou 40 ans⁴.
4. En 2009, EDF a exprimé le souhait d'exploiter les réacteurs au-delà de 40 ans⁵.
5. EDF doit obtenir les autorisations de prolongation auprès de l'Autorité de Sûreté Nucléaire⁶.
6. Suite à l'accident de Fukushima, la sûreté doit évoluer (~55 Milliards d'Euros)⁷.
7. La loi énergie-climat prévoit de faire baisser la part du nucléaire dans le mix français à 50%⁸.
8. Le plan publié en avril 2020 prévoit l'arrêt de 14 réacteurs d'ici à 2035⁹.
9. 2 premiers réacteurs ont été arrêtés définitivement en février et juin 2020 (Fessenheim)¹⁰.
10. Arrêt de tranche : arrêt de fonctionnement pour recharge de combustible pendant 1 mois tous les 12 ou 18 mois, intercalés avec arrêt pour opérations de maintenance pendant 2 mois. Autres arrêts pour les visites décennales.

¹ source : <https://opendata.reseaux-energies.fr>

² [edf](https://www.edf.fr)

³ [wikipedia.org](https://fr.wikipedia.org)

⁴ [irs.fr](https://www.irs.fr)

⁵ [asn.fr](https://www.asn.fr)

⁶ [asn.fr](https://www.asn.fr)

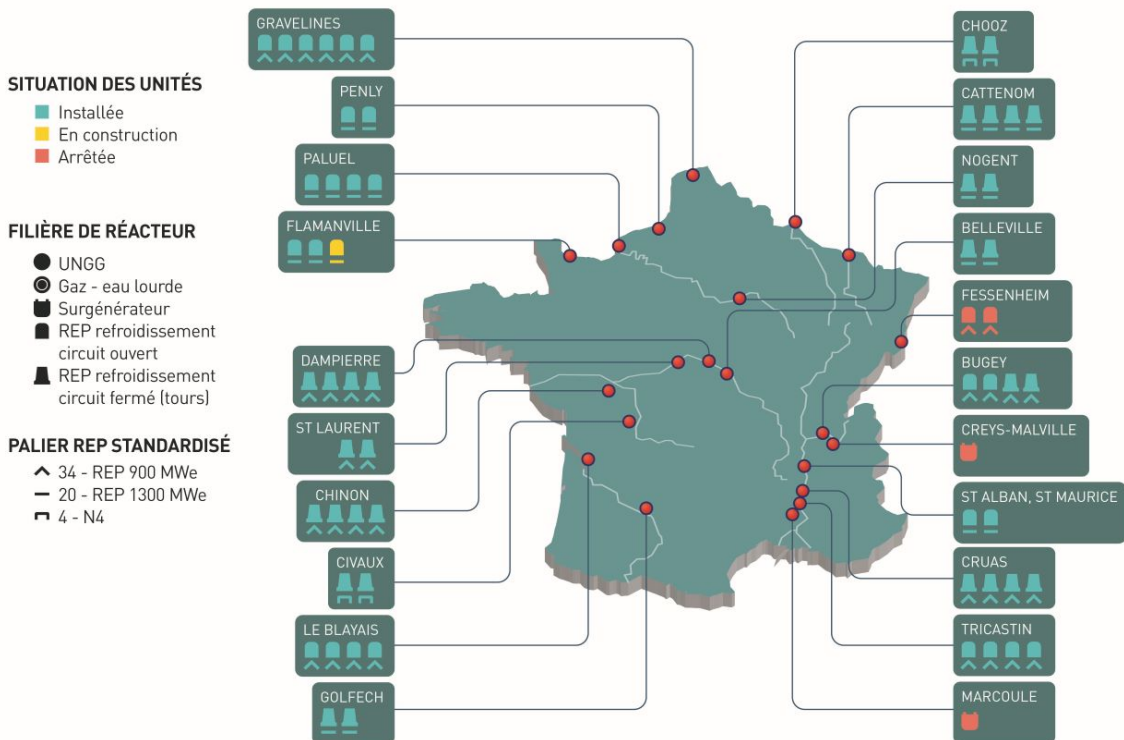
⁷ [irs.fr](https://www.irs.fr)

⁸ [ecologie.gouv.fr](https://www.ecologie.gouv.fr)

⁹ [LOI n° 2019-1147 du 8 novembre 2019 relative à l'énergie et au climat](#) & réponse d'EDF

¹⁰ [asn](https://www.asn.fr)

CARTE DES UNITÉS ÉLECTRONUCLÉAIRES EN FRANCE AU 01/01/2021



56 réacteurs sur 18 centrales¹¹

En octobre 2021, lors de la visite décennale sur le réacteur 1 de la centrale de Civaux, des anomalies ont été détectées. Aussitôt le 2^{ème} réacteur du site a été arrêté ainsi que les 2 réacteurs de Chooz qui sont de même génération. Ce sont les 4 plus gros réacteurs (1450 MW) qui sont à l'arrêt.

Par ailleurs, l'énergie thermique est également en déclin de nos jours du fait des politiques environnementales. Il ne reste que 3 centrales à charbon en 2020. On note notamment les 1250 MW stoppés dans le Grand Est en 2013, 2014 et 2015. Pour mémoire, Le Président de la République Française, Monsieur Emmanuel Macron, s'était engagé à arrêter cette filière pendant son premier quinquennat. La dernière s'arrêtera entre 2024 et 2026.

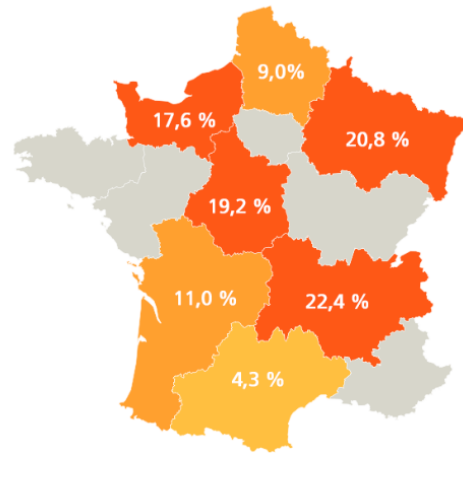
En résumé, le risque de black-out est très présent. Il est également une source de préoccupation pour les mois et années à venir étant donné l'arrêt évoqué jusqu'à récemment des filières principales actuelles, nucléaire et thermique. Cependant, les derniers événements internationaux à l'Est de l'Europe pourraient amener légèrement ces évolutions promises.

¹¹ cea.fr

Analyse des données

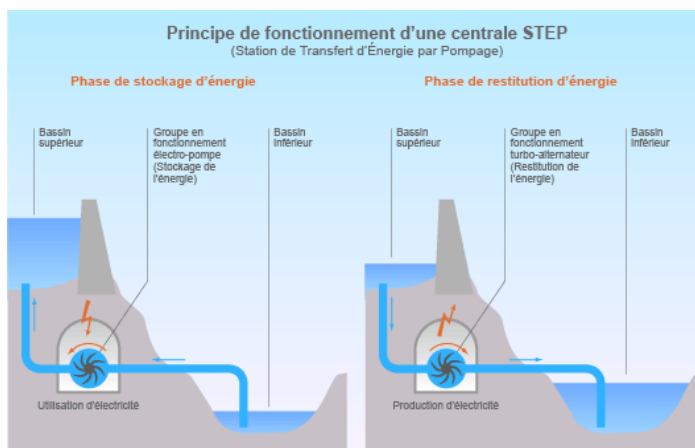
Production nucléaire en France

Les données manquantes pour la production nucléaire concernent les régions qui n'ont pas de production d'énergie nucléaire¹²



Production en 2019¹³

Production hydraulique par pompage



Sur france-hydro-electricite.fr on apprend ce qu'est une STEP¹⁴ et leur localisation en France.

Elles sont au nombre de 6, et permettent de comprendre l'absence de données dans certaines régions.

Certaines valeurs non nulles étant présentes dans des régions où il n'y a pas de STEP et afin de vérifier la cohérence des données, nous avons contacté RTE¹⁵ qui nous a confirmé l'existence de possibilités de pompage sur 3 stations identifiées, mais n'étant pas des STEP.

Production éolienne

Le Centre Val de Loire présente un défaut d'enregistrement de données pour la production d'énergie éolienne sur la journée du 27 décembre 2013. Nous avons vérifié les bulletins météorologiques correspondant¹⁶ qui peuvent expliquer l'anomalie.

Par observation des données de la veille et du lendemain, nous avons déterminé un pattern équivalent à la journée du 28 décembre 2013, inversé, et appliqué à la journée manquante.

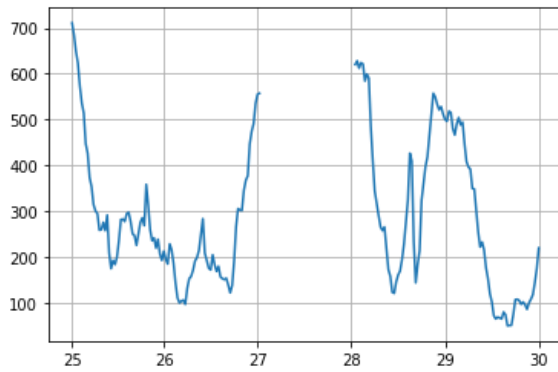
¹² source edf.fr

¹³ source edf.fr

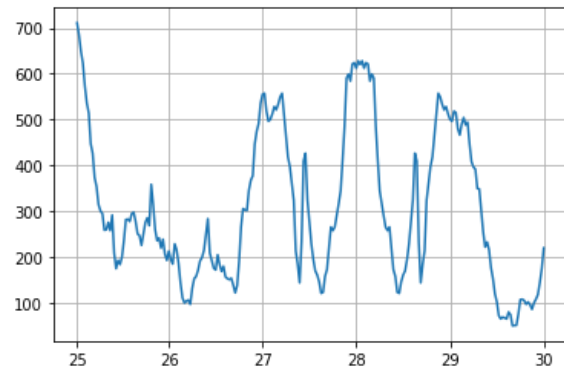
¹⁴ STEP : station de transfert d'énergie par pompage. Système en place entre 2 barrages. L'énergie est produite par gravité puis remontée par pompage. C'est un circuit fermé de production.

¹⁵ échanges avec RTE présents dans l'annexe [Échanges avec RTE](#)

¹⁶ [Bulletins météorologiques Val de Loire](#)



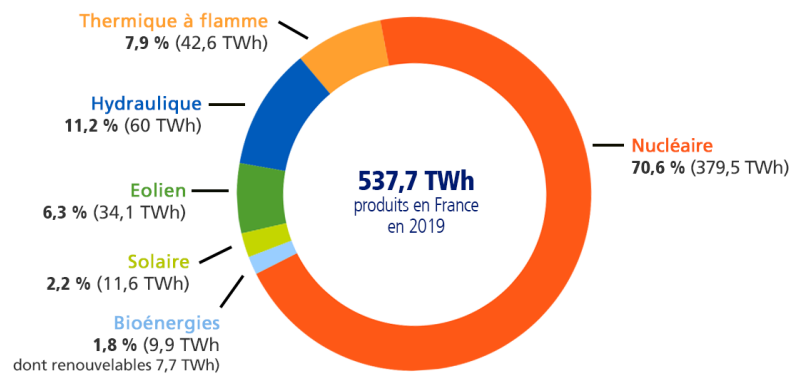
Données manquantes sur la journée du 27/12/2013.



Correction apportée sur la journée du 27/12/2013

Vérification de la réalité des données

Afin de tester la validité des données après nettoyage nous allons opérer une vérification sur la cohérence des puissances exprimées.



La production française d'électricité en 2019

Source RTE - bilan électrique 2019

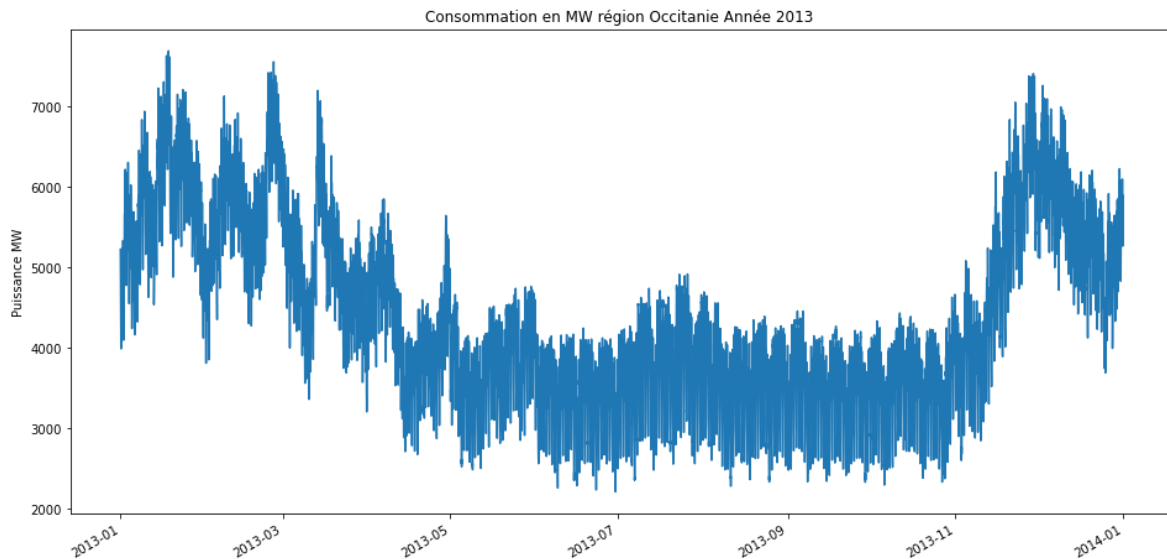
© EDF

Production d'électricité 2019 par filière de production¹⁷

Nous avons ainsi déterminé l'écart entre une intégration mathématique et la somme par pas de 30 minutes. Pour chacune des régions la différence ne dépasse pas 0.1 %.

¹⁷ source [edf.fr](https://www.edf.fr)

Exemple sur la consommation de la région OCCITANIE sur 2013:



Méthodes de calcul

1. Sommes sur 30 minutes de chaque puissance (méthode des rectangles)
2. Intégrale (méthode de trapèzes)
3. Intégrale de "composite Simpson's rule"

Quantité d'énergie totale consommée

1. area rectangles = 76 031 301
2. area trapèzes = 76 087 600
3. area Simpson = 76 036 803

0.0667%

Intégration mathématique vs somme / 30'

Vérification sur le bilan annuel 2019.

Production nucléaire de 2019 (source RTE)	379.5 TWh
Somme 2019 des productions / 30 minutes	758 459 526 MW
/2 ⇒ valeur exprimée par heure	
Production nucléaire de 2019 (calculée)	379.2 TWh

De la même manière, sur d'autres filières de productions nous retombons bien sur les données RTE 2019. Nous validons ainsi la consistance de nos données et la méthode de somme par pas de 30 minutes comme équivalente à une intégration mathématique de type trapézoïdale.

Ajout des variables nécessaires à l'étude

Le dataset¹⁸ initial nettoyé nécessite d'être optimisé du point de vue des libellés et de la construction des dates. Aussi nous ajoutons les variables production et balance :

- Dates
- Consommation
- Thermique
- Nucléaire
- Eolien
- Solaire
- Hydraulique
- Pompage
- Bioénergies
- Production : Thermique + Nucléaire + Eolien + Solaire + Hydraulique + Pompage + Bioénergies
- Balance : Production - Consommation

On notera que la colonne "Dates" à la qualité d'index.

Choix de l'unité de travail

Le jeu de données est fourni en MW. Le mégawatt correspond à l'énergie maximale qu'une installation électrique peut produire en 1 seconde. Le MWh exprime la quantité d'énergie produite par l'installation pendant une heure.

Nous avons sollicité RTE¹⁹ pour clarifier ce point. Ainsi, nous pouvons conclure que :

- Le MW est une unité de puissance
 - Focus : consommation instantanée.
- Le MWh est l'unité de l'énergie
 - Focus : capacité à produire par unité de temps.

Nous pouvons donc appréhender notre projet sous les 2 aspects :

1. En utilisant les puissances (MW) qui sont des pics de puissance maximale sur 30 minutes.
2. En utilisant les énergies (MWh)

Il convient alors d'être vigilant sur les agrégations de données. Il est possible de sommer les puissances régionales pour obtenir la puissance nationale. Il est également possible de sommer les puissances par filière pour avoir la production totale. En revanche, les agrégations sur l'axe temporel doivent être moyennées. En effet, la puissance obtenue sur 2 jours est une puissance moyenne et non cumulée.

La limite de cette approche est que l'énergie est estimée avec une puissance maximale et constante sur 30 minutes. Il s'agit en réalité du maximum d'énergie théoriquement disponible sur cette période.

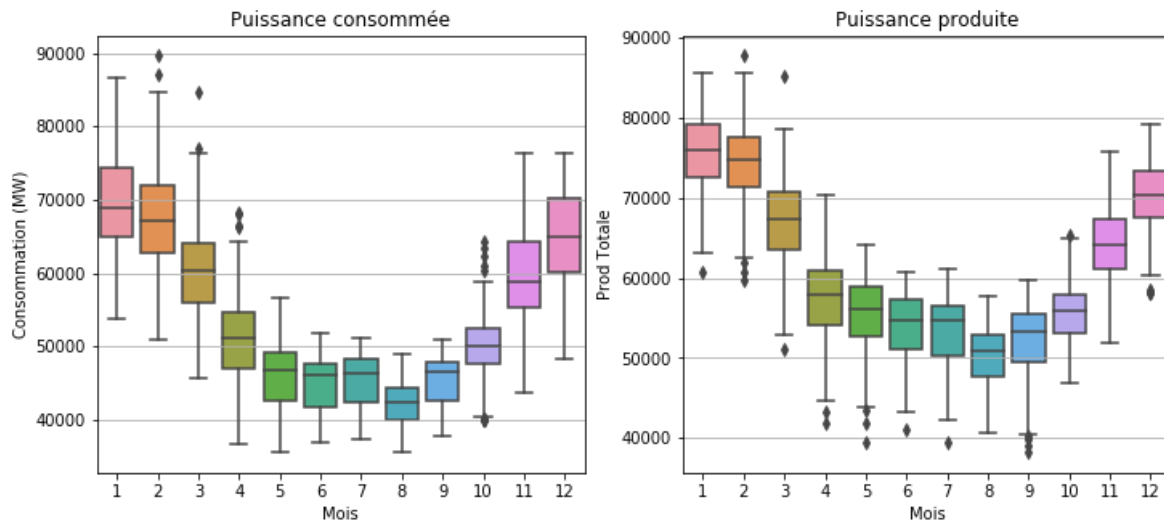
¹⁸ [Listing des colonnes et détails du jeu de données sur opendatasoft.com](#)

¹⁹ voir en annexe : [Échanges avec RTE](#)

Visualisation des datas brutes et transformées

Cette partie expose les visuels les plus marquants. L'ensemble de la production visuelle (part [I](#), [II](#) et [III](#)) est consultable sur github.com.

Visualisation des productions/consommations



Dispersion de la consommation et de la production sur l'ensemble du dataset

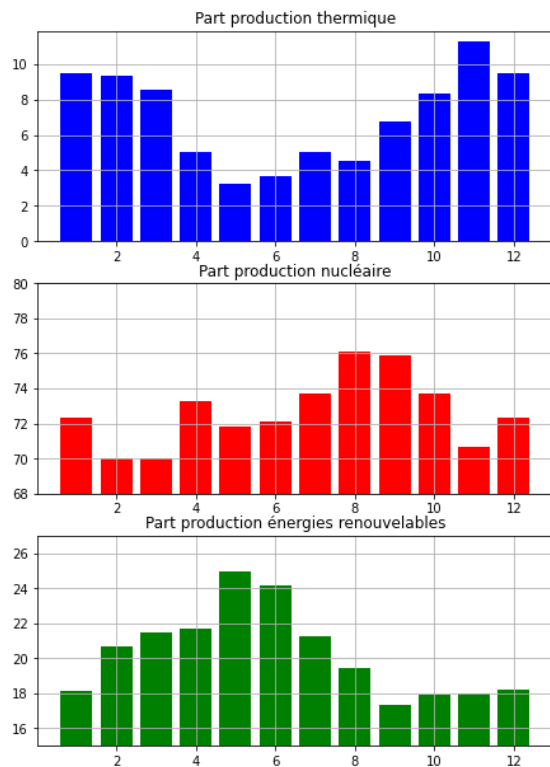
Sans surprise, on observe une demande de puissance supérieure en hiver, de fin octobre à mi-avril environ. En été, la puissance consommée est en moyenne de 55 000 MW, et d'environ 70 000 MW en hiver. Le chauffage, suivant les baisses de température, représente une hausse conséquente de près de 30% (les experts estiment que 1° de température en moins nécessite 2 400 MW de plus)

De fait, la variabilité est plus importante sur les mois d'hiver.

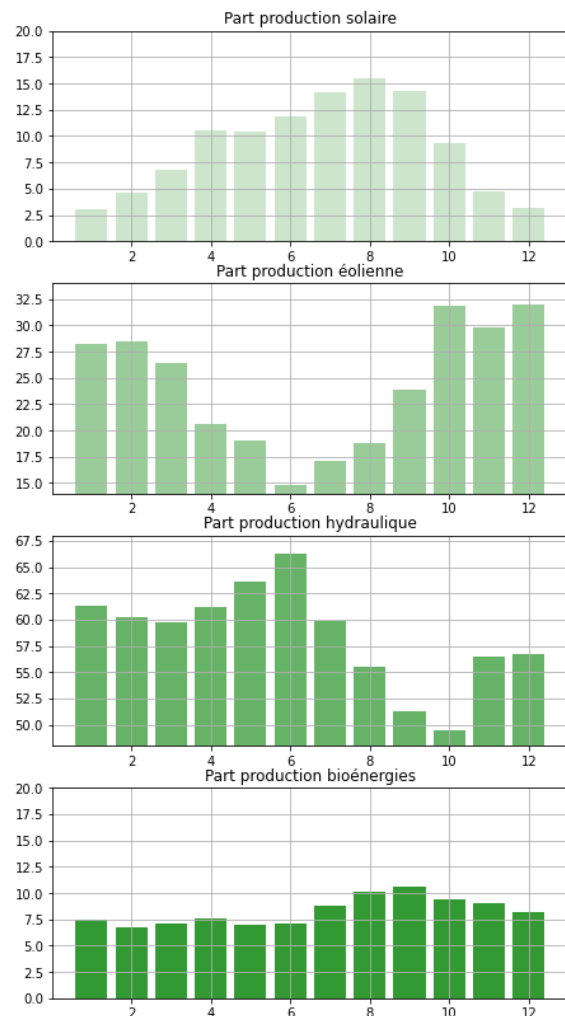
La puissance produite est une réponse à la consommation. Le pattern est logiquement semblable.

On peut donc estimer que la consommation énergétique est relativement reproductible d'une année sur l'autre et est très corrélée aux cycles météorologiques.

Part respectives des 3 types de productions



Part respectives au sein de la production d'énergies renouvelables



A gauche, les représentations des parts des 3 types de productions montrent une amplitude légèrement supérieure pour la filière thermique qui semble être la variable d'ajustement. On peut imaginer que cette filière a moins d'inertie de mise en route.

Faiblement contrôlables, les énergies renouvelables dépendent pour beaucoup des conditions météorologiques. On constate ainsi qu'à la fin du printemps et au début de l'été, périodes plus favorables à la production d'énergie renouvelable, se présente un pic de production permettant un creux sur la production d'énergie thermique. En écho à cette dépendance, on remarque, à la sortie de l'été, une forte baisse de la production d'énergies renouvelables, compensée par l'énergie nucléaire plutôt que par l'énergie thermique. De manière générale, les ressources provenant de production d'énergie renouvelable sont rapidement consommées tout au long de l'été et peinent à se reconstituer avant le cycle annuel suivant. De plus, hormis les bioénergies, les productions d'énergies renouvelables sont très disparates d'un mois sur l'autre.

On peut également ajouter que les énergies renouvelables sont en opposition de phase avec la demande. Leur pic est à l'été alors que le pic de la consommation est l'hiver.

Type de production	Synthèse des enseignements
Thermique	Production plus forte en automne et hiver avec une tendance à la décroissance au cours des années
Nucléaire	Légère décroissance de la production au cours des années, sans compter l'année 2020 au profil atypique vraisemblablement lié à la crise sanitaire.
Eolien	Croissance régulière de la production démontrant un réel investissement dans les installations, et variance de plus en plus notable. Production plus significative en automne et en hiver.
Solaire	Croissance régulière de la production démontrant un réel investissement dans les installations. Production plus significative au printemps et en été.
Hydraulique	Production stable au cours des mois avec un creux régulier sur la fin d'été et les mois exposés à la sécheresse.
Bioénergies	Croissance notable de la production au cours des premières années couvertes par le dataset, puis une stabilisation de cette production. Cette filière de production ne semble pas connaître de saisonnalité franche.
Échanges physiques	Sans réelle saisonnalité et avec une variance importante, on distingue tout de même une croissance régulière des échanges au cours des années

Visualisations régionales

Nous disposons du détail des données sur les 12 régions de France Métropolitaine, hors Corse.

Les visualisations concernant chacune de ces régions sont disponibles [github.com part_I](#), [github.com part_II](#), [github.com part_III](#)

En résumé, on note :

L'éolien se démarque dans le Grand Est et dans les Hauts de France où les capacités se sont développées plus qu'ailleurs en France.

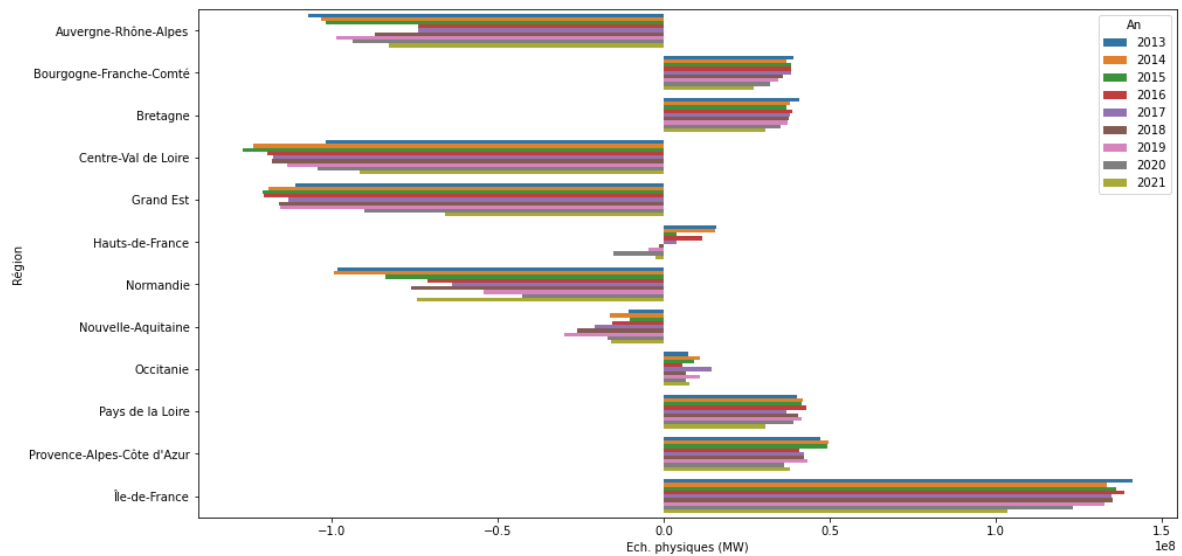
Les capacités d'énergie solaire ont significativement augmenté dans la moitié sud de la France, avec une différence marquée également entre l'ouest et l'est.

L'énergie hydraulique est relativement stable avec la région Auvergne-Rhône-Alpes (AURA) qui fournit logiquement plus de 50% de la capacité nationale. Cette région bénéficie de possibilités d'implantation de barrages, d'opportunités d'implanter des stations de transfert d'énergie par pompage (STEP) avec de forts dénivelés.

En 2013, l'Île de France (IDF) et la Nouvelle Aquitaine étaient en avance sur les autres régions pour ce qui concerne les bioénergies. En 2020, l'IDF n'a pas augmenté sa capacité alors que la Nouvelle Aquitaine a continué à progresser pour se démarquer sur cette filière d'énergie. L'IDF a été rejointe et dépassée par AURA, Provence-Alpes-Côte d'Azur (PACA), Hauts de France, Grand Est. Les autres régions restent en retrait.

En définitive, outre les cycles de productions et de consommations détaillés plus hauts, la vue par région permet surtout de repérer les régions importatrices des régions exportatrices au sein du mix national et d'identifier les spécialités en termes de production.

Ces différences composent l'ensemble du mix énergétique national et permet donc d'identifier les possibilités d'échanges d'énergie au sein du territoire.



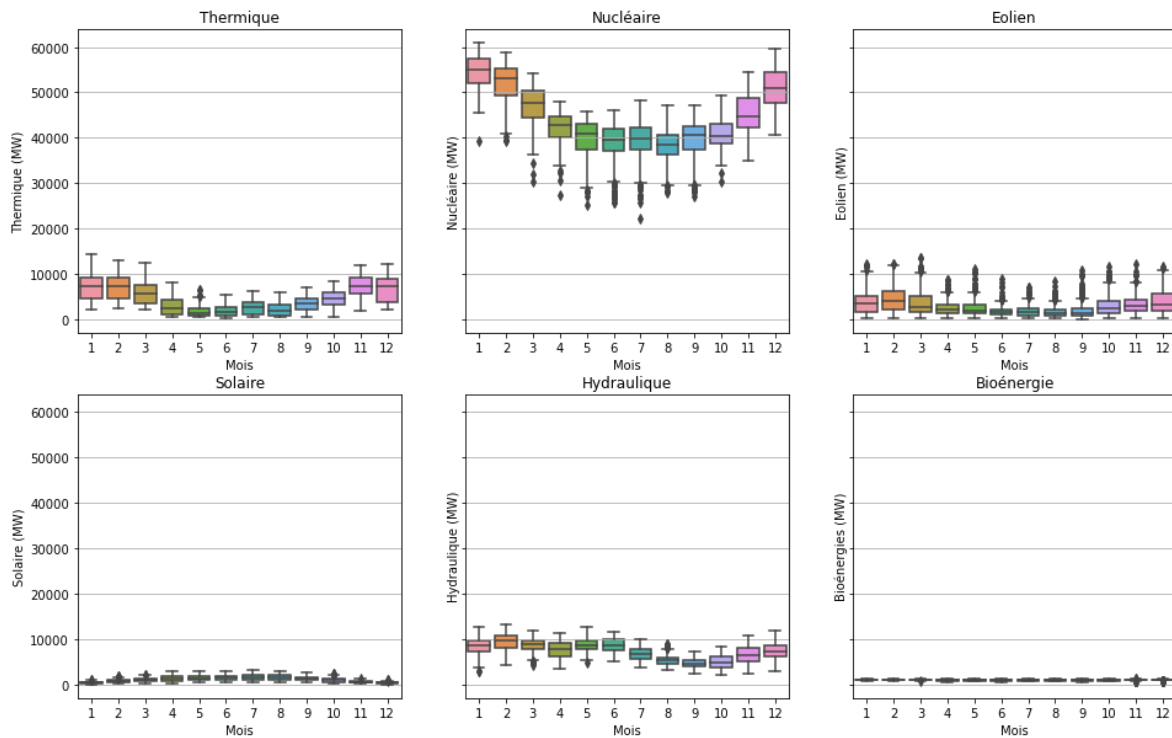
Échanges physiques entre les régions sur l'ensemble du dataset

L'intérêt du calcul de la balance au niveau des régions n'a que peu de sens car les échanges à l'intérieur du pays sont établis et ne répondent pas à une loi de marché.

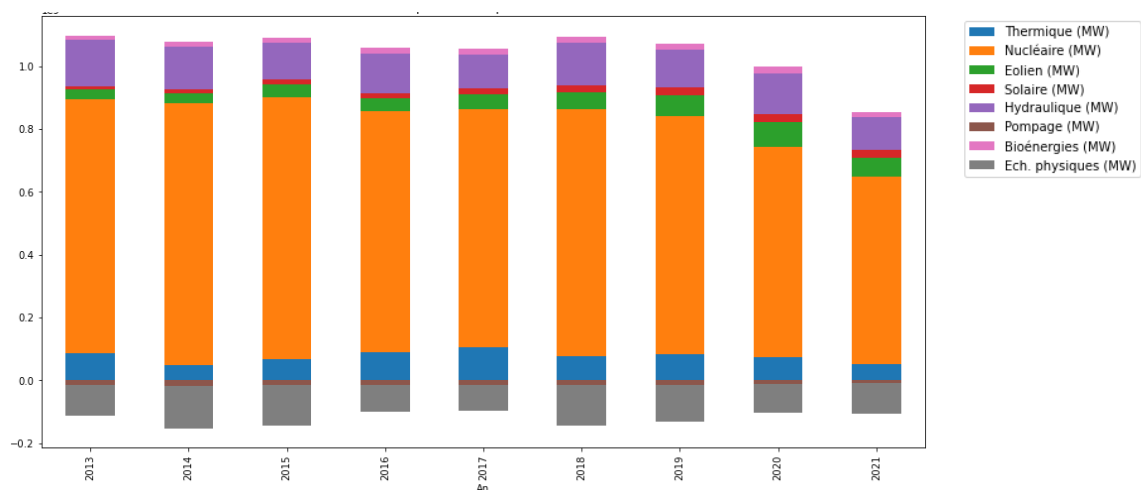
Notre objectif de définir la balance au niveau national devra néanmoins s'appuyer sur les forces et les faiblesses de chaque région.

Visualisation des types de production

L'objectif de la classification des types de production est de distinguer les données non maîtrisables et significatives, des données maîtrisables et non significatives. La visualisation des données par type de production permet d'avancer sur cette classification.



Dispersion annuelle des productions nationales des différentes sources d'énergie



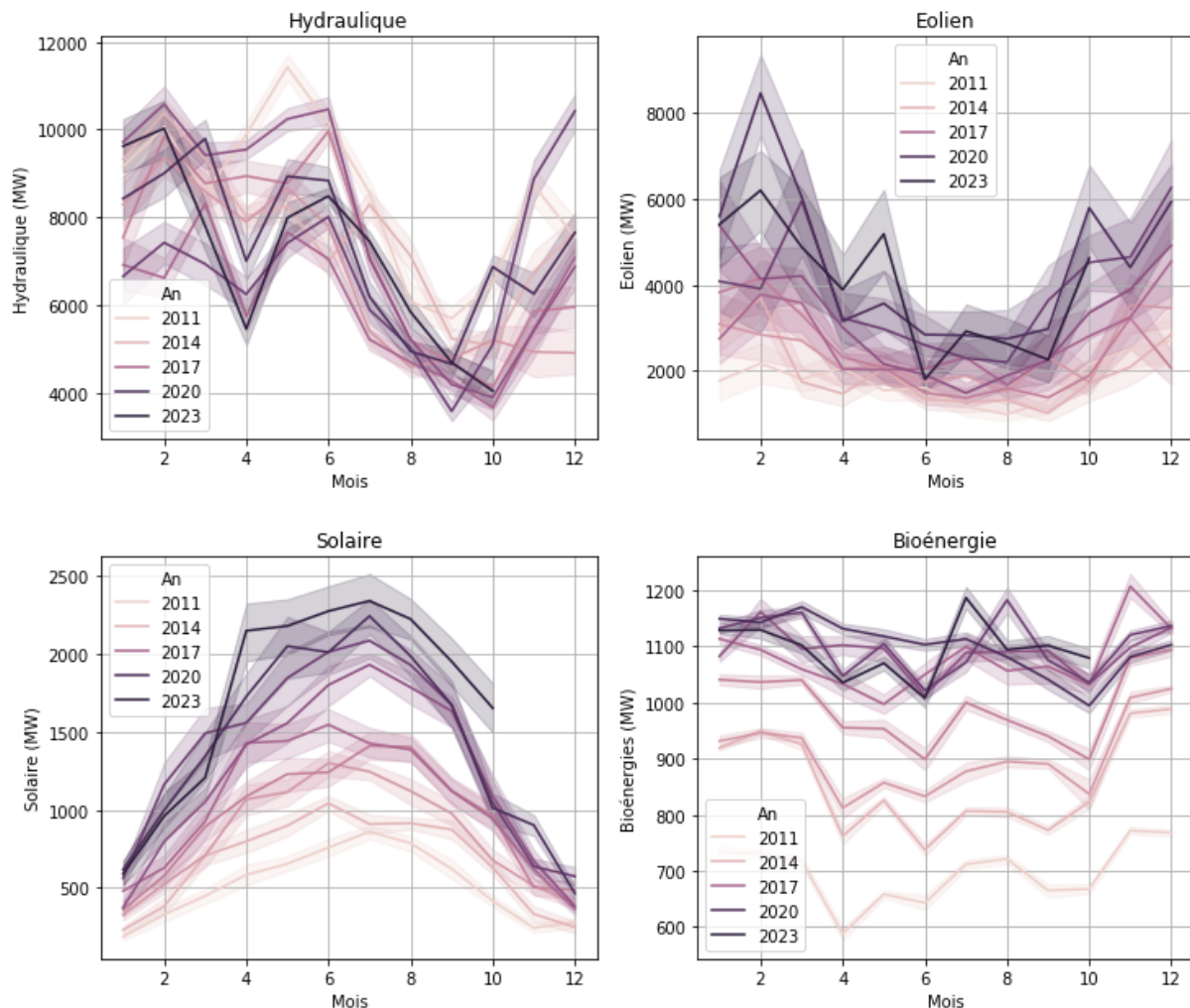
Les données dont nous disposons, ici au niveau national, semblent indiquer que la part de la production d'énergie éolienne ainsi que solaire a augmenté notablement. Dans ce cadre, la part de la production d'énergie nucléaire apparaît comme celle ayant la plus diminué.

Energies nucléaire et thermique

Ces productions d'énergie sont pilotées par l'homme. En ce sens, elle est soit une donnée d'entrée pour nos modèles, soit une variable explicative. La prédire n'a pas véritablement de sens.

On pourra définir le pattern de pilotage en construisant la moyenne de la production sur une année.

Energies renouvelables



Contrairement aux énergies nucléaires et thermiques qui sont a priori le résultat d'une planification humaine ("choix de l'Homme"), les énergies renouvelables dépendent de facteurs naturels météorologiques. Elles ne semblent pas maîtrisables sur le moment et constituent une variable explicative de nos modèles.

On se rappellera cependant que le pilotage prévisionnel des énergies nucléaires et thermiques de 12 à 24 mois en amont et que les énergies renouvelables pilotées au travers de la capacité progressivement installée rend plus flou cette notion de maîtrise des productions.

Ainsi, de 2013 à 2021, on observe très nettement le développement des énergies renouvelables. (cf "Bioénergie" sur le graphique précédent).

Energie hydraulique

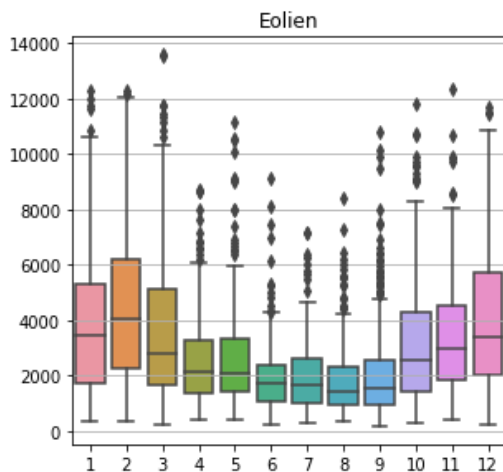
L'énergie hydraulique a un pattern identique d'une année sur l'autre. Un premier semestre fort mais chaotique en corrélation avec les précipitations : fonte des neiges et pluies. Un point bas quasi

identique toutes les années à la sortie de l'été et une remontée soutenue à l'automne. Le premier semestre doit laisser entrevoir les années de "sécheresse".

L'utilisation de ces données est plus ambiguë. La capacité disponible dépend de la météo et peut en ce sens être une variable à prédire. D'un autre côté, l'ouverture des vannes pour la production de l'énergie est une action de l'Homme qui donne à cette donnée une dimension de variable explicative.

En fonction du modèle de prédiction choisi, l'une ou l'autre de ces notions devra être considérée.

Energie éolienne



Dans le graphique précédent et ci-contre, il est intéressant de constater que l'énergie éolienne n'est pas totalement aléatoire. Le vent semble être une composante dépendant de la présence ou non de courants froids sur le pays.

Ci-contre, on note également la présence de nombreux phénomènes exceptionnels (outliers) sur cette filière.

- Les données de cette production ont vocation à intégrer nos modèles de prédiction.

Energie solaire

L'énergie solaire est en constante progression. A noter qu'elle ne baisse pas avec le solstice d'été. On peut estimer que la durée d'ensoleillement n'est pas le seul facteur influent. L'inclinaison des rayons du soleil ou la couverture nuageuse ont certainement un rôle à jouer.

- Les données de cette production ont vocation à intégrer nos modèles de prédiction.

Bioénergies

Après avoir fortement progressé, les bioénergies semblent avoir atteint une limite sur les 5 dernières itérations. Par ailleurs, les bioénergies paraissent "négligeables" et constantes à l'échelle des autres filières de production au niveau national.

- Pour ces raisons, cette variable n'a pas vocation à intégrer nos modèles de prédiction.

Table de classification des types de production

Les modèles sélectionnés ultérieurement devront établir des prédictions. Afin d'être les plus pertinentes possibles, ces prédictions devront prendre en données d'entrée, les données propres aux types d'énergies "non maîtrisables" et aux valeurs "significatives" dans le mix énergétique. Les visualisations précédentes permettent d'élaborer la classification suivante des types d'énergie :

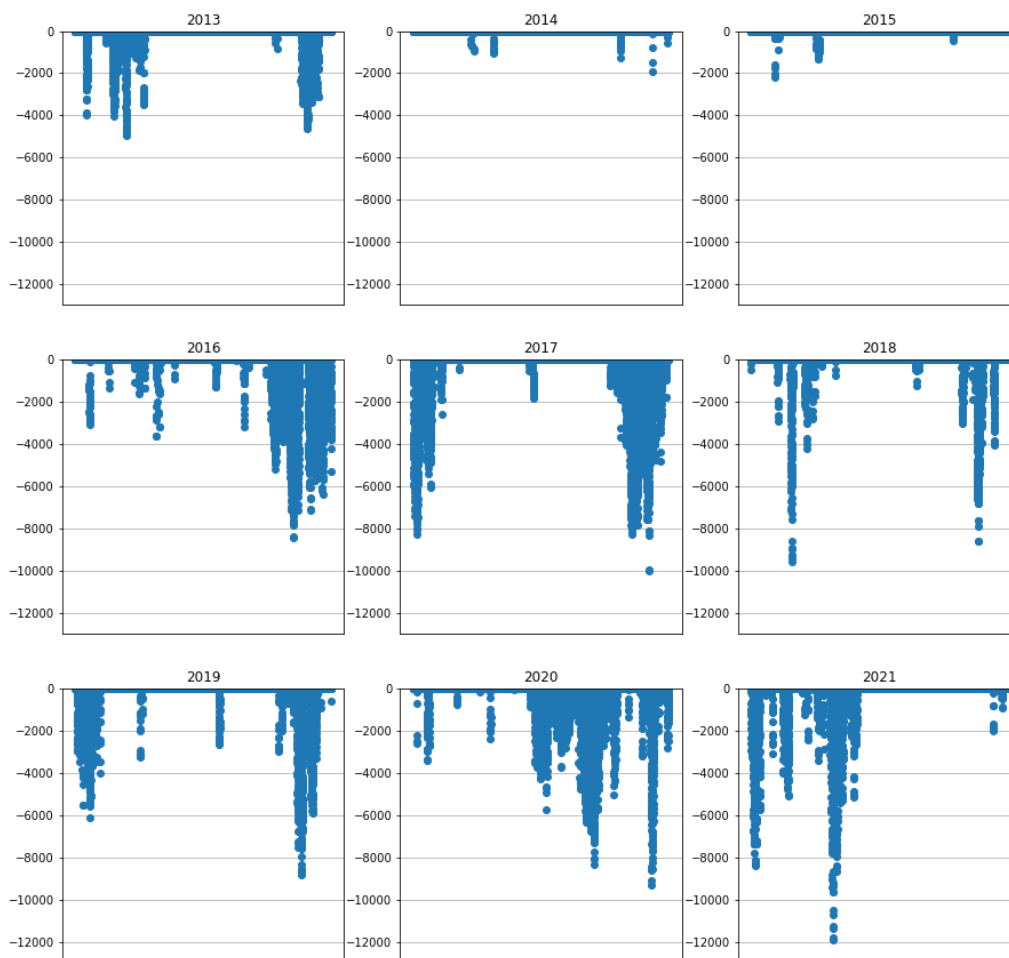
Type d'énergie	Maîtrisable	Valeurs significatives	Candidat à la prédiction
Nucléaire	oui	oui	non
Thermique	oui	oui	non
Hydraulique	oui/non	oui	oui
Eolien	non	oui	oui
Solaire	non	oui	oui
Bioénergie	non	non	non

Black-out

Définition et notion de balance

Le Black-out sera donc une notion nationale compensée par des rééquilibrages internationaux. Dans ce cadre, on opte donc pour une disparition des données régionalisées, privilégiant une vue France entière. Ainsi, la recherche étant pour le moment focalisée sur la différence entre Production et Consommation, on fera appel à la Balance entre ces 2 valeurs nationales pour chacun des relevés. Cette Balance calculée dans le paragraphe [Ajout/calcul de datas](#) sera au centre de notre démarche de prédiction. Dans ce cadre, on pourra parler indistinctement de Black-out et de Balance négative.

Visualisation des Black-outs au niveau national



● = puissance fournie par la production < puissance demandée par la consommation

Ces graphiques confirment l'autonomie énergétique passée de la France. En effet, depuis l'hiver 2016, les périodes avec une consommation supérieure à la capacité de production se multiplient.

Pour compenser ces déficits, le recours à l'énergie importée est nécessaire.

Depuis mi-2020, on observe une dépendance plus forte et plus régulière tout au long de l'année. Cela est vraisemblablement lié à la baisse des capacités thermiques et nucléaires déjà évoquée.

Méthodologie

Approches pour prédire la balance

Prédiction directe

Nous disposons de données initiales comprenant la consommation et les productions. De ce dataset initial nettoyé, a été déduit une Balance (production - consommation). Déclinée sur un pas de 3 heures, journalier et hebdomadaire, la Balance peut être directement utilisée dans un modèle comme données d'entrée pour la prédiction.

Prédiction indirecte

Contrairement à la “prédiction directe” précédemment décrite, la Balance calculée ne sera pas la donnée d'entrée du modèle. l'objectif de cette démarche alternative est de prédire les données individuelles de productions et de consommations puis d'en déduire la Balance consécutive. Ainsi, le dataset initial nettoyé sera découpé en autant de données d'entrée qu'il y a colonnes pertinentes de productions et de consommation.

Dans ce cadre, les premières visualisations du dataset ont permis d'élaborer une [Table de classification des types de production](#) et de leur pertinence en tant que données d'entrée dans les modèles de prédictions.

Premiers essais de modélisation

Les premières modélisations de séries temporelles ont été menées sur diverses agrégations de temps aussi bien en puissance (MW) qu'en capacité (MW/h), à l'échelle régionale et nationale. Pour parvenir à obtenir un résultat, nous devons parfois fortement lisser le signal (représentation graphique des données d'entrée) en agrégeant les données sur 1 mois pour pouvoir faire des prédictions mensuelles.

Les résultats obtenus sur une agrégation mensuelle ne permettent malheureusement pas de réaliser l'objectif que nous nous étions fixé, nous avons cherché des solutions pour pouvoir aider les modèles à converger avec un pas temporel inférieur à la semaine.

Aussi, cette recherche approfondie nous a permis d'identifier des pistes d'amélioration potentielles.

Modèles utilisés

SARIMAX

$$y_t^* = \Delta^d y_t$$

$$y_t^* = \mu + \underbrace{\sum_{i=1}^p \phi_i y_{t-i}^*}_{\text{AR}} + \underbrace{\sum_{i=1}^q \theta_i \epsilon_{t-i}}_{\text{MA}} + \epsilon_t$$

Les séries temporelles peuvent être prédites en utilisant un modèle SARIMAX à la condition que ladite série soit stationnaire. Si ce n'est le cas, un processus de (d) différenciation(s) (première équation) permet de l'obtenir par retranchement des éléments les uns avec les autres.

AR : les processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents (p), plus un terme aléatoire d'erreur.

MA : les processus de moyenne mobile supposent que chaque point est fonction des erreurs entachant les points précédents (q), plus sa propre erreur.

L'une des limites du modèle SARIMAX à laquelle nous nous sommes rapidement trouvés confrontés est le nombre de saisonnalités que l'on peut paramétrer : une seule alors que la consommation par exemple, présente plusieurs saisonnalités très marquées : jour, semaine, année.

Régression linéaire

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Nous avons pu vérifier que la consommation ou la production d'énergie avait une forte corrélation avec les phénomènes météorologiques.

Aussi, il semble opportun de s'appuyer sur ces liaisons pour modéliser les prédictions sur les énergies.

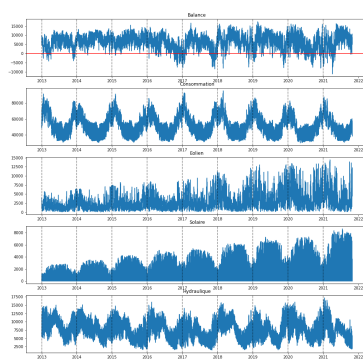
Aussi, naturellement, la régression linéaire est une alternative qu'il nous faut explorer dans l'optique d'identifier la meilleure prédiction possible.

En ce sens, il sera nécessaire de faire une recherche approfondie des variables météorologiques explicatives ayant un impact significatif sur les variables à expliquer.

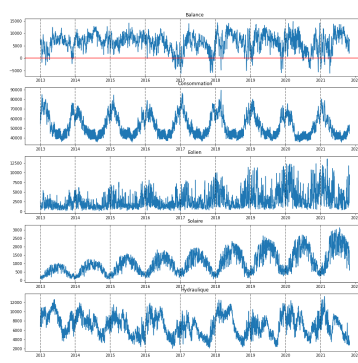
Il sera parfois nécessaire de faire preuve de créativité pour construire ces "features" étant donné l'évolution et le développement observés pour les énergies renouvelables notamment.

Pas temporels testés

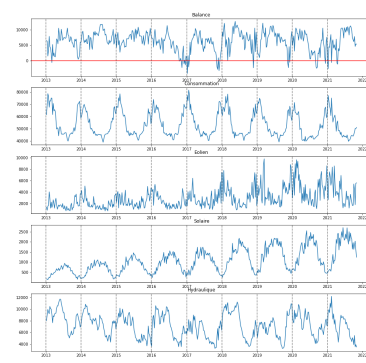
Deux approches, décrites dans le paragraphe [Approches pour prédire la balance](#), permettent d'envisager la prédiction. Une approche "directe" qui se base sur la donnée calculée de la balance, et une approche "indirecte" qui se base sur les données de consommation, de production éolienne, solaire et hydraulique, comme indiqué dans le paragraphe [Grille de classification des types de production](#). Ces deux approches se baseront sur 3 pas temporels construits en amont et présentés ci-après.



Pas temporel : 3 h



Pas temporel : 24 h



Pas temporel : 7 jours

La modification du pas temporel permet de lisser les données et donc de s'affranchir d'une partie du bruit que nous avons dans le signal initial.

Le pas minimum utilisé n'est pas celui du dataset de l'énergie (30 minutes). En effet, les données météorologiques en libre accès et disposant de suffisamment de variables significatives ont un pas temporel minimum de 3 heures. C'est donc le pas minimum considéré pour l'étude.

Le besoin d'un pas temporel suffisamment fin pour être pertinent a exclu le pas temporel de 1 mois ayant pourtant des résultats respectables.

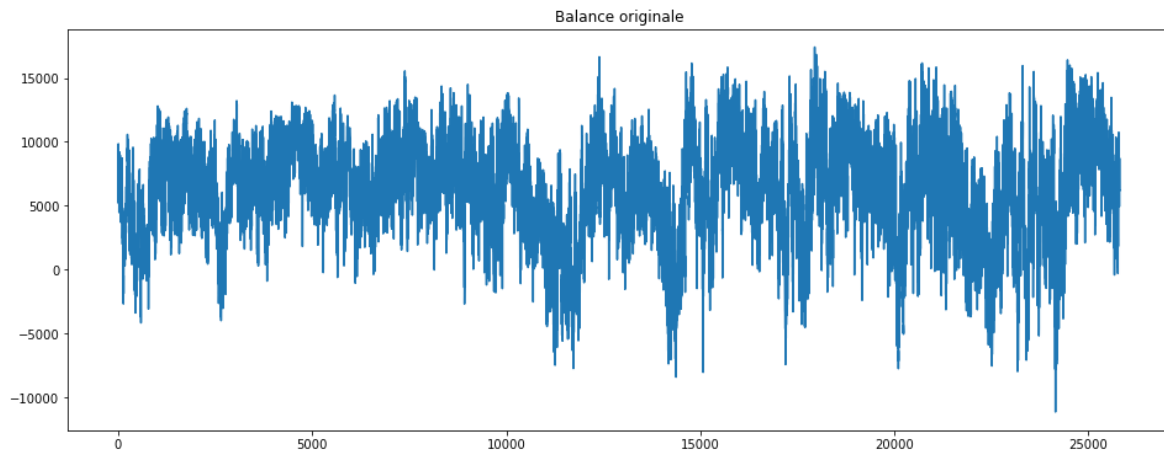
Les pas temporels de 3 heures, 1 jour et 1 semaine ont été retenus.

Etat des signaux donnés aux modèles

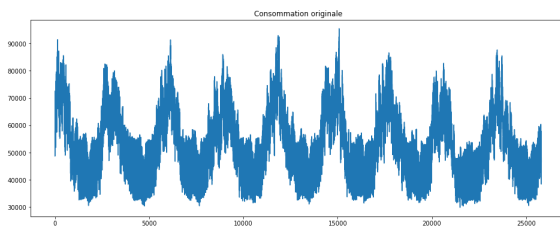
Signaux originaux

Les signaux dont nous disposons après nettoyage nous permettent de partir d'une base saine de données s'étirant sur la période de temps allant du 1^{er} janvier 2013 au 31 octobre 2021.

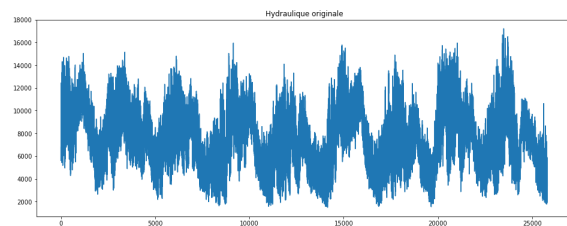
Les deux approches de prédictions directes ou indirectes se basent donc sur les spectres suivants :



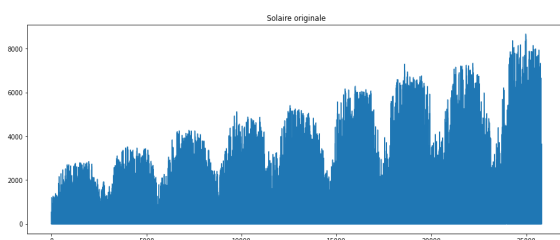
Signal original de la Balance (production - consommation)



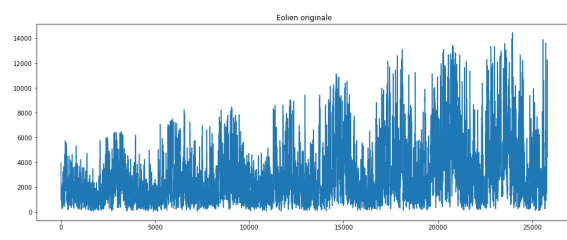
Signal original de la Consommation



Signal original de la production Hydraulique



Signal original de la production Solaire



Signal original de la production Éolienne

Signaux filtrés par transformation de Fourier

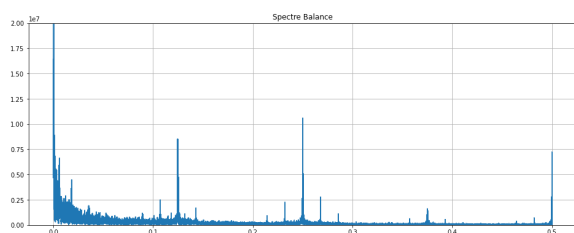
Les signaux agrégés sur un pas temporel de 3 heures et de 1 jour présentent une complexité que les modèles ont des difficultés à considérer. Plus que le sur-apprentissage, on constate que les modèles, SARIMAX notamment, ne parviennent pas à décomposer et exploiter correctement le signal.

Aussi, une deuxième approche permettant de filtrer le signal et donc de tenter d'améliorer la prédiction de nos modèles est la transformation de Fourier²⁰.

Cette méthode, que l'on peut considérer comme une opération de réduction de dimension au sens du Machine Learning, a l'avantage de lisser la représentation graphique des données, tout en conservant le pas temporel.

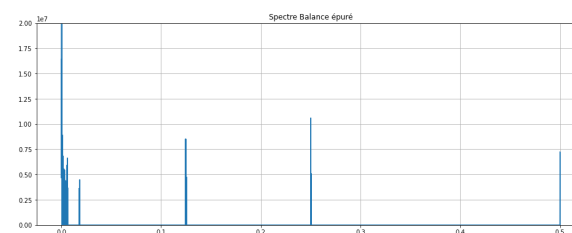
C'est une démarche de décomposition du signal en une somme d'une multitude d'harmoniques de type sinusoïdal de différentes fréquences et amplitudes. Ce traitement est fréquemment utilisé sur les ondes électromagnétiques ou le son. Il est entendu que tout signal peut ainsi être décomposé.

étape 1 : décomposition spectrale du signal



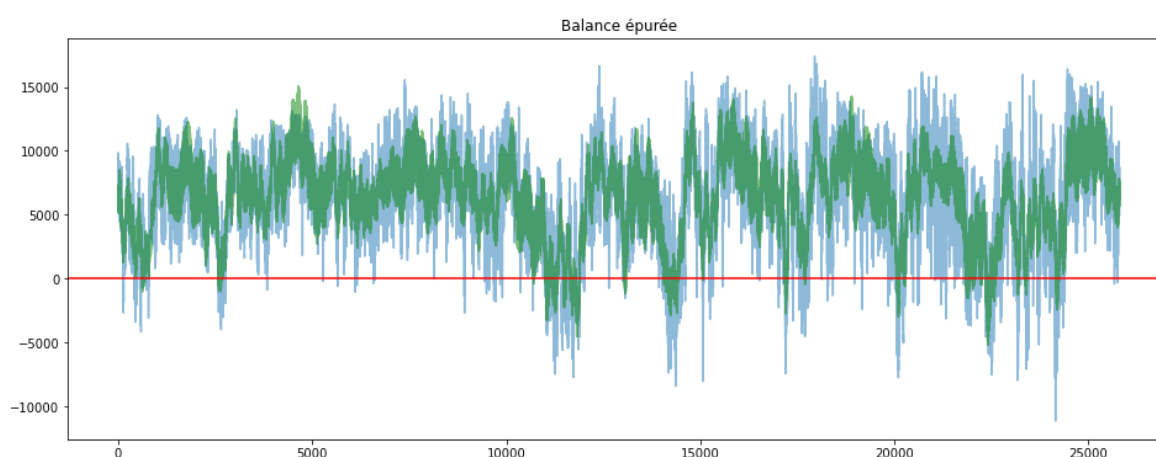
Le spectre fait ressortir les composantes du signal avec l'amplitude correspondante. La Transformée de Fourier met donc en évidence le "bruit" noyé dans le signal.

étape 2 : Élimination du bruit



On supprime du spectre les artefacts dus à l'algorithme de FFT, qui produit lui aussi des parasites. Ainsi, on se focalise sur les fréquences composites du signal.

étape 3 : Reconstitution du signal



Le signal filtré est représenté par la courbe verte et comparé au signal original (bleu). Sur cet exemple de la balance, le signal filtré permet de conserver les black-out principaux, élément que

²⁰ Retrouvez les détails de la mise en oeuvre de la transformée de fourier dans l'annexe Transformée de Fourier

nous allons chercher à prédire. En revanche, nous noterons que le "signal" représentant la série temporelle une fois traité présente une perte d'amplitude parfois significative.

Cette démarche aboutissant à des données en entrée aidant les modèles a donc été déclinée sur l'ensemble des inputs²¹ permettant d'emprunter les deux chemins de prédictions directes et indirectes. Les données respectant un pas de 3 heures et journalier ont donc suivi ce processus en 3 étapes détaillées ci-dessus. À l'étude, nous remarquons que la déclinaison du processus sur les données à la semaine n'apporte pas de gain notable tant les bruits y sont alors minimisés par construction même.

²¹ Ensemble des datasets traités par transformés de Fourier consultable sur [github.com](#) : [dataset \(3h\)](#), [dataset \(Jour\)](#), [dataset exogène \(3h\)](#), [dataset exogène \(jour\)](#)

Variables explicatives ajoutées aux modèles

Sinusoïdales ayant des périodes correspondantes aux saisons

Cette méthode est la première qui nous a permis de trouver une solution à l'échec de nos premières tentatives de modélisation. Nous avons 2 interrogations :

- le modèle SARIMAX n'aboutit pas avec un pas temporel de 3h ou 1 jour
- le modèle SARIMAX ne permet pas de paramétrer plusieurs saisonnalités

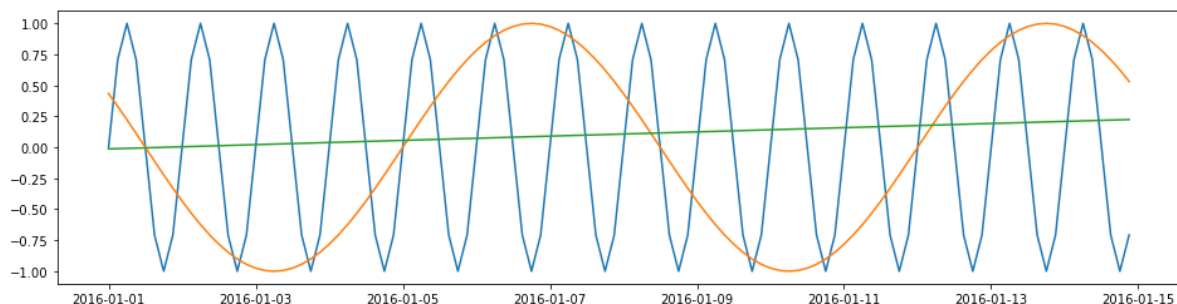
La documentation statsmodel du modèle SARIMAX propose un paramètre `exog` qui permet de donner des variables explicatives en entrée.

En première approche, nous avons trouvé qu'il était possible de construire des sinusoïdales de période correspondant à la saisonnalité.

L'utilisation de ce paramètre `exog` a été la clé pour obtenir un résultat de prédiction.

Comme pour la transformation de Fourier, on s'appuie sur le concept que tout signal suffisamment continu et périodique peut se décomposer en somme de courbes sinusoïdales.

A la différence de la transformation de Fourier qui altère le signal de l'étude, le paramètre `exog` permet de fournir les éléments permettant de décomposer le signal sur lequel s'appuie le signal de l'étude, sans modification.



3 courbes sinusoïdales correspondant au saisonnalité : jour, semaine, année

Données métiers

L'utilisation du paramètre `exog` pour aider le modèle à obtenir une prédiction, nous avons vu l'opportunité de construire des données métiers plus précises que les courbes sinusoïdales avec l'espoir d'obtenir de bons résultats.

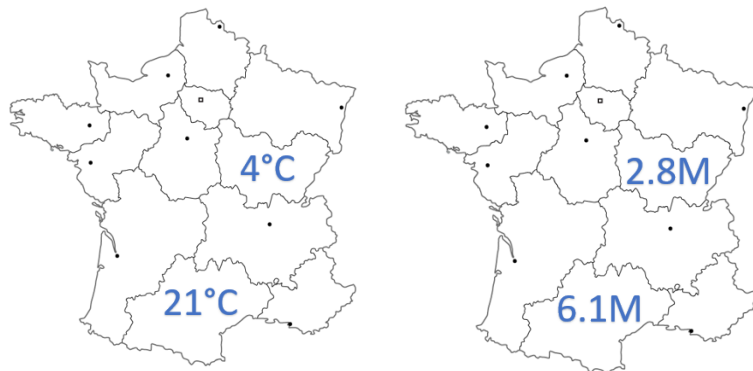
La construction de ces variables explicatives a été un très long travail de documentation, d'analyse, de réflexion et de construction.

*Principe de construction***Construction d'une variable nationale**

L'observation montre que la consommation est fonction de:

- température
- région

Nous avons donc pondéré la température par des données de population par région pour simuler une notion de "chauffage".



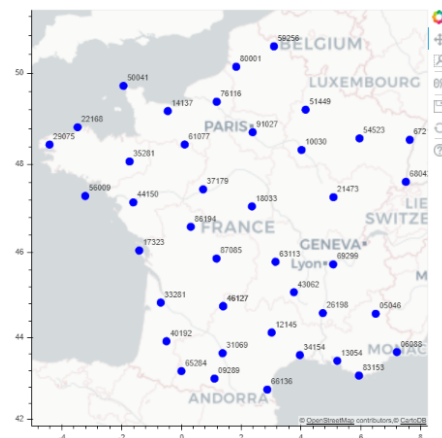
Chauffage = Consommation / Population
Avec 4°C, les 2.8 Millions d'habitants de Bourgogne Franche Comté devront plus chauffer leur habitation que les 6.1 Millions d'habitants d'Occitanie à 21°C.

Modération des données généralisées

Les éoliennes ne sont pas distribuées de manière homogène sur l'ensemble du territoire métropolitain. Par contre, les stations météorologiques le sont plus comme on peut l'observer ci-contre.

Ainsi, croiser les données de production éoliennes avec les données météorologiques sans explorer la proximité géographique serait une erreur générant des informations imprécises.

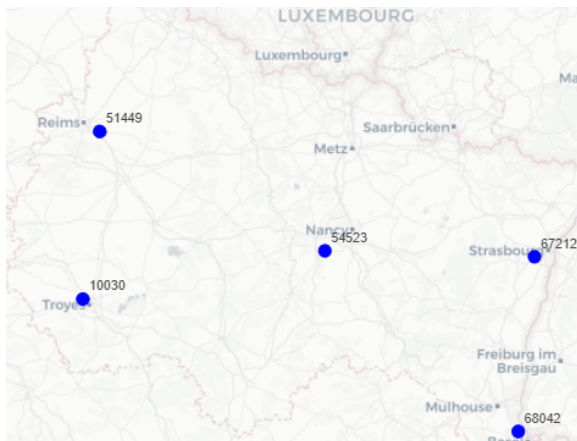
La localisation est donc importante à considérer pour réduire l'effet de généralisation des données.



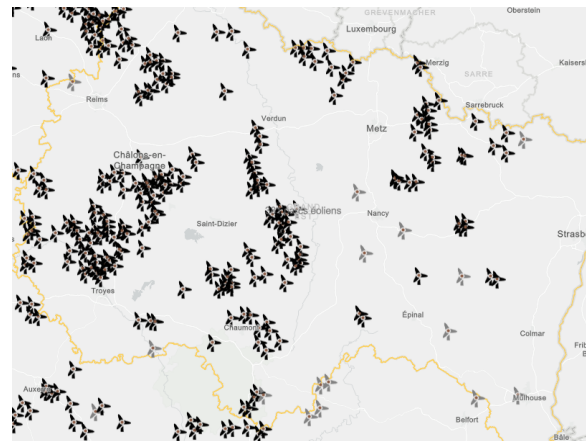
40 stations météorologiques métropolitaines²²

Dans ce cadre, par exemple, pour la région Grand Est, les 2 stations météorologiques (67212 & 68042) se trouvant à l'Est ont été exclues car on n'y trouve pas d'éoliennes :

²² Dataset du domaine public contenant les Longitude et Latitude : opendatasoft.com - Météo

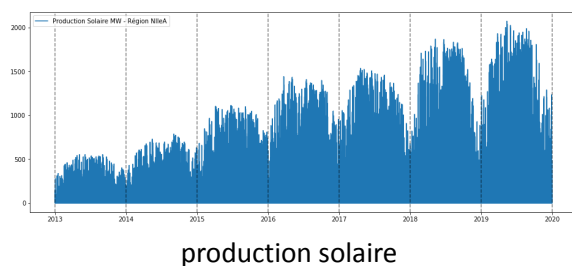


Stations météorologiques

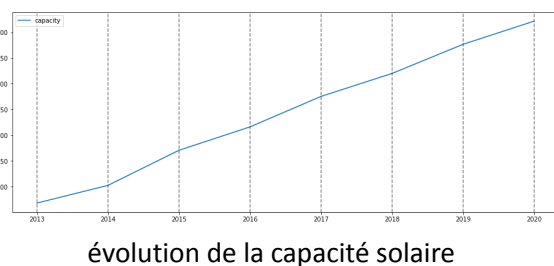
Eoliennes²³

Annulation de l'effet développement

L'affichage de l'évolution de la production d'énergie solaire montre un minimum constant et une amplitude qui augmente. C'est le fait de la capacité de production qui se développe au fil des années sur cette filière d'énergie renouvelable.

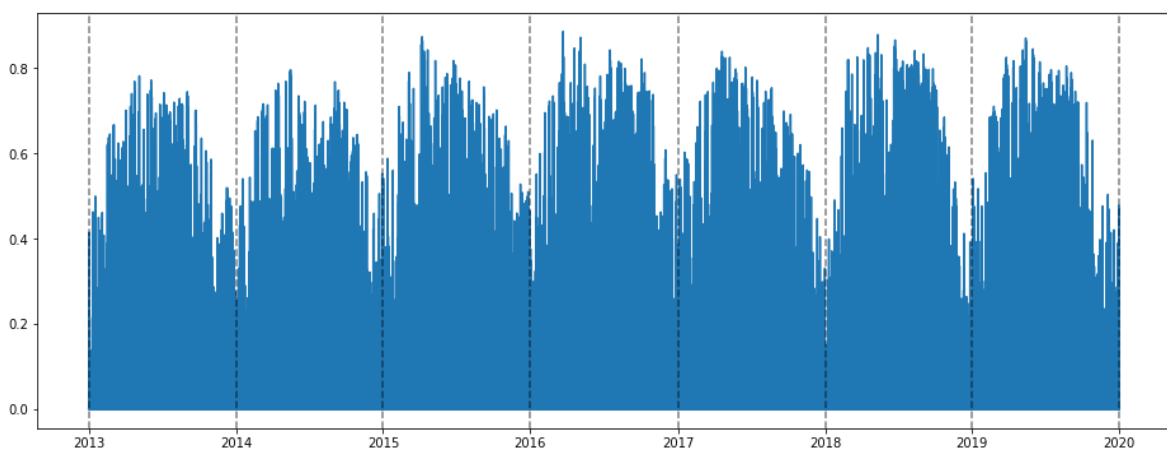


production solaire



évolution de la capacité solaire

Afin d'annuler l'effet du développement des énergies renouvelables et de stabiliser les données dans le temps, pour chaque filière, nous avons donc introduit la notion de taux de charge, en divisant la production par la capacité à chaque instant²⁴. On obtient le taux de charge de la production solaire :

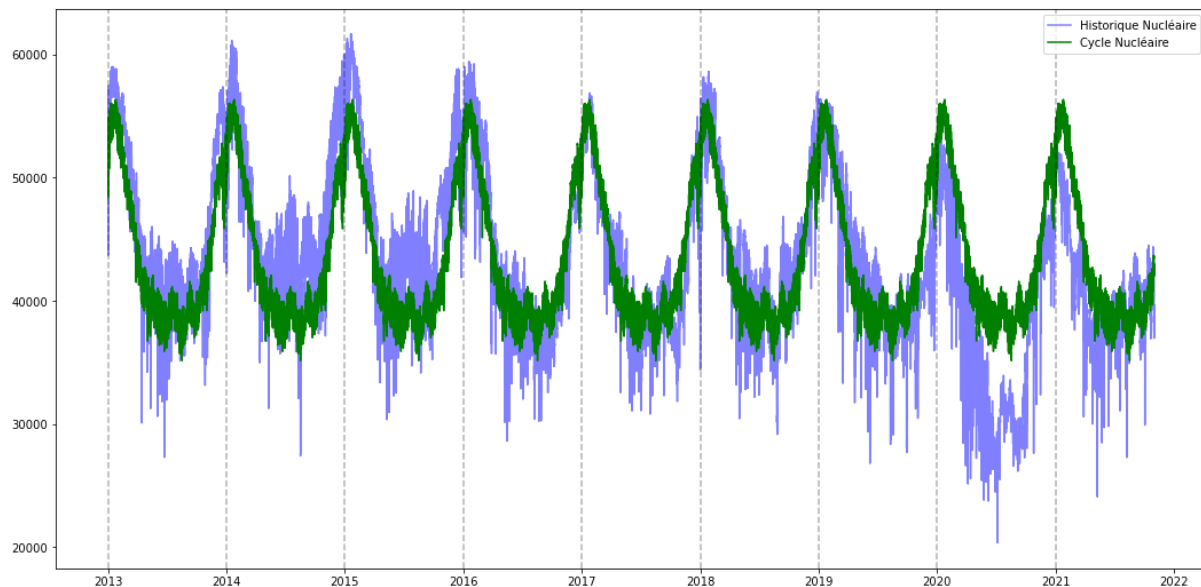


²³ <https://mapthenews.maps.arcgis.com> donnant la position des éoliennes sur le territoire français

²⁴ Capacités installées pour l'ensemble des filières d'énergie : Wikipédia

Variables explicatives utilisées pour la prédiction directe

Les données métiers en régresseurs exogènes utilisées pour la prédiction directe de la balance seront les Taux de CHarge (TCH) pour les productions considérées comme “pilotes” : **nucléaire, thermique et hydraulique**. Pour ces productions, nous avons vérifié l’hypothèse d’un cycle répétable d’une année sur l’autre, à savoir, une sollicitation forte des installations pendant l’hiver et des périodes de non disponibilité pendant l’été, permettant notamment la planification des opérations de maintenance. Ces patterns ont été construits suivant la moyenne de ces productions sur le dataset.



De même, il est considéré une hypothèse de stabilité de la production solaire qui pourra également jouer un rôle de variable explicative.

Enfin, on considérera également que la consommation se répète de manière habituelle d’une année sur l’autre, élément que nous avons également pu constater en analysant les signaux.

Dit de manière différente, on exclut des variables explicatives, les variables liées aux énergies renouvelables instables.

Variable explicative utilisée pour la prédiction indirecte

Ainsi, suivant les différentes méthodes de construction présentées, nous avons entrepris la construction d’un dataset de variables explicatives à partir de données météorologiques²⁵.

Au cours de la reconstruction des données météorologiques au format de notre base de données énergétiques au niveau France, nous avons écarté, sur la période correspondante aux [Dats de base](#), les variables inconsistantes. Typiquement, la qualité des données de nébulosité et solaire librement à disposition ont été écartées.

De même, nous avons écarté les données métiers ne présentant pas une corrélation significative avec les variables cibles à prédire.

²⁵ Dataset du domaine public “AROME” : opendatasoft.com - [Météo](#)

Les données métier en régresseur exogène utilisée pour la prédiction sont les suivantes :

Variables cibles	Régresseurs exogènes
Consommation	Chauffage ²⁶
Production Éolienne	Vent (m/s) ²⁷ Humidité (%) Pluie (mm) ²⁸ Température (°C) Différence de pression (Pa)
Production Solaire	Durée d'ensoleillement ²⁹
Production Hydraulique	Précipitation (mm)

Des jeux de données de variables explicatives, construites suivant les méthodes présentées en amont à partir de données météorologiques seront utilisés comme données exogènes de nos modèles de prédiction (github.com).

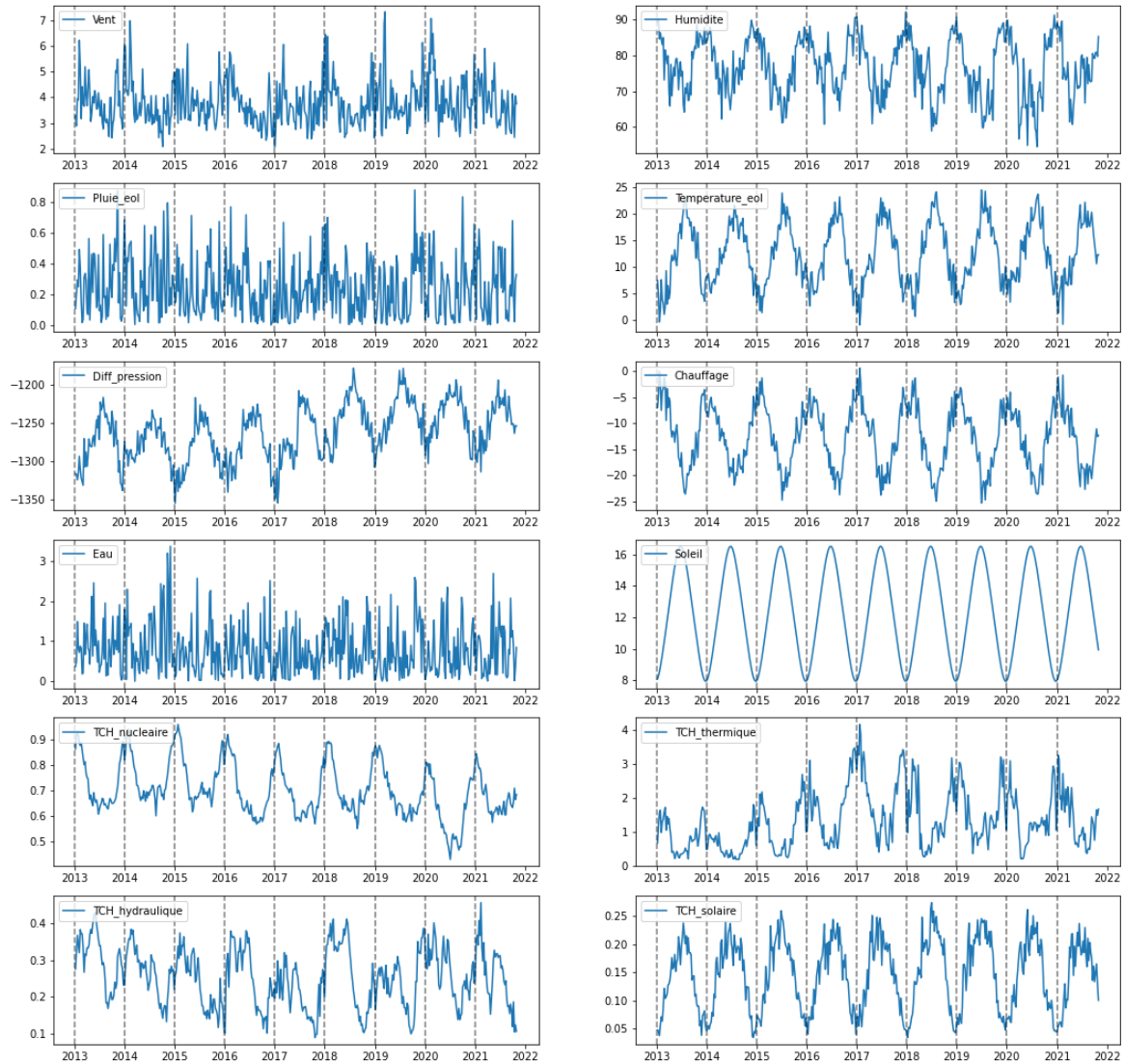
Ces données ont été considérées suivant leur échelle et après une méthode de standardisation afin de vérifier l'incidence sur la performance du modèle de prédiction.

²⁶ Description de la donnée dans le paragraphe [Variables explicatives ajoutées aux modèles](#)

²⁷ wikipedia.org

²⁸ wikipedia.org

²⁹ wikipedia.org



Visualisations des datasets à la semaine de données métiers
(TCH : Taux de Charge)

Résultats et interprétations

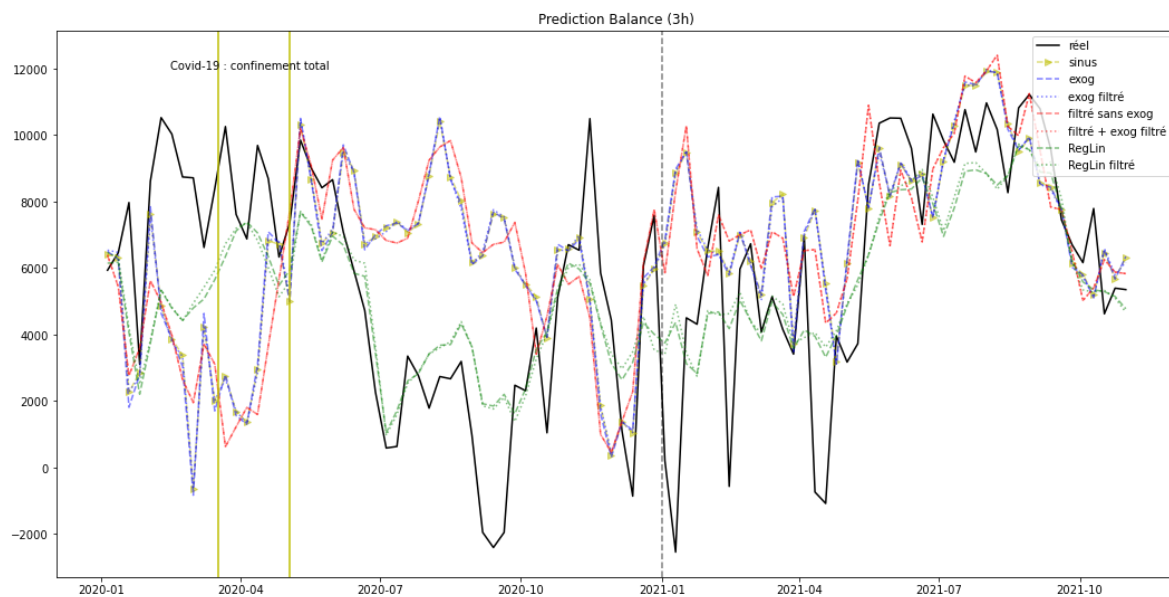
Rappel des modélisations effectuées

Les **modèles** ayant été appliqués à chacune de ces données ayant vocation à être prédites sont donc :

	Données exogènes			
Prédiction de la Balance	Aucune	Sinusoïde	donnée métier	donnée métier filtrée par Fourier
3h				
Directe	Echec	• Sarimax	• Sarimax	• Sarimax
Directe sur signal transformé par Fourier	• Sarimax	Redondant avec la TF	• Linéaire	• Sarimax • Linéaire
Indirecte	Echec	• Sarimax	• Sarimax	• Sarimax
Indirecte sur signal transformé par Fourier	• Sarimax	Redondant avec la TF	• Linéaire	• Sarimax • Linéaire
Jour				
Directe	Echec	• Sarimax	• Sarimax	• Sarimax
Directe sur signal transformé par Fourier	• Sarimax	Redondant avec la TF	• Linéaire	• Sarimax • Linéaire
Indirecte	Echec	• Sarimax	• Sarimax	• Sarimax
Indirecte sur signal transformé par Fourier	• Sarimax	Redondant avec la TF	• Linéaire	• Sarimax • Linéaire
Semaine				
Directe	• Sarimax	Signal d'entrée suffisamment lissé	• Sarimax • Linéaire	Echec
Indirecte	• Sarimax	Signal d'entrée suffisamment lissé	• Sarimax • Linéaire	Echec

Résultat de modélisation : balance directe

La compilation des résultats des prédictions élaborées comparées au réel constaté, permet d'identifier les modèles et données exogènes les plus pertinentes.



Prédiction directe de la balance sur un pas temporel de 3 heures

Evaluation des modèles

L'erreur calculée est la MAE, Mean Absolute Error qui est souvent utilisée pour l'évaluation des séries temporelles. Elle permet de mesurer l'écart en valeur absolue avec le signal à prédire. Elle permet donc une comparaison efficace entre les modèles.

MAE	original	original	original	original	signal TF	signal TF	R.Lin.	R.Lin. TF
	sinus	métier	métier TF	sans	sans	métier TF	métier	métier TF
3h	4109	4151	4105		3987	3982	2828	2833
Jour	3703	3803	3710		3766	3746	2379	2385
Semaine		3367		2999			1380	

Interprétation des résultats

De l'étude précédente concernant la modélisation de la prédiction directe, que l'on peut observer sur l'ensemble des modélisations que nous avons effectuées, on constate que :

1. Le modèle SARIMAX est insensible à la qualité des variables exogènes, le résultat avec des courbes sinusoïdales étant équivalent au résultat avec des données métiers liées.
2. Filtrer les données exogènes n'apporte pas d'amélioration, que ce soit pour le modèle SARIMAX ou la régression linéaire. Cependant, il n'y a pas de dégradation notable non plus, notamment pour la régression linéaire, ce qui signifie que cette réduction de dimension est pertinente et que l'information principale est conservée.
3. Filtrer le signal d'entrée à prédire n'apporte pas non plus d'amélioration significative. Le signal reste trop complexe pour une exploitation efficace par le modèle SARIMAX.
4. Lisser les données en augmentant le pas temporel est l'amélioration la plus significative de la performance des modèles.

Bilan de l'étude

Pour chaque composante à prédire de nos modèles, nous avons considéré les modèles les plus performants afin de construire notre résultat final.



MAE	Prédiction directe	Prédiction indirecte
3h	2828	5919
Jour	2379	5751
Semaine	1380	4738

La prédiction indirecte qui est une somme de prédiction n'est pas satisfaisante. Les modèles étant incapables de modéliser des signaux complexes, l'erreur cumulée devient très importante.

La prédiction directe permet une modélisation plus aboutie avec une performance accrue lorsqu'on lisse le signal d'entrée.

Conclusion

Le dataset disponible est une version gratuite ne donnant pas d'informations sur les DOM TOM et La Corse. Une extension géographique des données pourrait parfaire l'étude avantageusement.

Par ailleurs, les enregistrements de données météorologiques sur un pas de 3h n'était pas forcément synchronisés avec l'enregistrement du Dataset initial de consommation et productions disposant d'une acquisition toutes les 30mn. La finesse des données météorologiques disponibles étant inférieure à notre jeu de données nous a imposé de réduire le Dataset initial par lissage avec perte d'information.

De plus, lors de notre travail de pondération (chauffage, population, positions) nous n'avions pas les localisations exactes correspondant à nos points d'enregistrements.

Certainement, la géolocalisation de chaque enregistrement météorologique, calé sur les enregistrements des productions et consommations du réseau RTE permettrait une pondération plus exacte et un gain de précision des prédictions de nos modèles.

Ainsi, des données plus complètes et alignées les unes avec les autres, permettraient un gain notable de qualité de rendu final. L'optimisation de ces éléments nécessite, sans doute des investissements dans des sources plus fournies et un travail en amont dans la normalisation des process d'acquisition de données.

Concernant spécifiquement le modèle SARIMAX, l'étude montre que celui-ci n'est pas adapté à un signal complexe. Notamment, il ne prend en compte qu'une seule saisonnalité alors que les signaux de la réalité sont constitués d'une multitude de composantes (fréquences, amplitudes, tendance, etc). Le modèle semble donc incapable d'offrir une réponse satisfaisante, y compris lorsque on lui attribue des régresseurs exogènes.

De son côté, la régression linéaire semble en revanche prometteuse mais à la condition que l'on puisse construire des variables explicatives de qualité. Cela passe notamment par la localisation des enregistrements météorologiques à l'endroit des sites de production. Le revers de cette modélisation est qu'elle s'appuie sur des variables explicatives qui doivent elles-mêmes être prédites.

En définitive, le travail entrepris ici n'offre pas de résultat satisfaisant et représente de fait une première étape avant la recherche et la mise en œuvre d'autres modèles comme le Deep Learning.

Les travaux menés avant la mise en œuvre possible de modèles de Deep Learning permettent malgré tout de cadrer les conclusions.

1. La prédiction à quelques heures voire quelques jours est envisageable par régression linéaire avec l'amélioration des données en entrée.
2. La prédiction à une semaine voire quelques mois peut être rendue difficile du fait des prévisions météorologiques elles-mêmes perdant en pertinence en projection temporelle.
3. La prédiction sur plusieurs années peut être envisageable si elle est considérée du point de vue de la tendance et non de l'exactitude ponctuelle de la donnée prédite car nous avons pu constater la répétabilité des variables et avons su identifier les paramètres influents.

Ces cadrages de prédiction permettent de dégager en conclusion de l'étude, 2 grands axes à disposition des décideurs :

Pilotage

Le pilotage de l'offre d'énergie à court terme est envisageable après amélioration de la qualité des données d'entrées. Cette situation permet de bénéficier d'une maîtrise à court terme et donc d'une capacité d'anticipation d'événements prochains. Le black-out des heures prochaines, hors accident, est donc à portée.

Par contre, le pilotage permettant la planification sur plusieurs mois n'est aujourd'hui, avec les modèles mis en œuvre dans cette étude et avec les données à disposition, pas accessible. "L'engineering maintenance" des moyens de production n'apparaît pas comme maîtrisable avec les modèles et données ici utilisées. Outre l'évolution concernant la qualité des données, une poursuite de la démarche vers le deep Learning apparaît pertinente.

Stratégie

Si les prédictions élaborées dans l'étude montrent une distance notable avec le réel, on remarque que la tendance est respectée.

Dans ce cadre, les décideurs disposent ici d'un moyen complémentaire permettant de statuer et d'élaborer une stratégie sur le long terme. En effet, décider d'implanter un atelier de production d'énergie prend plusieurs années afin de réaliser les études de bases, le FEED (Front End Engineering Design), la construction, le commissioning (parfois très long dans le cadre d'atelier nucléaire).

Ainsi la planification de la gestion du parc peut s'appuyer sur la tendance prédite permettant de rendre éligible des projets de construction sur certaines filières novatrices tel que l'hydrogène mais surtout cela permettrait de décider des capacités futures à devoir installer.

Demain

Si la maîtrise de l'horizon proche n'est pas certaine, l'étude montre que le pilotage de demain et les choix pour le futur sont à portée de la main des datascientists.³⁰

³⁰ prendre [contact avec les auteurs](#)

Opportunités

Dans ce cadre, considérant aussi bien les *Multi Layer Perceptrons* que les réseaux de neurones convolutifs ou récurrents , il serait pertinent de vérifier jusqu'à combien de pas de temps les prédictions s'écartent des valeurs réelles.

Pour de premiers modèles nous proposons :

1. Le *Multi Layer Perceptrons*
 - Avec deux couches *Dense* de 32 et 64 neurones et une activation *relu*
 - Une couche de sortie avec un neurone
 - Compiler le modèle avec un *Adam optimizer*
2. Réseau de neurones convolutif (CNN)
 - Avec deux couches *Conv1D* de 16 et 32 neurones, une fonction d'activation *relu* .
 - Une couche de *MaxPooling1D* avec une taille de *pool* de 2
 - Une couche *Flatten*
 - Une couche de sortie avec un neurone
 - Compiler le modèle avec un *Adam optimizer*
3. Réseau de neurones récurrents (RNN)
 - Avec deux couches LSTM (Long Short Term Memory) de 32 et 64 neurones et une activation *relu*
 - Une couche de sortie avec un neurone
 - Compiler le modèle avec un *Adam optimizer*

Annexes

Bulletins météorologiques Val de Loire

Publié le 27/12/2013 à 09h27³¹ : Pays de la Loire : pluie et vent partout : « Nous sommes le 27 décembre, le ciel sera très gris, très venteux et très pluvieux aujourd'hui, des températures douces, 11 à 13°, la pluie va tomber toute la journée, 20 à 30 mm, prévoyez des parapluies solides, 45 km/h de vent en moyenne, 85 en rafale, et nous fêtons Jean. »

Publié le 28/12/2013 à 09h24³² : Pays de la Loire : pluie, vent, grêle, le ciel reste très agité avant un dimanche plus agréable : « Nous sommes le 28 décembre 2013, le ciel sera très agité encore aujourd'hui, de la pluie, du vent en rafale jusqu'à 65 km/h, des averses de grêle cet après-midi, un temps à ne pas mettre le nez dehors, demain dimanche, retour des éclaircies et d'un temps calme, nous fêtons les Innocents. »

³¹ francetvinfo.fr/pays-de-la-loire - 27/12/2013

³² francetvinfo.fr/pays-de-la-loire - 28/12/2013

Echanges avec RTE

Afin d'éclaircir des éléments en notre possession et disposer de nouvelles données, nous avons échangé avec le contact RTE grand public.

2^{ème} échange

De : **CONTACT.RTE** <contact.rte@rte-france.com>

Date: ven. 17 déc. 2021 à 10:06

Subject: Votre demande du 15/12

To: cyril.bruyere@gmail.com <cyril.bruyere@gmail.com>

Bonjour,

Voici les éléments de réponse à vos deux questions :

1) Lorsque les données sont en MW, il s'agit effectivement des données en puissance instantanée. Vous pouvez trouver les données en énergie dans l'onglet [Données en énergie](#). Ces données sont téléchargeables.

Vous pouvez extrapoler des énergies à partir des puissances selon le mode que vous proposez : lorsque vous multipliez une puissance par un pas d'une ½ heure vous obtenez une estimation de l'énergie sur cette ½ heure. Vous considérez en faisant cela que la puissance a été constante pendant 30 minutes. Il ne faut pas multiplier par 1800 mais par 0,5 car 30 minutes c'est ½ heure et l'énergie que vous obtenez est en MWh. En théorie, pour calculer l'énergie à partir de la courbe des données en puissance, il vous faut calculer l'intégrale, c'est-à-dire l'aire sous la courbe. En faisant les calculs tels que vous proposez, vous aurez une estimation englobante.

2) Il existe dans les trois régions que vous mentionnez des centrales pouvant à l'occasion faire du pompage : Ste Croix du Verdon (PACA), usine marémotrice de La Rance (Bretagne), Vouglans (Bourgogne Franche-Comté).

Espérant avoir répondu à votre attente,

Bien cordialement,

L'équipe web RTE

RTE sur le web : [rte-france.com](https://www.rte-france.com) | [rte-et-vous.com](https://www.rte-et-vous.com)

Rejoignez le réseau ! [twitter](#) | [facebook](#) | [linkedin](#)

De : noreply@swi.cloud-rte-france.com [mailto:noreply@swi.cloud-rte-france.com]

Envoyé : mercredi 15 décembre 2021 20:30

À : CONTACT.RTE <contact.rte@rte-france.com>

Objet : Nouveau message : Formulaire Grand public

Un nouveau message vient d'être envoyé sur le site institutionnel, via le formulaire Grand Public (mer 15/12/2021 - 20:30)

Email : cyril.bruyere@gmail.com

Message : Bonjour, Nous avons 2 questions concernant les données mises à disposition sur le site par tranches de 30' pour l'ensemble des régions de France : 1- les données sont exprimées en MW. Sauf erreur, cela correspond à une puissance. Aussi comment devons-nous interpréter les données présentées pour 30'. Pour avoir l'énergie consommée, faut-il multiplier la puissance donnée par 1800 = (60 secondes x 30 minutes) ? 2- Dans le fichier des données de 2013 à 2021, on observe des données de pompage dans des régions où il n'y a pas de STEP : Bourgogne - Franche - Comté, Bretagne et PACA. Dans ces régions, à quoi correspondent les données de pompage SVP ? D'avance, merci. Bonne journée, Cyril

1^{er} échange

De : **CONTACT.RTE** <contact.rte@rte-france.com>

Date: ven. 10 déc. 2021 à 15:52

Subject: TR: Nouveau message : Formulaire Grand public

To: cyril.bruyere@gmail.com <cyril.bruyere@gmail.com>

Bonjour

1) Les données alimentant le diagramme Sandkey sont disponibles en téléchargement sur le site web d'ÉCO₂mix ([téléchargez ici](#)). Il faut télécharger le fichier des données annuelles en énergie.

2) Pour les données en température :

a. Si vous évoquez le graphe de comparaison des pics journaliers de la consommation à la température moyenne, nous utilisons un jeu de données disponible dans l'open data réseaux énergies ([lien ici](#)). Dans ce cas-là, les températures proviennent du partenaire de l'open data réseaux énergies (Weathernews).

b. Si votre question porte sur quelles données sont utilisées pour les prévisions de consommation alors nos modèles exploitent des données fournies par Météo France.

Bien cordialement,

L'équipe web RTE

RTE sur le web : [rte-france.com](https://www.rte-france.com) | [rte-et-vous.com](https://www.rte-et-vous.com)

Rejoignez le réseau ! [twitter](#) | [facebook](#) | [linkedin](#)

De : noreply@swi.cloud-rte-france.com [mailto:noreply@swi.cloud-rte-france.com]

Envoyé : lundi 6 décembre 2021 22:12

À : CONTACT.RTE <contact.rte@rte-france.com>

Objet : Nouveau message : Formulaire Grand public

Un nouveau message vient d'être envoyé sur le site institutionnel, via le formulaire Grand Public (lun 06/12/2021 - 22:12)

Email : cyril.bruyere@gmail.com

Message : Bonjour, Serait-il possible de disposer SVP des 2 informations suivantes : 1- répartition des données de consommation que l'on trouve dans le diagramme Sankey (consommation des particuliers, PME/PMI, etc) 2- quelle est la source que vous prenez en référence pour les données météorologiques D'avance, merci. Et félicitations pour ce site qui est particulièrement bien réussi. Cyril

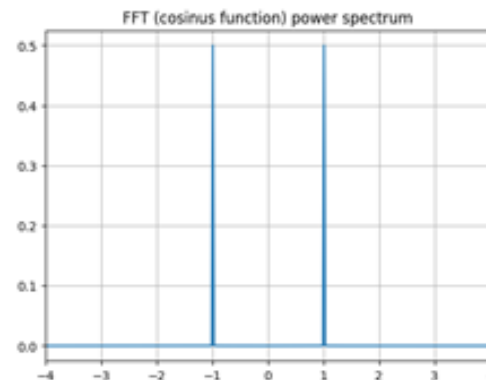
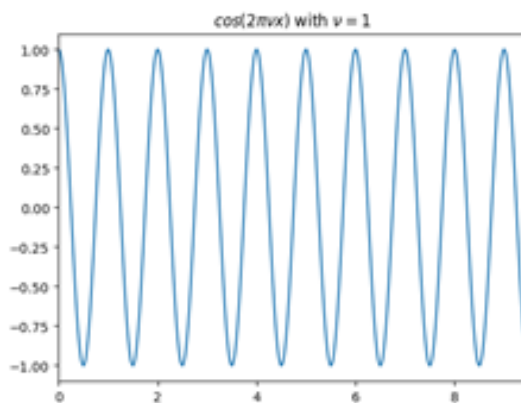
Transformée de Fourier

La transformée de Fourier est une [opération](#) qui permet de représenter en [fréquence](#) (développement sur une base d'exponentielles) des signaux qui ne sont pas périodiques. Il s'agit de l'analogue des séries de Fourier pour les fonctions périodiques (développement sur la base de fonctions sinusoïdales). Une fonction non périodique pouvant être considérée comme une fonction dont la période est infinie. Ce passage à la limite nous fait passer des séries aux intégrales.

$$TF[x(t)] = X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt$$

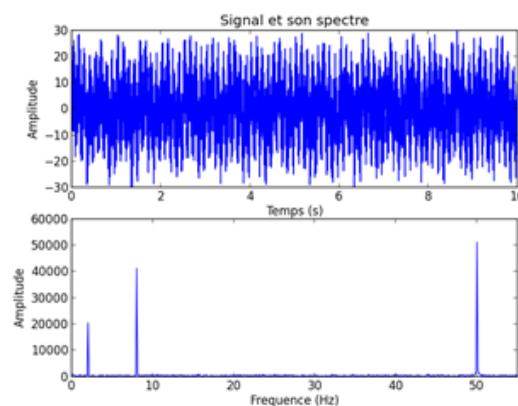
Caractériser un signal par son [spectre](#) de fréquence permet notamment de mettre en évidence l'importance de l'harmonique fondamental ainsi que la décroissance plus ou moins rapide de l'amplitude des harmoniques de rang plus élevé. Il peut aussi servir à déterminer le nombre d'harmoniques nécessaires pour transmettre la quasi-totalité de l'[énergie](#) du signal ([bande passante](#), coupe-bande...) ainsi que caractériser le bruit.

La transformée de Fourier, ou plus généralement l'analyse fréquentielle ou spectrale, est un outil fondamental pour la compréhension et la mise en œuvre de nombreuses techniques numériques de traitement des signaux et des images.



Sur le tracé de ce signal, il est pratiquement impossible de distinguer ses différentes composantes. L'amplitude varie aléatoirement. Mais le spectre fait très clairement ressortir les trois composantes à 2, 8 et 50 Hz, avec leur amplitude correspondante. La TFD met donc en évidence le bruit noyé dans le signal.

Sur le tracé du spectre, on notera de très petites fluctuations pour les fréquences autres que les trois fréquences composantes du signal. Leur amplitude semble aléatoire. Il s'agit d'artefacts dus à l'algorithme de FFT, qui produit lui aussi des parasites.



Le signal ainsi que sa décomposition spectrale