**1/ Introduction**

Le 28 février 2022, le Groupe d’experts intergouvernemental sur l’évolution du climat (GIEC) a publié un rapport “Climate Change 2022 - Impacts, Adaptation and Vulnerability”.

Ce rapport attire l’attention et alerte sur l’impact de nos sociétés et modes de vie sur la planète.

A l’intérieur de ce rapport, qui souligne le danger des effets du changement climatique, la question de la production et de l'approvisionnement en énergie apparaît comme un risque majeur.

**Avec la multiplication des catastrophes naturelles, l’accroissement de la population et des activités humaines toujours plus énergivores, quel est le niveau de pression que doit supporter le dispositif de production d’énergie en France? Existe-t-il un risque important de black-out?**

Inversement, le rapport a révélé les impacts négatifs de nos modèles de production d’énergie sur notre environnement.

En France, la loi n° 2015-922 relative à la transition écologique pour la croissance verte, en vigueur depuis 2015, a énoncé entre autres les objectifs suivants :

* Augmenter la part des énergies renouvelables dans notre consommation énergétique
* Réduire la part de nucléaire dans la production d’électricité

**Les évolutions des parts de la production d’énergies renouvelables et de nucléaire sont-elles alignées avec les objectifs de transition écologique? Sont-elles suffisamment marquées pour attester de la réelle prise en compte des objectifs définis ci-dessus?**

Au niveau de la France, la transition écologique vers un mix énergétique beaucoup plus décarboné est au cœur de la politique environnementale et représente un défi majeur aussi bien ambitieux que nécessaire.

**Aussi, la production d’énergies renouvelables, portée par la transition écologique française, pourra-t-elle soutenir le besoin énergétique français dans les années à venir?**

**2/ Présentation des données**

**A/ Données de consommation et production énergétique**

Le Dataset initial présente les données régionales consolidées définitives des consommations et de production d’énergie par filières de janvier 2013 à Novembre 2021. Les données sont obtenues à partir des comptages et complétées par des forfaits.

**Les variables de région et de date**

* Les colonnes Code INSEE région et Région fournissent des informations sur les régions concernées par les relevées énergétiques ;
* Les colonnes Date, Heure et Date - heure indiquent le moment où les relevés énergétiques ont été réalisés.

**La variable Nature**

La variable Nature correspond aux Données consolidées et Données définitives.

* Données consolidées : les données ont été vérifiées et complétées (depuis Janvier 2021 au 30 Novembre 2021) ;
* Données définitives : tous les partenaires ont transmis et vérifié l'ensemble des comptages (Janvier 2013 à Décembre 2020).

**La variable de consommation d'énergie**

* La colonne 'Consommation' indique la consommation d'énergie réalisée selon la région relevée à un moment précis. Le comptage se réalise chaque 1/2H.

**Les différentes variables de production d'énergie - mix énergétique**

Le comptage se réalise chaque 1/2H.

*Nucléaire (MW)*

Il s’agit de l’énergie produite par le parc nucléaire Français

*Thermique (MW)*

Il s’agit de la filière relative aux moyens de production d’énergie avec du gaz, fioul et énergies fossiles (charbon)

*Eolien (MW)*

Il s’agit de l’énergie produite par le parc éolien Français

*Solaire (MW)*

La filière solaire représente la puissance électrique produite par l’ensemble des panneaux photovoltaïques (parcs de production et productions diffuses).

*Hydraulique (MW)*

La filière Hydraulique comprend le fil de l’eau et l’éclusé, les lacs et les stations de transfert d’énergie pour pompage/turbinage (STEP turbinage).

*Bioénergies (MW)*

La filière Bioénergies comprend les biogaz, les biocombustibles solides (bois énergie et autres biocombustibles solides) et les déchets (ménagers et papetiers)

**TCO (Taux de Couverture) et TCH (Taux de Charge)**

*TCO (Taux de Couverture)*

Le Taux de Couverture est la part d'une filière de production d'énergie dans la consommation d'une région.

Il correspond à la production d'énergie / la consommation d'énergie d'une région.

*TCH (Taux de Charge)*

Le Taux de Charge est le volume de production de la filière de production d'énergie thermique par rapport à sa capacité de production installée et en service.

Il correspond au volume de la production d'énergie / la capacité de production d'énergie installée et en service.

**Pompage/STEP turbinage (MW)**

Il s’agit des installations hydroélectriques qui puisent aux heures creuses de l'eau dans un bassin inférieur afin de remplir une retenue en amont (lac d'altitude). L'eau est ensuite turbinée aux heures pleines pour produire de l’électricité.

Le comptage se réalise chaque 1/2H.

**Ech. physiques**

Il s'agit du solde des échanges avec les régions limitrophes.

Le comptage se réalise chaque 1/2H.

\*À noter que les variables 'Pompage' et 'Ech. physiques' peuvent prendre des valeurs négatives.

En effet, (1) une station de pompage peut consommer de l'électricité en heure creuse et (2) une région peut transférer plus d'énergie qu'elle en reçoit.

**B/ Données météorologiques (températures)**

Ce jeu de données présente les températures minimales, maximales et moyennes quotidiennes (en degré celsius), par région administrative française, du 1er janvier 2016 à Janvier 2022

**Les variables de région et de date**

* Les colonnes Code INSEE région et Région fournissent des informations sur les régions concernées par les relevées de températures ;
* La colonne Date indique la date à laquelle les relevés de températures ont été réalisés.

**Températures moyennes, minimales et maximales**

* Il s’agit des températures minimales, maximales et moyennes observées et obtenues à partir de l’ensemble des relevés de températures sur un jour donné

**3/ Visualisation des données**

**A/ Fixation du Dataset**

On note de nombreuses valeurs manquantes sur les relevés de façon générale.

Plus particulièrement, un important nombre de relevés de production nucléaire et de pompage sont absents.

Au regard du très faible nombre de valeurs manquantes pour les colonnes "Consommation", "Thermique", "Éolien", "Solaire", "Hydraulique", "Bioénergies" et "Ech. physiques", nous avons tout simplement supprimé les lignes concernées par les valeurs manquantes sur les colonnes concernées.

En ce qui concerne les colonnes "Nucléaire" et "Pompage", nous remarquons des régions avec des productions nulles ce qui peut être expliqué par l'absence d'unité de production nucléaire ou station de pompage dans les régions concernées.

De plus, pour les régions avec des valeurs totales de production non nulles, les valeurs détaillées sur les différentes années sont globalement du même ordre de grandeur et nous n’identifions pas de valeurs aberrantes.

Les valeurs manquantes ont alors été remplacées par 0 suite à l’analyse décrite ci-dessus.

Les colonnes TCO et TCH ont été mises de côté de par leur faible utilité dans le cadre de notre projet.

**B/ Echelle temporelle**

Le Dataset présente 1875348 lignes après retraitements des valeurs manquantes : en effet, nous avons ici les relevés énergétiques de production et consommation pour chacune des régions présentées chaque jour, chaque 30 minutes et sur une période de presque 9 ans.

Pour la visualisation, il nous a semblé peu pertinent de conserver les heures et jours de relevés dans la mesure où la périodicité des relevés(chaque 30 minutes) allaient être difficilement visualisables sur une période de 9 ans, réduisant le caractère interprétable des visualisations et leurs évolutions (notamment les courbes) sur ce même intervalle.

**Nous avons fait le choix de nous baser sur les années et mois en fonction de la nature du graphique:**

Les diagrammes en bar et les camemberts nous permettent de réaliser une comparaison de totaux de production/consommation entre différentes années. L'objectif étant de pouvoir identifier et caractériser les évolutions générales sur de longue période, il n'est pas ici pertinent de s'intéresser aux mois.

Pour les courbes, la définition des mois apporte un degré de précision qui permet d'apprécier les variations de consommations et de productions au sein de chaque année dans l'idée d'identifier des périodes à risque (production < consommation). On observe aussi une saisonnalité des variations avec un pic qui semble être atteint durant les périodes hivernales.

**C/ Analyse des graphiques**

**Phasage entre la consommation et la production énergétique au niveau national et au niveau départemental**

Au niveau national, nous remarquons que la production nationale toutes filières confondues couvre la consommation énergétique. Nous notons tout de même des périodes de tension où la courbe de la consommation reste très proche de la courbe de production.

On peut obtenir plus de détails au niveau régional où 3 groupes de régions se distinguent :

* les régions qui présentent un excédent de production, c'est-à-dire que leur niveau de production est très supérieur à leur niveau de consommation. On pourra voir dans la suite que la production d'énergie dans ces régions est majoritairement portée par la filière du nucléaire.
* les régions en auto-suffisance, c'est-à-dire que leur niveau de production d'énergie couvre leur consommation globale. On notera que des pics de consommation ou des faiblesses de production peuvent intervenir au fil des années. Le phasage énergétique de ces régions requiert un suivi déjà plus poussé afin de s'assurer du bon approvisionnement constant en énergie.
* les régions qui présentent un déficit de production, c'est-à-dire que leur niveau de production est inférieur à leur niveau de consommation et qui nécessitent des flux physiques interrégionaux. Le faible niveau de production énergétique est essentiellement soutenu par les énergies thermiques et hydrauliques.

On note que des mécanismes de compensation existent : les régions en excédent de production fournissent de l'énergie à d'autres régions. En effet, la valeur des échanges physiques est significativement négative pour les régions concernées (Grand-Est et Auvergne-Rhône-Alpes).

A l'inverse, les régions en déficit (Île-de-France par exemple) reçoivent d'importants volumes d'énergie afin de pallier la faible production régionale.

**Analyse par filière de production : énergie nucléaire / renouvelable**

Le nucléaire est la principale filière productrice d'énergie en France. Cependant, on note un essor timide des énergies renouvelables entre 2013 et 2021 au détriment des filières thermique et nucléaire.

L'augmentation de la part des énergies renouvelables dans le mix énergétique français se confirme au niveau régional à travers une tendance à la hausse de la production des énergies vertes. On note tout de même que le nucléaire reste la source d'énergie dominante par rapport aux énergies vertes et les régions avec un niveau de production d'énergie bas se démarquent par l'absence d'unité de production nucléaire.

Aussi, les énergies vertes restent aujourd'hui insuffisantes pour assurer l'indépendance énergétique des régions qui misent sur cette filière.

**Focus sur les énergies renouvelables**

Entre 2013 et 2021, les régions principales productrices d'énergie renouvelable restent les mêmes : Auvergne-Rhônes-Alpes, Occitanie, Nouvelle-Aquitaine, Provence-Alpes-Côte d'Azur et Grand Est. On peut noter un léger rééquilibrage de la part de production d'énergie verte au profit de plus petites régions productrices comme la Normandie, la Bourgogne-Franche-Comté ou encore le Centre-Val de Loire, l'augmentation la plus importante étant la région Hauts-de-France (+5.5 points).

En se basant sur les 6 plus grandes régions productrices identifiées précédemment, on note que l'énergie renouvelable est majoritairement hydraulique avec un faible volume de bioénergies. Les énergies solaires et éoliennes représentent un volume quasi négligeable et ne concernent que certaines régions (Grand Est et Nouvelle-Aquitaine).

**4/ Itération de modèles**

# **4a/ Série temporelle**

**A/ Fixation du Dataset**

Nous avons choisi d’étudier la consommation et nous avons converti la colonne date en tant

qu’index agrégé en minutes d’un pas 1⁄2 heure.

**B/ Echelle temporelle**

Dans cette section ,nous allons tout d’abord nous concentrer sur l’analyse de la consommation nationale puis ensuite choisir trois régions pour mesurer l’impact des régions déficitaires, excédentaires et équilibrées pour prévoir le risque de black-out.

**C/ Analyse et conclusion**

Nous avons utilisé les données mensuelles de consommation d’électricité en France, pour tester des modèles de prévisions de cette consommation à 12 mois.

L’exploration de ces données nous a révélé des caractéristiques telles que la non stationnarité de la série de consommation d’électricité et sa très forte inertie (autocorrélation forte et longue). Deux processus de différenciation ont été nécessaires pour rendre la série stationnaire.

Les prévisions de SARIMA sont globalement satisfaisantes. Le modèle SARIMA est bien adapté pour prédire la consommation d’électricité plus ponctuelle (à court terme). Il sera plus performant et plus robuste dans des prévisions devant prendre en compte de forts impacts saisonniers.

Le risque de black-out n’est pas à l'ordre du jour. Cependant, Il ne faut pas écarter ce risque pour trois raisons au moins:

* L’impact très fort de la fluctuation des 4 saisons de l’année et l’inertie de ce dernier;
* La croissance démographique et l’augmentation constante de la consommation des activités humaines (industrie etc.)
* La menace de changement climatique de plus en plus inquiétante et qui impacte nos habitudes.

Il serait également intéressant d’explorer d’autres modèles comme l’ajout des variables exogènes (températures) ou les modèles à mémoire longue de type ARFIMA.

# **4b/ Régression Linéaire**

**A/ Fixation du Dataset**

Pour la fixation du dataset, nous nous sommes appuyés sur le dataset fixé précédemment, auquel nous avons ajouté les données météo disponibles dans un second dataset.

Les 2 datasets ont été joints sur les colonnes de date et de région, ce qui a permis d’obtenir un nombre de 1244160 lignes, correspondant aux valeurs communes entre les 2 datasets.

Le différentiel de lignes correspond à la perte des données entre 2013 et 2015 et en 2022.

Nous avons converti toutes les heures en minutes de 00h à 23h30, par pas de demi heures, ceci afin d’obtenir des valeurs numériques permettant d'entraîner le modèle de régression linéaire.

Nous n’avons conservé que les mois car les entraînements avec des dates complètes ne permettaient pas d'obtenir de score concluant.

**B/ Echelle temporelle**

1/ A l’échelle nationale :

Les données ont été agrégées par mois et par heures parce qu'il s'agissait de l’association la plus adéquate afin d’obtenir un nombre de points suffisant et un score satisfaisant.

Les données de consommation ont été sommées et nous avons considéré la moyenne des données de température.

2/ A l’échelle régionale :

Le même raisonnement a été appliqué, et un filtre par régions a été ajouté selon les régions sélectionnées.

Nous nous sommes basés sur les mêmes groupes de régions que celles qui ont été présentées précédemment : Grand Est, Hauts de France, Ile de France.

**C/ Analyse et conclusion**

* Consommation en tant que variable cible

1/ A l’échelle nationale :

Après ajustement du modèle sur les données historiques, nous pouvons observer des scores satisfaisants sur les échantillons d'entraînement et de test.

L’analyse de la corrélation des variables explicatives (températures) avec la variable cible (consommation) laisse apparaître que la consommation est très fortement négativement corrélée aux variables de température, autrement dit, plus la température augmente et plus la consommation diminue.

L’observation des variables de mois et d’heure permet pour sa part de constater une saisonnalité avec la consommation.

Le modèle ainsi entraîné sur les données historiques a permis de prédire la consommation au 1er semestre 2023, en considérant une hypothèse de réchauffement climatique de 10% par rapport aux températures moyennes des 6 dernières années.

L’observation du comportement des données permet de déceler une amplitude de variation de la consommation, corrélée aux demi-heures, avec une tendance baissière selon la température moyenne mensuelle.

2/ A l’échelle régionale :

Nous nous sommes basés sur les mêmes groupes de régions que celles qui ont été présentées précédemment : Grand Est, Hauts de France, Ile de France.

Le même constat peut être réalisé que l’on se situe en région Grand Est, Hauts de France ou Île de France.

Cela nous permet de conclure qu’en l’état, selon les variables explicatives dont nous disposons, le modèle de régression linéaire permet de confirmer qu’il existe une forte corrélation entre la température et la consommation, que que l’on se place au niveau national ou au niveau régional.

* Production en tant que variable cible

Par ailleurs, une modélisation du comportement dans le temps de la production selon le mix énergétique (énergie nucléaire, énergie thermique, énergies renouvelables) a également été réalisée.

Le but étant d’essayer de prédire, selon une hypothèse mathématique, l’évolution de la part de chaque filière de production entre 2025 et 2070.

Selon les hypothèses calculées, il en est ressorti que la part de production nucléaire tendait à diminuer régulièrement avec le temps, que la part de production renouvelable suivait la tendance inverse et donc augmentait avec le temps.

L’évolution dans le temps de la part d’énergie thermique restait difficile à déterminer étant donné qu’il n’existe pas de corrélation entre ces deux variables.

**5/ Conclusion**

Le risque de black-out reste limité en France.

Les graphiques permettent d’affirmer que sur les dernières années, la capacité de production est globalement suffisamment robuste et adaptée pour répondre à l’évolution de la demande.

Aussi, les modèles de prédictions (série temporelle) ne laissent pas présager d’augmentation significative ou brute de la consommation sur les années à venir. On visualise des courbes modélisées avec une saisonnalité forte mais une tendance pas très prononcée.

Cependant, le risque de black-out n’est pas inexistant et des pics de tension peuvent s’observer sur les courbes (la courbe de la consommation est au même niveau ou au-dessus de la courbe de production).

Les causes de black-out sont à rechercher soit:

* dans une augmentation de la demande en énergie comme dans les périodes de grands froids;
* dans une baisse de la capacité de production en énergie avec des fermetures de centrales ou encore des niveaux d’eau très bas (pénalisant la production d’énergie hydraulique)

Au niveau régional, les régions faiblement productrices sont sous-tension et bénéficient de mécanismes de compensation inter régionaux via les échanges physiques. Il n’en demeure pas moins que le faible niveau de production en énergie des régions en font des zones à risque en cas de délestage régional. Il faut pouvoir s’assurer que l’approvisionnement en énergie depuis les autres régions soit maintenue et constamment ajustée en fonction de la demande.

\_\_\_\_\_\_\_\_\_\_\_\_

S’il est clair que la part des énergies renouvelables augmente dans le temps, la part de l’énergie nucléaire reste bien trop importante en comparaison.

Au niveau régional, les régions en déficit de production par rapport à la demande sont majoritairement celles qui ne produisent pas d’énergie nucléaire.

De plus, la production d’énergie renouvelable reste faiblement diversifiée en proportion, avec une forte présence d’énergie hydraulique ce qui implique un risque de sous production en cas d’aléa climatique.

Les prévisions indiquent une diminution significative de la production nucléaire au profit des énergies renouvelables dans les années à venir. Cependant, au regard de la vitesse de développement des sources d’énergie durable, et sauf changement de politique claire au niveau du développement massif des énergies renouvelables, la France restera encore bien dépendante des filières du nucléaire et du thermique pendant plusieurs années avant que les énergies renouvelables puissent assurer l’autonomie énergétique.

**6/ Ouverture**

1/ Si la régression linéaire nous permet de réaliser une prédiction de la consommation d’énergie en se basant sur les températures relevées, il reste difficile d’établir une corrélation avec la saisonnalité qui reste un paramètre important dans l’appréciation de la consommation.

A l’inverse, le modèle SARIMA permet d’observer la saisonnalité des variations de la consommation dans le temps.

2/ Il serait intéressant d’avoir les données météorologiques plus poussées et de les intégrer dans l’analyse afin de mieux appréhender l’impact des phénomènes météorologiques et naturelles sur la consommation et la production :

* niveau d’eau bas
* tornade détruisant des lignes hautes tensions
* données d'ensoleillement etc.

3/ Si des données ont été obtenues sur les échanges physiques entre régions françaises, on pourrait compléter l’analyse en obtenant des données des échanges physiques avec les pays frontaliers. En effet, des mécanismes de compensation de manque d’énergie existent au niveau national et obtenir les données permettrait de considérer les importations d’énergie dans l’analyse.

4/ Enfin, les variables de production restent fortement dépendantes de la stratégie énergétique au niveau national. Il serait pertinent de s'intéresser à l’orientation stratégique du gouvernement par rapport aux différentes filières énergétiques, à l’évolution de la réglementation ou encore plus concrètement aux différentes décisions et projets à venir afin de pouvoir mieux analyser l’évolution de la production.