

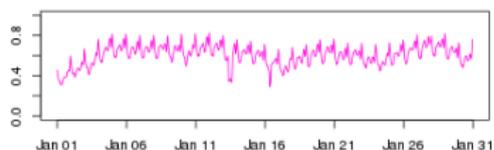
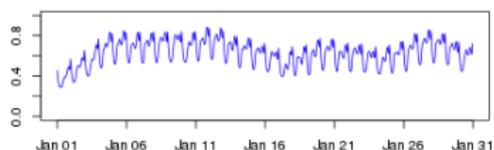
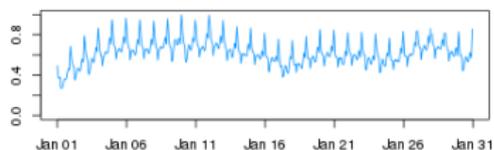
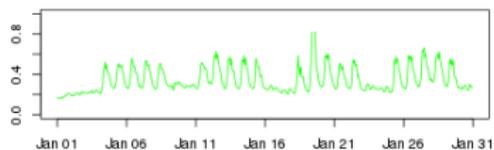
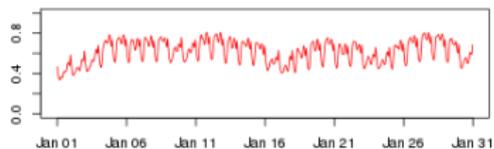
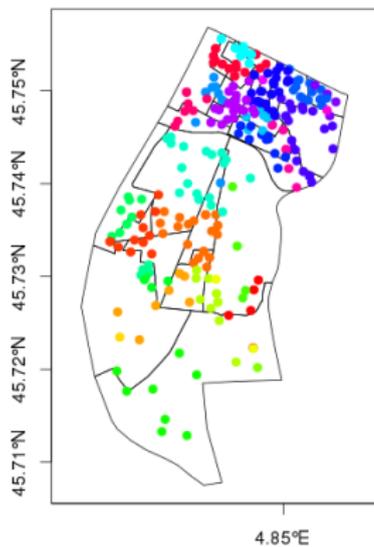
Estimation spatio-temporelle de consommation électrique

Jiali Mei Yannig Goude Georges Hébrail

Département Innovation Commerciale, Analyse des Marchés et de leur Environnement
Département Optimisation, Simulation, Risque et Statistiques pour les Marchés de l'Energie
EDF R&D

2 Avril 2015

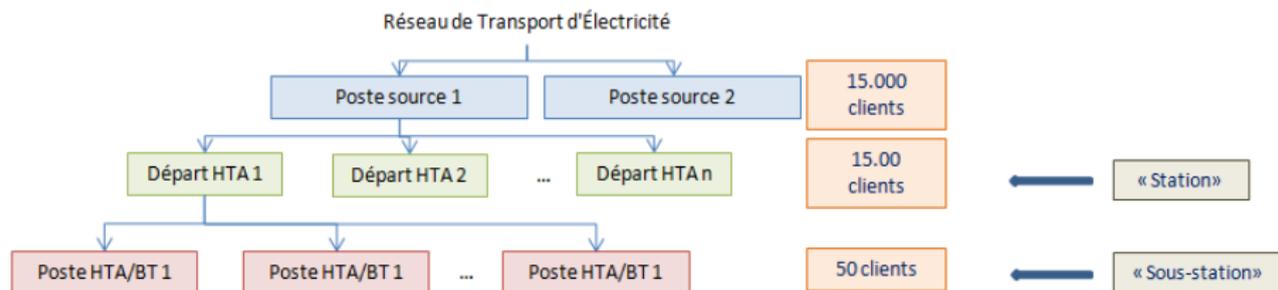
Problématique



Pourquoi a-t-on besoin de connaître la consommation d'électricité localement ?

- ▶ Maîtriser l'équilibre d'offre et demande de l'électricité, dans un contexte de production décentralisée ;
- ▶ Croiser des données de différentes sources (télécommunication, eau, transport) pour modéliser l'activité économique.

Problème : obtenir les consommations d'électricité d'une maille géographique et temporelle fine.



- ▶ Consommation électrique horaires pour plus de 600 départs HTA (“station”) des années 2010, 2011, et 2012
- ▶ Structure du réseau et sa localisation
- ▶ Socio-démographie : recensement de l’année 2010, la base SIRENE, données carroyées. Plus de 1000 IRIS
- ▶ Température et variables calendaires

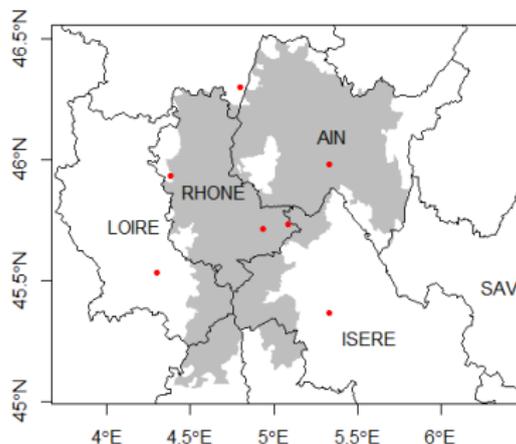


FIGURE: Zone desservie par la partie de réseaux étudiée

Motivation

Méthodes d'estimation

- Méthode d'interpolation

- Modèle de consommation locale

Évaluation

- Évaluation dans un contexte de prévision

- Jeu de données simulées

Conclusion

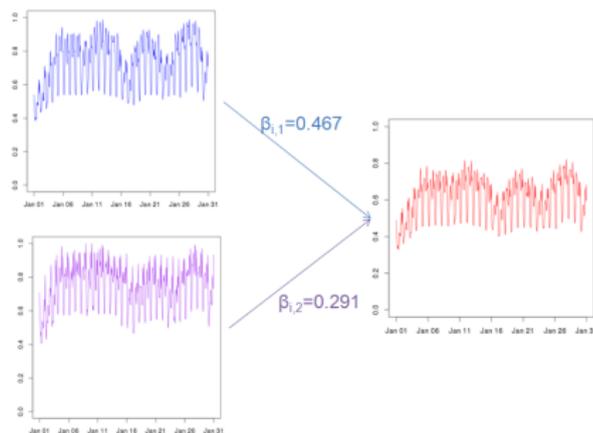
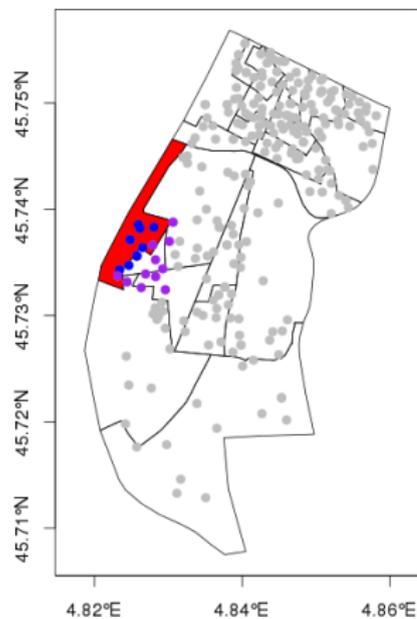
Interpolation de la consommation basée sur la population

Interpolation au prorata de la population

$$P_i = \begin{pmatrix} P_{i,1} \\ P_{i,2} \end{pmatrix} = \begin{pmatrix} 1160 \\ 2321 \end{pmatrix}$$

$$\begin{pmatrix} P_{.1} \\ P_{.2} \end{pmatrix} = \begin{pmatrix} 3973 \\ 4973 \end{pmatrix}$$

$$\begin{pmatrix} \beta_{i1} \\ \beta_{i2} \end{pmatrix} = \begin{pmatrix} 1160/3973 \\ 2321/4973 \end{pmatrix}$$



Avantages :

- ▶ Simple et toujours physiquement possible (pas de valeur négative, on retrouve la somme de la consommation).
- ▶ Permet d'estimer la forme et le niveau de la consommation au même temps.

Inconvénients :

- ▶ Les poids ne dépendent que de la population. Difficile à généraliser avec plusieurs variables démographiques.
- ▶ Pas de prise en compte de l'aspect série temporelle de la consommation.

Motivation

Méthodes d'estimation

Méthode d'interpolation

Modèle de consommation locale

Évaluation

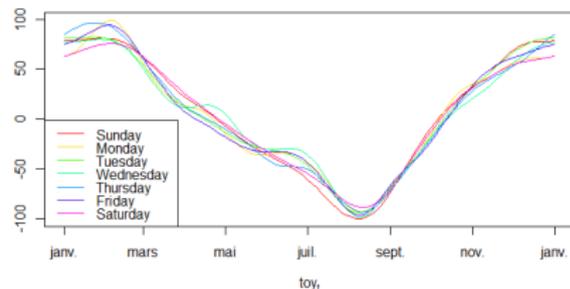
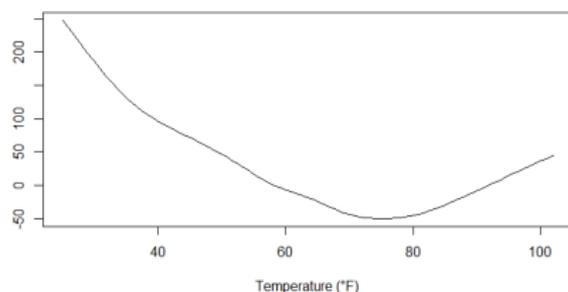
Évaluation dans un contexte de prévision

Jeu de données simulées

Conclusion

Modèle de prévision de consommation

Quand les données sont disponibles, l'aspect temporel de consommation peut être pris en compte par un modèle de régression, en prenant compte de la température et la saisonnalité journalier, hebdomadaire et annuel :



On spécifie des modèles additifs généralisés (GAM)

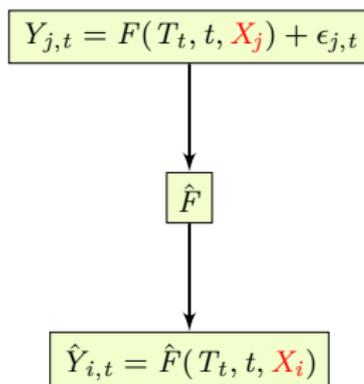
$$Y_t = g(T_t) + \sum_{d=1}^7 h_d(\text{time of year}_t) \mathbf{1}_{\text{day type}_t=d}.$$

Chaque heure de la journée est modélisée indépendamment.

Les modèles de prévision proposent une estimation à échelle temporelle fine pour chaque série.

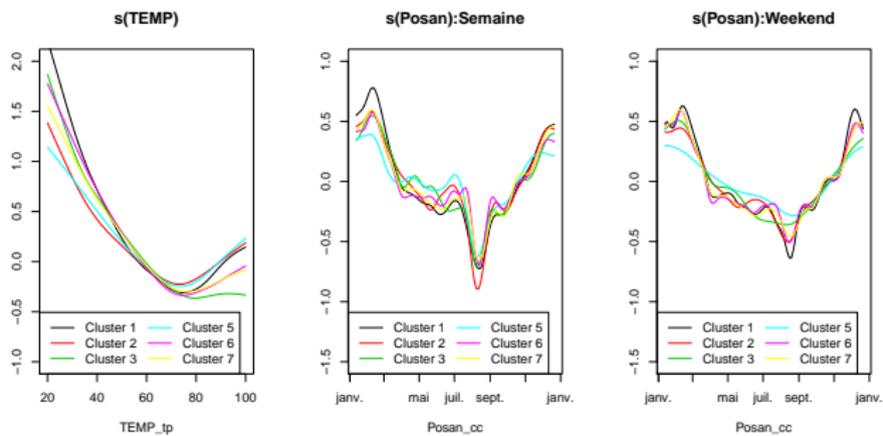
Le stratégie est de passer par des modèles applicables pour les IRIS sur les données de consommation des stations.

- ▶ But : intégrer les variables socio-démographiques dans le modèle de consommation.
- ▶ Nous proposons une approche de clustering sur ces variables.



Modèle de consommation locale

1. Interpoler les variables socio-démographiques sur les stations.
2. Clustering des IRIS sur les variables socio-démographiques, application sur les stations. Choix possible sur les variables à utiliser.
3. Estimer un modèle de consommation (GAM) par cluster sur les données des stations.
4. Application des modèles sur les IRIS.



Motivation

Méthodes d'estimation

Méthode d'interpolation

Modèle de consommation locale

Évaluation

Évaluation dans un contexte de prévision

Jeu de données simulées

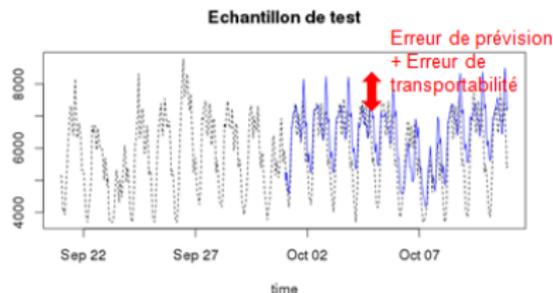
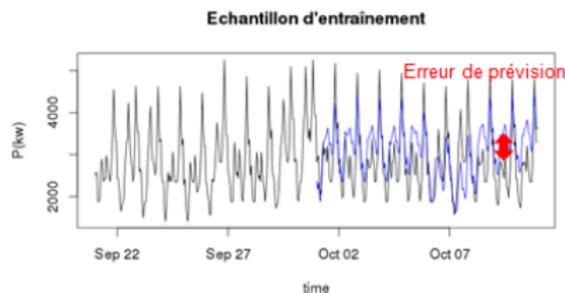
Conclusion

Validation du modèle par prévision

Les données ne sont pas disponibles pour les IRIS.

⇒ Tester si les modèles se transportent en prévision pour les stations gardées en test.

- ▶ Estimer les modèles sur le début de série temporelle des stations d'entraînement.
- ▶ Prédiction sur les 3 derniers mois pour toutes les stations.
- ▶ On distingue deux types d'erreur :
 1. erreur de prédiction
 2. erreur de transportabilité
- ▶ Sur les stations d'entraînement : erreur de prédiction
- ▶ Sur les stations de test : erreur de prédiction + erreur de transportabilité.
Sous l'hypothèse du modèle, l'erreur de transportabilité évalue la transportabilité du modèle sur les IRIS.



Résultats sur données réelles

On évalue les prédictions par l'erreur moyenne en pourcentage (MAPE) :

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \times 100\%.$$

			ACP	15 variables	Carroyé	Type client
Observed level (training data)		(training data)	18.90	18.83	18.70	17.59
Observed level (test data)			19.38	19.05	19.19	18.28

TABLE: MAPE du modèle avec différents groupes de variables en clustering : composantes principales des variables de recensement, quinze variables choisies manuellement, données carroyées, pourcentage des différents types de client.

- ▶ Une détérioration par rapport au taux d'erreur moyenne des modèles spécifiques à chaque station, 11.7%.
- ▶ Les différents groupes de variables de clustering ont des résultats similaires.
- ▶ L'information de type de client a un apport.

Motivation

Méthodes d'estimation

Méthode d'interpolation

Modèle de consommation locale

Évaluation

Évaluation dans un contexte de prévision

Jeu de données simulées

Conclusion

Simulation réaliste de la consommation

Pour compléter l'évaluation indirecte par la prévision, on propose une simulation au niveau des sous-stations.

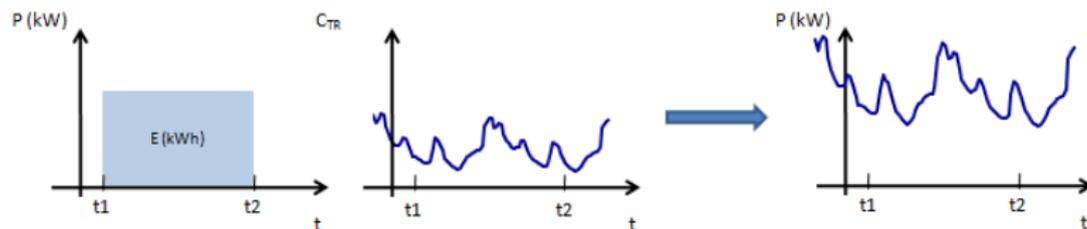
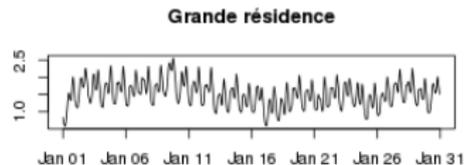
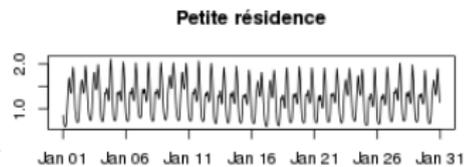
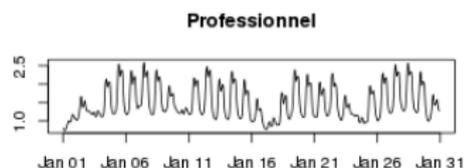
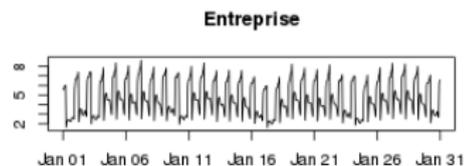
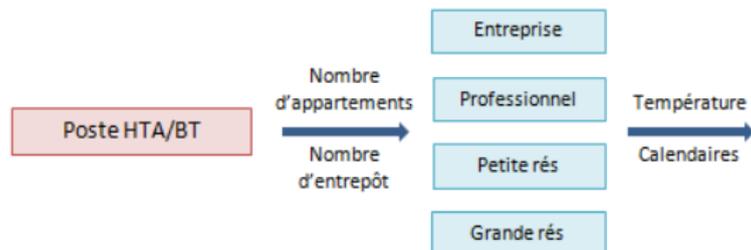


FIGURE: Méthode de profilage

- ▶ ERDF utilise des profils pour reconstruire la courbe de consommation pour les sites non munis de compteurs télé-relevés.
- ▶ Ces profils sont disponibles pour chaque catégorie de clients.
- ▶ On peut utiliser ces profils pour simuler la consommation.

Simulation réaliste de la consommation

1. Allouer une étiquette de type de client à chaque sous-station par les variables démographiques.
2. Simuler la courbe de consommation, utilisant la température et les variables calendaires.
3. Agréger la consommation sur les IRIS et les stations.

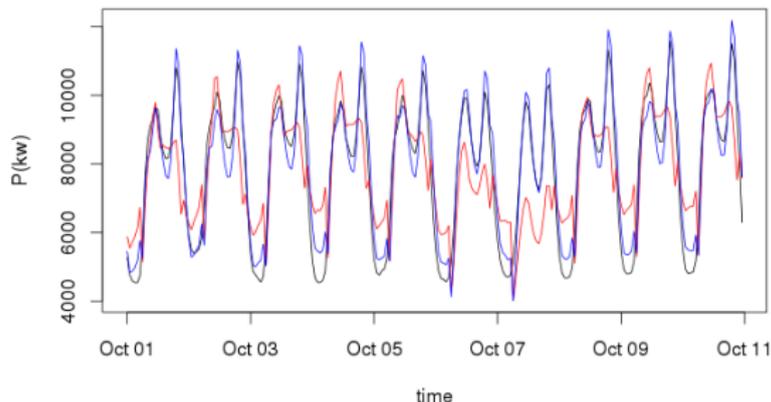


Evaluation de modèle sur les simulations

La même procédure d'estimation est appliquée sur la consommation simulée des stations. On évalue à la fois sur les stations et sur les IRIS.

	Appartement & entrepôt	Type client	Interpolation population
Départs HTA d'entraînement	13.20	11.22	0
Départs HTA de test	13.87	11.45	0
IRIS	16.93	14.09	49.57

Prédiction



Conclusion

On a traité un problème de “changement de support” dans cet exposé, en proposant

- ▶ deux méthodes d'estimation de la consommation IRIS à échelle temporelle fine ;
- ▶ une validation des méthodes par prévision ;
- ▶ simulation réaliste de la consommation.



Gotway, C. A. and Young, L. J. (2002).

Combining incompatible spatial data.

Journal of the American Statistical Association, 97(458) :632–648.



Goude, Y., Nedellec, R., and Kong, N. (2013).

Local short and middle term electricity load forecasting with semi-parametric additive models.

IEEE transactions on smart grid, 5 , Issue : 1 :440 – 446.



Hastie, T. and Tibshirani, R. (1986).

Generalized additive models.

Statistical science, pages 297–310.