

EDF R&D : PROPOSITION DE STAGE

Estimer la part d'autoconsommation solaire dans la consommation électrique résidentielle à l'aide de méthodes d'inférence causale

Le contexte :

Environnement de travail :

Pour répondre aux besoins des directions métiers (et notamment EDF Commerce), le département SEQUOIA (Services, Economie, Questions hUmanes, Outils innovants et IA) met en œuvre des méthodes d'apprentissage basées sur la statistique, le Machine Learning et le Deep Learning

Au sein de ce département, le groupe E7C est spécialisé en analyse de données de consommation, et en modélisation statistique dans le but de concevoir des nouvelles offres de fournitures ou des services.

Ce stage s'inscrit dans le projet « Smart Customer Data », dont l'objectif est de recueillir et valoriser les données clients. Par ailleurs, un volet méthodologique du projet identifie et développe des traitements innovants et créateurs de valeur pour la relation client et les offres à partir des courbes de charge au pas de temps des compteurs communicants (Linky et Gazpar), mais aussi à pas de temps fin (seconde).

Contexte du stage :

Afin d'éviter les risques de blackout du réseau électrique, il est crucial d'assurer en temps réel un équilibre entre la production et la consommation d'énergie. Dans un contexte d'intégration croissante d'autoconsommation solaire dans le mix énergétique, il est nécessaire de mettre à disposition des solutions de flexibilité à nos clients particuliers : produire davantage lorsque l'autoconsommation est faible et produire moins lors des pics de consommation solaire. Ainsi, il devient essentiel d'estimer de manière fine et individuelle l'autoconsommation solaire des clients résidentiels.

L'inférence causale peut être envisagée pour répondre à cette problématique. Dans ce cadre, la consommation électrique est la variable de résultat, la pose d'une installation photovoltaïque correspond à la variable de traitement, l'effet causal s'apparente à la part d'autoconsommation solaire.

Les méthodes dynamiques quasi-expérimentales (contrôle synthétique, différence de différences, two-way fixed effects...) nécessitent la connaissance de la consommation électrique avant traitement. En pratique, récupérer cette donnée s'avère être fastidieux.

Il conviendra donc de développer une méthode robuste au cas où aucune donnée avant traitement n'est disponible. L'un des enjeux du stage consiste à savoir si les méthodes statiques, conçues pour des variables aléatoires i.i.d, peuvent être facilement transposées au cadre dynamique. En particulier, sous quelles conditions est-il possible de construire des mesures de précision fiables ?

En parallèle, la consommation électrique ainsi que la décision d'installer un panneau photovoltaïque dépendent de certains facteurs, observés ou non. Les négliger rend l'identification de l'effet causal impossible. On parle de « biais de confusion ». Il sera important d'identifier des variables de contrôle pertinentes, issues des données de consommation individuelles ou d'autres sources (météo / signaux tarifaires), afin de neutraliser ce biais.

Le stage :

Le stage comportera plusieurs phases :

Etat de l'art des méthodes envisagées et montée en compétences :

- Se familiariser avec la problématique de l'inférence causale, dans un cadre dynamique.
- Établir un état de l'art des méthodes d'inférence causale traditionnelles : matching, stratification, IPW, AIPW, contrôle synthétique...
- Explorer les méthodes à base de Machine Learning : causal forests, debiased Machine Learning, TMLE, meta-learners...
- Explorer les méthodes de Deep Learning : TARNET, CF, DragonNET...

- Comparer ces différentes catégories de modèles aux algorithmes plus récents. Bibliographie non exhaustive des modèles :
 - Dunipace, E. (2021). *Optimal transport weights for causal inference*. arXiv preprint arXiv:2109.01991.
 - Ouali, M. et al. (2025). *Balanced Twins: Causal Inference in Time Series with Hidden Confounding*.

Prise en main des données :

- Jeu de données comprenant :
 - Données de consommation mensuelles et journalières pour clients avec et sans installation photovoltaïque.
 - Courbes de charge au pas de temps 30 minutes pour les deux groupes.
 - Questionnaires de connaissance client associés.
 - Données météorologiques.
 - Signal tarifaire.
- Exploration des statistiques descriptives et nettoyage (valeurs aberrantes et manquantes).

Mise en œuvre :

- Implémenter les méthodes identifiées lors de la revue de littérature en Python ou en R.
- Adapter ces méthodes à notre usage et à nos données.
- Identifier les variables de contrôle pertinentes.
- Réaliser un benchmark (comparaison) des différentes méthodes.

Capitalisation des travaux effectués :

- Rédaction d'un rapport (revue de littérature, comparaison, conclusion des résultats obtenus).
- Code Python / R accompagné d'une documentation.

Informations complémentaires :

La R&D propose ce stage de fin d'études, à destination d'étudiants en écoles d'ingénieurs ou Master 2, spécialisés en Statistique / Data Science / Econométrie / Deep Learning / IA.

Compétences :

- Data management (gestion, nettoyage, exploration, SQL...).
- Connaissances en Machine Learning (supervisé / non-supervisé).
- Connaissances en statistique (estimation ponctuelle, mesures de précision, bootstrap...).
- Modélisation de séries temporelles.
- L'absence de cours ou d'expérience préalable en inférence causale n'est pas un frein. Nous valorisons avant tout la curiosité et la motivation de l'étudiant.
- En fonction des méthodes envisagées, l'étudiant sera amené à utiliser les langages de programmation Python et R. Ainsi, la connaissance d'un des deux langages est nécessaire, connaître les deux est un plus.

Dates : à partir de février / mars 2026 (flexible).

Lieu du stage : EDF Lab Paris-Saclay – Recherche et Développement, 7 Bd Gaspard Monge, 91120 Palaiseau.

Contacts :

Philippe CHARPENTIER (Ingénieur Chercheur Expert), mail : philippe.charpentier@edf.fr

Yanis REMMACHE (Ingénieur Chercheur), mail : yanis.remmache@edf.fr

Merci d'envoyer un C.V et une lettre de motivation sur ces e-mails.

Horaires : 35 h / semaine

Indemnité : en fonction des formations