



MODÉLISATION ET OPTIMISATION D'UN RÉSEAU
MULTI-ÉNERGIES DANS UNE VILLE BAS CARBONE

TAÏEB Zinédine
Encadrants : Ibrahim AL ASMI, Hamid BENHAMED, Roman LE GOFF-LATIMIER

Table des matières

1	Introduction et motivation du sujet	1
1.1	Contexte environnemental et les enjeux énergétiques	1
1.1.1	Un problème de gestion énergétique qui est multi-modale	4
1.2	Contexte scientifique du stage	4
1.2.1	L'entreprise et son champ d'action	5
1.2.2	Le laboratoire SATIE et son expertise dans le domaine des smarts grids	6
1.2.3	Etat de l'art, les connaissances actuelles	7
1.2.4	Les objectifs de la thèse	7
1.3	Objectif du stage et les enjeux par rapport au contexte de recherche	8
1.3.1	Définition de la chaleur fatale industrielle	8
2	Modèles de prédictions sur différents contexte de production industriel	10
2.1	Présentation des différentes données	10
2.1.1	Un problème multi-modal	10
2.1.2	Étude sur les critères caractéristiques d'un problème de chaleur fatale	10
2.2	Un exemple de modèle de prédition de production industrielle : le cas de l'industrie Tégulys	12
2.2.1	Origine des aléas dans le process industriel	14
2.2.2	Traitement des données	14
2.2.3	Modèle mis en place	16
2.2.4	Prévision sur le long-terme des cycles de marche : modèle clustering chaîne de Markov sur des profils de clusters	18
2.2.5	Prévision sur le long-terme des cycles d'arrêt : estimation de la loi aléatoire sur les temps d'arrêt	22
2.2.6	Résultats de prévision long-terme de série de production	24
2.2.7	Conclusions sur les différents modèles de prédition pour le cas de production industrielle	26
2.3	Un exemple de modèle de consommation de production industrielle : le cas d'un réseau de chaleur	28
2.3.1	Les différentes dynamiques temporelles du réseau de chaleur	28
2.3.2	Modèles de prédition	29
2.3.3	Conclusions et pistes d'améliorations pour la prédition de série temporelles de consommation thermique	33
3	Les conclusions concernant l'étude de modèle de prédition dans un contexte industriel	34

1

Introduction et motivation du sujet

“ Afin d’atteindre l’objectif ambitieux fixé par l’accord de Paris et de s’attaquer efficacement au problème du changement climatique, les principales économies du monde doivent rassembler leurs forces et leurs atouts technologiques pour apporter des solutions innovantes zéro carbone. Avec les États-Unis, nous visons à créer une dynamique internationale sur ces aspects dans la perspective de la prochaine COP26.

”

B. Pompili, ministre française de la transition écologique ²

1.1 Contexte environnemental et les enjeux énergétiques

Une situation environnementale de plus en plus alarmante selon la communauté scientifique

La communauté scientifique ne cesse d’alerter sur le désastre environnemental qui s’accélère et s’aggrave au fil du temps. Le GIEC ³, un des organismes intergouvernementaux à haute crédibilité scientifique, participe à inverser la tendance sur le plan environnementale à l’échelle mondiale. Ses missions sont de comprendre, d’anticiper les impacts humains sur l’environnement, dans le but de fournir ces constats aux états au vu de pouvoir agir en conséquence. En 2014, Le GIEC a dès lors souligné l’extrême importance d’endiguer le réchauffement climatique en deçà de 2°C par rapport au niveau préindustriel ⁴ dans un rapport de plus de mille pages datant de 2014 ⁵. C’est sur ce résultat que les fondements de l’Accord de Paris voté en 2015 ont été posé, à l’issue de la COP21 ⁶. Accord historique, il fut le premier depuis les protocoles de Kyoto en 1991 à ratifier une volonté majeure internationale commune sur la préservation de l’environnement et non pas un objectif différencié en fonction du niveau de développement. L’avancée majeure : contenir le réchauffement climatique à plus de 2°C par rapport à la période pré-industrielle, comme il a été précédemment stipulé par le GIEC.

Plus tard en 2018, dans une optique de connaître d’avantage les impacts des risques

2. Ministère de l’Europe et des affaires étrangères (2021)

3. Groupe d’experts intergouvernemental sur l’évolution du climat réunissant des experts scientifiques des pays membres des Nations Unis

4. par rapport à la température moyenne mondiale de la période pré-industrielle en 1850 1900

5. cinquième Rapport d’évaluation du Groupe d’experts intergouvernemental sur l’évolution du climat [Rajendra K. Pachauri (2015)]

6. 21e Conférence des Parties

climatiques ainsi que l'envergure des mesures d'atténuation requises et leur faisabilité, le GIEC a abouti à un rapport spécial sur les conséquences d'un réchauffement planétaire de 1,5 °C. Ledit rapport stipule que « la limitation du réchauffement planétaire à 1,5 °C est possible, mais qu'il faudrait, pour la réaliser, des transitions sans précédent à tous les niveaux de la société. Limiter le réchauffement à 1,5 °C et non à 2 °C, voire plus, présente des avantages évidents. Chaque fraction de degré en plus compte. »¹

En écho à ces alarmes, une prise de conscience internationale, tant citoyenne que gouvernementale, résonne de plus en plus dans les débats actuels.

2021 sera alors plus que jamais l'année des grands bouleversements sociétaux. Bien que les états sont décideurs des politiques communes, les actions citoyennes ne sont pas pour autant moins importantes. Il faut donc fournir l'effort conséquent afin d'accélérer la transition énergétique qui transformera sans précédent nos modes vies actuels.

Quelques chiffres clés pour cibler les freins à la transition énergétique en France

Avec 308 TWh, les énergies renouvelables représentent 17,2 % de la consommation finale brute d'énergie en 2019, soit une progression de 8 points depuis 2005. Cette hausse résulte, d'une part, de l'augmentation importante de la consommation finale brute d'énergies renouvelables, conséquence des investissements réalisés pour en favoriser le développement, et, d'autre part, d'une baisse globale de la consommation finale brute d'énergie.

La directive 2009/28/CE fixe pour la France un objectif de 23 % d'énergie produite à partir de sources renouvelables dans la consommation finale brute d'énergie à l'horizon 2020. Cet objectif a été porté à 33 % à l'horizon 2030 par la loi relative à l'énergie et au climat de 2019.

Les principales filières renouvelables restent la biomasse solide pour la chaleur et l'hydraulique pour l'électricité.

Les actions internationales pour la transition énergétique

Depuis le décret en 2015 de la loi relative à la transition énergétique pour la croissance verte, la France s'est engagée à entamé la mise en route d'engagements forts dans le secteur de l'énergie. Le gouvernement a mis en place plusieurs actions de transition dont trois principaux documents formant un tryptique de la stratégie nationale concernant la transition écologique à l'échelle du territoire.

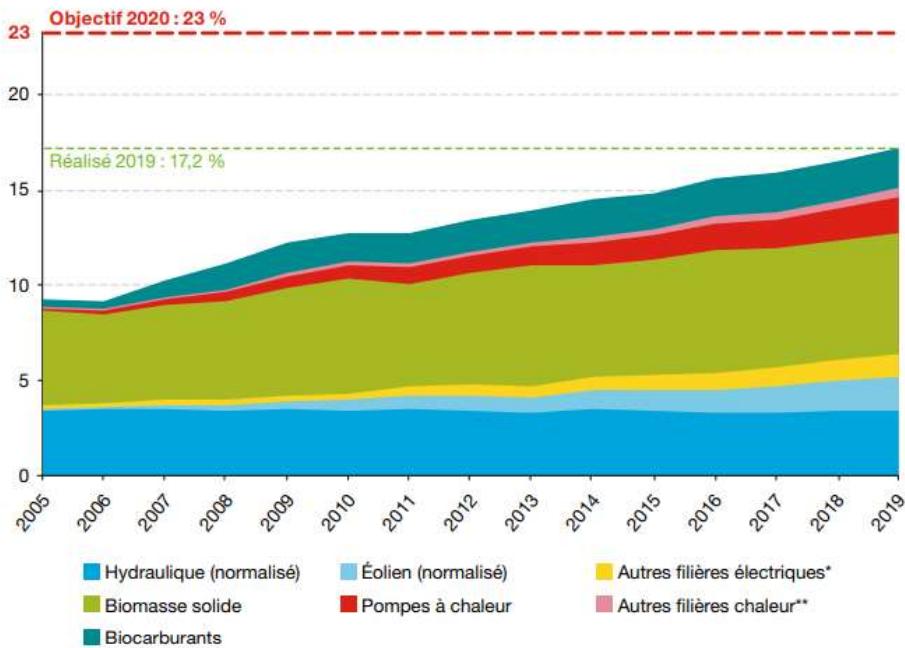
En signant l'Accord de Paris en 2015, les pays se sont engagés à limiter l'augmentation de la température moyenne à 2°C, et si possible 1,5°C. Pour cela, ils se sont engagés, conformément aux recommandations du GIEC, à atteindre la neutralité carbone au cours de la deuxième moitié du 21ème siècle au niveau mondial. Les pays développés sont appelés à atteindre la neutralité le plus rapidement possible. La France s'était engagée, avec la première Stratégie Nationale Bas-Carbone (SNBC) adopté en 2015.

La SNBC se définit donc comme feuille de route nationale pour réduire ses émissions de gaz à effet de serre (GES). vise à « réduire les émissions de gaz à effet de serre de 40 % (de 1990 à 2030) et diviser par quatre les émissions nationales de gaz à effet de serre de 1990 à 2050 ». Pour atteindre la neutralité souhaitée en 2050, elle s'appuie sur des quotas

1. [V. Masson-Delmotte, P. Zhai, H-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani,W. Moufouma-Okia, C. Péan, R. Pidcock, Sarah Connors,J. B. R. Matthews,Y. Chen, X Zhou, M I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield (2018)]

PART DES ÉNERGIES RENOUVELABLES DANS LA CONSOMMATION FINALE BRUTE D'ÉNERGIE PAR FILIÈRE

En %



* Solaire photovoltaïque, énergies marines et électricité à partir de biomasse et de géothermie.

** Solaire thermique, géothermie et biogaz.

Lecture : dans le cadre de la directive 2009/28/CE, la France a pour objectif de parvenir en 2020 à une part de 23 % d'énergies renouvelables dans la consommation finale brute d'énergie. En 2019, cette part s'élève à 17,2 %. Les productions hydraulique et éolienne sont normalisées après lissage sur respectivement quinze et cinq ans.

Source : calculs SDES

Source : <https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2020-07>

d'émissions de GES à ne pas dépasser par an dit "budgets carbone", et d'autre part sur le développement de méthodes d'absorption de carbone, par le biais de procédés industriels et par l'écosystème. Ces budgets carbones sont progressivement plus restrictifs au fil des ans et se veut être réaliste quant à la pertinence de ces seuils.

Dans la continuité de ces mesures, la programmation pluriannuelle de l'énergie (PPE) a été instituée dans le cadre de la loi à la transition énergétique pour la croissance verte. La PPE fixe les priorités d'action de la politique énergétique du Gouvernement pour les dix prochaines années. L'ensemble des piliers de la politique énergétique et l'ensemble des énergies sont traités dans une même stratégie : maîtrise de la demande en énergie, maîtrise des coûts des énergies, promotion des énergies renouvelables, garantie de sécurité d'approvisionnement et indépendance énergétique, etc.

La PPE a identifié plusieurs leviers qui participerait à la transition énergétique :

— Réduire notre consommation d'énergie

Cela passe à la fois par le développement de technologies sobres en énergie, mais aussi par la modification des comportements de tous les acteurs économiques. Le programme prévoit en 2023 une réduction de 12,3 % la consommation finale d'énergie en vue d'atteindre l'objectif d'une réduction haute de 20 % en 2030 par rapport à 2010. Des moyens concrets sont mis en place tel que la rénovation des bâtiments résidentiels dans le secteur du logement. En effet, le secteur du bâtiment repré-

sente 44 % de la consommation d'énergie et près de 25 % des émissions de CO₂. Des actions concrètes par exemple dans le secteur du transport sont en déploiement telle que l'incitation à des modes de transport plus écologique, ou de voitures électriques ou hybrides pour se détacher d'avantage des énergies fossiles. Parmi tout ces moyens, le gouvernement compte aussi sur la recherche et l'innovation de nouvelles technologies qui avancerait dans le sens de la transition écologique.

— **Diversifier notre mix énergétique**

Accélérer le développement des énergies renouvelables et leur déploiement en fonction des ressources locales sur le territoire national est un moyen fort de diminuer la part d'utilisation des énergies fossiles et nucléaires, à ce jour représentant quasiment 70 % de la production d'énergie primaire en France.

1.1.1 Un problème de gestion énergétique qui est multi-modale

De nombreuses problématiques sont identifiées pour la décarbonation de l'énergie, l'efficacité énergétique, le stockage de l'énergie, la gestion intelligente des réseaux de transport et de distribution, ainsi que les solutions de capture, stockage et réutilisation du carbone. Cela suscite ainsi un intérêt certain de développer la recherche et innover sur ces quelques points parmi d'autres.

Par conséquent, face à une transformation inédite du mode de production et de consommation, une adaptation du système actuel est nécessaire.

L'intérêt du smart grid et de la recherche dans un contexte de réseaux multi-énergie

Un réseau électrique intelligent, ou smart grid en anglais, est un réseau de distribution d'électricité qui favorise la circulation d'information entre les fournisseurs et les consommateurs. Faciliter les informations entre les différents agents du réseau, tant consommateurs que producteurs, permettrait donc d'améliorer l'efficacité de la gestion des ressources énergétiques. Ce concept peut être généralisé pour d'autres formes d'énergies autre qu'électrique, tel que thermique ou chimique. La faisabilité et la pertinence de ce type de réseau "multi-modale" repose sur la capacité à stocker l'énergie sous ses différentes formes possibles ainsi que de pouvoir la convertir. Par conséquent, un réseau multi-énergie serait pertinent de par à sa flexibilité. Pour y parvenir, il est nécessaire de modéliser les différents sous-systèmes physiques et de prédire mathématiquement les fluctuations des gisements énergétiques étudiés à savoir les chaleurs fatales industrielles et les énergies renouvelables. Le caractère stochastique des gisements et des besoins énergétiques identifiés est l'un des éléments centraux dans la gestion de ce type de réseaux efficace de réseaux.

1.2 Contexte scientifique du stage

Les missions de mon stage s'écrivent dans le cadre de la thèse de mon tuteur, M. Ibrahim Al Asmi actuellement en deuxième année. Ses travaux de thèses gravitent essentiellement autour des problématiques sur la mise en place d'un mix énergétique à travers un réseau multi-énergies smart-grid ¹, de sa faisabilité à sa gestion. La thèse en question s'inscrit sous une convention CIFRE, entre l'entreprise Eco-Tech Ceram, start-up spécialisée dans la revalorisation d'énergie thermique industrielle, et le laboratoire SATIE de l'ENS-Rennes, se consacrant aux systèmes et applications des technologies de l'information et de l'énergie.

1. ici réseau bi-énergie : thermique et électrique

1.2.1 L'entreprise et son champ d'action

L'entreprise Eco-Tech Ceram (ETC) est spécialisée dans la revalorisation d'énergie industrielle. Elle propose ainsi des solutions d'optimisation de l'efficacité énergétique dans le milieu industriel, en commercialisant notamment des machines de stockage d'énergie thermique dédiées aux agents producteurs d'énergie. Issue du CNRS PROMES, ETC développe et commercialise une solution de stockage de chaleur innovante : l'Eco-Stock. Cet accumulateur de chaleur fournit une énergie décarbonée à un prix inférieur à celui de l'énergie fossile (jusqu'à 80%). Cette énergie disponible peut ensuite être facilement réutilisée directement sous forme de chaleur, ou convertie en électricité, froid ou vapeur.

Ces machines de stockage améliorent l'efficacité énergétique des industriels d'une part. D'autre part, la flexibilité du stockage permettrait d'intégrer dans un réseau énergétique les énergies de sources renouvelables ; ce qui s'inscrit complètement dans une dynamique propice à la transition énergétique.

La solution Eco-stock, une batterie de stockage thermique pour valoriser l'énergie thermique

Ces batteries thermiques Eco-Stock remplies de billes de céramique, issues de circuit court, peuvent être utilisées dans un contexte de réseau multi-énergie en valorisant la chaleur émise par les usines thermiques. En effet, au lieu de délester cette chaleur dite fatale dans l'atmosphère, ces usines peuvent stocker l'énergie thermique captée dans les batteries afin de pouvoir la réutiliser. Les possibilités d'applications sont multiples.

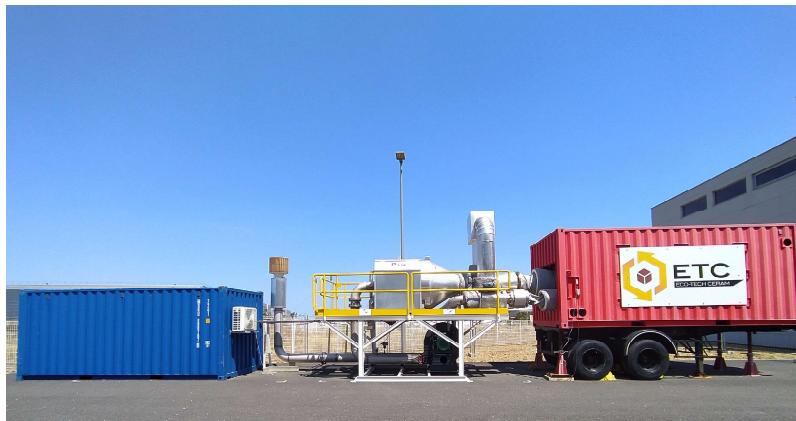


FIGURE 1.1: Un dispositif Eco-stock (container rouge) commercialisé par l'entreprise, connecté à un convertisseur chimique (container bleu) comme source de chaleur d'entrée

Ses possibilités d'utilisation dans un contexte smart-grid

L'accumulateur Eco-Stock est un mode de stockage flexible aux différents contextes de production et de consommation possible. Cette flexibilité ouvre le champ des possibilités dans l'utilisation complète des vecteurs d'énergies et ainsi s'adapter aux contraintes de demandes du réseaux. Avec un stockage thermique de l'ordre de quelques jours, il est possible de valoriser la chaleur fatale à l'aide de convertisseurs énergétiques :

- 1) “Heat to Heat” : stockage thermique et valorisation de chaleur fatale pour les besoins en chaleur dans le secteur industriel et domestique (exemple : chauffage, séchage matières, préchauffage d'air de combustion...).

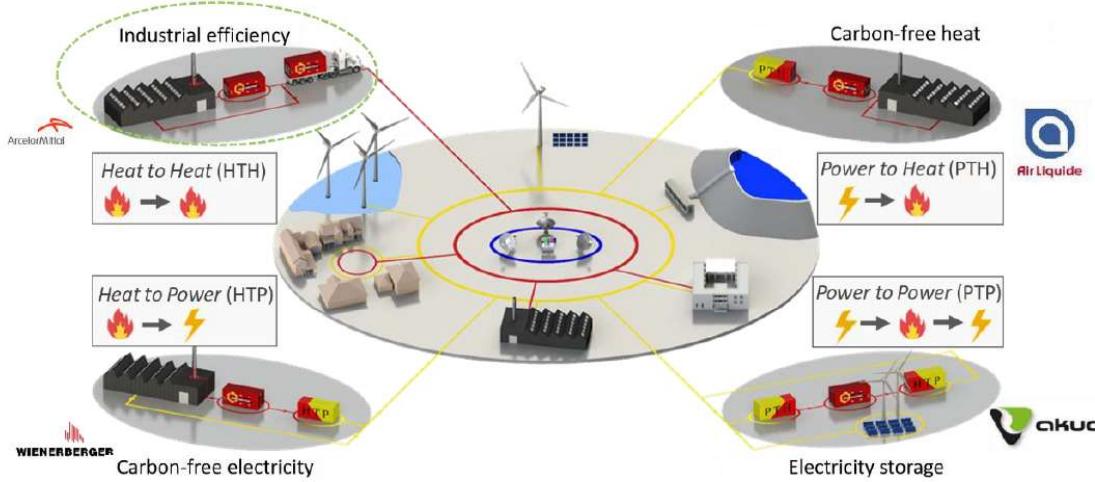


FIGURE 1.2: Illustration des champs d'applications de la solution Eco-Stock

- 2) “Heat to Power” : stockage thermique et valorisation de chaleur fatale en électricité dans le secteur industriel et domestique.
- 3) “Power to Heat” : stockage thermique pour l’effacement et l’engagement de consommation électrique dans le secteur industriel et domestique pour un substitut aux équipements de chauffage, de cuisson...
- 4) “Power to Power” : stockage massif d’électricité sous forme thermique pour répondre aux besoins “énergivores” insatisfaits par les systèmes conventionnels de stockage d’électricité, (exemple : Pompe à chaleur couplée avec stockage thermique et un système de “Heat to Power”)

Cette vision intègre le couplage possible entre le stockage thermique et électrique, du fait de leur complémentarité du point de vue de capacité de stockage et de la puissance délivrée

1.2.2 Le laboratoire SATIE et son expertise dans le domaine des smarts grids

Le SATIE (Systèmes et Applications des Technologies de l’Information et de l’Energie) est un laboratoire de recherche en sciences appliquées. C'est une unité mixte de recherche du CNRS comptant 200 personnes. Plus spécifiquement, l'équipe avec laquelle j'ai collaboré est spécialisée dans les systèmes d'énergies pour les transports et l'environnement sous la direction de Hamid BEN AHMED.

Concernant des réseaux électriques dits intelligents – Smart Grid –, le SATIE a développé depuis plusieurs années un savoir-faire reconnu dans le domaine des méthodes de coopération pour un réseau électrique avec stockage. Les principaux apports méthodologiques de l'équipe sont la modélisation des composantes aléatoires de la production et de la consommation d'électricité, l'optimisation stochastique des stratégies de gestion prenant en compte le vieillissement des composantes et le couplage dimensionnement-gestion sur le cycle de vie de ces réseaux. Ceci en intégrant à la fois des aspects stochastiques¹, mais aussi des approches de cycle de vie (vieillissement par stockage électrochimique)².

De même, ce travail de thèse et cette collaboration avec la société ETC représentent une formidable opportunité pour le laboratoire d'étendre les problèmes traités aux réseaux

1. P. Haessig and Bondon (2013)
 2. A. Arteconi and Polonara

hybrides dont les paramètres de contrôle sont plus riches et d'améliorer les connaissances dans le domaine de la modélisation, de la gestion et du dimensionnement des smart (flexibles) multi-énergies.

1.2.3 Etat de l'art, les connaissances actuelles

L'étude d'un réseau multi-énergies se décompose en plusieurs problèmes allant du contrôle en temps réel à la planification de la construction d'infrastructures¹.

L'étude des réseaux multi-énergies est tout d'abord considérée par la littérature comme un problème de contrôle qui associe deux marchés distincts. Le couplage entre ces deux marchés passe par un carrefour énergétique – energy hub – qui permet la conversion entre les formes diverses d'énergie². Bien que ces travaux permettent des développements très avancés quant à l'optimisation de la performance globale, plusieurs aspects nécessitent d'y être développés. En effet le problème considéré – le plus souvent déterministe – concerne un seul pas de temps, ce qui exclut de fait l'optimisation de la stratégie de gestion d'un stockage, qu'il soit électrique ou thermique. De surcroît ce type d'étude est fait à notre connaissance systématiquement à une infrastructure donnée et n'est donc pas couplée en général avec l'optimisation du dimensionnement des équipements sous-jacents³.

Les travaux qui s'intéressent à l'optimisation de ce dimensionnement considèrent quant à eux des lois de gestion à chaque instant le plus souvent simplifiées⁴. L'analyse de l'état de l'art montre ainsi que le couplage avec les travaux décrits dans le paragraphe précédent est très faible alors que l'optimisation globale d'un système requiert de prendre en compte simultanément les deux horizons temporels. Cette prise en compte des décisions à chaque instant dès l'étape d'optimisation des infrastructures est fondamentale dans une approche de Co-optimisation. Or cette dernière n'a jamais à notre connaissance été traitée dans la littérature suivant une démarche sur cycle de vie qui requiert des éléments de connaissance sur les coûts de fabrication, sur le vieillissement et sur le recyclage des composants. De tels modèles n'existent pas à notre connaissance à l'heure actuelle. Ils sont en effet très difficiles à développer sans une collaboration étroite avec des industriels permettant un retour sur expérience efficace. Enfin les cas d'étude considérés jusqu'ici dans l'état de l'art considère des associations intéressantes mais néanmoins incomplètes des diverses configurations permettant d'interfacer un réseau de chaleur avec un réseau d'électricité⁵.

1.2.4 Les objectifs de la thèse

L'objectif final de la thèse de M. Al Asmi est d'implémenter un algorithme d'intelligence artificielle capable de gérer en temps réel le réseau multi-énergie. A ce jour, les algorithmes commandant les réseaux ne proposent qu'une approche déterministe. La thèse a pour objectif d'étudier la pertinence de considérer ce problème d'optimisation par des méthodes stochastiques. Les travaux réalisés serviront à l'élaboration de modèles et d'outils nécessaires à une Co-optimisation sur cycle de vie d'un réseau hybride multi-énergie. Le but ultime étant d'analyser les conditions de pertinence d'un réseau hybride avec stockage.

1. N. Neyestani and Catalao (2015)

2. Favre-Perrod (2005)

3. Kienzle and Andersson (2009)

4. Geidl and Andersson (2007)

5. M. Geidl and Fröhlich

1.3 Objectif du stage et les enjeux par rapport au contexte de recherche

Pour pouvoir proposer un contrôleur de réseau de type stochastique délivrant une série de commandes adaptée au réseau , plusieurs données d'entrées sont nécessaires. Pour entraîner l'algorithme, nous avons besoin de connaître les données de production de consommation d'énergies, ainsi que les modèles de stockages et de conversion des différentes infrastructures. Comme le problème traite d'un réseau hybride électrique et thermique, les trois types d'entrées doivent être identifiés pour les deux types d'énergies, thermiques-électriques. Les modèles de stockage et de conversion ont été caractérisé durant la première année de thèse. Les données tant sur la production que la consommation, dans le milieu de l'énergie électrique du moins, sont largement accessibles , grâce à des institutions comme EDF qui centralise la gestion des ressources électriques. Elle publie de manière périodique des bilans et analyses : passage de l'hiver, bilan électrique de l'année, évolutions nécessaires du réseau, enjeux autour du développement de la mobilité électrique, etc. Le verrou principal dans la conception du contrôleur est l'accès aux données sur la production et consommation thermique en chaleur fatale. Il est difficile d'accéder à ce type de données, car d'une part privée par les industriels, et d'autre part la revalorisation des énergies thermiques issus des chaleurs fatales n'est pas très développée à ce jour.

Pour palier à ce manque de données, un de mes objectifs et de générer des scénarios de productions et de consommations de puissances de chaleur fatale. Ces données prédictives substitueront le manque d'accès directe aux données réelles. De ce fait, le but est de générer des prédictions et de quantifier la fiabilité associée, qui seront utilisées dans le contrôleur du réseau multi-énergies.

Prédire des séries de données issus des chaleurs fatales est assez nouveau, et peu de résultats sont connus. Le second objectif de mon stage est donc d'apporter des éléments de réponse quant aux bonnes familles de modèles adaptées à des séries temporelles de chaleur fatale, dans différents contextes industriels.

1.3.1 Définition de la chaleur fatale industrielle

La chaleur fatale est la chaleur résiduelle issue d'un procédé et non utilisée par celui-ci. La récupération de cette ressource thermique représente un potentiel d'économies d'énergie à exploiter pour les entreprises industrielles. Ainsi l'Eco-Stock permet de revaloriser une certaine classe de chaleur fatale en sortie de circuit industriel. La quantité de chaleur fatale valorisable dépend de trois facteurs : la taille et le type de l'industrie, ainsi que la qualité de la chaleur, en grande partie liée à, la température en sortie de process. En effet, dans le domaine énergétique, au-delà de la quantité d'énergie produite, la qualité de l'énergie mise en jeu doit être pris en compte : de la chaleur à 500°C n'a pas la même "qualité" que de la chaleur à 2000°C qui a une grande rentabilité énergétique que des chaleurs à basses températures. Cette qualité est liée à la notion d'exergie en thermodynamique. Pour avoir une utilisation plus vaste de la chaleur fatale, on ne stocke que des chaleurs ayant une température au de-là d'un certain seuil pour maximiser la qualité des énergies thermiques en question.

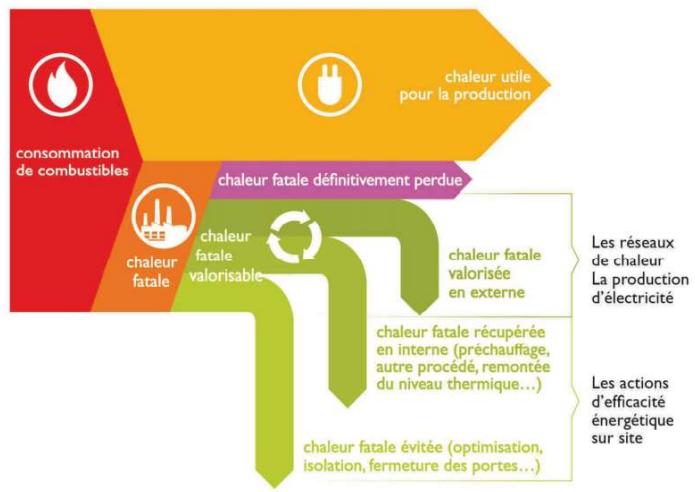


FIGURE 1.3: Origines de la chaleur fatale industrielle

2

Modèles de prédictions sur différents contexte de production industriel

2.1 Présentation des différentes données

2.1.1 Un problème multi-modal

L'existence de différentes modalités de production et de consommation rend complexe la construction d'un seul modèle général de prédition de production et de consommation énergétique. Est-il possible d'inférer un modèle qui généraliserait la prédition de productions et de consommations dans un réseau complexe multi-énergie ? A ce jour, nous avons peu de résultat sur la prédition de série temporelle de chaleur fatale, d'énergie globalement ; et encore moins sur des modèles généralisant le comportement plusieurs architectures d'industrie à la fois. Avant de pouvoir répondre à cette question, l'objectif est de fournir des premiers résultats sur le choix de familles de modèles de prédictions pertinents, adaptées aux séries temporelles de chaleur fatale, en fonction du contexte industriel et de consommation. Le but est donc apporter des premières réponses sur la modélisation de réseau multi-énergie, puis ouvrir le sujet sur des problématiques connexes. Comme le problème est de nature multi-modale, une des premières choses à faire avant tout est de pouvoir regrouper le plus de données représentatives des différents types d'industries. Cette tâche est assez fastidieuse au vu du peu de données libre de droits ou bien répondant à notre problème. On a pu récolter des données de type industriels, et également sur la consommation dans un contexte de réseau de chaleur en France.

2.1.2 Étude sur les critères caractéristiques d'un problème de chaleur fatale

L'aléa d'un problème

On pourrait caractériser un problème par sa nature et son type d'industrie. Or ce choix n'est pas intéressant car cela ne mettrait pas en évidence les mécanismes du problème à modéliser. Par exemple, deux usines dans deux domaines différents, peuvent partager les mêmes process de production et donc le même type de modèle pourrait être satisfaisant. Un choix de caractérisation des problèmes plus judicieux serait d'identifier les mécanismes aléatoires ou mal connues. De ce fait, les données seraient triées en fonction de leurs aléas rencontrés dans le milieu industriel. Une fois cela identifié, il sera plus simple de proposer des familles modèles qui sont en adéquation aux inconnues d'un problème donné, ce qui facilite de la généralisation à tout type d'industrie. Ces modèles devront bien sûr être mieux adapté dans un second temps en fonction du problème physique en question.

Par abus de langage on considère comme aléa, un paramètre aléatoire, ou bien mal connu

bien qu'il peut être déterministe. En analysant les différentes données récoltées, plusieurs aléas ont pu être identifié dans un contexte de stockage thermique :

— **Aléas sur le process :**

Il se peut qu'il y ait un aléa sur la quantité, le type de matériaux, type de programme, que la durée du cycle soit inconnue etc...

— **Aléas sur le besoin :**

L'aléa dans ce cas peut être contenu dans des mécanismes extérieurs aux process industriels, comme par exemple le carnet de commande inconnu à l'avance, ou encore le départ du cycle de production dépendant du marché etc...

Ces types d'aléas dans un problème donné vont donc conditionner le type de modèle de prédiction à appliquer.

Type de consommation ou de production

Pour dépeindre un champ exhaustif des cas de figures possible, la suite du chapitre sera alors dédié aux modèles construits sur deux cas de figures rencontrés :

- Un cas de production d'énergie avec de l'aléa sur le process et sur le besoin.
- Un cas de consommation d'énergie avec de l'aléa sur le besoin et avec de la périodicité.

Outre l'accès à des données, l'exploitation et la pertinence de ces données est une autre contrainte non négligeable, et ne pourrait être véritablement résolu qu'a posteriori en confrontant les modèles issus sur ces données précises, sur des cas réels. On pourra également mesurer la performance de ces modèles.

2.2 Un exemple de modèle de prédiction de production industrielle : le cas de l'industrie Tégulys

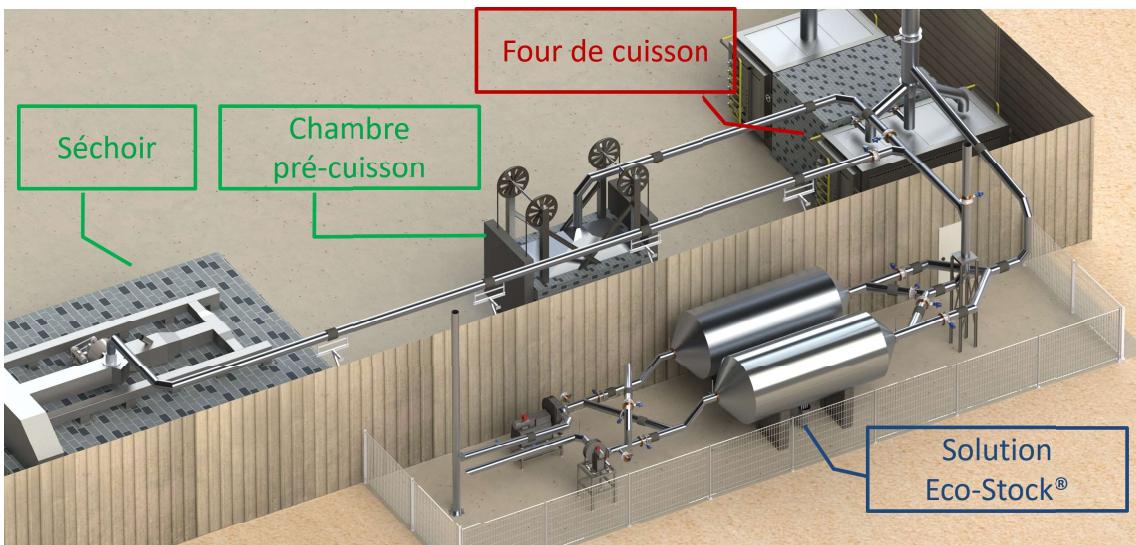
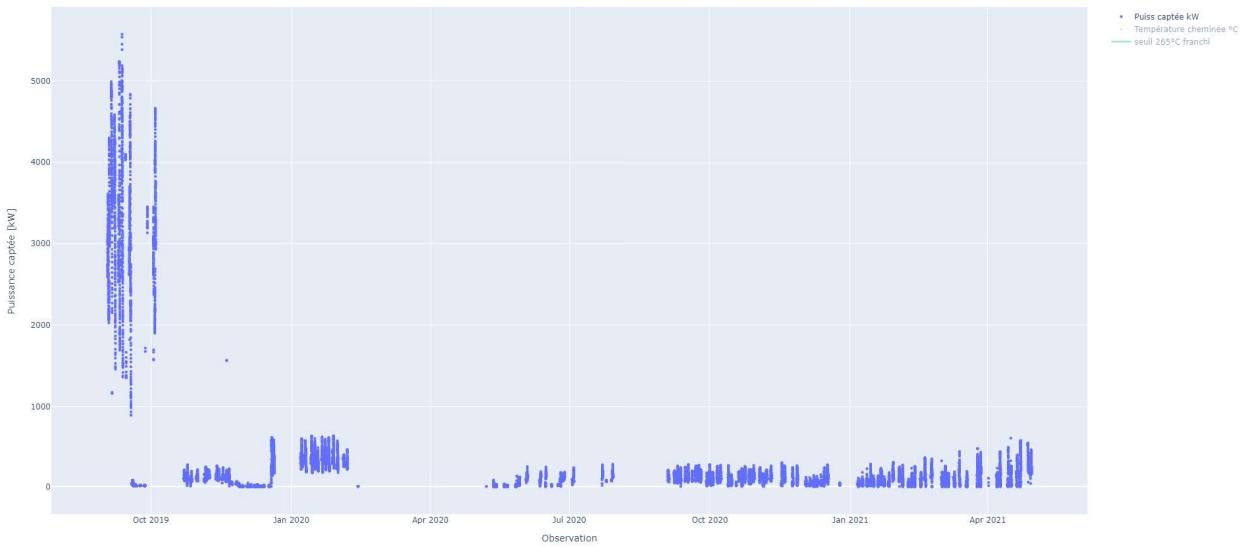


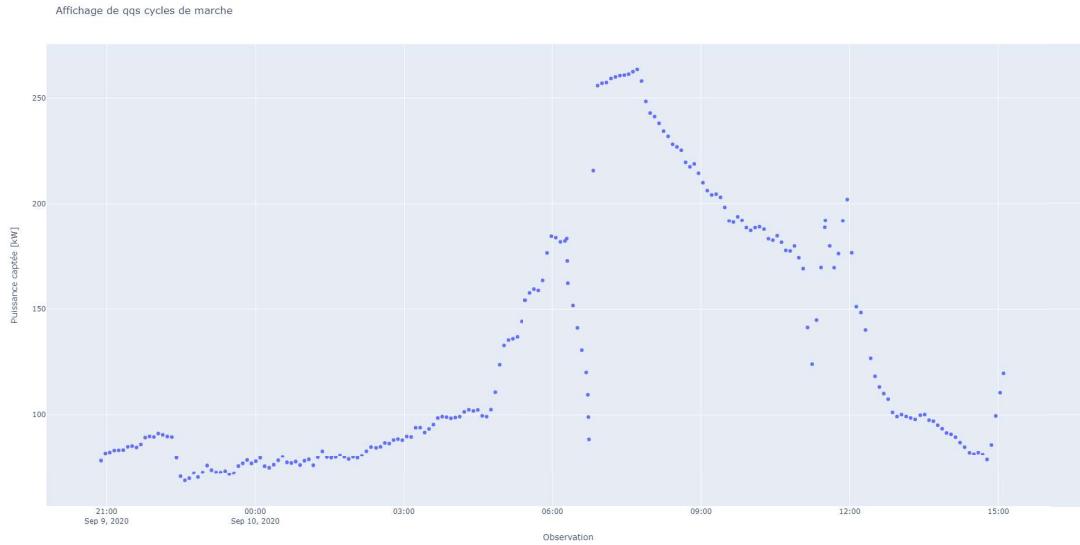
FIGURE 2.1: Schéma de principe du fonctionnement de l'entreprise et valorisation

Tégulys est une petite industrie familiale basée en Corrèze spécialisée dans la production de tuiles. Son process de cuisson consomme à l'année 1 GWh, et la chaleur fatale représente 58 % de l'énergie consommée sur le process. La température des fumées dans la cheminée du four de cuisson peuvent atteindre 600°C et sont stockées en sortie du four dans des stockages thermiques. La chaleur fatale est alors revalorisée dans le séchoir en début de process avant la cuisson des matériaux.

Filtrage des cycles de marche



(a) Données entières



(b) Détail d'un cycle de fonctionnement

FIGURE 2.2: Série de chaleur fatale Tégulys pour les mois décembre 2020 – avril 2021, observé toutes les 5 minutes

Les données suivantes représentent la puissance thermique captée durant la production d'une fabrique de tuile. Cette usine familiale effectue des cycles de chauffe en fonction de la demande reçue. Les cycles de productions sont donc de fréquence et de durée de fonctionnement variables. Comme on peut le constater en examinant la série temporelle de puissance de chaleur produite , les épisodes de chauffes sont intermittents, souvent étalement sur un voire deux jours en fonction de la demande. L'arrêt des fours est également variable, et peut se produire sur plusieurs jours jusqu'à une semaine entière. On peut également noter l'absence de données sur la période de février 2020 à mai 2020, en raison de l'arrêt des industries forcé par la crise sanitaire.

De plus dû à un problème de capteurs, les données ne sont pas toutes exploitables, elles ne le sont qu'à partir de septembre 2020, ce qui divise par deux le temps d'observation. La base de données est donc moins fournie et cela peut influer sur un mauvais rendement

de méthodes de machine learning. Une des améliorations certaine sera donc de compléter la base par des mesures précises à l'avenir.

En outre les données ne sont pas directement utilisables, due à l'existence de données manquantes et du caractère variable des cycles de marche et d'arrêt. Des données manquantes peuvent subsister pendant un cycle de fonctionnement ou entre deux cycles.

2.2.1 Origine des aléas dans le process industriel

2.2.2 Traitement des données

En raison de la notion d'exergie correspondant à la "qualité" d'une énergie thermique , la puissance thermique captée n'est issue que de hautes températures. En dessous d'une certaine température, la puissance thermique des cheminées est délestée sans être stockée. Le seuil d'activation des capteurs du stockage thermique est à **265°C** : à partir de cette température, une vanne est ouverte et la chaleur est stockée dans la batterie Eco-stock.

Extraction des cycles de marche et d'arrêt

Nous allons donc commencer par traiter les données en extrayant les cycles en fonction de la température de la cheminée. On considère qu'en dessous de 265°C, le système de production de puissance est à l'arrêt ; au dessus, on est à l'état de fonctionnement.

La première chose à faire est de bien distinguer les cycle de marche. Exploiter les cycles de marche depuis la série des puissances ne peut pas se faire directement en partie causé par des données manquantes qui entravent à l'extraction des cycles. Les données peuvent être manquantes lié à des problèmes de mesure. Dans ce cas-là on ne peut pas savoir si le cycle se termine réellement ou si cela est dû à une température en deçà du seuil d'activation des capteurs de puissances.

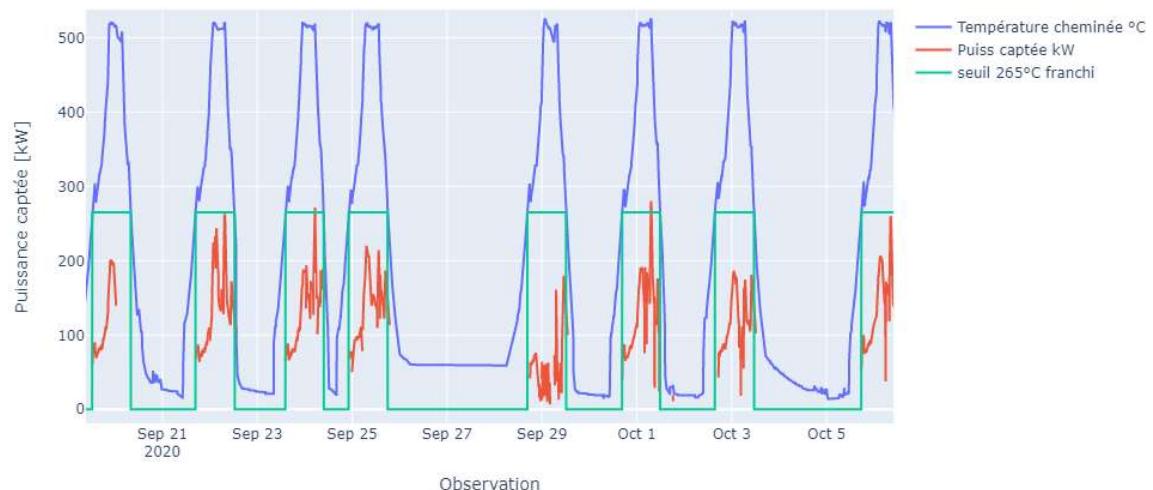


FIGURE 2.3: Filtrage des cycles de marche et d'arrêt en fonction de la température des gaz dans les cheminées

Au lieu de découper approximativement les cycles à partir de la série temporelle de puissance, on s'est appuyé sur le lien entre température et activation du stockage de chaleur. La série des températures des cheminées étant plus régulière, cela a pu garantir une extraction des cycles de marche et d'arrêt de meilleur qualité.

Choix du seuil de filtrage

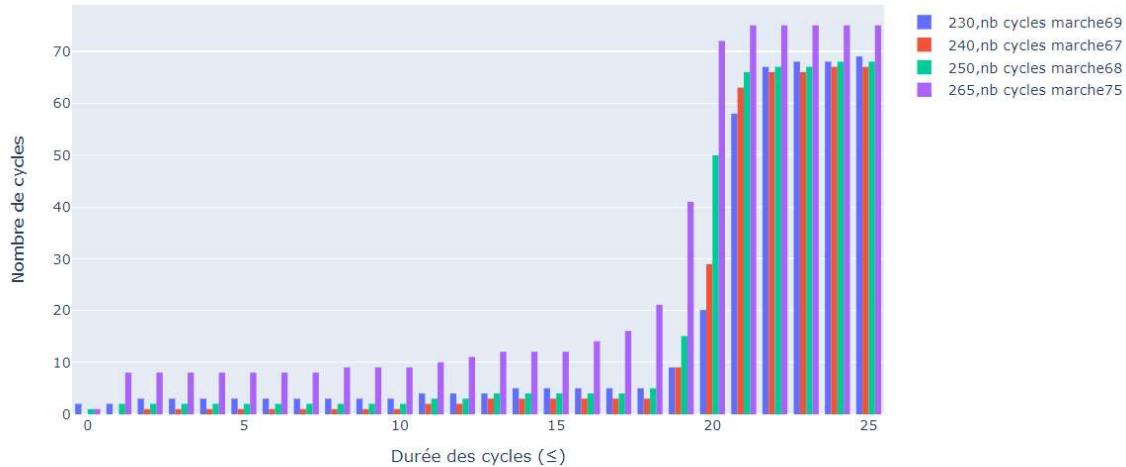


FIGURE 2.4: Répartition des temps de cycle en fonction du seuil d'extraction (en cumulé)

Nous savons par l'entreprise que les cycles ont une période en moyenne de 20h, qui peut varier en fonction de la charge de demande. Or, choisir le seuil théorique de 265°C ne permet pas d'obtenir en majorité des cycles de 20h. Cela est dû aux intermittences des process du fonctionnement des fours, par exemple une ouverture du four en plein cycle qui peut faire chuter de quelques degrés la température en dessous du seuil. Le seuil à fixer vérifiant le maximum de cycles d'une durée moyenne de 20h est donc à discuter.

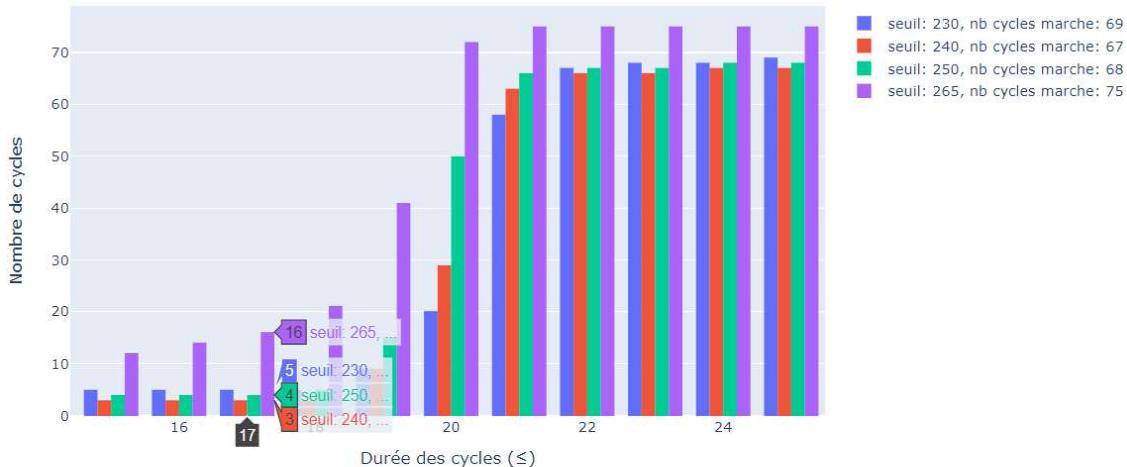


FIGURE 2.5: détail sur les temps de cycles

En prenant un seuil à 240°C au lieu de 265°C, la distribution des temps des cycles est beaucoup plus homogène autour de 20h qui est la valeur moyenne de fonctionnement en principe. En effet pour le seuil théorique de 265°C, 21.33% des données ont une durée en dessous de 20h, contre 4.48% pour un seuil corrigé à 240 d'après les statistiques de temps de cycles ci-dessus. On a pratiquement réduit par cinq le nombre cycles de durées anormalement faibles en modulant le seuil de température.

En prenant un seuil à 240°C, on diminue les erreurs liés aux intermittences de process et obtient des cycles de production de durées approximatives de 20h plus réalistes. Les quelques cycles de durées anormales, ne sont pas pris en compte.

2.2.3 Modèle mis en place

L'objectif est de prédire la puissance thermique stockée de l'industrie de manière fiable, à la fois en temps réel mais aussi sur le long-moyen terme. Par conséquent le modèle permettrait de donner des informations sur des horizons de temps court et long. Ainsi, en fonction de ces prévisions de production, et en prenant en compte les capacités de stockage thermique dans le réseau, l'opérateur pourra prendre les décisions de commande de stockage adéquats dans le but d'anticiper les fluctuations de productions et de consommations à l'avenir. Les prévisions courtes permettront de répondre à des besoins ponctuels et gérer les fluctuations rapides ; tandis que les prévisions sur le moyen-terme serviront à la planification donneront des tendances hebdomadaires ou mensuel, ce qui dégrossira finalement l'ensemble des directives de commandes à prendre. Restera alors à définir quantitativement les temps d'horizon "longs" de prévisions en fonction du système de réseau mis en place.

Le modèle devra également prédire la puissance captée puis stockée et le temps de fonctionnement, dans le cas d'état de marche, mais aussi le temps d'arrêt dans le cas de l'état de repos du four, pour pouvoir simuler des états de marches et d'arrêt et déduire des scénarios de production de puissance sur un horizon de temps. Lors d'un cycle de fonctionnement, on fait également l'hypothèse que le four ne peut prendre qu'un nombre discret d'états de marche lorsque les fours fonctionnent, par exemple, régime à pleine puissance, moyen-régime, bas-régime etc... Entre deux états de marche existe toujours un étape d'arrêt des fours. Par ailleurs, pour simplifier le problème, on va considérer que les cycles de marche sont indépendants des cycles d'arrêts.

On commence par modéliser un modèle qui prédira à quel moment les cycles vont se produire, et quels seront leur tendance générale. Ce modèle aura pour objectif de fournir des planifications de production. Ce modèle n'aura pas l'ambition de prédire fidèlement les variations durant un cycle, mais de fournir une première idée de profils types.

Sur cette base dans un second temps, on révisera ce modèle long-terme afin de fournir des prévisions plus fiables sur du court terme. Les prévision sur le moyen-terme émises précédemment seront améliorées : sachant ce qui s'est passé dans une période antérieure d'une, deux voire plusieurs heures, le modèle s'adaptera sur le passé afin de prédire ce qui pourrait vraisemblablement se passer dans les prochaines heures.

Il restera donc à définir ensuite à valider ce modèle complet éventuellement caractériser sa performance.

Prévisions sur le long terme

On décide de résoudre le problème de prévision long-moyen terme en deux sous-modèles distincts : un modèle de prédiction pour les cycles de marche, et un modèle de prédiction des temps d'arrêt des fours. Ainsi nous proposons un modèle de prédiction par le biais d'une chaîne de Markov sur les cycles de fonctionnement en état de marche, couplée à un tirage aléatoire d'une durée d'arrêt suivant une loi probabiliste établie sur les données de temps d'arrêt.

a) Prédiction des cycles de marche : chaîne de Markov discrète sur des profils de régimes Selon une matrice de transition et un état de marche initial, on simulera une séquence de succession d'états de marche. Parmi les états possibles un sous-état de fonctionnement, déterministe et de durée fixe, sera tiré selon sa probabilité d'apparition.

A partir des données de cycles \mathbf{X} , la première étape consiste donc à extraire des motifs représentatifs de fonctionnement de la série temporelle par clustering, correspondant à

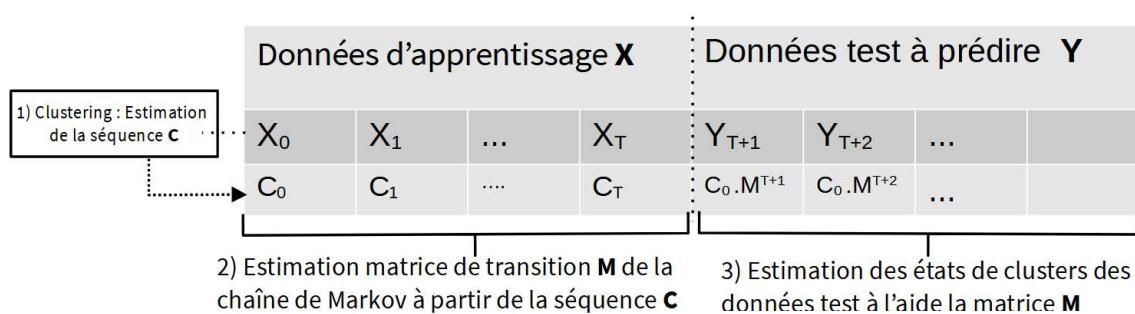


FIGURE 2.6: Modèle Cluster Markov pour estimer les cycles de marche

la séquence de données de cycle \mathbf{C} . Ainsi, ces motifs constitueront nos sous-états de fonctionnement qui seront assignés selon une distribution aléatoire pour chaque cycle issu de la chaîne de Markov.

Ensuite nous estimerons la matrice de transition à partir des cycles de marche des données d'apprentissage. Les probabilités de transition d'un état de marche vers un autre seront approchés par les fréquences de transition. Enfin les données estimées seront comparées aux données tests \mathbf{Y} .

Amélioration avec un modèle de Markov caché Une amélioration est possible est de simuler les transitions de cycles de marche par un modèle de Markov caché ou Hidden Markov Model (HMM) :

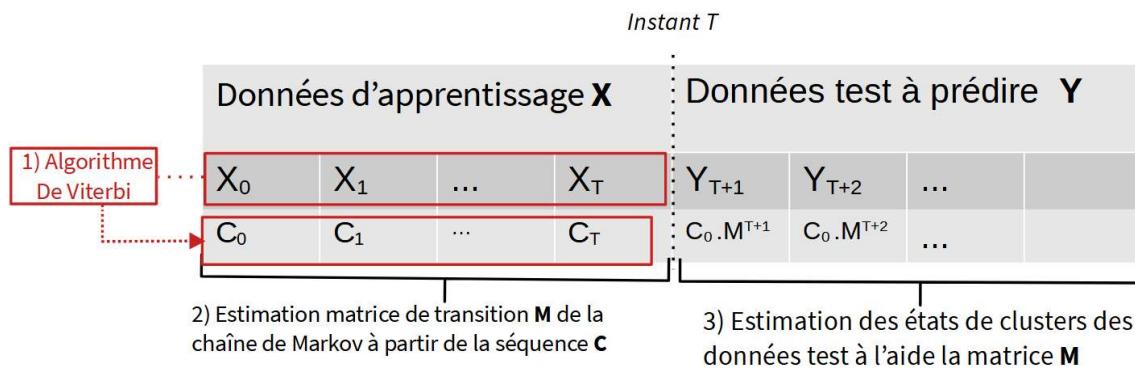


FIGURE 2.7: Modèle Cluster HMM pour estimer les cycles de marche

Dans un modèle de Markov caché (HMM), on ne peut pas observer directement les états du processus, mais des symboles émis par les états selon une certaine loi de probabilité. On considère qu'une observation, ie un cycle de marche, est un symbole émis par un état de cluster qui représente la classe des états cachés. Chaque symbole est donc issu d'un état caché dont on n'a pas accès à ce dernier. On fait l'hypothèse que les observations suivent une loi normale multidimensionnelles, de paramètres vecteur moyenne et une matrice de covariance :

Soit :

$\mathbf{X} = X_1, X_2, \dots, X_T$:les états du HMM inconnu

$\mathbf{O} = O_1, O_2, \dots, O_T$:suite des symboles émis par \mathbf{X} tel qu'un symbole $O_i \in \mathcal{C}_j$, sa classe de cluster et $O_i \sim \mathcal{N}(m_j; \Sigma_j)$

Au vu d'une séquence d'observation **O** on ne peut pas savoir par quelle séquence d'états cachés **X** le processus est passé.

L'objectif est donc de trouver la séquence d'états cachés **X** la plus vraisemblable à partir des observations **O**. Pour ce faire, nous allons utiliser l'algorithme de Viterbi qui n'est d'autre qu'un algorithme qui maximise la vraisemblance de la séquence d'états.

Les paramètres des états cachés peuvent être estimés.

b) Prédiction des temps d'arrêt En parallèle, entre deux cycles de marche, on tirera selon une loi aléatoire établie, une durée d'arrêt. La loi de sera déterminée à partir de la distribution des temps d'arrêt sur l'échantillon d'apprentissage. Parmi plusieurs familles de loi, on affectera la famille de loi la plus vraisemblable aux données selon un test de Kolmogorov Smirnov. A ce modèle sera assigné une p-valeur qui traduit la significativité de l'échantillon des durées de temps d'arrêt à suivre cette loi.

Prévisions sur le court terme

2.2.4 Prévision sur le long-terme des cycles de marche : modèle clustering chaîne de Markov sur des profils de clusters

Clustering sur les états de marche

Dans les espaces de très haute dimension, les distances euclidiennes ont tendance à se gonfler dû « fléau de la dimension » ce qui diminue la performance des algorithmes.

On va dans un premier temps tenter de ressortir des motifs temporels types parmi les cycles de marche.

Clustering par k-means On va dans un premier temps appliquer une méthode clustering par k-means. Cette méthode essaye de former des groupes homogènes en séparant les échantillons en k groupes, en minimisant l'inertie à l'intérieur du cluster. L'inertie peut être vue comme une mesure de la cohérence interne des clusters. L'algorithme k-means divise un ensemble d'échantillons en clusters disjoints, chacun décrit par la moyenne des échantillons dans le cluster. Les moyens sont communément appelés le cluster « centroïdes » vivant dans le même espace que les éléments, ici les cycles de marche. L'algorithme K-means vise à choisir des centroïdes qui minimisent l'inertie à l'intérieur du cluster.

- inconvénients Cet algorithme requiert que le nombre de clusters soit spécifié au préalable.

Bien que le k-means est simple à mettre en place et à expliquer, elle n'est pas adaptée à des profils de longueur variable. Néanmoins on sait que les cycles ont une période en moyenne de 20h. Dans ce cas, on pourra appliquer le k means en transformant les cycles sur un temps de fonctionnement moyen de 20h théorique. Avant d'appliquer le clustering par k-means, on resample les tailles des cycles de sorte d'avoir la même taille de séries. Le clustering s'applique sur de séries numériques sans dimension temporel , que nous retrouvons ensuite en profils temporelles de durées de 20h.

L'avantage principal de cette méthode comparé à d'autres méthode de clusterisation, est la possibilité d'avoir facilement des profils de clusters, ce qui nous intéresse avant tout.

Comparer des séries temporelles selon une métrique euclidienne n'est pas le meilleur des choix car on ne prend pas en compte d'éventuels décalages temporels et ne permettrait pas de généraliser des comportements en s'affrichissant des différences de variations d'un

motif de cycle à un autre. Nous allons utiliser une métrique basée sur les algorithmes de déformation temporelle dynamique (dynamic time warping DTW). Elle est en effet adaptée à un problème de séries temporelles en étant plus apte à mesurer la similarité entre deux suites qui peuvent varier au cours du temps.

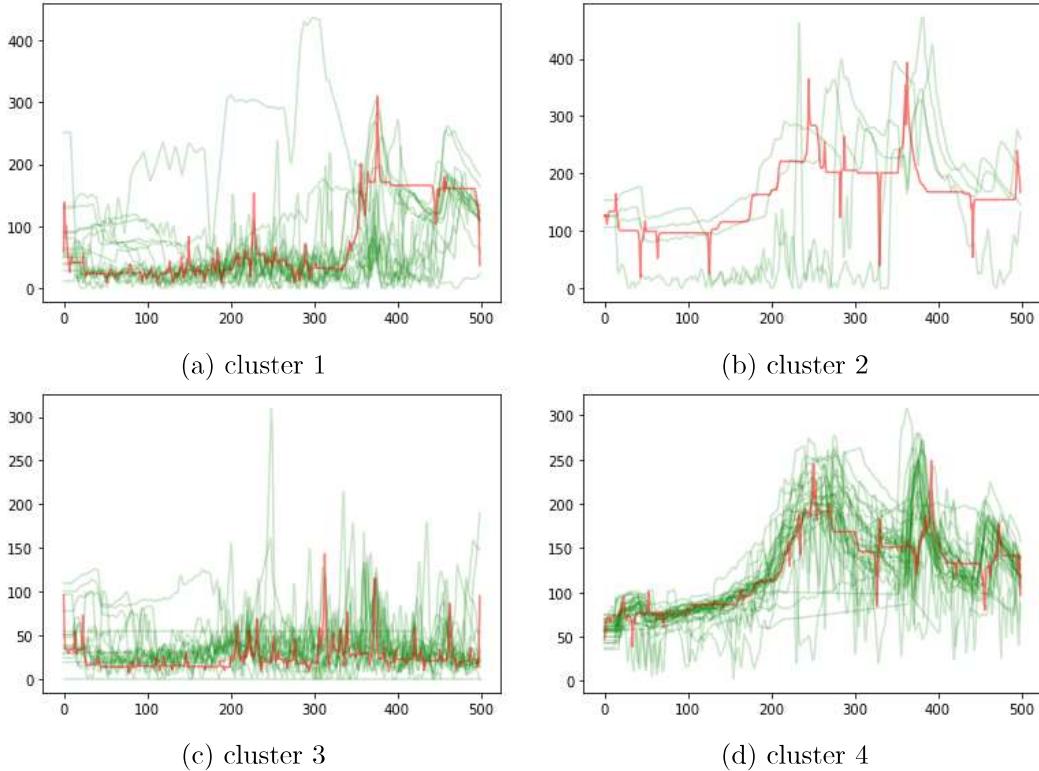


FIGURE 2.8: Clustering à 4 profils de cycles de marche

On a effectué une méthode de clustering k-means avec $k = 4$. On a représenté les centroïdes en rouge parmi les cycles de marche de la même classe. Ces centroïdes sont des classes représentent le profil moyen des cycles d'une même famille : ils formeront la famille discrète profils d'états de la chaîne de Markov.

Les clusters 1 et 3 représentent des cycles de puissance de faible régime ou ponctuel sur de courtes durées. Le cluster 4 est le plus représentatif et contient 45% de l'ensemble des cycles de marche : c'est donc le motif type de fonctionnement le plus commun et représente un régime de fonctionnement dans des conditions normales. On remarque donc sur un clustering k-means, avec $k=4$ choisi arbitrairement, que des motifs de réguliers et distincts se démarquent du jeu de données.

n° clusters	1	2	3	4
fréquence (%)	24	6	25	45

Répartition des profils de clusters

Validation du clustering

- **Qualité du clustering et choix du nombre de clustering : le "silhouette score"**
L'analyse silhouette peut être utilisée pour étudier la distance de séparation entre les agrégats résultants, qui forment des classes plus ou moins distinctes et éloignées des unes des autres. On calcule ce coefficient pour chaque cycle auquel a été affecté

un cluster. Cette mesure a une plage de valeurs [-1, 1] : Les coefficients de silhouette proches de +1 indiquent que le cycle est "loin" des autres , et donc vraisemblablement bien classé. La valeur 0 indique que la classification du cycle au sein d'un cluster n'est pas distinctement départagée, il peut par exemple se trouver très près de la limite de décision entre deux clusters voisins. Finalement les valeurs négatives indiquent que ces échantillons ont peut-être été affectés à un mauvais cluster.

On a vu que le nombre de cluster à choisir est un paramètre à fixer au préalable, ce qui peut ne pas être évident. En effet, plus on augmente le nombre de clusters, plus les éléments constituant les clusters seront plus homogènes. On réduirait à zéro l'homogénéité si on effectuait un k-means avec autant de clusters qu'il y a de cycles, mais le modèle serait alors sur-ajusté. Il y a donc un équilibre à trouver entre généralisation des données en ayant peu de clusters, et bonne classification des éléments dans les clusters, Le silhouette score peut être le juge de paix en choisissant le nombre de clusters selon ce critère de performance.

nb clusters	2	3	4	5	6	10	25
silhouette score	0.437	0.303	0.383	0.415	0.347	0.274	0.090

TABLE 2.1: Silhouette score moyen en fonction du nombre de cluster k pour k-means

- **Erreur d'apprentissage** Pour le k-means optimal calibré à k=2, on obtient les deux motifs suivants :

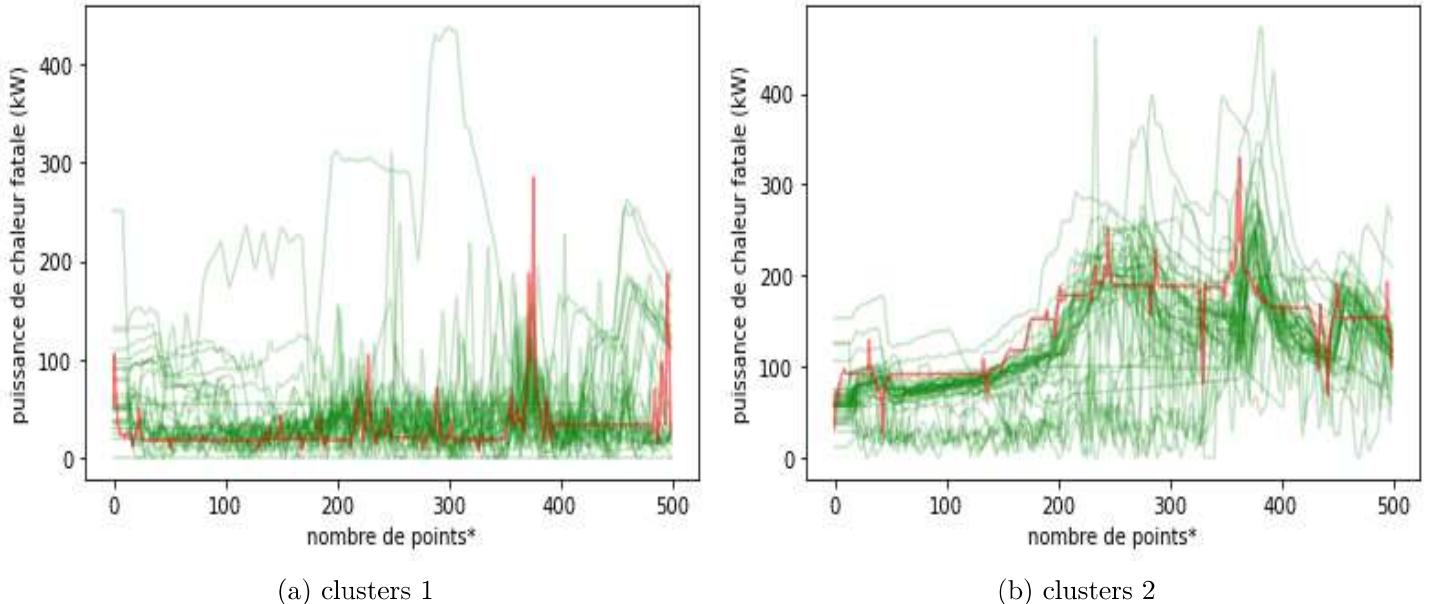


FIGURE 2.9: Profils de clusters (rouge) et les cycles d'états associés (vert)

* les cycles ont été resmplé et interpolé à 500 points pour avoir une taille homogène de cycles.

on calcule pour chaque clusters l'erreur moyenne commise sur la durée du fonctionnement de cycle :

n° clusters	1	2
fréquence (%)	42	58

Répartition des profils de clusters et erreurs commises avec le nombre de cluster optimal

Chaîne de Markov des profils de clusters : construction de la matrice de transition

On estime simplement la matrice à l'aide des fréquences de classes des clusters des données ayant servi à l'apprentissage on obtient cette matrice :

$$M = \begin{bmatrix} 0.64 & 0.36 \\ 0.26 & 0.74 \end{bmatrix}$$

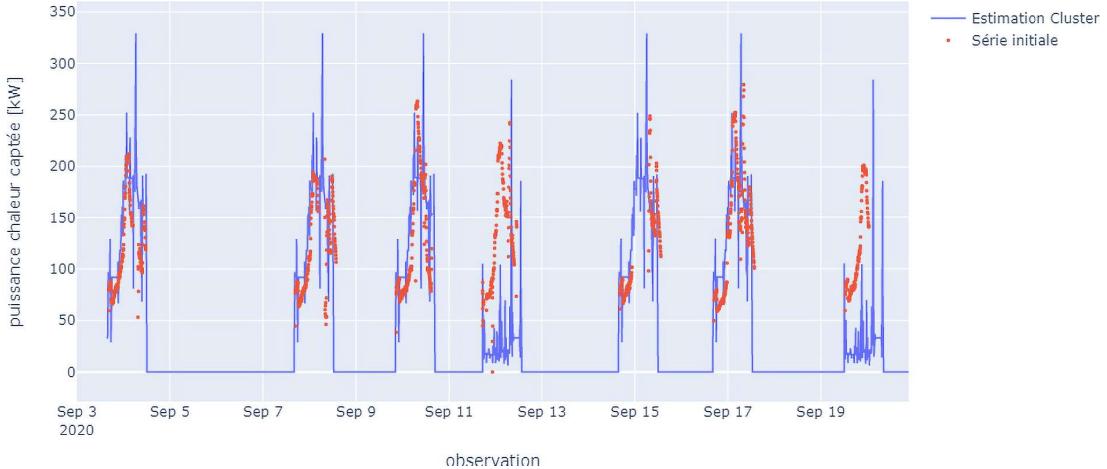
Validation du modèle complet : chaîne de Markov des profils de cluster, indice Energy Score

Une fois la matrice de la chaîne Markov estimée, on peut simuler des séries temporelles de profils de puissances captées durant le temps d'observation. On connaît les temps d'arrêts et on recalcule les cycles de marches sur les temps d'arrêts connus, afin de n'étudier que la performance lié à notre modèle de prédiction de cycle :

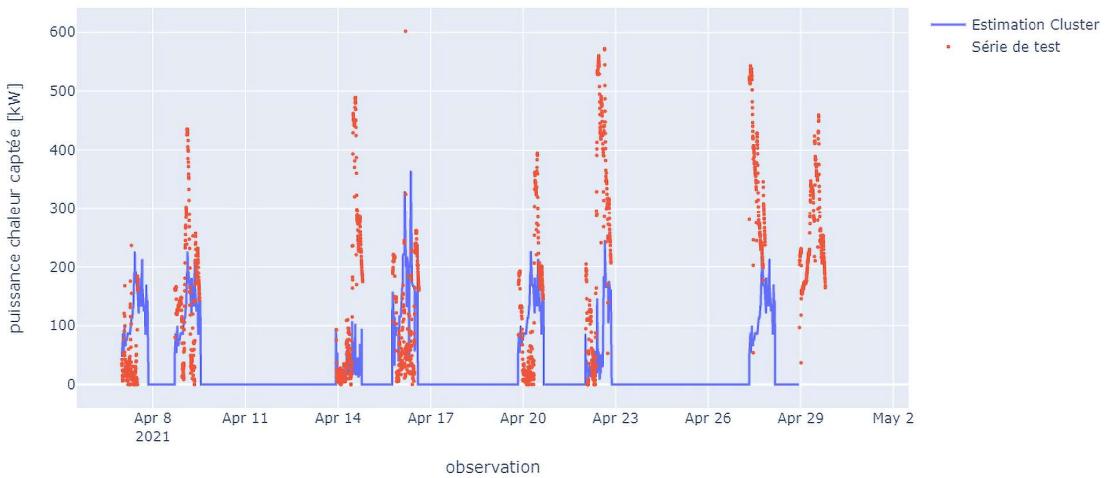
La première remarque que l'on peut faire, est que le modèle actuel n'approche pas les tendances des cycles de marche de l'échantillon dès qu'ils sont hors de la base de données d'apprentissage. Le modèle semble être assez performant sur les données d'apprentissage ce qui était attendu. Mais le modèle doit pouvoir généraliser les profils sur des données nouvelles en dehors du périmètre des données ayant servi à l'apprentissage. Cela est dû à la nature du modèle qui ne ressort pas de nouvelles tendances mais ne fait que reproduire les tendances des profils utilisés à l'apprentissage. Ainsi dès lors que les données d'apprentissage ne sont pas explicatives à l'égard des nouvelles données fournies, le modèle ne sera plus efficace.

Les cycles de production étant de faibles de amplitude dû au contexte de production lié à la crise sanitaire, les cycles étaient anormalement faibles et ne reflétaient pas la véritable tendance dans le cas de fonctionnement standard.

En effet, comparé aux données d'apprentissage les cycles sont de plus grande amplitude, on peut penser que cela est en grande partie due à des productions qui recouvrent leur tendance normale de fonctionnement, et que les cycles précédents furent fortement perturbées par le contexte d'épidémie actuel. Les données des mois précédents ne semblent donc pas être représentatifs des productions de chaleurs fatales des mois à venir. Les données d'apprentissage expliquent ainsi un régime de transition : elles-seules ne suffisent pas à expliquer des événements futures peu corrélés. Elles ne peuvent donc pas être utilisées comme référence pour la prédiction d'événements futures. Ceci est inhérent au modèle et est lié à la nature des données, ce qui néanmoins rend impossible une bonne performance de méthode de machine learning. Ainsi, le modèle ne peut être validé ni réfutée tant que l'ensemble des données d'apprentissage ne se stabilise pas sur des tendances invariantes au court du temps.



(a) sur les données d'apprentissage (mesures septembre-mars)



(b) sur des données test indépendantes (mesures avril)

FIGURE 2.10: Approximation de l'estimation cluster-markov sur des les données d'apprentissage et de test

2.2.5 Prévision sur le long-terme des cycles d'arrêt : estimation de la loi aléatoire sur les temps d'arrêt

Analyse des temps d'arrêts et loi associée

La modélisation des arrêts des fours est indépendante du modèle du prévision de cycles de marche. A partir de la distribution des durées d'arrêt, on va fitter la loi aléatoire la plus proche.

Comme on ne sait pas quelle loi suit la distribution, on va chercher parmi un lot de familles de loi celle qui correspondrait le plus. Les temps d'arrêt étant des valeurs strictement positives, on cherche des lois parmi les familles de loi à support positif telle que les lois exponentielles, de Pareto, et Gamma.

Validation de la loi aléatoire

Pour chaque familles, on calibre la loi la plus proche de la distribution observée dans un premier temps. Ensuite, parmi ces lois calibrées on va sélectionner la loi par un test de Kolmogorov, avec comme hypothèse $F = G_L$ avec F la fonction de répartition des temps d'arrêts observés, et G_L la fonction de répartition de la loi L parmi les familles de loi de

probabilités sélectionnées. La loi la plus vraisemblable étant celle de plus grande p valeur.

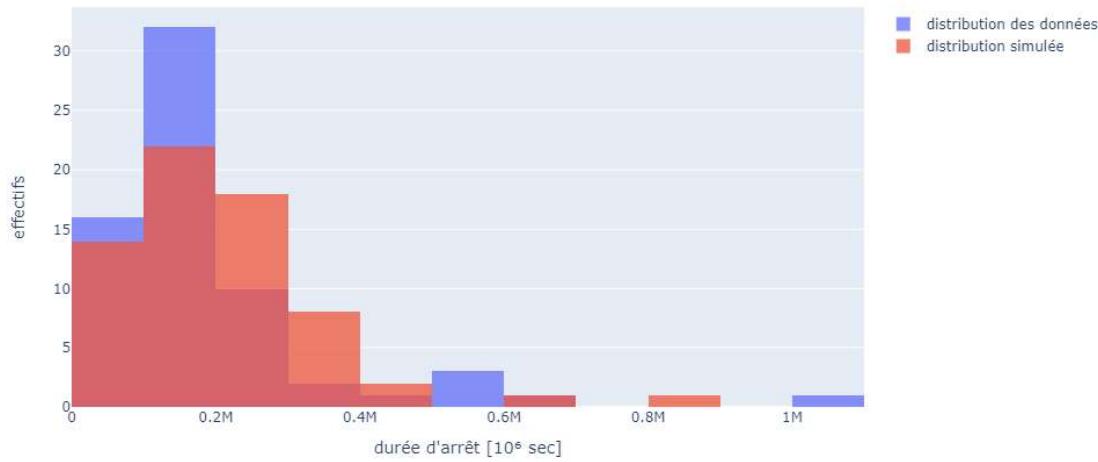


FIGURE 2.11: Approximation de la distribution observée par une loi hypo-exponentielle déduite du test de Kolmogorov, $p\text{valeur}=0.1754$

Les p valeurs sont supérieurs à 0.17. On ne rejette pas l'hypothèse que la durée des cycles suit la loi hypo-exponentielle en question.

2.2.6 Résultats de prévision long-terme de série de production

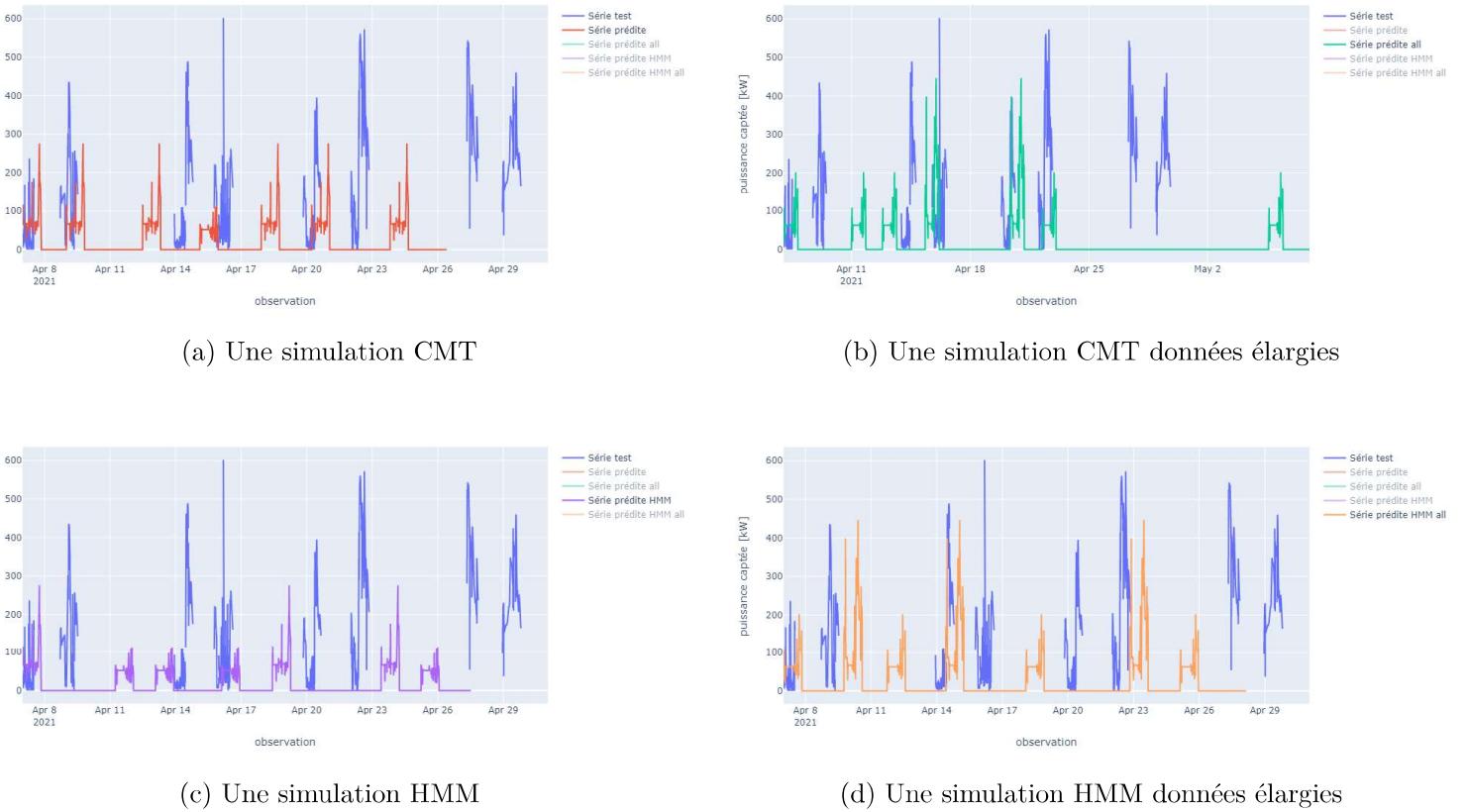


FIGURE 2.12: Comparaison qualitative entre les méthodes de prédictions

On a simulé plusieurs fois des séquences de productions à l'aide de notre modèle Cluster Markov pour le mois de mai, à partir des données des mois antérieurs. La première remarque que l'on peut faire, est que le modèle **a)** n'approche pas les tendances des cycles de marche de l'échantillon de test en amplitude. Les données des mois précédents ne semblent donc pas être représentatifs des productions de chaleurs fatales des mois à venir. Les données d'apprentissage expliquent ainsi un régime de transition : elles-seules ne suffisent pas à expliquer des évènements futures peu corrélés. Elles ne peuvent donc pas être utilisées comme référence pour la prédition d'évènements futures. Ceci est inhérent au modèle et est lié à la nature des données, ce qui néanmoins rend impossible une bonne performance de méthode de machine learning. Ainsi, le modèle ne peut être validé ni réfutée tant que l'ensemble des données d'apprentissage ne se stabilise pas sur des tendances invariantes au court du temps.

On a simulé une série avec la même méthode mais en élargissant les données (en incorporant les données test dans les données d'apprentissages) **b)**. On remarque une légère amélioration des cycles simulés, ce qui renforce l'importance d'avoir des données d'apprentissages représentatives.

Les simulations **c)** et **d)** réutilisent respectivement les mêmes jeux de données que **a)** et **b)**, en utilisant maintenant une méthode HMM. Tout comme le modèle **a)**, on remarque que le modèle **c)** ne semble pas être qualitativement adapté au jeu de test. En adaptant le jeu de données, la simulation **d)** semble mieux estimer les données test. Le modèle HMM étant une amélioration du modèle Markov Clusterisé, corrigéant le bruit pouvant

être transmis entre les transitions en déterminant la séquence la plus vraisemblable.

Cette analyse qualitative n'étant pas précise, on décide d'utiliser un critère pour déterminer les paramètres optimaux pour un modèle donné, et de comparer ensuite les modèles entre eux.

Energy Score Plan d'expérience sur les différents paramètres, paramètres optimaux :

Comme critère de comparaisons de méthodes de prédictions de série temporelles, nous utilisons l' Energy score ¹, qui est orienté négativement.

$$ES = \sum_{j=1}^J \|z - \hat{z}_j\|_2 - \frac{1}{2} \sum_{i=1}^J \sum_{j=1}^J \|\hat{z}_i - \hat{z}_j\|_2$$

Ce critère mesure à la fois la précision et la dispersion d'un modèle générateur de scénario. Pour une observation z , on génère un nombre J de simulations \hat{z} . On mesure à la fois l'écart entre l'observation et les simulations, mais aussi les simulations entre elles. Pour chaque observation, on obtient alors un energy score traduisant la performance du modèle utilisé. L'energy score peut donc être utilisé dans le choix des paramètres optimaux des modèles. Dans le cas du modèle CMT défini précédemment, on va déterminer les valeurs optimales des différents paramètres à l'aide de l'Energy Score. Puis dans un second temps, on comparera les méthodes CMT avec les paramètres optimales avec les méthodes HMM et définir un classement sous ce critère-là. On prend un nombre de simulation $J = 100$ pour un modèle donné.

métrique	données d'apprentissage	nombre de clusters	seuil température
{dtw, euclidienne}	{élargies, non élargies}	{2,4,6,8,10}	{230,235,240,...,265}

Plan d'expériences pour déterminer les paramètres optimaux du modèle Cluster-Markov-Temps aléatoire CMT

En calculant l'energy score associé à chaque cycle et en faisant la moyenne, on obtient un energy score moyen pour chaque set de paramètres du plan d'expérience ci-dessus. On a déduit que les energy scores étaient meilleurs pour des choix de seuil de filtrage de cycle à 265°C. La métrique dtw utilisée dans la construction des meilleurs donnait de meilleurs energy scores. De plus, comme on pouvait s'y attendre, les modèles utilisant comme données d'apprentissages toutes les données avaient également un meilleur energy scores. Les modèles Néanmoins, on remarque une absence de tendance forte quant au choix du nombre de cluster mais les modèles avec un nombre élevé de clusters (6 8) ont de meilleurs ES.

On compare maintenant les trois méthodes de prévision longue présentées, avec les paramètres optimaux pour chaque méthode. Seul la méthode des HMM présentée a été utilisée sans initialisation, les profils de clusters et les matrice de covariance ont été estimés à l'aide des profils des cycles de chaque clusters

1. Staid et al. (2017)

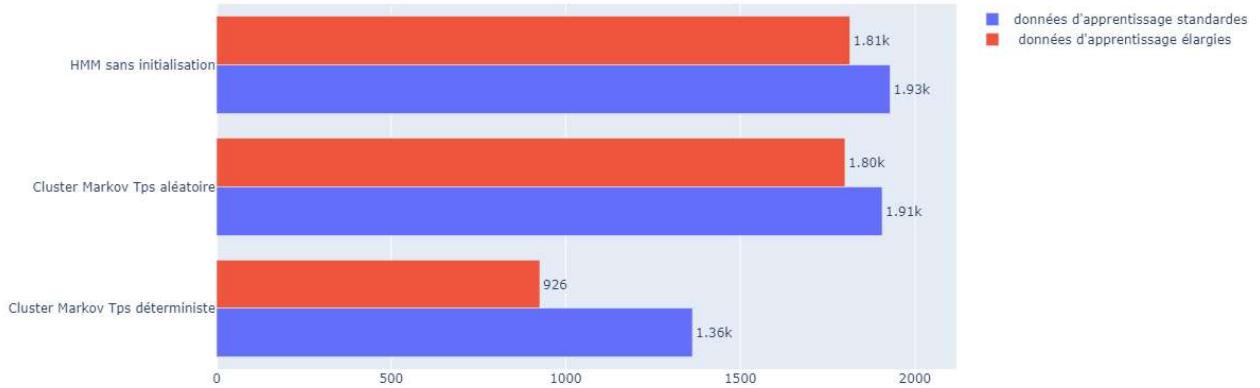


FIGURE 2.13: Comparaison des meilleurs modèles issus des méthodes précédentes

L'étude des energy scores confirme que la méthode des chaines de markov avec le temps d'arrêt connu commet moins d'erreurs que les autres méthodes simulant les temps d'arrêt. On peut estimer également l'apport d'incertitudes apportés par la simulation des temps d'arrêt en comparant les energy scores à partir des données d'apprentissage standardes (en bleu) de la méthode avec et sans simulation des temps d'arrêts. Ainsi en confrontant l'enrgy score des méthodes Cluster Markov à temps aléatoire et à temps déterministe, on remarque qualitativement qu'une grande partie de l'erreur est due à la simulation aléatoire temps d'arrêt, représentant pratiquement un tiers des erreurs dans le cas des données d'apprentissage non élargies (bleu). On remarque que cette part d'erreur lié à la simulation des temps d'arrêts n'est plus vraie pour le cas des données d'apprentissage élargies (rouge), elle est même plus grande. Ceci peut s'expliquer par le fait que les données de test ne sont plus indépendante des données d'apprentissage, il y a une corrélation entre les deux sets de données, ainsi les modèles de prédiction des états sont anormalement meilleurs.

De manière générale, on vérifie que l'energy score est plus petit dans le cas des données d'apprentissage élargies, ce qui est cohérent car less données test sont corrélées aux données d'apprentissage élargies.

Le modèle de Markov caché utilisé n'apporte pas d'améliorations nets. La correction apportée sur la séquence des états ne semble pas réduire les erreurs du modèle. Le modèle de Markov caché ici n'a été utilisé que dans une optique de correction de la séquence, en réutilisant les mêmes paramètres par le modèle précédent de markov clusterisé. Une piste d'amélioration est de ne pas ré-utiliser les paramètres du modèle précédent, mais de les ré-estimer par une méthode de espérance-maximisation tel qu'algorihme de Baum-Welch.

2.2.7 Conclusions sur les différents modèles de prédiction pour le cas de production industrielle

Pour répondre à la problématique de prédiction dans le cas de production industrielle, nous avons implémenté dans un premier temps une méthode de simulation de chaîne de Markov discrète sur des profils moyens obtenu par clustering.

Les méthodes de prédictions implémentées peuvent être utilisées pour de la planification sur des temps de prévisions sur un trimestre par exemple pour visualiser des possibles scénarii de production. Une limite rencontrée durant la conception des méthodes fut qu'au vu du caractère évoluant de l'entreprise et du peu de jeux de données en main, il est difficile de fournir des prédictions basées uniquement sur les données des temps précédentes. En effet, l'entreprise étant une petite entreprise, elle est potentiellement dans une période de

croissance plus ou moins grande, ce qui finalement peut se répercuter sur son rythme de production. Avec le peu de visibilité sur les données, il est difficile d'intégrer ces tendances de croissance pour avoir des prédictions justes. La qualité et un large intervalle d'observations pour construire une solide base de données d'apprentissage sont les principales limites des méthodes développées dans cette partie. La piste possible est donc de compléter le jeu d'apprentissage jusqu'à avoir un nombre conséquent de données assez stables en tendance. En vert figure la série simulée avec les données d'apprentissage antérieures. Tandis qu'en jaune, on a représenté les séries simulées en fusionnant les données de test avec les données d'apprentissage. Ainsi, en complétant la base d'apprentissage précédente par le jeu de données test, on obtient des estimations de profils de cycles de marche plus proche. Or, les dernières données de l'entreprise Tégulys ne seront pas exploitable. Actuellement la base de données ne sera pas étoffée d'avantage. Autre chose, le caractère transitoire est propre à l'entreprise car étant de petite taille, elle est fondamentalement en pleine évolution, on ne pourra attendre d'avoir des données stationnaires dans le temps et stables. Vouloir prédire uniquement à partir des observations passées des productions précédentes ne semble pas être intéressant étant donné du caractère variable des process. Les incertitudes du process sont causées par plusieurs facteurs aléatoires externes à l'entreprise identifiable. La fréquence variable de mise en marche des cycles dépend en grande partie du carnet de commande. Avoir accès aux demandes permettrait par exemple de rendre complètement déterministe le départ des cycles par exemple si les commandes sont données suffisamment tôt par rapport au dimension de stockages.

Lors de la comparaison des méthodes par energy score, on a pu conclure que l'étape de simulation d'un temps aléatoire peut représenter une partie non négligeable de l'incertitude.

Une des améliorations de l'étape d'estimation d'états discrets, serait d'utiliser d'autres méthodes de clustering. La méthode k-means est intéressante car on peut facilement déduire les centroïdes, les profils moyens du cluster, qui seront ensuite utilisés comme états de la chaîne de Markov. Mais cette méthode k-means exige d'avoir des éléments, ici les cycles de marche, de même taille, or cela n'est pas vérifié et une phase intermédiaire de renormalisation des cycles a due être nécessaire.

Dans un second temps, nous avons tenté d'améliorer cette méthode primaire par un chaîne de markov caché, ce qui n'a pas été concluant. Plus précisément l'amélioration résidait dans la réestimation de la séquence des états cachés simulés par l'algorithme de Viterbi. En effet, on n'a pas utilisé le modèle de Markov caché à son plein potentiel. Les paramètres du modèle de Markov caché, notamment les moyennes et matrice de covariances des états cachés, sont les mêmes paramètres du modèle de la chaîne de Markov des états clusters. Il faudra faire un calibrage plus fin des paramètres du modèle de Markov caché, en procédant par une étape de maximum de vraisemblance.

Même si cela n'a pas été détaillé, les méthodes de simulation par chaîne de markov sur des états de clusters peuvent être utilisées pour la prévision en temps court, pour prédire les tendances de cycles en cours par exemple. Parmi les profils de clusters, on pourrait trouver le cluster qui minimise l'écart avec le cycle en cours, et utiliser la suite du cycle comme prévision. Les méthodes de prévisions longues peuvent être adaptées pour répondre à des questions de prévisions courtes, mais ne sont pas les méthodes les plus performantes. Ainsi, une amélioration certaine serait de fournir des modèles pour des prévisions courtes.

2.3 Un exemple de modèle de consommation de production industrielle : le cas d'un réseau de chaleur

On a pu trouver dans la littérature des données opensources dans le cas de consommation d'énergie thermique s'inscrivant dans le cadre de notre étude .¹ La série temporelle ci-après représente à l'échelle nationale, la demande en chaleur totale, c'est-à-dire à la fois pour le chauffage sanitaire et spatiale de 2008 à 2013. La série temporelle mélange les consommations de plusieurs profils de consommateurs dont les maisons individuelles, les bâtiments multi-logements et les bâtiments de commerce. De plus, la demande thermique de ces profils dépendent de trois facteurs principaux : des conditions météorologiques, du type d'infrastructure et du comportement propre au consommateur [citer opendata]. Le calcul de ces demandes a été réalisé dans un premier temps comme moyenne pour des grilles géographiques 28km par 17km, afin de prendre en compte le caractère locale de la consommation, une personne habitant au sud de la France n'ayant pas les mêmes besoins thermiques qu'une personne au nord par exemple. Ces moyennes sont en suite remis à l'échelle en prenant en compte le nombre d'habitant pour chaque portion d'espace. La série temporelle suivante une fois traitée est l'équivalent d'un réseau de chaleur à très grande échelle, avec par conséquent ses propres dynamiques et ses inerties. En effet, la dimension d'un réseau de chaleur va influer sur les dynamiques et ainsi sur les paramètres du modèle de prédiction. Il est donc important de bien identifier les caractéristiques du réseau. En particulier, on pourrait penser que les fluctuations très courtes d'un réseau de chaleur de plus petite tailles seraient gommées .

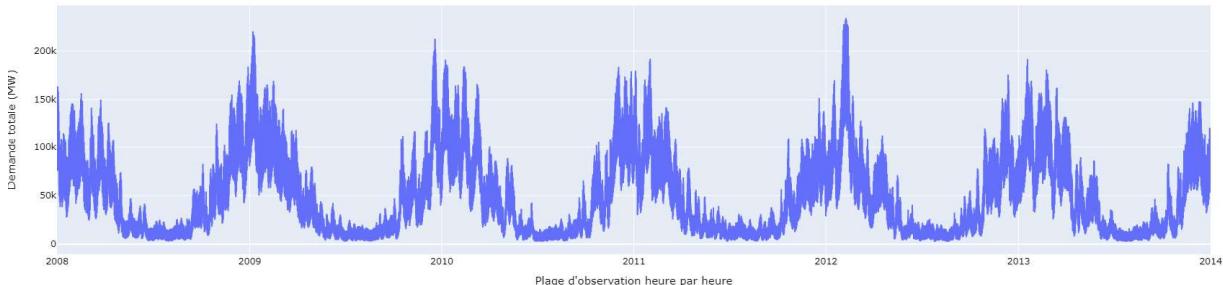


FIGURE 2.14: Consommation des réseaux de chaleur pour l'eau chaude sanitaire et le chauffage en France sur 6 ans

2.3.1 Les différentes dynamiques temporelles du réseau de chaleur

Cette série temporelle présente des saisonnalités à plusieurs échelles temporelles. Tout d'abord, la demande en chauffage étant induite par les conditions météorologiques elles-mêmes saisonnières, on retrouve cette périodicité annuelle dans la série temporelle a). En effet, en moyenne les demandes de chauffages en période d'été est inférieure comparé en période hivernale. Par ailleurs, des fluctuations dans la journée peuvent exister b), ce qui rajoute une deuxième périodicité journalière dans les demandes. On distingue facilement un motif de demande de chauffage qui se repète chaque jour . En outre e motif journalier de demande peut être modulé, de jour en jour, selon une tendance propre à des conditions extérieures. D'une semaine à une autre par exemple, des épisodes de vague de froid peuvent être à l'origine de ces fluctuations, lentes par rapport à la saisonnalité journalière, mais rapide comparé à la saisonnalité sur l'échelle de l'année.

1. <https://data.open-power-system-data.org/when2heat/2019-08-06>

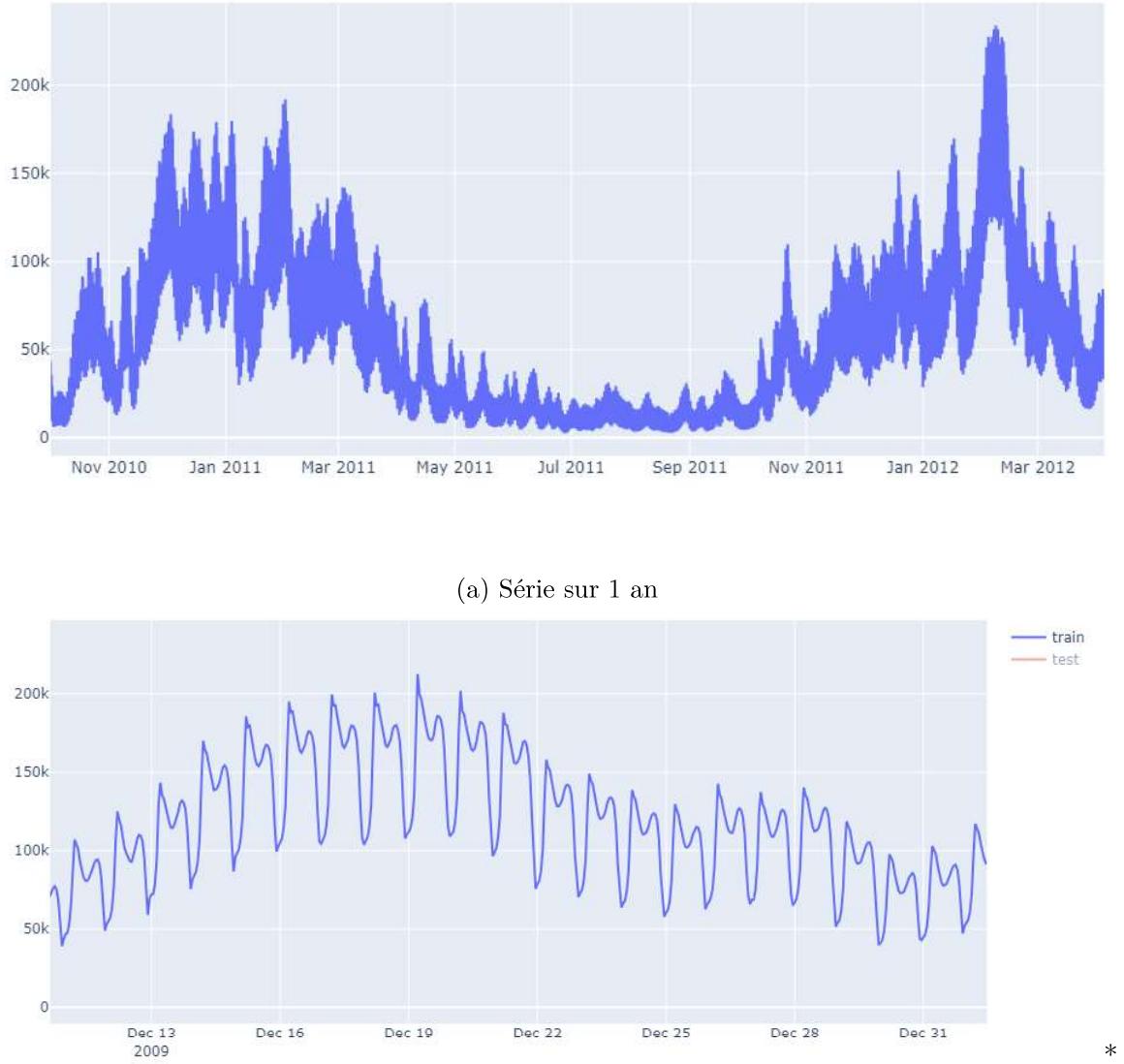


FIGURE 2.15: Dynamique du réseau de chaleur sur différentes périodes de temps

La série temporelle suivante présente donc à plusieurs échelle de temps de nombreuses périodicité, qui devront être nécessairement pris en compte dans le modèle de prédiction.

2.3.2 Modèles de prédiction

Méthode naïve : prédiction en répétant le dernier cycle journalier connu

Une idée de départ serait de simplement prédire l'horizon de prédiction par des cycles passées de N heures auparavant, sans ajouter de nouveautés. On estime ainsi la demande à l'instant t en reprenant l'instant $t - N$ heures. La question étant de bien fixer le décalage N , on part du principe qu'il n'est pas nécessaire de prédire sur un horizon trop large, car il doit rester cohérent par rapport aux capacités de stockages des batteries dans le réseau multi-énergies. Les capacités actuelles étant de l'ordre de la journée, on va tester des modèles de prédiction sur deux horizons de prédiction assez courts : de l'ordre de la journée et de l'ordre de la semaine.

La figure ci-dessus représente deux méthodes de prédiction naïve pour les deux horizons de prédictions présentées. L'estimation en vert représente une prédiction basée sur la semaine

Consommation des réseaux de chaleurs pour l'eau chaude sanitaire et le chauffage en France sur 6 ans

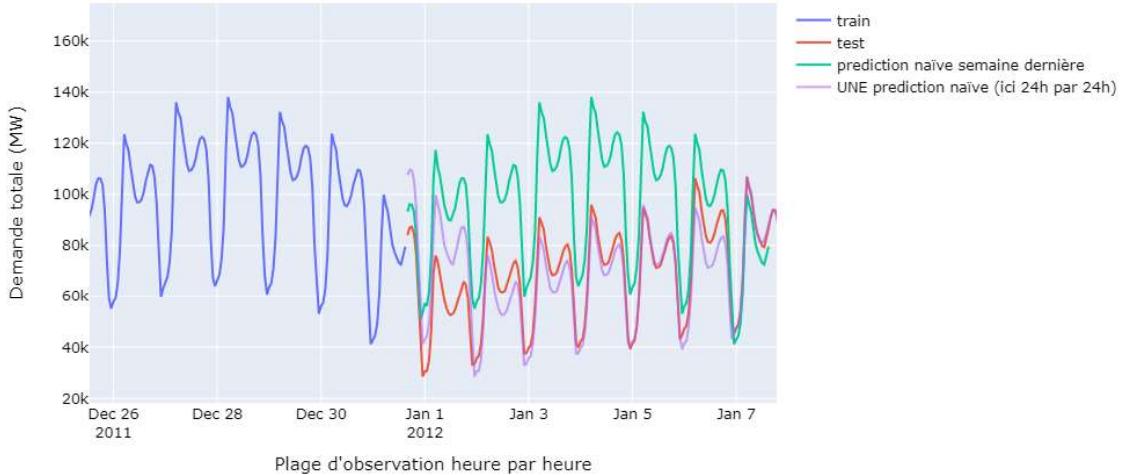


FIGURE 2.16: Prédiction par la donnée 24h auparavant

dernière, et l'estimation en mauve étant une prédiction naïve sur la journée. Pour cette dernière estimation, la prédiction étendue sur une semaine, en partant du principe qu'à chaque jour, la donnée test du jour précédent est accessible et réutilisée dans l'estimation.

On remarque que les estimations naïves, sur une période donnée, donnent des premiers résultats passables en ne prenant en compte que les données décalées. Les deux estimations captent bien la saisonnalité journalières mais présentes des décalages

L'estimation sur la journée est évidemment plus précise car elle capte davantage les fluctuations entre les jours. Le pas de temps étant plus grand pour l'estimation sur la semaine, le décalage entre les données test est bien plus grand.

Méthode basé sur les process ARMA(p,q) et SARMA(p,q,s)

En statistique, les modèles ARMA (AutoRegressive and Moving Average models), sont les principaux modèles de séries temporelles.

En se donnant une série temporelle X_t , le modèle ARMA est un outil pour comprendre et prédire, éventuellement, les valeurs futures de cette série. Le modèle est composé de deux parties : une partie autorégressive (AR) et une partie moyenne-mobile (MA). Le modèle est généralement noté ARMA(p,q), où p est l'ordre de la partie AR et q l'ordre de la partie MA. Un processus temporel X_t peut donc être modélisé par un processus ARMA(p,q) comme suit :

$$X_t = \varepsilon_t + \sum_{i=1}^p (\varphi_i \varepsilon_{t-i} + \sum_{i=1}^q (\theta_i X_{t-i}))$$

où les φ_i et θ_i sont les paramètres du modèle et les ε_i les termes d'erreur tel que $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. La première étape est de déterminer les ordres p et q du modèle. Le modèle SARMA(p,q,s) similaire au modèle ARMA(p,q) est plus adapté pour les séries avec une saisonnalité s. Il ne prend en compte que les termes antérieurs, multiple de s.

$$X_t = \varepsilon_t + \sum_{i=1}^p (\varphi_i X_{t-si} + \sum_{i=1}^q (\theta_i \varepsilon_{t-si}))$$

Une des grandes questions dans l'étude de séries temporelles est de savoir si celles-ci suivent un processus stationnaire. On entend par là le fait que la structure du processus sous-jacent supposé évolue ou non avec le temps. Si la structure reste la même, le processus est dit alors stationnaire. Si la série ne l'est pas, il est important de la stationnariser avant d'appliquer le modèle ARMA(p,q). Une manière de la stationnariser est d'estimer les différentes tendances non stationnaires, puis de les retrancher. Les modèles ARIMA(p,d,q) sont des variantes incluant un terme de différenciation d'ordre d entre les termes $X_t, X_{t-1}, \dots, X_{t-d}$, qui permet de diminuer la non stationnarité de la série.

Pour savoir si la série d'étude vérifie la propriété de stationnarité, on a recours au test augmenté de Dickey-Fuller. Ce test vérifie si l'hypothèse de stationnarité est valide, nous donnant une p-valeur p associée à la conviction que l'on peut porter sur l'acceptation de l'hypothèse. Plus la p valeur est petite, plus on peut accepter l'hypothèse avec conviction. A l'aide de la fonction python `statsmodels.tsa.stattools.adfuller`, on obtient $p = 5 * 10^{-6}$: on peut accepter avec conviction l'hypothèse de stationnarité de la série.

Un moyen pour déterminer approximativement les ordres p,q est de recourir aux graphes d'autocorrélation et d'autocorrelations partielles.

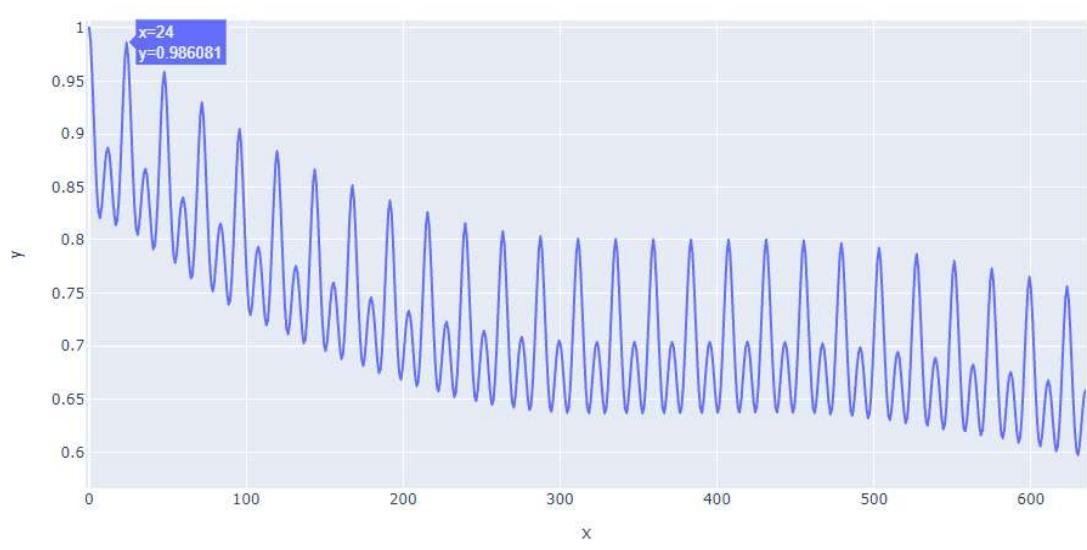
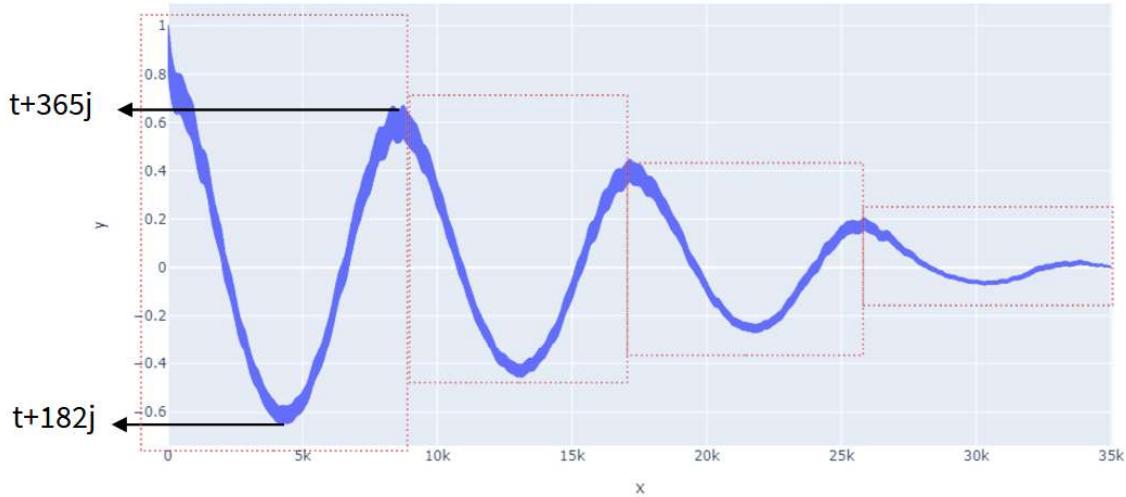


FIGURE 2.17: Graphe d'auto-corrélation

Ces graphes représentent la corrélation entre la série et la série décalée d'un certain ordre, dit "lag". A titre d'exemple, pour calculer le coefficient d'autocorrélation d'ordre k , on calcule la corrélation entre la série donnée, et elle-même décalée de k éléments. Pour $k=0$, on obtient une autocorrélation de 1 qui est le maximum, ce qui est logique car aucun décalage a été fait. L'autocorrélation pour un process MA(q), est nulle au-delà du décalage q . On remarque que les pics d'autocorrelations se situent à des multiples de 365 jours et 182 jours. On vérifie donc la corrélation positive de la demande entre un instant t et la demande antérieure à des multiples de 365 jours et la corrélation négative entre t et des multiples de 182 jours auparavant, dont cette dernière correspond à la corrélation négative entre la demande en période estivale et hivernale. Le modèle SARMA pourrait être pertinent car, pour une saisonnalité s fixée, il prendrait en compte les termes X_{t-s}, X_{t-2s}, \dots sans prendre en compte les termes intermédiaires, ce qui diminue considérablement la complexité du modèle. Les trois pics multiples de $t-365j$ - les trois pics multiples de $t-182j$ - contiennent le plus d'informations de corrélation, un choix serait donc de prendre $q = 6$

En regardant en détail le graphe pour les premiers termes, on retrouve bien une corrélation entre X_t et X_{t-24} qui correspond à la demande 24h auparavant. Un modèle SARMA avec une période $s = 24$ serait un modèle intéressant à tester

Le graphe d'autocorrélation nous confirme ainsi des hypothèses de choix d'ordres d'autorégression q et le paramètre de périodicité pertinents des process ARMA(p,q) et SARMA(p,q,s) avec $s=365j = 24 * 365h$.

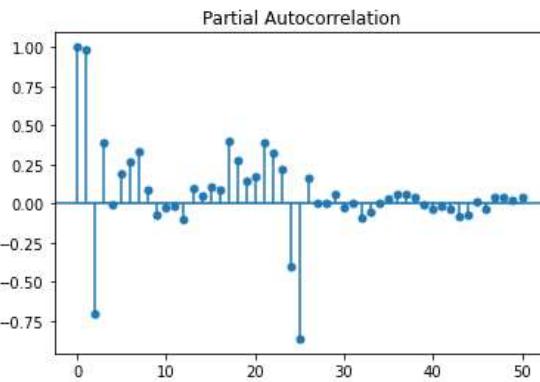


FIGURE 2.18: Graphe d'auto-corrélation partielle

L'autocorrélation partielle PACF (Partial AutoCorrelation Function) est une autocorrélation auquel on a enlevé l'effet de corrélation des décalages précédents. Le coefficient d'autocorrélation partielle d'ordre k , correspond à la corrélation entre les éléments X_t et X_{t+k} en ayant au préalable annulé les corrélations de X_t et des décalages strictement inférieur à k , $X_{t-1}, X_{t-2}, X_{t-3}$ etc...

L'autocorrélation partielle (PACF) sera préférée à l'autocorrélation classique pour choisir l'ordre de process AR(p), car des propriétés des process AR(p) est d'avoir une PACF nulle au-delà de l'ordre p .

D'après la figure ci dessus, on voit que la corrélation chute au delà du 24e pic, donc un choix empirique serait de prendre $p=24$ pour simuler la série par un process AR(p).

Ainsi, nous avons testé le modèle SARMA(24,6,24*365) sur les données tests : pour un horizon d'une semaine . Comparé au modèle de prédiction naïve d'horizon une semaine également, la prédiction par le modèle SARMA est de meilleure qualité.

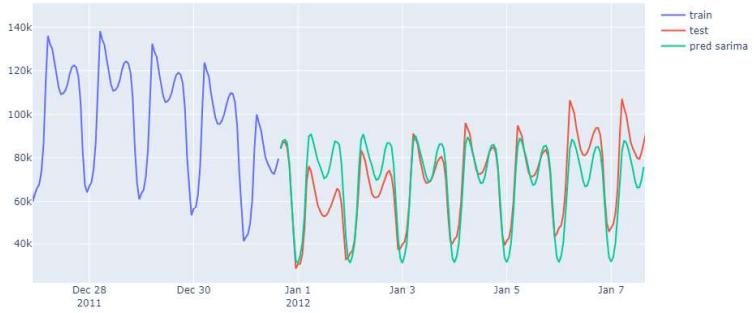


FIGURE 2.19: Prediction SARMA(24,6,24*365)

2.3.3 Conclusions et pistes d'améliorations pour la prédition de série temporelles de consommation thermique

Dans le cadre de cette partie, nous avons tenté d'établir des modèles de prédition idoine à la nature des données acquises. En analysant la forme des données, on a décidé de prédire les données à l'aide de modèles ARMA et SARMA, modèles pertinents dans le cadre d'étude de séries temporelles saisonnières. Pour un temps d'horizon fixé, on a implémenté un premier modèle SARMA après avoir fixer les ordres à l'aide des graphes d'autocorrélations. Néanmoins, ce choix un peu empirique doit être confirmé de manière plus précise.

Pour les améliorations futurs, il reste encore à choisir le meilleur modèle ARMA/SARMA et d'implémenter une méthode de choix des ordres de ces process, à travers par exemple un critère AIC pour trouver la meilleur combinaisons des ordres. On pourra alors affirmer ou réfuter le choix d'ordre pris précédemment et conclure. Ensuite, de la même manière dans le cas des modèles établis pour Tégulys, on pourra calculer les energy scores associés aux modèles en question. Un point important serait également de quantifier l'incertitude de ces modèles et d'attribuer un pourcentage de fiabilité des prédictions émises.

3

Les conclusions concernant l'étude de modèle de prédition dans un contexte industriel

Au cours de mon stage, j'ai dû fournir des modèles de prédition pour des séries temporelles de puissances des chaleur fatales dans le milieu industriel, tant dans un contexte de production que de consommation. Ces modèles permettraient de générer des scénarii prédisant les évolutions de puissance thermiques produites ou consommées à des fins de contrôle d'un réseau multi-énergies. Les prévisions issus de ces modèles seraient données en entrée à un algorithme ayant pour but de fournir les bonnes commandes sur les stockages thermiques et électriques sur la base de ces données recréées. Ces prédictions substitueraient alors les séries temporelles de puissance sur les chaleurs fatales qui sont à ce jour encore peu accessibles dans le milieu industriel thermique. On a pu voir que les données d'apprentissages utilisées constituent une part importante de la performance des modèles de préditions. Au vu de la situation actuelle et de l'importance des questions concernant la gestion de l'énergie, il est très probable que les problématiques liées à l'acquisition et le traitement des données prennent de plus en plus d'ampleur à leur tour. Ainsi de manière plus globale, développer l'accès et le traitement de ces données serait un levier d'amélioration des prédictions de ressources énergétiques, thermiques ou pour d'autres sources. A ce jour, le développement de modèles de prédition dans ce domaine est une question ouverte et des recherches sont faites dans ce sens.

Durant mon stage, j'ai pu apporter des premières réponses sur le choix des familles de prédition en fonction du problème et des contraintes associées. De plus, Il a été également souligné l'importance de la bonne compréhension physique du problème. Une part non négligeable de la physique du problème rentre en compte dans les modèles de prédition et doivent être intégrer pour tenter de saisir les dynamiques du problèmes. En effet, des premières pistes de méthodes de prédition à utiliser ont été présenté en fonction du problème, et plus précisément des aléas rencontrés dans le milieu industriel. Nous avons présenté deux modèles En étudiant plusieurs type de données en annexe, on a identifié deux aléas majeurs dans le problème de prédition de puissances : les aléas sur la process et/ou les aléas sur le besoin :

- **Existence d'aléa uniquement sur le process :**
Appliquer modèle de type Clustering Markov Caché.
- **Lorsque le process est déterministe (le déroulement est entièrement connu), mais les aléas dépendent que du besoin :**
On peut construire une loi aléatoire sur la variable en question. Par exemple, dans un process où le départ est inconnu mais que le profil de chauffe est déterminé, à partir d'estimation statistique, on estime la loi aléatoire qui simule la date et on peut tirer une date départ du cycle.
- **Aléa sur le besoin avec des saisonnalités et/ou de l'auto-corrélation :**

Utiliser un process type ARMA ou SARIMA.

J'ai présenté deux modèles pour deux cas d'études distincts, permettant de parcourir toutes les problématiques de modèles prédictifs ; et les cas de figures industriels possibles rencontrés à ce jour. Le premier modèle concernait un problème lié à de l'aléa sur le process mais aussi sur le besoin : le modèle appliqué fut un mélange des modèles présentées ci-dessus. Le second modèle fut développé dans le cas où les données présentent des périodicités et des autocorrélations dans la série temporelle de consommation dans un réseau de chaleur à l'échelle nationale. Les pistes d'améliorations possibles seraient d'améliorer les méthodes en étoffant les données d'apprentissages.

J'ai pu également les adapter, et implémenter un critère de performance basé sur l'energy score quantifiant la performance du modèle. De manière globale, ce critère de performance est une mesure pour comparer la performance des modèles entre eux. Ce critère peut à la fois renseigner sur la variabilité du modèle et donner une incertitude lié au modèle, ce qui est une amélioration à apporter.

Sur le plan personnel, ce stage m'a permis de m'apporter beaucoup de réponses quant aux choix de carrière et sur le type de poste auxquels je pourrais candidater. Avant d'effectuer ce stage, je n'avais pas d'expériences dans le milieu de la recherche bien que cela m'attirait particulièrement. J'ai pu à la fois travailler dans un milieu de recherche dans un laboratoire performant, et à la fois dans un milieu plus industriel ou ingénieur au sein de l'entreprise ETC. Ce stage a confirmé mon désir de travailler dans un milieu où la recherche a une place importante. J'ai pu découvrir de nouvelles cultures d'entreprises, ce qui nourrit ma perception du monde du travail et me permet de mieux construire mon projet professionnel. J'ai pu être témoin que la recherche peut s'inscrire profondément dans des problématiques concrètes issus de problèmes physiques et industriels d'actualité, ce qui m'a particulièrement plus. Enfin, pouvoir parler à plusieurs interlocuteurs de différentes spécialités, est une qualité que j'ai pu développer tout au long de mon stage.

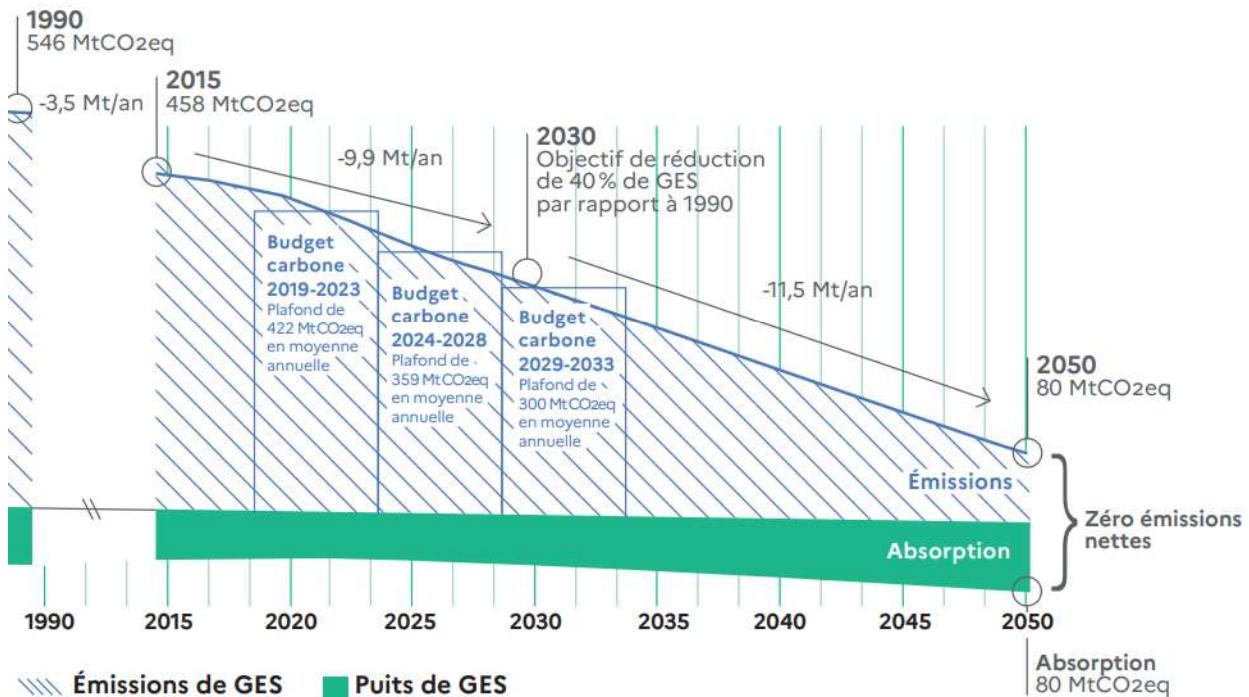
Bibliographie

- A. Arteconi, N. J. H. and Polonara, F. State of the art of thermal storage for demand-side management. 93 :371–389.
- Favre-Perrod, P. (2005). A vision of future energy networks. *n Proceedings of the Inaugural IEEE PES Conference and Exposition in Africa*, pages 13–17.
- Geidl, M. and Andersson, G. (2007). Optimal coupling of energy infrastructures. *IEEE Lausanne Power Tech Meeting*, pages 1398–1403.
- Kienzle, F. and Andersson, G. (2009). A greenfield approach to the future supply of multiple energy carriers. *Power Energy Society General Meeting*, pages 1–8.
- M. Geidl, G. Koeppel, P. F.-P.-B. K. G. A. and Fröhlich, K. The energy hub-a powerful concept for future energy systems. pages 13–14.
- Ministère de l'Europe et des affaires étrangères (2021). Déclaration conjointe des ministres de l'Énergie des États-unis et de la france sur la technologie énergétique et la résolution politique (28 mai 2021). <https://www.diplomatie.gouv.fr/en/french-foreign-policy/climate-and-environment/news/article/joint-statement-of-the-united-states-and-france-energy-ministers-on-energy>.
- N. Neyestani, M. Yazdani-Damavandi, M. S.-k. G. C. and Catalao, J. P. S. (2015). Stochastic modeling of multienergy carriers dependencies in smart local networks with distributed energy resources. *IEEE Trans. Smart Grid*, 6(4) :1748–1762.
- P. Haessig, B. Multon, H. B. A.-S. L. and Bondon, P. (2013). Energy storage sizing for wind power : impact of the autocorrelation of day-ahead forecast errors. *Wind Energy*, 18(1).
- Rajendra K. Pachauri, Leo Meyer, S. (2015). GIEC, 2014 : Changements climatiques 2014 : Rapport de synthèse. . GIEC, Genève, Suisse, 161 p. Technical report.
- Staid, A., Watson, J.-P., Wets, R. J.-B., and Woodruff, D. L. (2017). Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators. *Wind Energy*, 20(12) :1911–1925.
- V. Masson-Delmotte, P. Zhai, H-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Piranian, W. Moufouma-Okia, C. Péan, R. Pidcock, Sarah Connors, J. B. R. Matthews, Y. Chen, X Zhou, M I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield (2018). Réchauffement planétaire de 1,5°C. Technical report.

Annexes

Annexe 1

Évolution des émissions et des puits de GES sur le territoire français entre 1990 et 2050 (en MtCO₂eq). Inventaire CITEPA 2018 et scénario SNBC révisée (neutralité carbone)



Source : ecologique-solidaire.gouv.fr/snbc