

MODÉLISATION INDIVIDUELLE ET AGRÉGÉE DES CONSOMMATIONS ÉLECTRIQUES : VERS UNE AMÉLIORATION DU MODÈLE DE CHARGE POUR LE DIMENSIONNEMENT DU RÉSEAU

Problématique industrielle

Enedis est le gestionnaire du réseau public de distribution d'électricité en France, chargé d'exploiter et de moderniser plus de 1,4 million de kilomètres de lignes pour alimenter 40 millions de clients. Acteur clé de la transition écologique et opérateur de données mature, l'entreprise accompagne l'intégration massive des énergies renouvelables, le développement de la mobilité électrique et l'essor de nouveaux usages tout en garantissant la qualité et la sécurité de l'alimentation électrique. Dans ce contexte, disposer d'un modèle des séries temporelles de consommations électriques robuste, précis et scientifiquement fondé est essentiel pour dimensionner efficacement le réseau et anticiper les besoins futurs. La qualité d'un tel modèle est devenue un enjeu stratégique majeur : les évolutions rapides des usages (mobilité électrique, pompes à chaleur, autoconsommation), la variabilité croissante des comportements individuels et les incertitudes climatiques rendent le dimensionnement plus complexe. Enedis s'appuie sur une approche *bottom-up*. Afin de dimensionner correctement un poste ou un tronçon du réseau (entre une vingtaine et une centaine de clients raccordés), les consommations électriques individuelles sont d'abord estimées. Dans un second temps, elles sont agrégées afin de représenter la charge locale du tronçon ou du poste. Des indicateurs critiques comme les quantiles élevés de consommation (par exemple le 90e percentile annuel) sont utilisés pour calibrer les investissements. Dans ce contexte, une question devient centrale : comment mieux modéliser la variabilité individuelle et collective, en particulier liée aux jours d'absence ou de présence des consommateurs ?

Les modèles utilisés pour prévoir les consommations journalières mélangent souvent implicitement des jours de présence (activité normale, occupation du logement ou du local professionnel) et des jours d'absence (vacances, fermetures, week-ends prolongés...). Or l'absence modifie profondément la distribution des consommations : niveaux moyens plus faibles, variances réduites, dépendances calendaires fortes, effets saisonniers marqués. L'objectif de ces travaux sera donc d'introduire explicitement la présence/absence dans le modèle afin d'améliorer la structure statistique des résidus individuels (normalité, homogénéité de variance), de renforcer la qualité des agrégations de population et ainsi d'améliorer l'estimation des quantiles critiques utilisés pour le dimensionnement (notamment à 90%).

Modélisation mathématique

Les doctorant·es seront invité·es à explorer plusieurs axes où les apports mathématiques sont majeurs.

1. Modélisation paramétrique de la consommation individuelle

Actuellement, les compteurs Linky remontent chaque jour les consommations électriques individuelles journalières. Seuls quelques clients donnent accès à leur consommation au pas de temps

demi-horaire (l'accès aux données intrajournalières est soumise à consentement du client). De fait, la modélisation actuelle de la consommation électrique infra-jour pour un foyer i fait intervenir des quantités propres au foyer i estimées sur des données journalières ainsi que des “profiles” propres à la catégorie $\kappa(i)$ du clients (résidentiel, professionnel etc.) et estimés sur les clients mettant à disposition leurs données demi-horaires. Plus précisément, pour un jour t et une demi-heure h ; nous considérons le vecteur de variables calendaires c_t (qui contient, par exemple, l'information du jour de la semaine, de la saison etc.) ainsi que la température T_t^h . Avec $\Delta T_t^h = \max(0, T_{\text{seuil}} - T_t^h)$, où T_{seuil} est la température de déclenchement du chauffage, la modélisation actuelle que la consommation électrique du foyer i est :

$$Y_t^{h,i} = a_{\kappa(i)}^h(c_t)E^i(c_t) + b_{\kappa(i)}^h(c_t)s^i\Delta T_t^h + \varepsilon_t^{h,i}, \quad \text{avec} \quad \varepsilon_t^{h,i} \sim \mathcal{N}(0, V^i\sigma_{\kappa(i)}^h).$$

L'énergie journalière $E^i(c_t)$, la termo-sensibilité du foyer s^i et la variance V^i ne dépendent pas de la demi-heure h : elles sont estimées à partir de la courbe de charge journalière du foyer. Les profiles $a_{\kappa(i)}^h$ et $b_{\kappa(i)}^h$ permettent de déformer la courbe pour obtenir une modélisation infra-journalière. Enfin, $\sigma_{\kappa(i)}^h$ permet de modéliser la différence de variance entre les différentes demi-heures de la journée (car la variabilité est moindre la nuit).

Une première piste consisterait à enrichir le modèle en introduisant une variable aléatoire binaire A_t^i indiquant l'absence ou la présence d'individus dans le logement du foyer i (la distribution de A_t^i , qui dépendra, a minima, de c_t et $\kappa(i)$ étant à modéliser) et de considérer :

$$Y_t^{h,i} = a_{\kappa(i)}^h(A_t^i)E^i(A_t^i, c_t) + b_{\kappa(i)}^h(A_t^i)s^i\Delta T_t^h + \varepsilon_t^{h,i}, \quad \text{avec} \quad \varepsilon_t^{h,i} \sim \mathcal{N}(0, V^i(A_t^i)\sigma_{\kappa(i)}^h).$$

La consommation agrégée des foyers serait alors un mélange gaussien que nous espérons plus pertinent que la modélisation actuelle. Une attention particulière devra être portée aux résidus. Les distributions des consommations sont souvent asymétriques, parfois quasi-lognormales. Nous espérons que l'intégration de la variable ‘absence/présence’ améliore la normalité des résidus et stabilise leur variance.

D'autres modélisations paramétriques sont évidemment à envisager (modèles additifs, mélanges gaussiens, modèles hiérarchiques bayésiens, etc.).

2. Simulation Monte Carlo de trajectoires annuelles individuelles, agrégation et estimation des quantiles réseau

Le modèle permettra de générer des années synthétiques afin de comparer des approches naïves (sans absence/présence) et enrichies (avec absence/présence), d'analyser les dépendances temporelles et calendaires, et d'étudier l'impact de l'indépendance (ou non) entre individus (conditionnellement au calendrier et à la météo).

Les agrégations typiques d'Enedis à la maille d'un poste de distribution ou un tronçon du réseau (agrégats de 40 clients en moyenne, mais potentiellement entre 20 et 100 selon les contextes) seront étudiées pour déterminer :

- l'effet de la séparation présence/absence sur la variance agrégée,
- le comportement asymptotique vers la normalité,
- l'impact sur l'estimation du quantile 90% (paramétrique vs empirique).

3. Extensions possibles

Pour approfondir le sujet, plusieurs pistes optionnelles peuvent être envisagées telles que l'intégration de la météo au-delà de la température (nuages, précipitations), un *clustering* (définition des $\kappa(i)$) des foyers basés uniquement sur les séries temporelles, la pertinence d'une modélisation globale (sans segmentation par classe), la redéfinition de la variable ‘absence/présence’, etc.

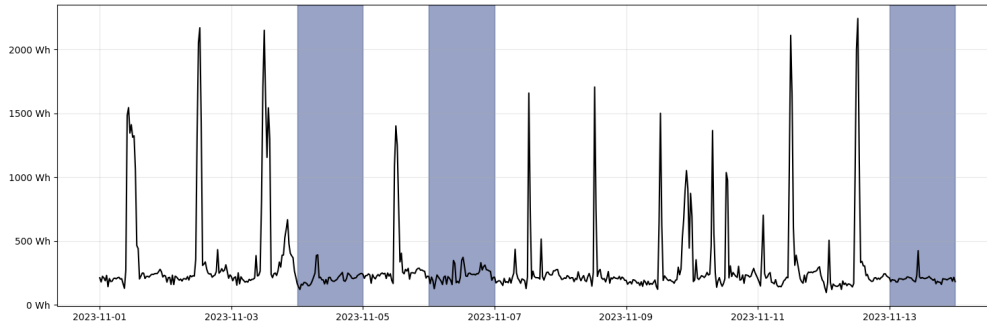


FIGURE 1 – Courbe de charge individuelle simulée au pas de temps demi-horaire avec, en bleu, les jours d’absence.

Jeu de données

Le projet s’appuie sur des courbes de charge simulées *open data*, de plusieurs milliers de clients, sur une année complète, au pas de temps demi-horaire, sur trois segments : Résidentiel (RES), Entreprises Basse Tension de puissance souscrite ≤ 36 kVA (PRO), Professionnels Basse Tension de puissance souscrite >36 kVA (ENT). Nous mettrons à disposition une variable estimée ‘absence/présence’ (ainsi que la méthode de labellisation qui pourra être remise en cause lors de la semaine, car Enedis n’a bien entendu pas de vérité terrain sur l’occupation d’un logement ou bâtiment). Une série temporelle individuelle synthétique est représentée sur quelques jours sur la Figure 1. Des données de températures France sont également fournies pour modéliser la thermo-sensibilité des foyers. Ces données comportent plusieurs limites (absence de localisation, simplifications), mais l’objectif principal est la rigueur méthodologique et la reproductibilité scientifique.