

**PLAN
D'APPROVISIONNEMENT GAZIER**

**PROPOSITION RELATIVE AU FACTEUR DJ*V
DANS L'ÉTABLISSEMENT DE LA PRÉVISION
DES BESOINS EN JOURNÉE DE POINTE**

HORIZON 2026-2029

TABLE DES MATIÈRES

INTRODUCTION	3
1 AUDIT DU MODÈLE DE PRÉVISION DE LA DEMANDE EN JOURNÉE DE POINTE	4
1.1 Validation de l'utilisation du facteur DJ * V	4
1.2 Recommandation d'allonger la période d'apprentissage	5
2 COMPARAISON DE LA MÉTHODE ACTUELLE ET DE LA MÉTHODE PROPOSÉE	6
2.1 Comparaison des deux méthodes	6
2.2 Les limites de la méthode proposée.....	9
CONCLUSION	11

INTRODUCTION

1 Dans la Cause tarifaire 2023-2024 (R-4213-2022), Énergir, s.e.c. (Énergir) a constaté que la
2 régression basée sur la demande de l'hiver 2021-2022 génère une variation de la pointe qui ne
3 suit pas la variation du volume projeté pour l'hiver 2023-2024. Énergir a également constaté une
4 augmentation importante du facteur $DJ * V^1$ par rapport aux années précédentes. En l'absence
5 d'explication plausible, Énergir n'a pas pu exclure une erreur d'échantillonnage. Ainsi, la solution
6 retenue pour le dossier tarifaire 2023-2024 a été d'utiliser l'hiver 2019-2020 comme année de
7 référence, soit le dernier hiver complet qui n'a pas été affecté par la pandémie de la COVID-19.

8 Dans sa décision D-2023-127², la Régie de l'énergie (Régie) a approuvé la proposition d'Énergir
9 quant à l'utilisation d'une année antérieure comme année de référence, mais lui a demandé de
10 produire une analyse permettant de valider le paramètre $DJ * V$ dans le cadre de la
11 Cause tarifaire 2024-2025.

12 Pour répondre à la demande de la Régie, Énergir a mandaté la firme de conseil externe Artelys
13 (Artelys) pour mener une étude sur la pertinence du facteur lié au vent. Compte tenu des délais
14 serrés, Énergir n'a pas pu obtenir le rapport d'étude final au moment du dépôt de la
15 Cause tarifaire 2024-2025. Énergir présente donc, dans une première section, les principales
16 conclusions du rapport d'étude³ susmentionné. Des changements à la méthode de régression
17 d'Énergir ont été proposés par Artelys et sont également présentés dans la section 1. Ensuite,
18 dans une deuxième section, Énergir compare les estimations entre la méthode actuelle et la
19 méthode recommandée par Artelys. Le rapport d'étude complet d'Artelys est déposé à l'annexe 1.

¹ Degrés-jours (en degrés Celsius) multiplié par le vent (en km/h).

² Dossier R-4213-2022.

³ Déposé à l'annexe 1.

1 AUDIT DU MODÈLE DE PRÉVISION DE LA DEMANDE EN JOURNÉE DE POINTE

1.1 VALIDATION DE L'UTILISATION DU FACTEUR DJ * V

1 Cette section présente les principales conclusions du rapport d'étude d'Artelys. Le mandat était
2 d'étudier la validité de l'intégration de l'effet DJ * V dans le modèle de régression utilisé par
3 Énergir afin de déterminer la demande en journée de pointe. Pour ce faire, Énergir a partagé les
4 données historiques reliées à la consommation journalière au cours de dix années passées,
5 couvrant la période du 1^{er} octobre 2012 au 30 septembre 2023. En partant de cette base, Artelys
6 a recréé le modèle de régression d'Énergir, puis a réalisé un ensemble de tests afin d'analyser la
7 performance du modèle. Malgré le constat d'une forte variabilité des paramètres de régression,
8 notamment du paramètre DJ * V au cours des derniers dossiers tarifaires, Artelys a pu confirmer
9 les éléments suivants :

- 10 • Par le biais d'une revue de littérature, l'utilisation d'un facteur croisé degrés-jours (DJ) et
11 de la vitesse du vent (V) est une pratique utilisée ailleurs par d'autres utilités publiques;
- 12 • Tous les facteurs du modèle actuellement utilisé par Énergir sont statistiquement
13 significatifs avec une *valeur-p* inférieure à 0,05;
- 14 • Le résidu d'un modèle qui n'intègre pas le facteur DJ * V, mais dont tous les autres
15 paramètres sont identiques au modèle actuellement utilisé par Énergir, est fortement
16 corrélé au facteur DJ * V;
- 17 • L'intégration du facteur DJ * V permet de mieux estimer les valeurs extrêmes, donc la
18 pointe;
- 19 • L'intégration du facteur DJ * V permet d'améliorer la performance générale du modèle en
20 termes d'erreur absolue moyenne.

21 **L'analyse d'Artelys a permis de répondre à la demande de la Régie, soit de confirmer la**
22 **pertinence du facteur DJ * V dans le modèle actuel d'Énergir.** Toutefois, depuis quelques
23 causes tarifaires, l'application de la méthode de régression usuelle semble avoir généré des
24 consommations projetées anormales en journée de pointe. Des traitements particuliers ont dû
25 être appliqués chaque année. Pour pallier ces traitements, Artelys a émis une opinion quant à la

1 taille insuffisante de l'échantillon d'observations et a émis comme recommandation l'allongement
2 de l'utilisation de trois hivers consécutifs comme base à la régression.

1.2 RECOMMANDATION D'ALLONGER LA PÉRIODE D'APPRENTISSAGE

3 Au constat de la forte variabilité des facteurs de régression d'année en année, Artelys s'est
4 penchée sur une piste de solution potentielle et recommande d'utiliser un historique plus long
5 comme base de référence pour la régression. Concrètement, l'utilisation de trois hivers successifs
6 comme intrant au modèle de régression est recommandée comme solution à la variabilité
7 annuelle des facteurs. Ceci viendrait multiplier par trois le nombre de points d'observation
8 journaliers utilisés par le modèle.

9 Artelys a proposé d'introduire une nouvelle variable « année » au modèle afin de tenir compte de
10 la tendance de consommation de l'année de l'observation⁴. Le modèle proposé est le suivant :

$$11 \quad \text{Conso}_t = \text{constante} + \alpha_0 DJ_t + \alpha_1 DJ_{t-1} + \alpha_2 DJ^*V_t + \beta_{\text{dimt=dim}} + \dots + \beta_{\text{vent=ven}} + \beta_{\text{an}} \text{Année}$$

12 La consommation journalière au jour t se calcule comme étant la somme de la constante, des
13 coefficients météo sensibles multipliés par les paramètres météorologiques projetés pour chaque
14 jour sur l'horizon du plan d'approvisionnement, et des coefficients liés aux variables calendaires,
15 c'est-à-dire le jour de la semaine de la journée concernée (dimanche à vendredi) et l'année de
16 l'observation. Les coefficients sont, quant à eux, issus de la régression linéaire utilisant
17 l'échantillon des données journalières de trois hivers consécutifs. Selon cette méthode, la
18 consommation de la journée de pointe se calculerait de la manière suivante :

$$19 \quad \text{Conso}_{\text{pointe}} = \text{constante} + \alpha_0 DJ_{\text{max}} + \alpha_1 DJ_{-1\text{max}} + \alpha_2 DJ^*V_{\text{max}} + \max(\beta_{\text{dimt=dim}} + \dots + \beta_{\text{vent=ven}}) + \beta_{\text{an}} \text{Année}$$

20 Les coefficients utilisés dans le calcul de la journée de pointe sont toujours ceux issus de la
21 régression linéaire. Les paramètres météorologiques correspondent aux paramètres qui
22 conduisent à la plus forte demande au cours des 30 dernières années. L'année correspond à la
23 dernière année des trois hivers utilisés dans la régression. Par exemple, si la régression est faite
24 sur les données journalières des années 2021, 2022 et 2023, la variable « année » serait alors
25 égale à celle de 2023.

⁴ Voir la section À retenir : mémo de la méthode proposée, p. 24 de l'annexe 1.

1 La méthode recommandée par Artelys est donc assez similaire à la méthode actuelle d'Énergir,
2 à deux exceptions près : l'utilisation d'un historique d'apprentissage plus long et l'introduction de
3 la variable « année » dans le modèle de régression linéaire. Selon Artelys :

- 4 • La modification au modèle est acceptable, car les coefficients de régression demeurent
5 statistiquement représentatifs; et
- 6 • Les modifications au modèle permettent de diminuer la variation annuelle des coefficients
7 de régression. Par conséquent, les variations annuelles de la pointe sont également moins
8 marquées d'une année à l'autre.

2 COMPARAISON DE LA MÉTHODE ACTUELLE ET DE LA MÉTHODE PROPOSÉE

9 Cette section compare la méthode actuellement utilisée par Énergir (méthode actuelle) à la
10 méthode recommandée par Artelys (méthode proposée). Les changements impliqués par
11 l'adoption de la méthode proposée seront comparés aux bénéfices escomptés.

2.1 COMPARAISON DES DEUX MÉTHODES

12 Afin de comparer les pointes entre les deux méthodes, certains ajustements ont été nécessaires :

- 13 • Les pointes calculées par la méthode proposée étant toutes calculées sur les paramètres
14 de la journée de pointe du dossier tarifaire 2023-2024, l'ensemble des pointes d'Énergir
15 ont été recalculées avec ces paramètres;
- 16 • Pour s'harmoniser avec la période de données disponibles dans le rapport d'Artelys,
17 Énergir a recalculé neuf pointes annuelles associées aux causes tarifaires 2016-2017 à
18 2024-2025. Un historique sur plusieurs années permet de bien évaluer la différence dans
19 l'estimation des deux méthodes;
- 20 • Les derniers hivers disponibles sont utilisés comme base de référence pour chaque pointe
21 calculée avec la méthode d'Énergir. Certaines pointes recalculées et présentées dans le
22 tableau 1 et le graphique 1 peuvent donc être différentes de celles déposées dans les
23 causes tarifaires précédentes;

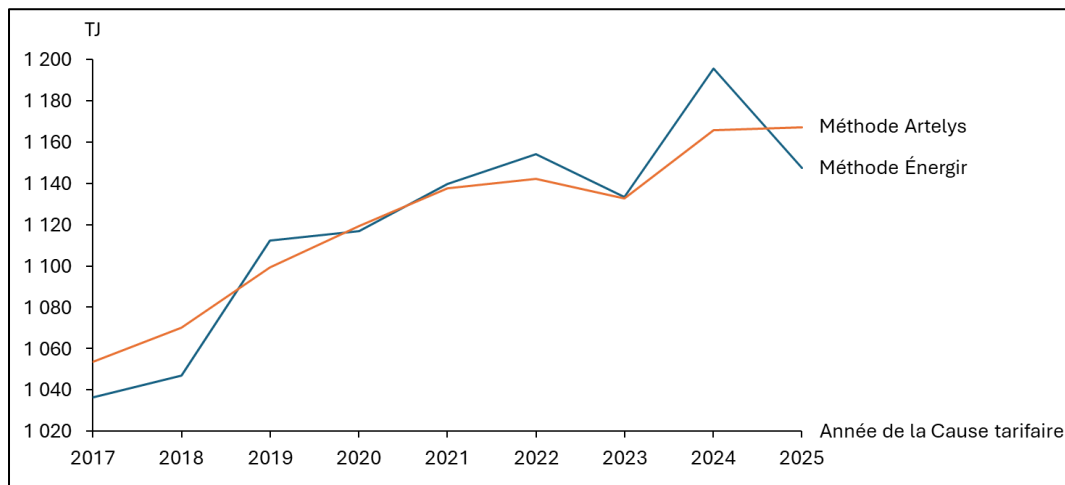
- 1 • Il est également à noter que la demande de pointe présentée dans cette section n'inclut
 2 que la portion de consommation de la clientèle au continu, avant l'ajout à la marge des
 3 volumes des clients en combinaison tarifaire et ceux des grands clients.
- 4 Pour bien comparer la performance des modèles en termes de calibration des paramètres, les
 5 pointes sont comparées avant l'application du facteur d'ajustement.

Tableau 1
Paramètres utilisés pour le calcul selon la méthode d'Énergir

Cause tarifaire (années 20**)	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	24-25
Hiver de référence utilisé (années 20**)	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23
Paramètres de la journée de pointe	DJ = 36,54 DJ ₋₁ = 39,13 DJ * V = 1 151,03								

Graphique 1

Comparaison des pointes annuelles avant l'application du facteur d'ajustement



- 6 Le graphique 1 illustre l'apport de la proposition d'Artelys. En effet, l'évolution annuelle de la
 7 pointe a une apparence plus lisse sur la période étudiée avec la méthode proposée.

1 Depuis quelques causes tarifaires, compte tenu du contexte spécial de l'année 2020 et de la
 2 reprise qui s'ensuit, Énergir a – avec l'approbation de la Régie – dérogé à la méthode habituelle
 3 en modifiant l'hiver utilisé comme base de référence. Des traitements au cas par cas ont été
 4 nécessaires, car il s'agissait d'un contexte exceptionnel. De plus, en utilisant l'hiver 2021-2022
 5 comme base de régression, Énergir a constaté que la demande en journée de pointe variait
 6 légèrement à la hausse, alors que les volumes de consommation hivernale variaient à la baisse,
 7 et ce, sans aucun changement significatif dans la composition de la clientèle d'Énergir. Les hivers
 8 2020-2021 et 2021-2022 ont été exclus et jugés non représentatifs de la consommation de la
 9 clientèle d'Énergir. L'hiver 2019-2020 a été utilisé comme base de régression pour les dossiers
 10 tarifaires 2022-2023 et 2023-2024.

11 En comparaison avec la méthode proposée, Artelys a également émis un doute sur l'utilisation
 12 de l'hiver 2020-2021 dans la régression. Une analyse sur le retrait des données affectées par la
 13 pandémie de la COVID-19 est étayée à la section 4.3.2 de l'annexe 1.

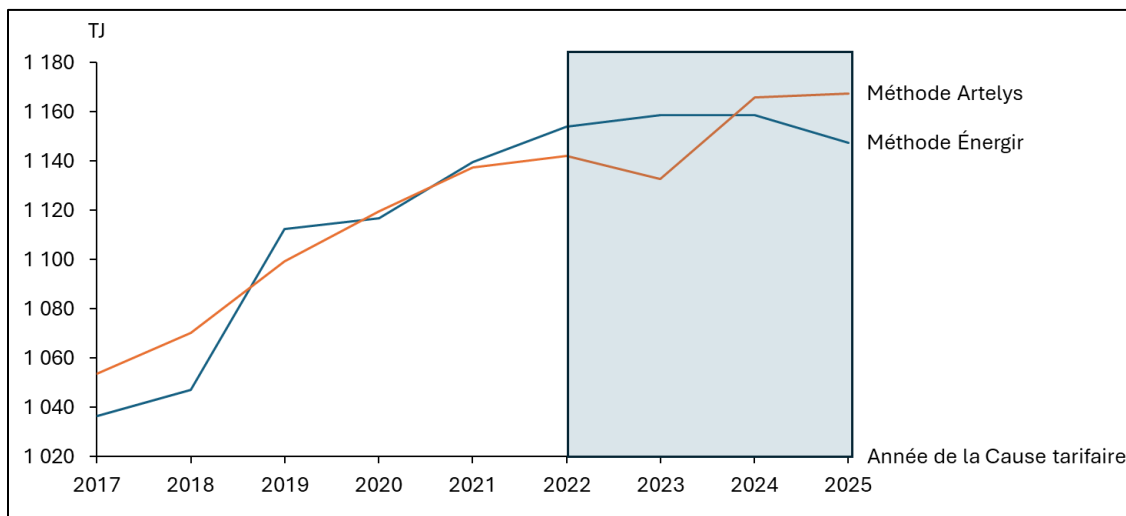
14 Un deuxième exercice est réalisé pour analyser l'impact du traitement au cas par cas par Énergir
 15 sur la projection de la demande en journée de pointe. Les prévisions sont comparées à nouveau
 16 entre les deux méthodes.

Tableau 2

Paramètres utilisés pour le calcul selon la méthode d'Énergir

Cause tarifaire (années 20^{**})	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	24-25
Hiver de référence utilisé (années 20^{**})	14-15	15-16	16-17	17-18	18-19	19-20	19-20	19-20	22-23
Paramètres de la journée de pointe	DJ = 36,54 DJ ₋₁ = 39,13 DJ * V = 1 151,03								

Graphique 2
Comparaison des deux méthodes



1 La proposition d'Énergir d'appliquer un traitement au cas par cas pour les deux hivers 2020-2021
 2 et 2021-2022 a permis de diminuer la variation annuelle de la pointe entre les causes tarifaires
 3 2022-2023 à 2024-2025. Au cours de cette période, la méthode d'Énergir a même mieux performé
 4 que la méthode proposée en termes de limitation de la variation annuelle de la pointe.

2.2 LES LIMITES DE LA MÉTHODE PROPOSÉE

5 L'allongement de la période d'apprentissage d'un à trois hivers permet de limiter les variations
 6 annuelles des paramètres et, en conséquence, de limiter la variation des pointes projetées d'une
 7 année à l'autre. Ceci est une solution potentielle au problème de la forte variabilité du facteur
 8 $DJ * V$ observée depuis quelques causes tarifaires. Bien qu'Énergir soit ouverte aux pistes
 9 d'amélioration continue de son modèle, il est important de tenir compte de la méthode proposée
 10 dans son ensemble, y compris ses bémols et le contexte dans lequel Énergir évolue.

11 Au moment de la préparation annuelle de la cause tarifaire, Énergir priorise autant que possible
 12 l'utilisation des données hivernales de la dernière année financière disponible, le but étant de
 13 refléter la composition la plus récente de la clientèle en termes de profil de consommation dans
 14 la régression linéaire. L'intégration de trois hivers passés comme intrant au modèle viendrait à
 15 l'encontre de cet objectif. Des comportements de consommation « obsolètes » seront
 16 continuellement pris en compte dans les paramètres de régression. La méthode proposée

1 exacerbe l'inexactitude de la prévision pour les dernières années du plan d'approvisionnement.
2 Ceci est d'autant plus problématique dans un contexte où Énergir s'attend à des modifications
3 plus importantes du profil de consommation de sa clientèle, compte tenu de l'adhésion croissante
4 prévue et souhaitée des clients au programme de la biénergie gaz naturel – électricité. Ainsi, le
5 lissage des variations annuelles des paramètres estimés viendrait fausser la représentativité du
6 modèle de régression dans ce contexte particulier.

7 Dans la section précédente, il a été démontré qu'un traitement au cas par cas est nécessaire
8 dans certains contextes, qui peuvent être considérés comme des années de « chocs ». Cette
9 dérogation est une exception à la méthode actuelle utilisée par Énergir et le résultat est
10 satisfaisant (voir graphique 2). Sur ce point, la méthode proposée n'apporte pas d'amélioration.
11 En effet, en présence d'une année « choc », la recommandation d'Artelys n'est pas concluante
12 quant au traitement à appliquer :

13 *« Quant à savoir s'il est préférable ou non de retirer l'année 2021 pour l'estimation de la*
14 *pointe en 2022 et 2023, les résultats présentés ici ne permettent pas objectivement d'en*
15 *juger ».*⁵

16 Considérant les éléments décrits dans les paragraphes précédents, la méthode proposée
17 représente une solution à la méthode actuelle, mais apporte d'autres enjeux. De plus, l'application
18 de la méthode proposée nécessiterait des changements aux fichiers de travail qui alourdiraient
19 ceux-ci, d'autant plus que potentiellement, l'horizon du plan d'approvisionnement peut être amené
20 à être allongé, dépassant la période de quatre ans.

⁵ Annexe 1, p. 22.

CONCLUSION

1 Le mandat confié au consultant externe Artelys a permis à Énergir de valider la pertinence de
2 l'utilisation du facteur $DJ * V$ dans son modèle de régression utilisé pour déterminer la demande
3 en journée de pointe. Toutefois, Énergir juge que la méthode proposée par Artelys est plus
4 complexe sans être plus précise que la réponse d'Énergir dans les précédentes causes tarifaires,
5 lorsque des facteurs atypiques viennent influencer le profil de la demande. Enfin, Énergir propose
6 de continuer à appliquer la méthode actuelle, tout en étant vigilant à d'autres variations atypiques.

7 **Énergir demande à la Régie :**

- | | |
|----|--|
| 8 | • de prendre acte de l'étude réalisée par Artelys en réponse à sa demande; |
| 9 | • de reconduire l'utilisation de la méthode actuelle pour le calcul de la journée de |
| 10 | pointe. |

Audit du modèle de prévision de la demande en journée de pointe

Rapport d'étude

Ce document est strictement confidentiel ; aucun élément ne doit être diffusé hors d'Artelys et d'Énergir.

AUTEURS

Le travail présenté dans ce document a été réalisé par :

- Arthur Bossavy, Consultant sénior en sciences des données, Artelys
- Aya Ankir, Consultante en sciences des données, Artelys
- Violette Berge, Direction de projet, Artelys

Synthèse

Au sein d'Énergir, la direction du Transport et de l'Approvisionnement Gazier est responsable d'établir la quantité d'outils d'approvisionnement nécessaire pour répondre à la demande en journée de pointe. Afin de réaliser cet exercice, la direction applique une méthode, approuvée par la Régie de l'Énergie du Québec, se basant sur les consommations observées de l'hiver précédent. La méthode comporte les 5 étapes suivantes :

- ▣ **Étape 1** : Régression linéaire de la consommation des clients continus purs sur l'hiver précédent en fonction des variables climatiques : degrés-jours, degrés-jours de la veille et degrés-jours vent ; et de variables calendaires : jours de semaine et jours fériés
- ▣ **Étape 2** : Projection de la partie de la régression liée aux variables climatiques sur chaque jour des 30 dernières années d'historique pour trouver le jour où les conditions climatiques conduisent à la plus forte demande
- ▣ **Étape 3** : Ajout des coefficients de régression calendaires maximaux à la pointe thermosensible
- ▣ **Étape 4** : Ajustement de la pointe obtenue pour prendre en compte les évolutions de portefeuille client
- ▣ **Étape 5** : Ajout des prévisions de pointe des clients non compris dans la régression (grands clients, combinaison tarifaire et cas particuliers)

Récemment, Énergir a pu constater une plus grande variabilité annuelle des coefficients issus du modèle de régression de la consommation des clients continus. Cette variabilité accrue des coefficients a pour conséquence immédiate une variabilité accrue de la demande de pointe estimée à partir des 30 années d'historiques météorologiques. La variabilité d'un coefficient en particulier a retenu l'attention d'Énergir : celui associé à **l'interaction des degrés-jours (DJ) et du vent (V)**. Cette observation a conduit à une remise en question du modèle en place, mettant en lumière le besoin d'une (ré)évaluation approfondie de sa formulation, et en particulier de celle de l'effet associé à la variable vent.

Dans une première phase, Artelys s'est attaché à étudier la validité de l'intégration de l'effet DJ x V dans le modèle. Pour ce faire, Artelys s'est appuyé à la fois sur l'étude des pratiques observées dans la littérature, ainsi que sur une analyse exhaustive de différents critères de performance et de diagnostic du modèle, étudiés à partir d'un historique de données couvrant une période d'une dizaine d'années (2013 – 2023). Cette étude a permis de **valider la légitimité de la formulation actuelle** de l'effet vent dans le modèle.

Dans une seconde phase, Artelys s'est attaché à démontrer qu'une part importante de la variabilité des coefficients de régression résidait, non pas dans la formulation du modèle, mais dans **l'échantillonnage des données**. En intégrant une tendance au modèle, dans le seul but d'accroître la taille de l'échantillon d'apprentissage, nous avons réussi à réduire la variabilité des coefficients, sans dégrader les indicateurs de performance d'estimation de la consommation. Ceci a permis de **réduire de moitié la variabilité annuelle de la pointe**.

Table des matières

SYNTHESE	3
TABLE DES FIGURES	6
1 INTRODUCTION	8
1.1 CONTEXTE	8
1.2 METHODOLOGIE	8
2 CONSTAT AUTOUR DE LA VARIABILITE DES ESTIMATIONS FOURNIES PAR LE MODELE ACTUEL	10
2.1 VARIABILITE DES COEFFICIENTS DU MODELE DE REGRESSION	10
2.2 VARIABILITE DE LA POINTE	11
3 ETUDE AUTOUR DE L'INTEGRATION DE L'EFFET DJ X V	13
3.1 PRISE EN COMPTE DE LA VARIABLE VITESSE DU VENT DANS LA LITTERATURE	13
3.2 ANALYSE A PARTIR DES DONNEES D'ÉNERGIR	14
3.2.1 RELATION ENTRE DJ X V ET LES RESIDUS D'UN MODELE N'INTEGRANT PAS LE VENT	14
3.2.2 SIGNIFICATIVITE DES COEFFICIENTS DE DJ X V	15
3.2.3 REDUCTION DU BIAIS LIE A V ET DANS L'ESTIMATION DES FORTES VALEURS DE CONSOMMATION	15
3.2.4 EVOLUTION DE LA MAE MENSUELLE	16
4 ETUDE AUTOUR DE L'ACCROISSEMENT DE LA PROFONDEUR D'HISTORIQUE D'APPRENTISSAGE	18
4.1 FORMULATION DU MODELE POUR UN APPRENTISSAGE SUR PLUSIEURS ANNEES	18
4.2 PERFORMANCES DE GENERALISATION	19
4.3 VARIABILITE DE LA POINTE	20
4.3.1 ANALYSE DE SENSIBILITE A LA PROFONDEUR D'HISTORIQUE D'APPRENTISSAGE	20
4.3.2 ANALYSE DE SENSIBILITE AU RETRAIT DES DONNEES IMPACTEES PAR LE COVID	22
4.3.3 ANALYSE DE SENSIBILITE SUITE A LA PROJECTION DE LA POINTE PAR COEFFICIENTS D'AJUSTEMENT	23
5 CONCLUSION	25

6	ANNEXE	26
6.1	VOLUME HIVERNAL DE CONSOMMATION DETHERMALISEE	26
6.2	RESIDUS MODELE SANS $D \times V$, DIVISES PAR DJ , EN FONCTION DE V	26
6.3	BIAIS ANNUEL LIE A L'ABSENCE DE MODELISATION DE LA CROISSANCE DE CONSOMMATION DE GAZ	27
6.4	EVOLUTION DES COEFFICIENTS DE REGRESSION LORS D'UN AJUSTEMENT DU MODELE DE CONSOMMATION SUR 3 ANNEES D'APPRENTISSAGE	28
6.5	EVOLUTION DES VALEURS DE POINTE AVEC LA PROFONDEUR D'HISTORIQUE D'APPRENTISSAGE	29
6.6	VOLUMES HIVERNAUX DETHERMALISES PREDITS PAR LE DEPARTEMENT DE PREVISION DE DEMANDE D'ÉNERGIR POUR LES CT 2021 A CT 2025.	29
6.7	VALEURS NUMERIQUES DE LA FIGURE 12 : PROJECTIONS DE POINTES AJUSTEES	30
7	BIBLIOGRAPHIE	31

Table des figures

FIGURE 1 - EVOLUTION DES COEFFICIENTS DE REGRESSION ENTRE LES HIVERS (2012-)2013 ET (2022-)2023. L'UNITE DE MESURE EN ENERGIE EST LE GJ.	10
FIGURE 2 – EVOLUTION DE LA POINTE ESTIMEE ET DE SA DECOMPOSITION SUIVANT LES DIFFERENTS EFFETS DU MODELE. LES VALEURS EN ORDONNEES SONT EXPRIMEES EN GJ.	12
FIGURE 3 – DIFFERENCES FINIES, CALCULEES PAR EFFET, REPRESENTANT LES VARIATIONS ANNUELLES A LA HAUSSE ET A LA BAISSSE DE LEURS CONTRIBUTIONS A LA POINTE. LE TOTAL DES VARIATIONS EST REPRESENTE PAR LA COURBE BLEUE. SES VALEURS EN 2021 ET 2022 SONT IMPACTEES PAR LE COVID, ET SONT DONC A CONSIDERER AVEC PRECAUTION. LES VALEURS EN ORDONNEES SONT EXPRIMEES EN GJ. 12	12
FIGURE 4 - RESIDUS DU MODELE SANS DJXV, EN FONCTION DE DJXV. LA LIGNE ROUGE REPRESENT LE RESULTAT D'UNE REGRESSION LINEAIRE. LA COURBE BLEUE CELUI D'UN LISSAGE PAR REGRESSION LINEAIRE LOCALE. LES POINTS EN ROUGE SONT DES VALEURS EXTREMES (E.G. PERIODES DE CONGES NON MODELISEES) QUI ONT ETE ECARTEES LORS DE L'AJUSTEMENT DES 2 REGRESSIONS. LES VALEURS EN ORDONNEES SONT EXPRIMEES EN GJ. CELLES EN ABSCISSE EN °C X M/S.	14
FIGURE 5 - P-VALUES DES TESTS STATISTIQUES DE NULLITE DES COEFFICIENTS DE REGRESSION.	15
FIGURE 6 - EVOLUTION DU BIAIS EN FONCTION DE LA VITESSE DU VENT V (A GAUCHE) ET DE LA CONSOMMATION MESUREE <i>OBS</i> (A DROITE) ; POUR UN MODELE N'INTEGRANT PAS DJ X V (ORANGE), ET CELUI L'INTEGRANT (BLEU). LES VARIABLES V ET <i>OBS</i> SONT EXPRIMEES EN RATIO DE LEUR VALEUR MAXIMALE. LE BIAIS EST EXPRIME EN RATIO DE LA CONSOMMATION HIVERNALE JOURNALIERE MOYENNE DETHERMALISEE (PAR ANNEE). LES ANNEES 2020 ET 2021 ONT ETE EXCLUES (COVID).	16
FIGURE 7 - EVOLUTION MENSUELLE DE L'ERREUR ABSOLUE MOYENNE (MAE), MESUREE SUR L'ANNEE PRECEDANT CELLE D'APPRENTISSAGE, POUR DES MODELES ENTRAINEES AVEC DJXV (BLEU) ET SANS DJXV (ROUGE).	17
FIGURE 8 : MAPE DU MODELE DE CONSOMMATION AJUSTE SUR 1, 2 OU 3 ANS, A PARTIR DE LA PROCEDURE PROPOSEE (REDRESSEMENT DES DONNEES PAR TENDANCE LINEAIRE). LA MAPE EST MESUREE A PARTIR DES DONNEES ISSUES DE L'ANNEE SUIVANT L'ANNEE D'EXERCICE (DERNIERE ANNEE D'APPRENTISSAGE, REPRESENTEE EN ABSCISSE ICI).	19
FIGURE 9 : EVOLUTION DE LA POINTE ESTIMEE ET DE SA DECOMPOSITION SUIVANT LES DIFFERENTS EFFETS DU MODELE. LE MODELE DE CONSOMMATION EST ICI AJUSTE SUR 3 ANNEES CONSECUTIVES (DONNEES EN ABSCISSES). 'CONST' SOMME LES EFFETS 'INTERCEPT' ET ' $\alpha_0 Y_n$ '. LES VALEURS EN ORDONNEES SONT EXPRIMEES EN GJ.	20
FIGURE 10 : COMPARAISON DES VALEURS DE POINTE ESTIMEES PAR LE MODELE DE CONSOMMATION AJUSTE SUR 1 ('1Y'), 2 ('2Y ') OU 3 ('3Y') ANNEES D'APPRENTISSAGE.	21
FIGURE 11 : EVOLUTION DE LA POINTE ESTIMEE A PARTIR D'UN MODELE DE CONSOMMATION AJUSTE SUR 3 ANNEES D'APPRENTISSAGE CONSECUTIVES, HORMIS LORS DU RETRAIT DE L'ANNEE 2021 DE L'HISTORIQUE D'APPRENTISSAGE (« NOCOVID »).	22
FIGURE 12 : PROJECTIONS DE Y+2 A Y+5 DE LA POINTE ESTIMEE EN Y, A PARTIR DES COEFFICIENTS DE REGRESSION ESTIMES SUR 1 AN (COURBES EN POINTILLEES) OU SUR 3 ANS (COURBES PLEINES), ET DES COEFFICIENTS D'AJUSTEMENT QUI EN RESULTENT (VOIR LE TEXTE POUR PLUS DE DETAILS). LES DIFFERENTS EXERCICES SUCCESSIFS SONT REPRESENTES PAR DES COULEURS (CT 2021 EN VIOLET, CT 2022 EN ROUGE, ETC.). LA CT 2023 S'APPUIE SUR LES RESULTATS DE CALIBRATION DE 2019-2020. LA CT 2024	

S'APPUIE TANTOT SUR CETTE MEME CALIBRATION (GRAPHE DE GAUCHE), TANTOT SUR CELLE DE L'ANNEE 2021-2022 (GRAPHE DE DROITE).	23
FIGURE 13- EVOLUTION ANNUELLE DE LA CONSOMMATION HIVERNALE DETHERMALISEE, FOURNIE PAR LE MODELE DE REGRESSION AVEC 1 ('1Y') OU 3 ('3Y') ANNEES D'APPRENTISSAGE, ET CALCULEE A PARTIR D'UNE METEO « NORMALE », MOYENNE DES 30 DERNIERES ANNEES.	26
FIGURE 14 - RESIDUS DU MODELE SANS DJXV DIVISES PAR DJ (EN GJ/°C), EN FONCTION DE V (M/S).....	26
FIGURE 15 : BIAIS ANNUEL DES MODELES INTEGRANT DJ X V (BLEU) OU PAS (ORANGE). LES ERREURS ASSOCIEES ONT ETE MESUREES SUR L'ANNEE PRECEDANT L'APPRENTISSAGE. LE BIAIS EST ICI EXPRIME EN RATIO DE LA CONSOMMATION HIVERNALE JOURNALIERE MOYENNE DETHERMALISEE.	27
FIGURE 16 : EVOLUTION DES COEFFICIENTS DE REGRESSION ENTRE LES HIVERS 2015 ET 2023. CES COEFFICIENTS ONT ETE AJUSTES A PARTIR DE 3 ANNEES CONSECUTIVES D'APPRENTISSAGE. L'UNITE DE MESURE EN ENERGIE EST LE GJ.....	28
FIGURE 17 : VOLUMES HIVERNAUX DETHERMALISES PREDITS PAR LE DEPARTEMENT DE PREVISION DE DEMANDE D'ÉNERGIR POUR LES CT 2021 A CT 2025.....	29

1 Introduction

1.1 Contexte

Au sein d'Énergir, la direction du Transport et de l'Approvisionnement Gazier est responsable d'établir la quantité d'outils d'approvisionnement nécessaire pour répondre à la demande en journée de pointe. Afin de réaliser cet exercice, la direction applique une méthode, approuvée par la Régie de l'Énergie du Québec, se basant sur les consommations observées de l'hiver passé ainsi que sur des prévisions de demande en volume annuel communiquées par le groupe Prévision de la demande. Cette méthode repose notamment sur une régression linéaire de la consommation en fonction de différents facteurs comme les degrés-jours, les degrés-jour de la veille, un facteur croisé degrés/jour et vent ($Dj*V$) et des variables calendaires.

Fin 2023, la direction en charge de la prévision de pointe a constaté des variations significatives de prévision de pointe de consommation en mettant à jour les données historiques permettant de faire la prévision. Artelys a alors été mandaté pour mener une étude détaillée de la méthode utilisée et des variations observées, en se focalisant sur l'utilité du facteur $DJ*V$ et son impact sur l'instabilité de la projection de pointe observée.

La présente étude a alors pour but d'interroger l'utilité du facteur $Dj*V$ pour expliquer la consommation gazière de pointe et d'émettre des recommandations quant à l'utilisation de ce facteur dans la méthode de régression d'Énergir.

1.2 Méthodologie

Dans une première phase, Artelys s'est chargé de prendre en main les données fournies par Énergir. Nous avons notamment pu reproduire les estimations des coefficients de régression du modèle de consommation, obtenues lors des derniers exercices annuels de causes tarifaires. Pour conduire son étude, Artelys a pu disposer d'un historique de 10 années des consommations journalières agrégées sur le périmètre des clients continus. Artelys a également pu se reposer sur un historique de 30 années des variables météorologiques nécessaires au calage du modèle d'estimation de la consommation journalière, et à l'estimation de la demande de pointe.

Lors de cette phase, les analyses produites par Artelys l'ont conduit à **confirmer le constat établi par la direction en charge de la prévision de pointe** d'Énergir, à savoir celui de la **variabilité importante des coefficients de régression**, et notamment celle du coefficient lié à l'effet $DJ \times V$, ainsi que la variabilité importante induite sur la demande de pointe.

Dans une deuxième phase, Artelys s'est concentré sur **l'étude de la formulation de l'effet intégrant la variable de vitesse du vent** dans le modèle. En effet, la 1ère phase d'étude ayant révélé une forte variabilité annuelle récente du coefficient associé, il s'est agi de vérifier l'adéquation de son intégration dans le modèle sous sa formulation actuelle. Artelys a pour cela réalisé une étude bibliographique afin de comparer cette formulation à celles trouvées dans la littérature. Artelys a ensuite réalisé une

analyse comparative du modèle intégrant ou non cet effet, à partir d'un ensemble d'indicateurs statistiques et des historiques de données fournis.

Enfin, dans une ultime phase, Artelys s'est intéressé à **l'hypothèse d'un surajustement systématique de l'ensemble des coefficients du modèle, à un historique potentiellement trop court** de données d'apprentissage. Dans son approche actuelle, Énergir cherche à garantir la bonne représentativité des usages de consommation présents par le modèle, à travers l'utilisation de données de calage ne couvrant que le dernier hiver. Artelys s'est intéressé à l'évolution d'une telle représentativité, ainsi qu'à celle de la variabilité des coefficients et de la pointe, lors de l'accroissement de la profondeur d'historique d'apprentissage du modèle. Sur la base de ces derniers résultats, Artelys a pu établir des recommandations concernant l'évolution de la méthode d'estimation de la demande de pointe.

2 Constat autour de la variabilité des estimations fournies par le modèle actuel

Dans cette section, nous étudions certains indicateurs permettant de visualiser et quantifier le problème de variabilité annuelle accrue, constaté par Énergir, dans les années les plus récentes. Nous en étudions ici, d’abord la manifestation dans les coefficients de régression du modèle de consommation, puis sa répercussion dans l’estimation de la pointe.

2.1 Variabilité des coefficients du modèle de régression

Dans la Figure 1 est représentée l’évolution des coefficients de régression lors des exercices annuels d’estimation successifs. Chaque exercice fait intervenir les données journalières météorologiques et de consommation, s’étalant entre le mois de novembre d’une année (représentée ici en abscisse) et le mois de mars de l’année suivante. Pour chaque variable du modèle, les valeurs moyenne, minimale et maximale obtenues lors des différents exercices ont été représentées à l’aide des droites horizontales noire et rouges. L’écart entre les valeurs maximale et minimale, rapporté à la valeur maximale, permet de se représenter l’amplitude de variation maximale observée sur la période considérée.

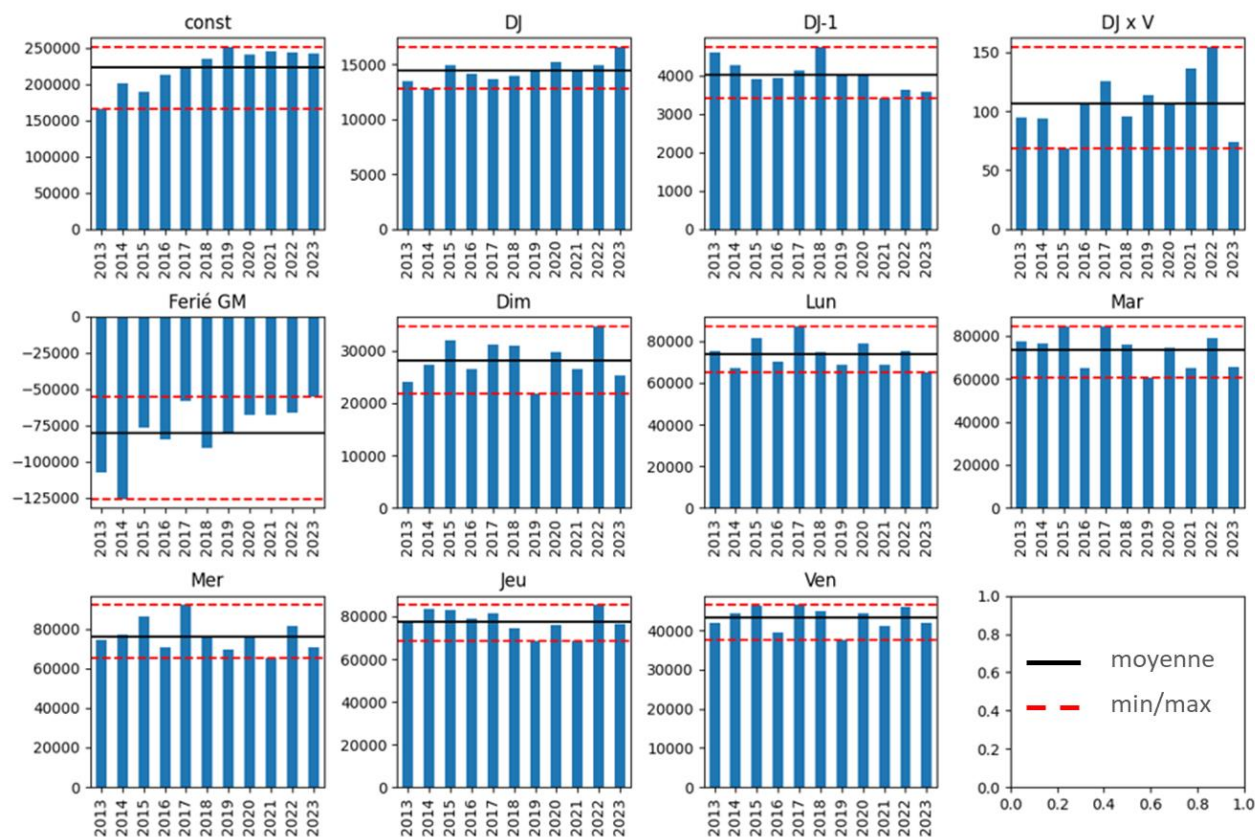


Figure 1 - Evolution des coefficients de régression entre les hivers 2013 et 2023. L'unité de mesure en énergie est le GJ.

L'examen des coefficients de régression de 2013 à 2023 indique une variabilité marquée pour $DJ \times V$, avec une variation annuelle entre 2022 et 2023 quasiment égale à l'amplitude de variation maximale observée sur 10 ans. Cette amplitude représente plus de 50% de la valeur maximale du coefficient sur 10 ans, ce qui est supérieure à celles observées, de manière générale, pour les autres variables du modèle. Certes, le coefficient des jours fériés présente une amplitude maximale proche, mais de l'aveu d'Énergir, il se pourrait que cela soit dû à un changement dans le processus de décompte des jours fériés au cours du temps. Il est également à noter que l'effet associé aux jours fériés n'intervient pas dans l'estimation de la pointe, donc sa variabilité n'impacte pas celle de la pointe.

Outre ces constatations, on pourra observer la présence d'une tendance à la hausse, suivie d'une stabilisation dans les dernières années, de l'intercept (*const*). D'autres tendances, moins nettes, peuvent être observées sur les coefficients des variables DJ et DJ-1.

2.2 Variabilité de la pointe

Dans la Figure 2, est représentée l'évolution de la pointe estimée au cours des exercices annuels successifs. Dans cette représentation, la pointe est décomposée suivant les effets des différentes variables du modèle. On notera l'absence de l'effet des jours fériés qui ne participe pas à la pointe. Le jour passé assurant les conditions météorologiques mesurées à la pointe, issu de la base d'historiques météo de 30 années, reste généralement identique entre les exercices d'estimation¹ (il s'agit du 15/01/2004). En 2023 toutefois, les conditions météorologiques à la pointe ont changé, du fait de la variation importante des coefficients (il s'agit alors du 15/01/1994). Autrement dit, la variabilité annuelle de la pointe montrée ici *n'est généralement que le reflet de celle des coefficients de régression*, et jamais celle des variables d'entrée, **hormis pour 2023**.

Sur une période de dix ans, on constate une tendance à la hausse de la pointe, qui se stabilise depuis quelques années. Cette observation peut également être faite sur le volume hivernal de consommation *déthermalisée* (i.e. à conditions météorologiques « normales » / moyennes, voir Figure 13 en annexe), et semble refléter ainsi un ralentissement de la croissance en consommation de gaz qui serait à confirmer par des analyses approfondies. L'année 2021, impactée par le COVID, constitue une anomalie dans cette évolution avec une consommation en retrait par rapport à celle de 2020, mais rattrapée dès 2022.

On constate l'impact majoritaire de l'effet lié à la température du jour courant seule (DJ), sur la valeur de la pointe (environ 50% du total). L'ensemble des effets météo-sensibles représentant approximativement 75% de la valeur de la pointe. L'effet $DJ \times V$ n'en est qu'un contributeur mineur en comparaison. En effet, sa contribution est équivalente ou légèrement inférieure à celle de DJ-1, et largement inférieure à celle des effets non météo-sensibles (*const + Jeu* ; 25% de la pointe environ).

¹ Cela peut rester vrai, malgré la mise à jour glissante de cette base. A noter que cette dernière n'est pas mise à jour dans notre étude. Il s'agit ici toujours de la période s'étalant entre 10/1990 et 10/2022.

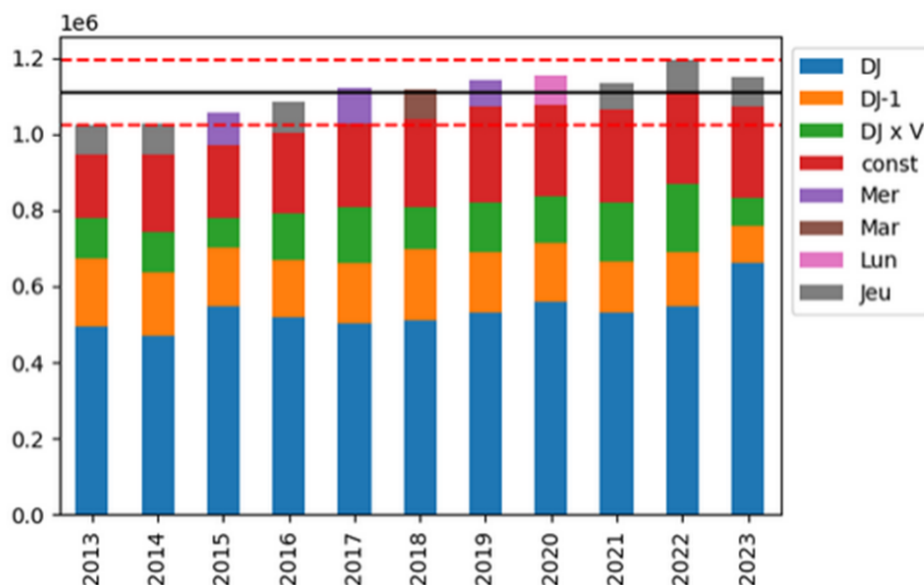


Figure 2 – Evolution de la pointe estimée et de sa décomposition suivant les différents effets du modèle. Les valeurs en ordonnées sont exprimées en GJ.

Malgré cette hiérarchie des influences, il est important de constater que la variabilité annuelle de l'effet lié à DJ x V constitue en revanche *un contributeur majeur de la variabilité de la pointe* (voir Figure 3). L'année 2023 est en cela symptomatique : la variation annuelle de la pointe y est plus importante que toutes les années précédentes (hors COVID) et la contribution de la variation de l'effet DJ x V y est dominante.

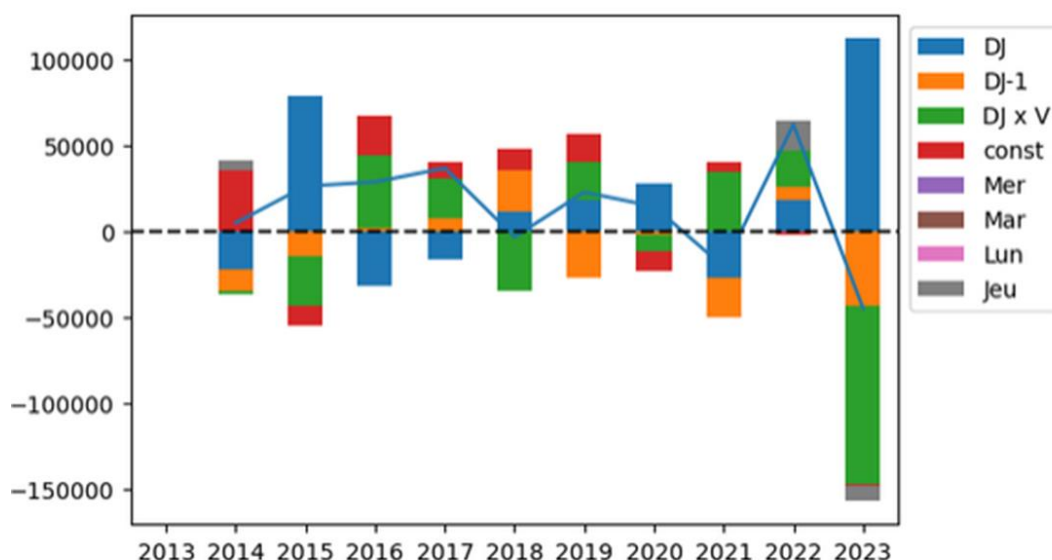


Figure 3 – Différences finies, calculées par effet, représentant les variations annuelles à la hausse et à la baisse de leurs contributions à la pointe. Le total des variations est représenté par la courbe bleue. Ces valeurs en 2021 et 2022 sont impactées par le COVID, et sont donc à considérer avec précaution. Les valeurs en ordonnées sont exprimées en GJ.

3 Etude autour de l'intégration de l'effet DJ x V

La section précédente a révélé une grande variabilité du coefficient de régression propre à l'effet DJ x V, responsable d'une variabilité récemment accrue de la pointe. Cette variabilité pose la question de l'adéquation du modèle. S'agit-il d'un problème de surajustement, de colinéarité ou plus généralement de mauvaise formulation ?

Dans cette section, nous étudions le bien-fondé de l'intégration de la variable de vitesse du vent dans sa formulation actuelle. Nous comparons d'abord cette dernière aux formulations présentes dans la littérature. Puis, nous étudions l'évolution des performances du modèle, à partir de l'historique de données Énergir, lors de l'intégration de cet effet.

3.1 Prise en compte de la variable vitesse du vent dans la littérature

Avant d'étudier les données fournies par Énergir, pour évaluer l'adéquation de la formulation de l'effet DJ x V dans le modèle de consommation, nous avons étudié quelles étaient les formulations présentes dans la littérature.

L'étude des rapports concernant la revue de la méthode de normalisation des revenus [1] [2] a montré que l'intégration de la variable vitesse du vent, sous différentes formulations, a déjà été envisagée. L'usage initial d'un effet additif DJ + V [1] provenait d'un choix par défaut, résultant du trop peu de données disponibles pour l'ajustement (réalisé à la maille mensuelle), et malgré l'observation d'un biais résiduel qui révélait que l'effet linéaire de V dépendait de la valeur de DJ. Ainsi, la dernière révision de la méthode de normalisation des revenus [2] a :

- Validé l'intégration de l'effet DJ x V
- Rejeté celle, additionnelle, de l'effet V (car statistiquement non significatif, lorsqu'associé à DJ x V)
- Rejeté l'usage de puissance 0.5 de DJ et/ou V, comme alternative dans DJ x V, car n'apportant pas de gains significatifs.

La littérature scientifique fait quant à elle parfois usage d'une « température ressentie » (« Wind Chill Factor »), généralement de la forme :

$$\begin{cases} a + bxT + cxV^\alpha + dxTV^\alpha, & \text{si } T \leq T_0 \text{ et } V \geq V_0, \\ T, & \text{sinon} \end{cases},$$

avec des valeurs de paramètres pouvant varier (voir [3] [4] [5]). RTE [6] a quant à lui prouvé l'intérêt d'utiliser un effet de la forme $DJxV^{4/5}$ (et plus encore celui d'un lissage exponentiel de ce dernier), dans la prévision court-terme de la demande électrique. On voit bien que, en l'absence de lissage, l'effet considéré par RTE est très proche de l'effet considéré dans notre modèle, ce qui semble donc valider son usage.

3.2 Analyse à partir des données d'Énergir

3.2.1 Relation entre DJ x V et les résidus d'un modèle n'intégrant pas le vent

Dans la Figure 4 sont représentés les résidus (erreurs) du modèle d'estimation *n'intégrant pas* l'effet DJ x V, vis-à-vis de DJ x V, sur son ensemble d'apprentissage. Il s'agit ici d'identifier la présence d'un biais résiduel vis-à-vis de la variable que l'on cherche à intégrer dans le modèle. Pour mieux s'assurer de la présence d'un biais, mais aussi de la forme qui le relie à la variable DJ x V, nous avons ajouté le résultat d'une régression en réalisant deux types d'ajustements : une régression linéaire et une non linéaire (lissage par régression linéaire locale). Dans cet exercice, nous avons pris soin d'exclure les valeurs « aberrantes »² (e.g. résidus élevés liés aux périodes de congés non modélisées), afin notamment d'éviter une trop grande sensibilité du lissage aux valeurs extrêmes, et pouvoir ainsi mieux comparer les résultats des 2 types d'ajustements.

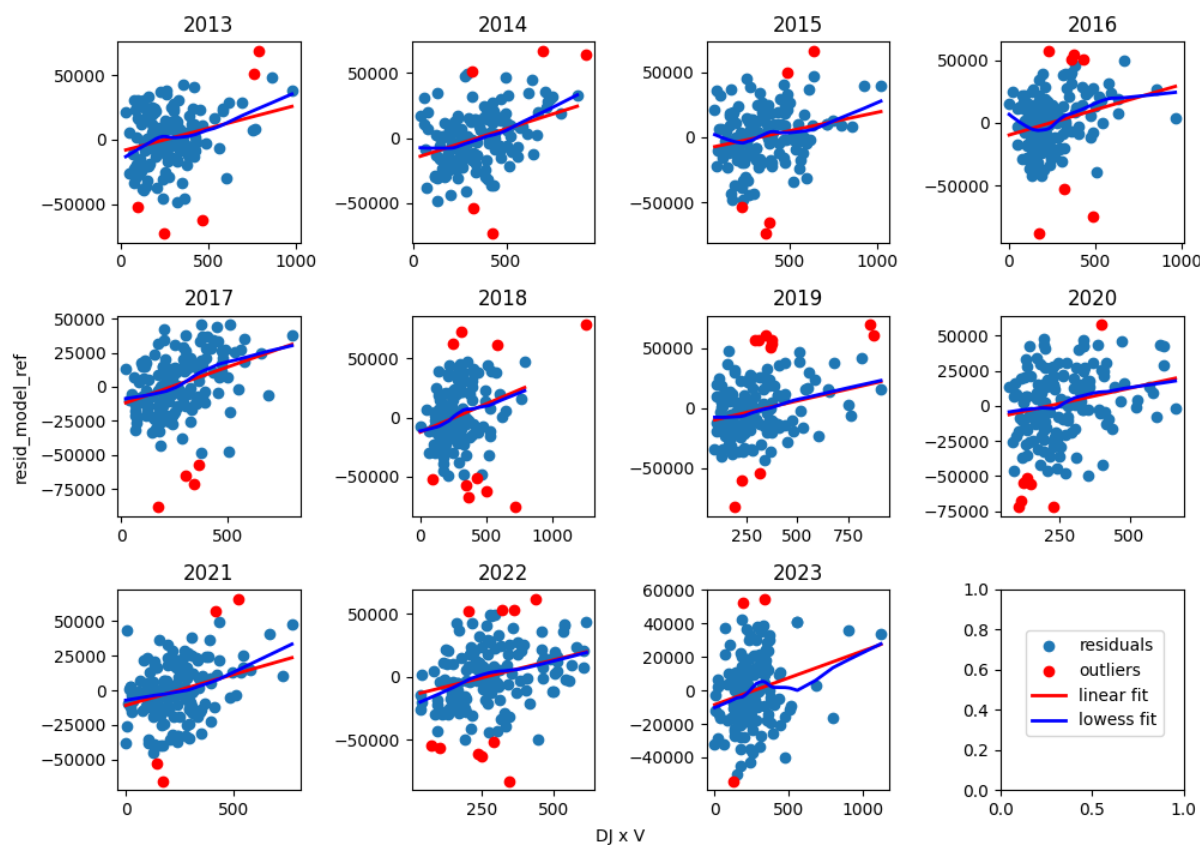


Figure 4 - Résidus du modèle sans DJxV, en fonction de DJxV. La ligne rouge représente le résultat d'une régression linéaire. La courbe bleue celui d'un lissage par régression linéaire locale. Les points en rouge sont des valeurs extrêmes (e.g. périodes de congés non modélisées) qui ont été écartées lors de l'ajustement des 2 régressions. Les valeurs en ordonnées sont exprimées en GJ. Celles en abscisse en °C x m/s.

² Les valeurs de résidus non comprises entre -50 000 GJ et 50 000 GJ, ont été retirées lors des *fits* linéaires et non linéaires, mais conservées dans la figure en rouge.

Les résultats révèlent un biais significatif et positivement corrélé à DJ x V, persistant à travers les exercices annuels. Le lissage par régression linéaire locale n'ayant pas fait l'objet d'une optimisation systématique dans l'ajustement de son hyperparamètre (proportion de données d'apprentissage à chaque valeur unique de DJ x V), on observe ce qui s'apparente à un léger surajustement certaines années. Pour autant, ces résultats semblent clairement indiquer qu'un effet linéaire relie le résidu à la variable DJ x V. Afin de confirmer l'usage de V et non d'une puissance de V dans l'effet à introduire, nous avons réalisé un exercice similaire reliant les résidus / DJ, vis-à-vis de la variable V (voir Figure 14 en annexe). Cette fois encore, l'ajustement par lissage montre qu'une hypothèse linéaire (puissance α égale à 1) semble appropriée.

3.2.2 Significativité des coefficients de DJ x V

Nous avons également évalué la pertinence d'intégration de l'effet DJ x V à travers la significativité des coefficients de régression associés. Dans la Figure 5 sont représentées les p-values des tests statistiques de nullité des coefficients de régression. Une p-value < 0.05 signifie qu'on peut rejeter l'hypothèse de nullité avec probabilité d'erreur de 5%. Comme on le voit ici, tous les effets sont jugés significatifs avec un risque d'erreur < 0.5%. Si on excepte l'effet dimanche (*Dim*), et donc en particulier pour l'effet DJ x V, le risque d'erreur est encore beaucoup plus faible.

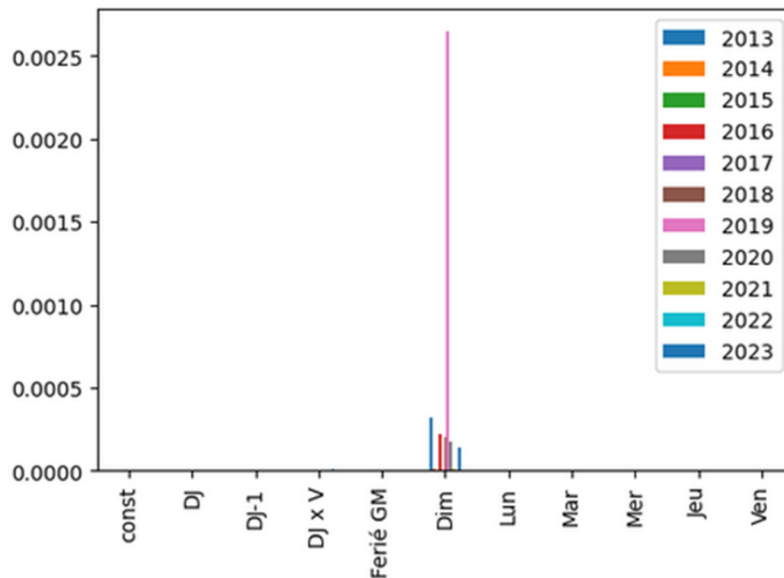


Figure 5 - P-values des tests statistiques de nullité des coefficients de régression.

3.2.3 Réduction du biais lié à V et dans l'estimation des fortes valeurs de consommation

Dans la Figure 6, est représentée l'évolution du biais en fonction de la vitesse du vent V (graphe de gauche), et de la consommation mesurée « *obs* » (graphe de droite), d'un modèle n'intégrant pas l'interaction DJ x V (« *_ref* » ; en orange), et du modèle l'intégrant (en bleu). Cette analyse est basée sur les erreurs des modèles mesurées l'année précédant l'apprentissage, permettant ainsi d'étudier

sur un échantillon de test, les résultats issus du calage sur le dernier hiver (2022-2023). Les variables V et obs sont exprimées en ratio de leur valeur maximale. Le biais est quant à lui exprimé en ratio de la consommation hivernale journalière moyenne déthermalisée (distincte pour chaque année). Les résultats des années 2020 (car issus du calage en 2021) et 2021, ont été exclus de l'analyse afin de neutraliser l'impact du COVID-19 sur les tendances de consommation.

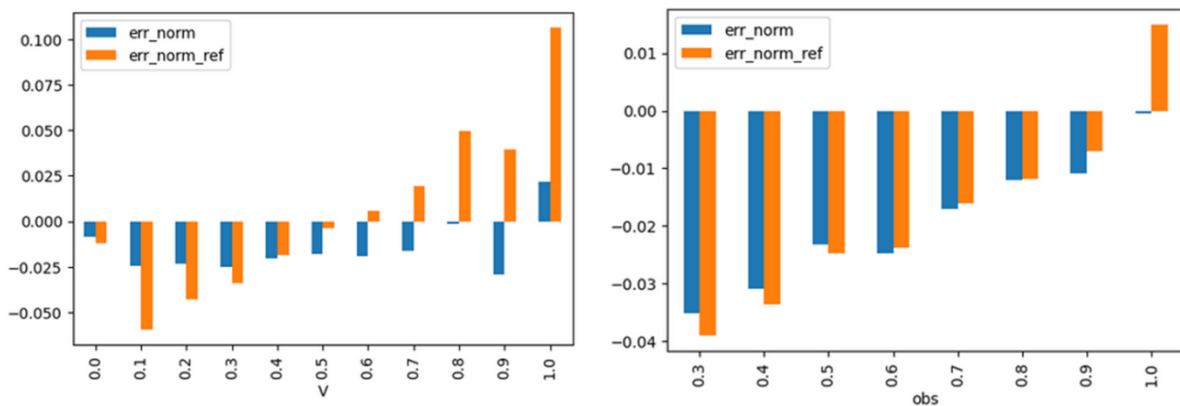


Figure 6 - Evolution du biais en fonction de la vitesse du vent V (à gauche) et de la consommation mesurée obs (à droite) ; pour un modèle n'intégrant pas $DJ \times V$ (orange), et celui l'intégrant (bleu). Les variables V et obs sont exprimées en ratio de leur valeur maximale. Le biais est exprimé en ratio de la consommation hivernale journalière moyenne déthermalisée (par année). Les années 2020 et 2021 ont été exclues (COVID).

L'introduction de l'effet $DJ \times V$ dans le modèle a eu un impact notable sur la réduction du biais vis-à-vis de la variable V . La dépendance du biais avec V tend ainsi à disparaître. Il est important de noter que la consommation en gaz connaissant une croissance structurelle lors des 10 dernières années (en diminution récente), et aucun des modèles considérés ici n'intégrant une représentation de cette croissance, l'évaluation de ces derniers sur une année de test (précédent l'apprentissage en l'occurrence), provoque sans surprise l'apparition d'un biais systématique (négatif ; voir Figure 15 en annexe). Un biais, non nul, mais relativement constant en fonction de V est donc un résultat tout à fait attendu et satisfaisant.

L'observation d'un léger biais négatif dans l'estimation des consommations faibles, et positif dans l'estimation des consommations fortes, est aussi un résultat attendu de la régression. Cependant, l'absence de modélisation de la croissance de la consommation tend à décaler les valeurs de biais vers les valeurs négatives ici (graphe de droite). Il est surtout intéressant de constater que l'amplitude entre biais négatif et biais positif dans l'estimation des valeurs faibles et fortes de consommation, tend à s'atténuer après l'introduction de $DJ \times V$ dans le modèle. Cet effet permet donc en particulier de mieux estimer les fortes consommations, et par extension la pointe.

3.2.4 Evolution de la MAE mensuelle

La Figure 7 ci-dessous représente l'impact de l'introduction de $DJ \times V$ dans le modèle, sur la performance globale agrégée du modèle à travers l'évolution de l'erreur absolue moyenne (MAE) mensuelle. Cette dernière est toujours mesurée ici sur l'année précédant celle d'apprentissage. Elle

est exprimée en ratio de la consommation hivernale journalière moyenne déthermalisée. Les résultats montrent que les gains de performance obtenus par l'intégration de DJ x V dans le modèle sont robustes et persistent dans le temps. Cela reste en particulier valable pour les résultats issus de l'année 2021-2022, dont les coefficients de régression ont été estimés sur l'année 2022-2023.

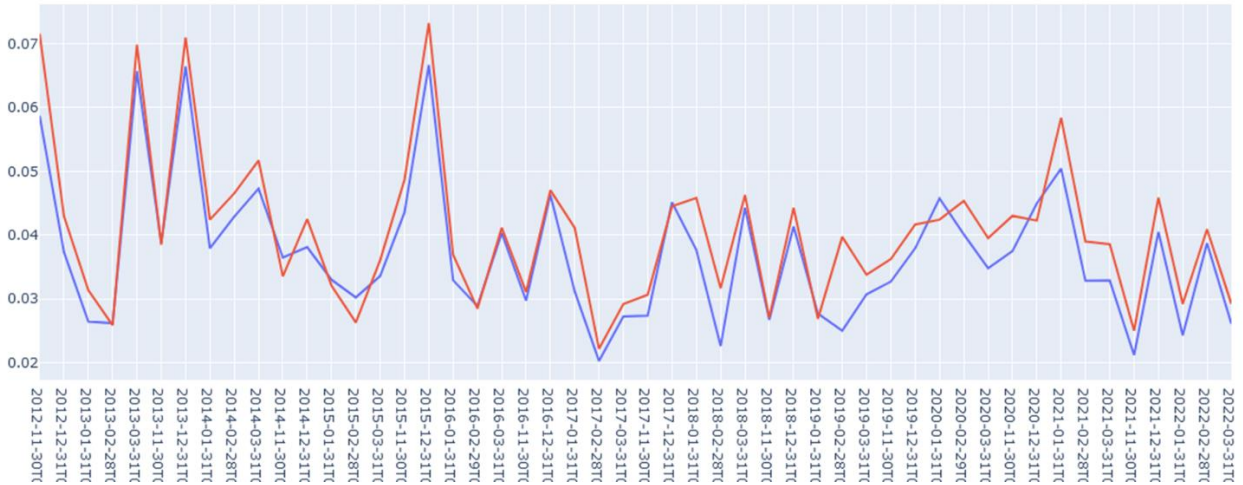


Figure 7 - Evolution mensuelle de l'erreur absolue moyenne (MAE), mesurée sur l'année précédant celle d'apprentissage, pour des modèles entraînés avec DJxV (bleu) et sans DJxV (rouge).

4 Etude autour de l'accroissement de la profondeur d'historique d'apprentissage

Suite aux constatations liant variabilité des coefficients, et en particulier celle du coefficient de l'effet DJ x V, et variabilité de la pointe (Section 2), nous avons étudié, et écarté, l'hypothèse d'une mauvaise formulation de l'effet en question dans le modèle (Section 3).

Une autre hypothèse concernant la variabilité de ce coefficient est celle d'un surajustement à des données d'apprentissage trop peu nombreuses. Ceci entraînerait une forte variabilité annuelle qui toucherait en réalité l'ensemble des coefficients du modèle. Dans cette section, nous démontrons cette hypothèse, et proposons une procédure visant à permettre un apprentissage sur un historique plus long, et ainsi lisser l'estimation des coefficients dans le temps.

4.1 Formulation du modèle pour un apprentissage sur plusieurs années

Nous faisons l'hypothèse que la variation des effets introduits actuellement dans le modèle, est suffisamment lente, pour autoriser leur estimation à partir de plusieurs années consécutives de relevés de consommation. Une hypothèse corollaire est que la croissance de la consommation de gaz, constatée chaque année, affecte avant tout les usages non météo-sensibles et n'altère pas le profil hebdomadaire de consommation.

Ceci nous permet d'envisager modéliser cette croissance par une simple tendance (linéaire). Cette tendance a pour (unique) but de *redresser les mesures de consommation des années précédentes*, et ainsi d'autoriser l'estimation des coefficients des autres effets du modèle à partir de relevés couvrant 2 ou 3 années d'apprentissage.

De manière plus explicite, le modèle se définit de la manière suivante :

Phase d'apprentissage en année d'exercice Y_n

Lors de l'apprentissage, le modèle intègre l'année de mesure comme variable supplémentaire (représentée en rouge dans la formule ci-dessous). Ce terme représente une tendance linéaire modélisant la croissance (ou décroissance) annuelle de la consommation. Elle permet le redressement des mesures de consommation effectuées lors des années Y_i ($i < n$), lors d'un apprentissage effectué sur un ensemble d'années $Y_1 < \dots < Y_n$:

$$\text{Conso}_t = \text{Intercept} + \alpha_0 \text{year}(t) + \alpha_1 DJ_t + \alpha_2 DJ_{t-1} + \alpha_3 DJxVent_t + \beta_{dim} I_{t=dim} + \dots + \beta_{vend} I_{t=vend}$$

Phase d'inférence ou d'estimation de la pointe en année d'exercice Y_n

Grâce au redressement de données, les coefficients estimés lors de la phase d'apprentissage restent ainsi représentatifs de l'année d'exercice Y_n . Ils peuvent donc être employés lors de l'estimation de la pointe à partir du modèle de consommation suivant :

$$\text{Conso}_t = \text{Intercept} + \alpha_0 Y_n + \alpha_1 DJ_t + \alpha_2 DJ_{t-1} + \alpha_3 DJxVent_t + \beta_{dim} I_{t=dim} + \dots + \beta_{vend} I_{t=vend}$$

4.2 Performances de généralisation

Pour valider le modèle proposé dans la section précédente, et notamment l'hypothèse de variation lente des effets du modèle, nous avons étudié ses performances de généralisation sur l'année qui suit l'année d'exercice (dernière année d'apprentissage). Il est à noter que la tendance linéaire modélisée n'est pas projetée lors de ces estimations. Nous postulons que des coefficients permettant une meilleure estimation de la consommation de l'année suivante, alors même qu'ils ont été ajustés sur plusieurs années passées, sont plus représentatifs de l'année en cours.

Dans la Figure 8, est représentée la MAPE (Mean Absolute Percentage Error – erreur absolue relative moyenne) du modèle de consommation ajusté sur 1, 2 ou 3 années d'apprentissage, à partir de la procédure proposée de redressement des données par tendance linéaire. La MAPE est ici mesurée sur l'année suivant l'année d'exercice. On peut observer une stabilité, voire une amélioration, des performances attribuées aux modèles ajustés sur 2 voire 3 années d'apprentissage. Ceci prouve que les coefficients estimés sur davantage de données passées restent tout autant, voire davantage, représentatifs de la consommation de l'année immédiatement suivant l'apprentissage, et donc certainement de l'année d'exercice également. L'évolution des coefficients obtenus à travers les différentes années d'exercice est fournie en annexe 6.4.

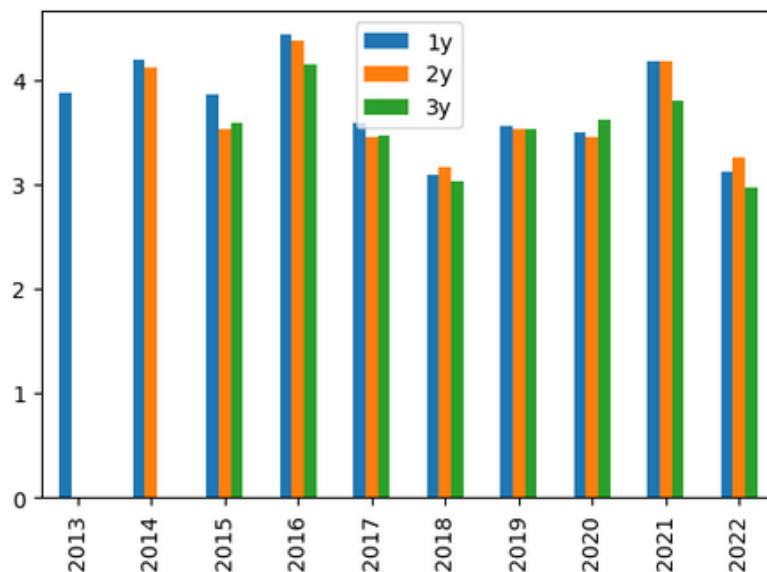


Figure 8 : MAPE du modèle de consommation ajusté sur 1, 2 ou 3 ans, à partir de la procédure proposée (redressement des données par tendance linéaire). La MAPE est mesurée à partir des données issues de l'année suivant l'année d'exercice (dernière année d'apprentissage, représentée en abscisse ici).

4.3 Variabilité de la pointe

Disposer d'une bonne représentativité de l'année d'exercice par les coefficients de régression est un préalable requis pour valider l'apprentissage sur plusieurs années. Maintenant qu'il est acquis (section précédente), nous allons valider le choix d'un apprentissage sur historique plus long, à travers l'étude des indicateurs métiers qui nous intéressent, i.e. ceux propres à la variabilité de la pointe.

4.3.1 Analyse de sensibilité à la profondeur d'historique d'apprentissage

Pour commencer, nous donnons en Figure 9, l'évolution de la pointe estimée et de sa décomposition, obtenues à partir du modèle de consommation ajusté sur 3 années d'apprentissage. Les effets d'Intercept et de tendance linéaire (voir la formule en *phase d'inférence* de la section 4.1), sont ici sommés dans le coefficient 'const'. La comparaison avec le résultat obtenu à partir d'un modèle ajusté sur 1 année (voir Figure 2) est frappante. On observe ici une évolution beaucoup plus lisse (monotone si on excepte l'année d'exercice du COVID en 2021) de la pointe, avec des contributions issues de chaque effet beaucoup moins variables d'une année sur l'autre. Du fait de la moindre variabilité des coefficients, les conditions météorologiques sous-jacentes restent ici fixes (celles du 15/01/2004, même pour l'exercice 2023).

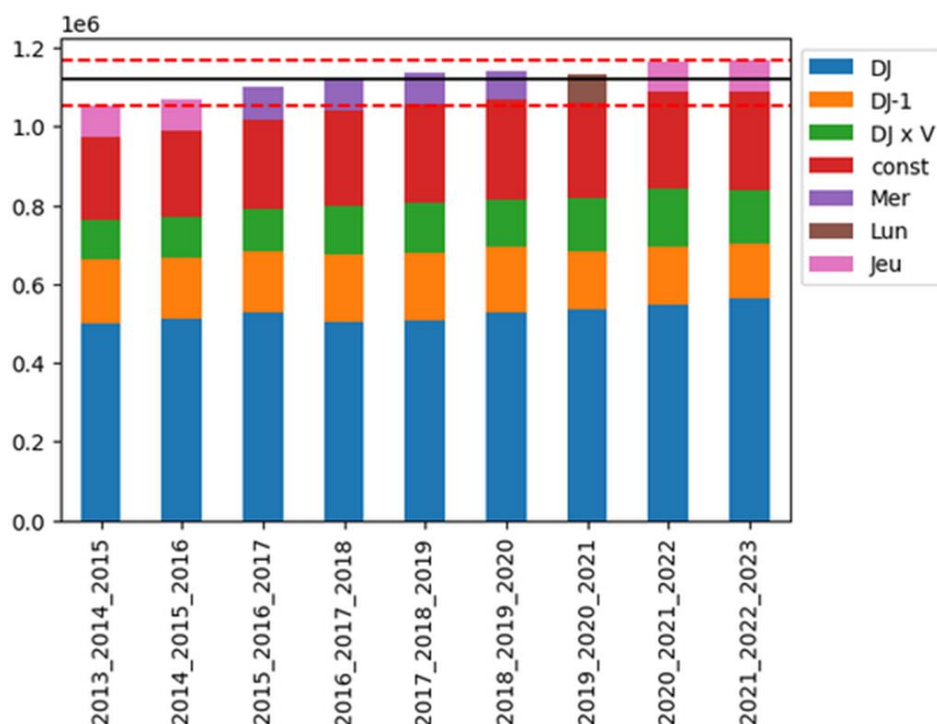


Figure 9 : Evolution de la pointe estimée et de sa décomposition suivant les différents effets du modèle. Le modèle de consommation est ici ajusté sur 3 années consécutives (données en abscisses). 'const' somme les effets 'Intercept' et ' $\alpha_0 Y_n$ '. Les valeurs en ordonnées sont exprimées en GJ.

Pour mieux visualiser les différences de résultats selon la profondeur d'historique d'apprentissage, nous avons représenté en Figure 10, l'évolution de la valeur de la pointe estimée grâce au modèle de

consommation ajusté respectivement sur 1, 2 ou 3 années d'apprentissage. On constate bien l'évolution monotone de la pointe estimée à partir d'un modèle ajusté sur 3 ans (hormis en 2021). De manière générale, la pointe est plus souvent inférieure à celle obtenue à partir d'un modèle ajusté sur 1 an. C'est notamment le cas en 2022 où la pointe estimée à partir d'1 an d'historique avait atteint le niveau record de 1195 TJ (contre 1166 TJ avec 3 ans d'apprentissage). En 2023, l'estimation de la pointe s'était « effondrée » à 1151 TJ (contre 1167 TJ avec 3 ans d'apprentissage). Ainsi, l'énorme variabilité annuelle de la pointe, auparavant enregistrée entre 2022 et 2023, disparaît dès lors qu'on utilise 3 années d'apprentissage. La réduction de la variabilité annuelle, alors qu'augmente la profondeur d'historique d'apprentissage, est un constat qui se généralise (voir Table 1). Les valeurs numériques des estimations de la pointe sont fournies en annexe (Table 3).

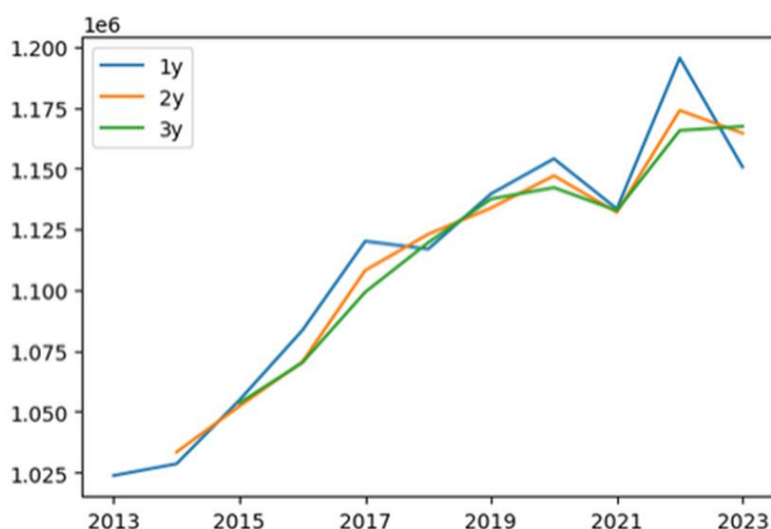


Figure 10 : Comparaison des valeurs de pointe estimées par le modèle de consommation ajusté sur 1 ('1y'), 2 ('2y') ou 3 ('3y') années d'apprentissage.

Table 1 : Valeurs moyennes ('mean') et maximales ('max') des variations annuelles (en valeur absolue), des pointes estimées depuis 2015 (hormis celles des années 2021 et 2022, impactées par le COVID), pour différentes profondeurs d'historique d'apprentissage du modèle de consommation. Les valeurs sont exprimées en GJ.

	1y	2y	3y
mean	25259	17570	15035
max	44815	37632	29046

4.3.2 Analyse de sensibilité au retrait des données impactées par le COVID

Nous avons cherché à étudier la résilience de l'approche d'estimation de la pointe (à partir d'un modèle de consommation ajusté sur 3 années), en étudiant l'impact sur ses résultats de la présence de données associées à un choc ponctuel de demande (COVID). Les résultats illustrant l'impact du retrait de l'année 2021 des historiques d'apprentissage sont présentés en Figure 11.

On peut constater 2 choses sur cette figure. D'une part, le retrait de l'année 2021, lors de l'exercice d'estimation de 2021, s'apparente à un exercice dans lequel la consommation est modélisée à partir d'un apprentissage sur 2 années consécutives (2019 et 2020), et dont l'estimation de la pointe s'appuie sur une projection de la tendance à 2021 (ce qui ne contredit pas la formule d'inférence donnée plus haut, 2021 étant l'année d'exercice ici). L'estimation de la pointe poursuit dès lors sa tendance à la hausse, ce qui semble être une hypothèse assez réaliste. Ainsi, cette approche pourrait fournir une estimation de la pointe satisfaisante une année de choc de demande.

Par ailleurs, il est intéressant de constater que le retrait de l'année 2021 de l'historique d'apprentissage, lors des exercices de 2022 et 2023, provoque une réapparition d'une décroissance de la pointe mesurée entre 2022 et 2023. S'agit-il d'une décroissance structurelle de la consommation et de sa pointe qui s'entame ici ? L'estimation sur 2 années seulement (non consécutives pour l'exercice 2022) est peut-être trop variable pour permettre de l'affirmer. Seul l'exercice d'estimation conduit sur de nouvelles années permettra de le confirmer. Quant à savoir s'il est préférable ou non de retirer l'année 2021 pour l'estimation de la pointe en 2022 et 2023, les résultats présentés ici ne permettent pas objectivement d'en juger. L'aspect plus lisse des estimations conservant l'année 2021 dans l'historique d'apprentissage (et redressant en partie les données par tendance) pourrait être préférable.

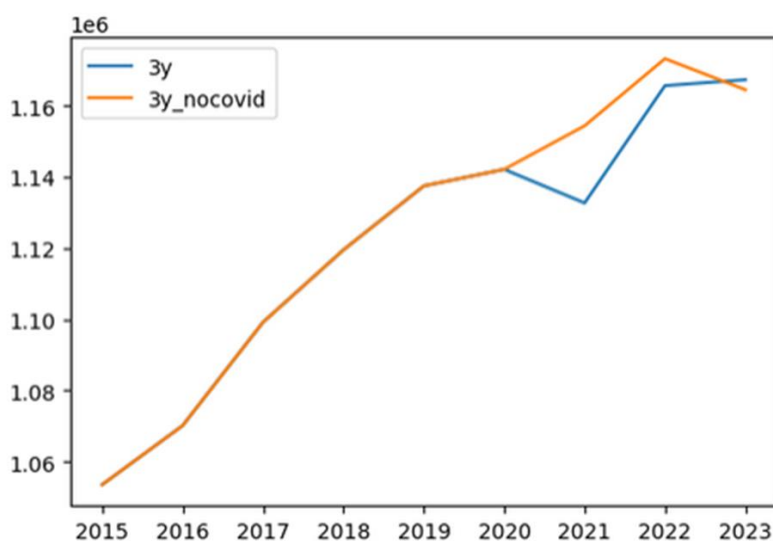


Figure 11 : Evolution de la pointe estimée à partir d'un modèle de consommation ajusté sur 3 années d'apprentissage consécutives, hormis lors du retrait de l'année 2021 de l'historique d'apprentissage (« nocovid »).

4.3.3 Analyse de sensibilité suite à la projection de la pointe par coefficients d'ajustement

Pour finir, nous étudions l'impact de l'apprentissage réalisé sur 3 années, sur la projection des pointes entre Y+2 et Y+5 (horizons concernés pour l'approvisionnement en gaz), suivant l'année d'exercice Y. La méthode de projection de la pointe estimée se base sur des coefficients dits d'« ajustement » qui représentent *le ratio entre les volumes hivernaux déthermalisés*, fournis par le département de prévision de la demande d'Énergir d'une part, et par le modèle de régression d'autre part. Ainsi, l'estimation des coefficients de régression sur 3 années plutôt qu'une impacte-t-elle les projections de pointes, à la fois à travers l'estimation de la pointe à l'année Y, et à travers les coefficients d'ajustement. Les variations de la pointe estimée en Y, et celles du volume hivernal déthermalisé, ont tendance à se suivre (comparer Figure 10 et Figure 13), ce qui entraîne a priori un sens de variation opposé entre estimation de la pointe en Y et coefficients d'ajustement. Les résultats présentés ici ont pour but de voir si une forme de compensation est observée entre ces 2 phénomènes ou si l'un des 2 l'emporte.

Dans la Figure 12, dont les valeurs sont fournies en annexe 6.7, sont représentées les projections obtenues à partir des coefficients de régression estimés à l'aide d'une année (lignes pointillées) ou 3 ans (lignes pleines) d'historique d'apprentissage. La CT 2023 s'appuie sur les coefficients de régression issus de la calibration 2019-2020, et non 2020-2021, pour éviter l'impact du COVID. La CT 2024 s'appuie tantôt sur cette même calibration (graphe de gauche), tantôt sur celle de l'année 2021-2022 (graphe de droite). Les premiers résultats sont destinés à reproduire ceux communiqués à la Régie. Les seconds sont plus adaptés à la comparaison des approches utilisant différentes profondeurs d'historique d'apprentissage.

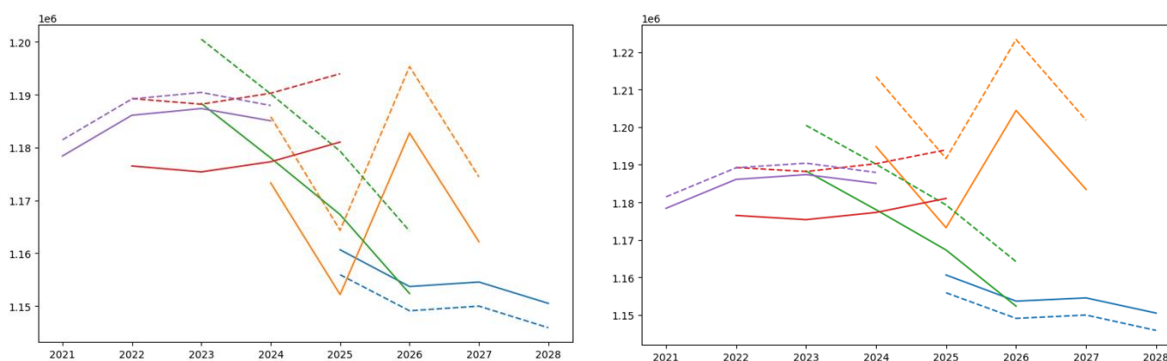


Figure 12 : Projections de Y+2 à Y+5 de la pointe estimée en Y, à partir des coefficients de régression estimés sur 1 an (courbes en pointillées) ou sur 3 ans (courbes pleines), et des coefficients d'ajustement qui en résultent (voir le texte pour plus de détails). Les différents exercices successifs sont représentés par des couleurs (CT 2021 en violet, CT 2022 en rouge, etc.). La CT 2023 s'appuie sur les résultats de calibration de 2019-2020. La CT 2024 s'appuie tantôt sur cette même calibration (graphe de gauche), tantôt sur celle de l'année 2021-2022 (graphe de droite).

On constate principalement 2 choses dans ces résultats. D'une part l'ordonnement des projections, entre apprentissages réalisés sur 1 ou 3 ans, suit celui de la pointe estimée en Y (voir Figure 10). Ainsi, la stabilisation de l'estimation de la pointe entre 2022 et 2023, grâce à un apprentissage réalisé sur 3

ans, permet d'obtenir des projections de pointes plus basses qu'auparavant pour la CT 2024, et plus hautes qu'auparavant pour la CT 2025. D'autre part, si une chute des projections observée entre les CT 2024 et CT 2025 demeure, c'est principalement dû à celle des volumes prédits par le département de prévision de la demande (voir Figure 17). Les estimations de la pointe entre 2022 et 2023, réalisées à partir de 3 années d'apprentissage, sont stables et n'accroissent pas cette chute, alors que celles réalisées à partir d'1 année d'apprentissage l'accroissent (figure de droite). Les projections obtenues à partir de 3 années d'apprentissage (figure de droite, lignes pleines), sont ainsi plus proches de celles qu'Énergir a choisi de communiquer à la Régie (figure de gauche, lignes pointillées), ce qui prouve une évolution positive des résultats.

Un point d'attention demeure vis-à-vis de la chute des prévisions de demande fournies par le département de prévision de demande, entre les CT 2024 et CT 2025. Énergir nous a confirmé que cette chute était le résultat d'une hypothèse prospective assumée, et pas celui a priori d'un problème de surajustement similaire à celui rencontré dans le modèle de régression étudié ici. Il n'y a donc a priori pas de problème de mise en cohérence entre ces 2 modèles, et la formule de projection de la pointe s'appuyant sur les coefficients d'ajustement demeure ainsi valide.

A retenir : mémo de la méthode proposée

Exemple ici de la projection de la pointe pour l'exercice CT 2025.

1. Augmentation de l'échantillon d'apprentissage aux 3 derniers hivers (2020-2021, 2021-2022, 2022-2023) au lieu du dernier hiver uniquement (2022-2023)
2. **Ajout d'une variable « Année »** dans la calibration du modèle qui prend donc 3 valeurs différentes en fonction de l'année considérée (2021, 2022, 2023)
3. Obtention des coefficients de la régression (GJ, Année, DJ, DJ-1, DJV, Férié, Dim, Lun, Mardi, Mercredi, Jeudi, Vendredi)
4. Projection sur 30 ans uniquement avec les variables météo (DJ, DJ-1, DJV) pour obtenir la journée de pointe hivernale
5. Projection du modèle avec les variables correspondant à la journée de pointe (DJ, DJ-1, DJV) + le maximum des coefficients calendaires (Lundi-Vendredi) + le coefficient constant (GJ) + le **coefficient Année * la dernière année de calibration (ici 2023)**

On obtient alors la nouvelle valeur de consommation de pointe hivernale.

5 Conclusion

Au cours de ce mandat, Artelys a été sollicité afin d'auditer la méthodologie mise en place par Énergir afin d'estimer la demande gazière en journée de pointe. Cette sollicitation faisait suite au constat d'une variabilité récemment accrue des coefficients de régression, et des estimations de la demande de pointe en résultant, lors des exercices des causes tarifaires de 2024 et 2025. Ce document présente les résultats d'analyses conduites par Artelys pour répondre à ce mandat. Ces analyses ont reposé sur l'étude de l'évolution du modèle implémenté par Énergir, à partir de 10 années d'historiques de consommation journalière, ainsi que de 30 années d'historiques de données météorologiques.

La première étape fut d'être en mesure de reproduire les résultats et constatations d'Énergir. Artelys en a ainsi pu **réitérer le constat d'une forte variabilité annuelle** des coefficients de régression, et plus particulièrement de celui associé à l'effet DJ x V, ainsi que de la pointe entre les années 2022 (avec une pointe estimée à 1195 TJ, pour les clients couverts par le périmètre de la régression, sans projection dans le futur) et 2023 (1151 TJ).

La piste d'un surajustement ou d'une mauvaise formulation du modèle liée à l'effet DJ x V avait été envisagée par Énergir. Artelys a ainsi conduit une étude exhaustive aboutissant à la **confortation des conclusions issues des études précédentes** (sur la méthode de normalisation des revenus), qui **confirmaient l'intérêt d'intégrer l'effet DJ x V**, ainsi formulé, dans le modèle. Pour ce faire, Artelys s'est appuyé sur l'état de l'art de la littérature scientifique, ainsi que sur une analyse de données approfondie visant à comparer les performances d'un modèle intégrant ou pas l'effet DJ x V.

Enfin, Artelys s'est intéressé à la **mise en évidence d'un surajustement touchant actuellement l'ensemble des coefficients** de régression du modèle. Pour y remédier, Artelys a proposé de **réaliser le calage du modèle à partir d'un historique d'apprentissage profond** de 3 ans. Afin d'obtenir malgré tout une bonne représentativité des coefficients de régression, vis-à-vis des habitudes de consommation les plus récentes, Artelys a proposé l'intégration d'une tendance dans le modèle lors de l'estimation (afin de tenir compte de la croissance, bien qu'aujourd'hui réduite, de la consommation). Artelys a prouvé la validité de l'approche proposée par l'analyse de données. Cette approche a notamment permis de réduire significativement la variabilité annuelle des coefficients de régression et de la pointe associée.

6 Annexe

6.1 Volume hivernal de consommation déthermalisée

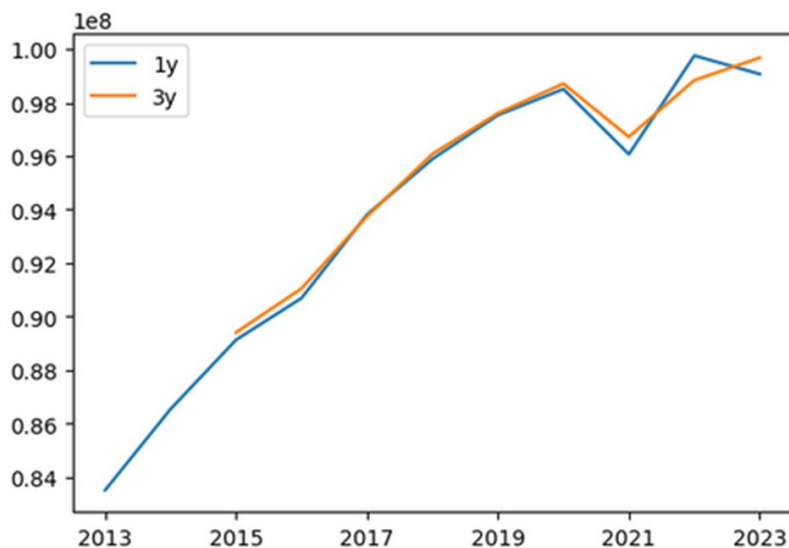


Figure 13- Evolution annuelle de la consommation hivernale déthermalisée, fournie par le modèle de régression avec 1 ('1y') ou 3 ('3y') années d'apprentissage, et calculée à partir d'une météo « normale », moyenne des 30 dernières années.

6.2 Résidus modèle sans $D \times V$, divisés par DJ, en fonction de V

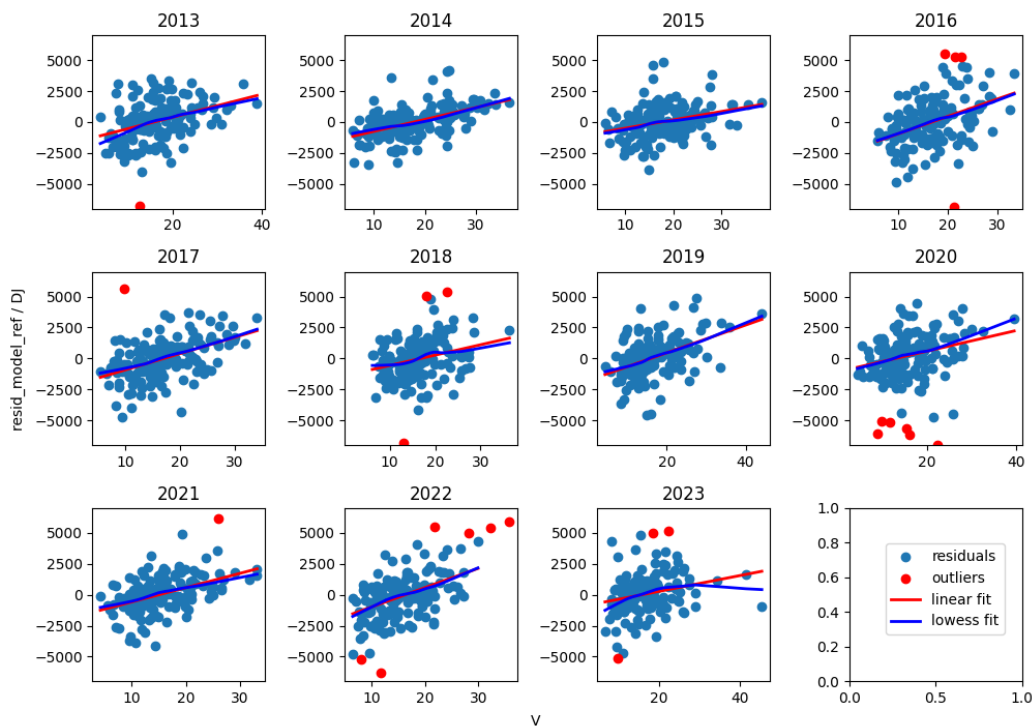


Figure 14 - Résidus du modèle sans $D \times V$ divisés par DJ (en $GJ/^\circ C$), en fonction de V (m/s)

6.3 Biais annuel lié à l'absence de modélisation de la croissance de consommation de gaz

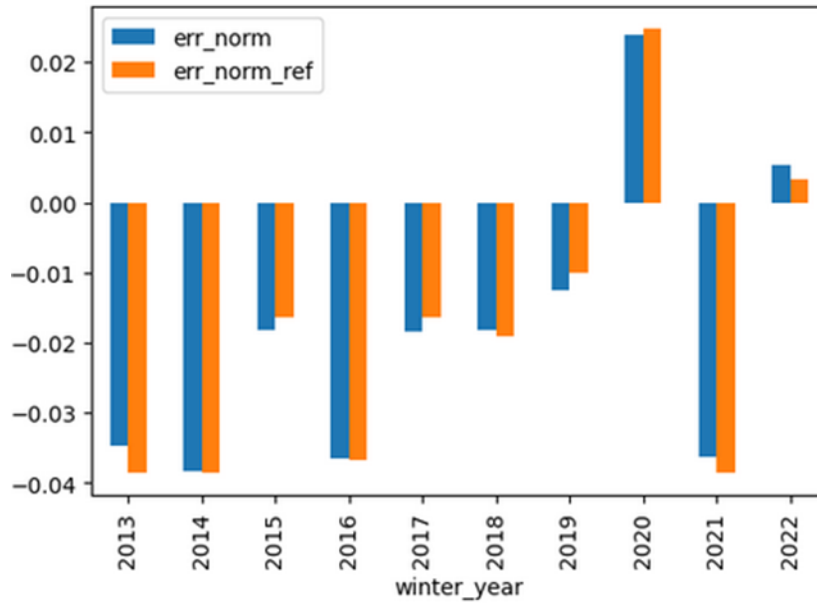


Figure 15 : Biais annuel des modèles intégrant DJ x V (bleu) ou pas (orange). Les erreurs associées ont été mesurées sur l'année précédant l'apprentissage. Le biais est ici exprimé en ratio de la consommation hivernale journalière moyenne déthermalisée.

6.4 Evolution des coefficients de régression lors d'un ajustement du modèle de consommation sur 3 années d'apprentissage

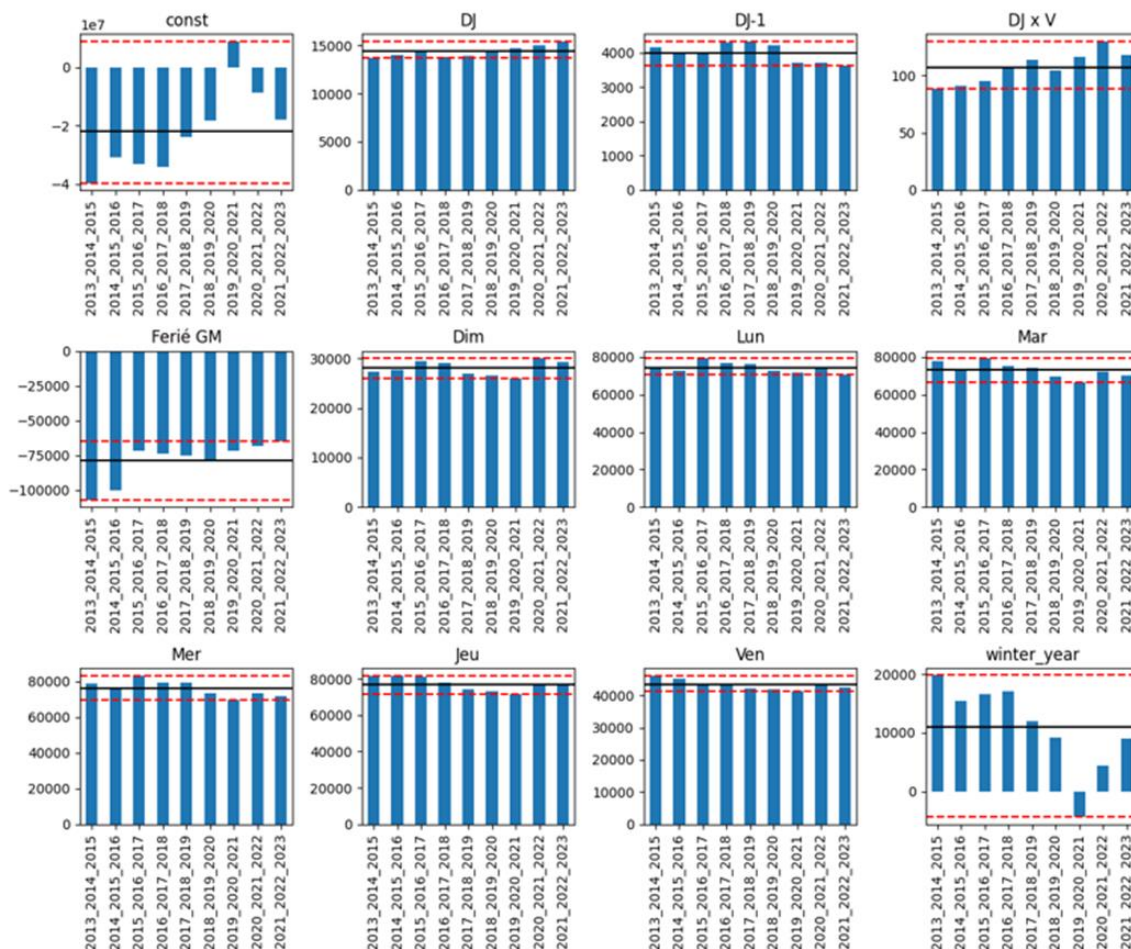


Figure 16 : Evolution des coefficients de régression entre les hivers 2015 et 2023. Ces coefficients ont été ajustés à partir de 3 années consécutives d'apprentissage. L'unité de mesure en énergie est le GJ.

Table 2 : Valeurs numériques associées. 'const' somme ici les effets 'Intercept' et ' $\alpha_0 Y_n$ '

	2015	2016	2017	2018	2019	2020	2021	2022	2023
const	207568,7	216886,4	223344,3	241825,9	250083,1	253898,6	243922,7	245835,9	251001,6
DJ	13658,37	13965,53	14394,42	13833,52	13893,22	14491,76	14716,49	15012,23	15399,78
DJ-1	4174,853	3989,297	4008,252	4314,936	4336,867	4215,122	3718,946	3727,613	3622,193
DJ x V	88,32249	91,26498	95,24174	107,6071	113,6303	104,5199	116,2142	129,6264	117,6637
Ferié GM	-106723	-99828,2	-71596,9	-73435,9	-74783,4	-78677	-71664,9	-68063,4	-64696,9
Dim	27336,87	27678,43	29478,06	28966,71	27012,87	26538,74	26019,81	29958,74	29277,9
Lun	74417,2	72712,14	79247,84	76637,91	76439,99	72867,12	71699,35	74447,32	70712,45
Mar	77926,35	73536,3	79160,6	75245,23	74046,08	69806,79	66366,28	72316,75	70227,87
Mer	78802,52	76413,49	83372,91	79458,23	79201,51	73421,17	69773,35	73210,71	71786,21
Jeu	81872,88	81775,2	80968,8	78013,65	74345,67	73171,91	71408,35	76159,46	76384,78
Ven	46160,52	45144,17	43382,56	42953,88	42034,89	41809,39	41277,37	43781,04	42518,33

6.5 Evolution des valeurs de pointe avec la profondeur d'historique d'apprentissage

Table 3 : Valeurs de la pointe estimée lors des différents exercices annuels, à partir du modèle de consommation ajusté sur 1 ('1y'), 2 ('2y') ou 3 ('3y') années d'apprentissage. Les valeurs sont exprimées en GJ.

	1y	2y	3y
2013	1023675		
2014	1028522	1033383	
2015	1054737	1052214	1053540
2016	1083400	1070460	1070111
2017	1120069	1108092	1099157
2018	1116801	1122936	1119461
2019	1139551	1133700	1137434
2020	1153981	1146988	1142090
2021	1133377	1131935	1132652
2022	1195384	1173855	1165608
2023	1150569	1164469	1167267

6.6 Volumes hivernaux déthermalisés prédits par le département de prévision de demande d'Énergir pour les CT 2021 à CT 2025.

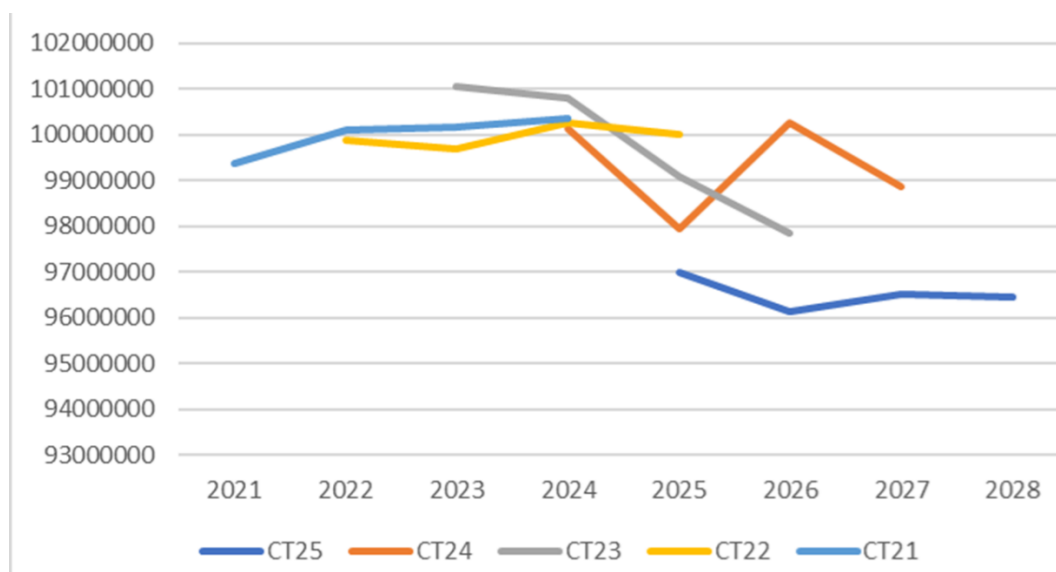


Figure 17 : Volumes hivernaux déthermalisés prédits par le département de prévision de demande d'Énergir pour les CT 2021 à CT 2025.

6.7 Valeurs numériques de la figure 12 : Projections de pointes ajustées

CT 2025 (calibration 2022-2023)				
Année de projection	2025	2026	2027	2028
Méthode Énergir	1 155 924	1 149 108	1 149 992	1 145 899
Méthode Artelys	1 160 648	1 153 695	1 154 550	1 150 512
CT 2024 (calibration 2021-2022)				
Année de projection	2024	2025	2026	2027
Méthode Énergir	1 213 410	1 191 646	1 223 327	1 201 851
Méthode Artelys	1 194 813	1 173 236	1 204 440	1 183 423
CT 2024 (calibration 2019-2020)				
Année de projection	2024	2025	2026	2027
Méthode Énergir	1 185 827	1 164 340	1 195 306	1 174 432
Méthode Artelys	1 173 298	1 152 175	1 182 732	1 162 170
CT 2023 (calibration 2019-2020)				
Année de projection	2023	2024	2025	2026
Méthode Énergir	1 200 480	1 190 129	1 179 292	1 164 203
Méthode Artelys	1 188 317	1 178 022	1 167 302	1 152 367
CT 2022 (calibration 2019-2020)				
Année de projection	2022	2023	2024	2025
Méthode Énergir	1 189 259	1 188 189	1 190 262	1 193 954
Méthode Artelys	1 176 503	1 175 378	1 177 297	1 181 020
CT 2021 (calibration 2018-2019)				
Année de projection	2021	2022	2023	2024
Méthode Énergir	1 181 449	1 189 178	1 190 413	1 187 941
Méthode Artelys	1 178 391	1 186 097	1 187 363	1 185 047

7 Bibliographie

- [1] G. Metro, «Modifications proposées à la méthode de normalisation des revenus,» Cause tarifaire 2008, R-3630-2007.
- [2] G. Metro, «Modifications proposées à la méthode de normalisation des revenus,» Cause tarifaire 2009, R-3662-2008.
- [3] a. T. H. Jingrui Xie, «Wind Speed for Load Forecasting Models,» *Sustainability*, pp. 9(5), 795, 2017.
- [4] e. a. Francesco Apadula, «Relationships between meteorological variables and monthly electricity demand,» *Applied Energy*, p. (98) 346–356, 2012.
- [5] e. a. M. Sperandio, «Probabilistic demand forecasting to minimize overtaking the transmission contract,» *Electric Power Systems Research*, pp. (112) 27-36, 2014.
- [6] S. T. e. al., «Residuals modeling with wind data to improve short-term load forecast,» chez *IEEE Trondheim PowerTech*, 2011.