

CLASSIFICATION DE COURBES INDIVIDUELLES ET PRÉVISION DÉSAGRÉGÉE DE LA CONSOMMATION ÉLECTRIQUE

Jairo Cugliari ¹ & Yannig Goude ² & Jean-Michel Poggi ³

¹ *Université Lumière Lyon 2, 5 av. Pierre Mendès-France, Bât. K, 69676, Bron, France, Jairo.Cugliari@univ-lyon2.fr*

² *EDF R&D, 1 avenue du Général de Gaulle, 92141 Clamart Cedex, France, Yannig.Goude@edf.fr*

³ *Université Paris Descartes & Université Paris Sud, Bât. 425, 91405 Orsay Cedex, France, Jean-Michel.Poggi@math.u-psud.fr*

Résumé. La disponibilité de données électriques, à des résolutions temporelles et géographiques plus fines met l'industrie face à de nouveaux défis et opportunités. Nous nous intéressons à la prévision non paramétrique d'une série par désagrégation et le profilage des consommateurs individuels à des fins commerciales. On se propose de construire des outils de classification utiles pour les deux tâches en général considérées séparément. L'idée est de décomposer le signal global de telle sorte que la somme des prévisions désagrégées améliore considérablement la prédiction du signal global. La stratégie est en trois étapes : à partir d'une première classification de courbes par partitionnement on définit des super-consommateurs, on construit ensuite une hiérarchie de partitions à l'intérieur de laquelle on sélectionne la meilleure partition pour de la prévision désagrégée.

Mots-clés. Classification, désagrégation, prévision

Abstract. New challenges and opportunities facing industry the availability of electrical data at multiple time and spatial resolutions. We are interested in nonparametric forecasting of a time series by disaggregation and profiling of individual customers for business. It is proposed to build classification tools useful for the two tasks simultaneously considered generally in a separate way. The idea is to disaggregate the global signal in such a way that the sum of disaggregated forecasts significantly improves the prediction of the whole global signal. The strategy is in three steps: from a first clustering of curves defining super-consumers, then we built a hierarchy of partitions within which the best one is selected with respect to a disaggregated forecast criterion.

Keywords. Clustering, disaggregation, forecasting

1 Introduction

La disponibilité de données électriques, à des résolutions temporelles et géographiques plus fines met l'industrie face à de nouveaux défis et opportunités. Nous nous intéressons

à deux questions : la prévision d’une série agrégée et le profilage des consommateurs individuels à des fins commerciales. Ces deux tâches sont en général considérées séparément la première étant plutôt liée à la prévision de la production d’électricité et la seconde à la commercialisation de l’électricité. Nous nous proposons de construire des outils de classification utiles pour les deux. En effet des premiers travaux optimisaient la classification à l’objectif de prévision alors que nous présentons une hiérarchie de partitions compatible avec la multiplicité des objectifs, c’est seulement la sélection d’une partition dans cette hiérarchie qui est effectuée sur des bases liées à l’objectif de prévision. On distinguera donc deux aspects : la classification non supervisée de courbes en vue de la prévision et l’adaptation des modèles de prévision à ces classes.

Nous disposons d’un historique de courbes de charge électrique individuelles pour un segment de clients du producteur EDF. A l’instant t la demande agrégée (ou *courbe synchrone*), $S(t)$, est simplement la somme $S(t) = \sum_i X_i(t)$ de toutes ces consommations individuelles $X_i(t)$. Nous avons alors un premier predicteur de la demande agrégée construit à partir du passé de S : $\hat{S}_{\text{agr}}(t) = \hat{S}(t)$. Pour une partition de clients en K groupes $g = 1, \dots, K$, nous considérons la demande du groupe g , i.e. $S_g(t) = \sum_{i \in g} X_i(t)$. De manière alternative $\hat{S}_{\text{agr}}(t)$ nous pouvons calculer la prévision désagrégée à partir de la prévision de chaque groupe $\hat{S}_{\text{des}}(t) = \sum_{i \in g} \hat{S}_g(t)$.

Zhou *et al.* (2013) présentent un panorama des techniques de classification pour les données électriques. Peu de références optimisent la classification des clients en vue de la prévision, mentionnons Misiti *et al.* (2010) et Alzate et Sinn (2013). Les références dédiées à la classification en vue du profilage de clients sont de loin les plus nombreuses. Dans ce travail nous cherchons à améliorer la qualité de la prévision de la courbe de charge agrégée à partir des données individuelles.

2 Les données

Nous considérons des données de consommation d’un segment de 25011 clients professionnels, enregistrées au pas demi-horaire sur les années 2009 et 2010, et une partie de 2011. Une année complète de données compte environ 438 millions d’enregistrements et occupe 3.25 Go d’espace de disque. La Figure 1 montre pour l’année 2010, une courbe de consommation pour un client individuel à droite et la courbe synchrone à gauche, après filtrage des phénomènes infra-journaliers.

Après plusieurs prétraitements (détection d’incohérences, de valeurs aberrantes, de valeurs manquantes) sur les fichiers d’origine afin de construire trois matrices de données complètes et après filtrage des valeurs aberrantes, nous conservons 24500 clients.

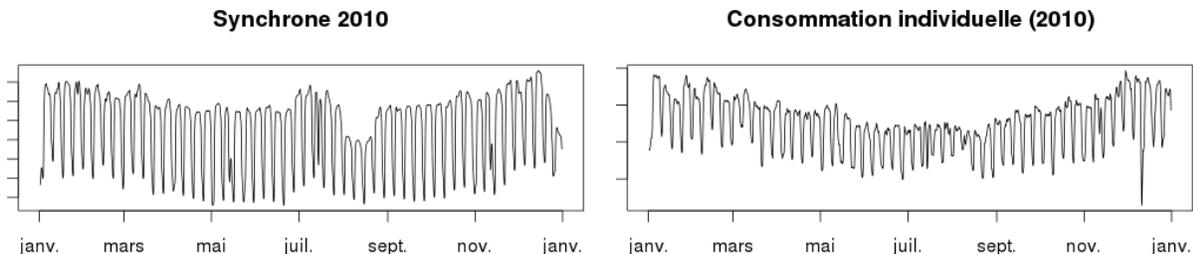


Figure 1: Courbe synchrone (à gauche) et courbe individuelle (à droite) pour l’année 2010 filtrées pour éliminer les phénomènes infra-journaliers

3 Les méthodes

Représentation des courbes. Chaque client est décrit par sa demande en fonction du temps. Pour prendre en compte la nature fonctionnelle des données à classer, nous les décomposons en ondelettes et les résumons par les contributions énergétiques relatives des échelles (voir Antoniadis *et al.* (2013)). Grâce au caractère dyadique de la transformée en ondelettes discrète nous réduisons ainsi le nombre de descripteurs à 15 ($\sim \log_2(17520)$). Nous disposons alors à la fin de cette étape d’une matrice de descripteurs de taille 24500×15 que nous normalisons par colonnes en utilisant la transformée proposée par Steinley & Brusco (2006). Un premier indice dit de *clusterabilité* est calculé pour déterminer quelles sont les variables qui portent de l’information pour révéler la structure de classes. Cette procédure présente plusieurs avantages : la dimension du problème est encore une fois réduite, les descripteurs jusqu’à maintenant indépendants du problème sont désormais adaptés, et elle est enfin porteuse de connaissance car elle rend la classification interprétable.

Une première classification. Étant donné le grand nombre d’individus à classer et la contrainte informatique que nous nous imposons, en vue d’un passage à l’échelle, nous utilisons l’algorithme CLARA (Clustering LARge Applications) de Kaufman and Rousseeuw (1990), de classification séquentielle qui travaille sur des sous-échantillons tirés de manière aléatoire dans l’ensemble des données. Plus précisément, pour le premier tirage, on sélectionne (selon un tirage simple sans remise) m individus qui sont traités par un algorithme PAM (*Partitioning around Medoids*, une version robuste des k -means) pour obtenir K centres, encore appelés super-consommateurs, avec K très supérieur au nombre de classes final plausible. Ensuite, $m - K$ individus (différents des m précédents) sont choisis et combinés aux K centres dans une nouvelle classification PAM. Ainsi, l’algorithme est répété un nombre de fois défini par l’utilisateur, conservant à chaque étape les centres de l’étape précédente. Les centres de la dernière itération sont les centres finaux.

Dans les expériences réalisées, nous avons choisi d’itérer la procédure 4 fois, $m = 4000$ observations, et un nombre de super-consommateurs égal à $K = 200$.

Une hiérarchie de partitions. Nous utilisons une classification ascendante hiérarchique (CAH) pour révéler la structure de classes des super-consommateurs. La CAH est définie par une matrice de distances, ou de dissimilarités, et une fonction de lien. Nous utilisons deux notions de distance différentes proposées dans Antoniadis *et al.* (2013). La première est la distance euclidienne calculée sur les descripteurs de l'espace de représentation des trajectoires. La deuxième distance est basée sur la notion de cohérence en ondelettes (on appelle cette distance WER pour son acronyme en anglais: *Wavelet Extended R*). Dans les deux cas, l'algorithme CAH utilisera la matrice des distances (de taille 200×200) avec la fonction de lien de Ward.

Modèles de prévision. Antoniadis *et al.* (2013, 2014) proposent un modèle de prévision (KWF) pour des séries chronologiques fonctionnelles en présence de non stationnarités. Le principe général consiste à rechercher dans le passé des contextes similaires à la situation présente et prévoir le futur par une combinaison linéaire des futurs des passés les plus semblables au présent. La notion de similarité est basée sur les ondelettes et plusieurs stratégies sont mises en oeuvre pour prendre en compte les diverses sources d'instationnarités.

En complément, on pourra s'intéresser aux techniques de prévision de la consommation récemment développées par EDF. Dans Pierrot et Goude (2011) on trouve les principes de la méthode et l'application aux données nationales EDF et dans Goude et al. (2013) l'application aux données des postes sources.

4 Résultats

Le premier résultat à commenter concerne les descripteurs sélectionnés lors de l'étape de choix des variables significatives. L'algorithme de Steinley & Brusco (2006) élimine de l'analyse les échelles liées aux fréquences infra-journalières. On conserve alors les échelles d'ondelettes 6, 7, 8, 9 et 13 qui représentent des plages de 16 heures, 1.5, 3 et 6 jours, et 3 mois. A ce stade, la dimension de chaque observation est passée de 17520 demi-heures à 5 descripteurs normalisés.

Avec l'ensemble de descripteurs retenus, nous obtenons une classification non supervisée des observations avec CLARA. Nous fixons un nombre de classes égal à 200 qui vont représenter les super-consommateurs.

Nous construisons alors la consommation des super-consommateurs en agrégeant les trajectoires observées des consommateurs individuels, donc nous revenons au 17520 points de consommation mais cette fois-ci, le nombre des courbes est de 200.

Nous suivons ensuite suivre deux voies. D'une part nous allons utiliser les 200 super-consommateurs pour déterminer une hiérarchie de partitions, ainsi nous n'aurons pas à déterminer un nombre de classes vraisemblable mais à étudier la structure globale des classes. D'autre part, nous allons faire le lien entre la construction des classes et l'objectif

de prévision.

Nous commenterons dans l'exposé uniquement les résultats concernant la distance basée dans la notion de cohérence en ondelettes au moyen de la transformée continue en se concentrant sur la hiérarchie de partitions issue de l'algorithme CAH.

Parallèlement à la construction d'une hiérarchie nous cherchons maintenant à évaluer l'impact de la désagrégation de la synchrone sur la qualité de la prévision. Pour ce faire, nous construisons les classes de consommateurs à l'aide de la classification en deux étapes précédente sur les données de l'année 2009. Ensuite, nous mesurons la qualité de la prévision avec le modèle KWF au pas journalier sur toute l'année 2010. Deux stratégies pour la prévision de la synchrone sont mises en compétition, la première est la prévision de la synchrone à partir des données passées de la synchrone même, $\hat{S}_{agr}(t)$, et pour la seconde, nous obtenons des prévisions $\hat{X}_g(t)$ pour chaque classe issue de la classification non supervisée, puis nous agrégeons les prévisions $\hat{S}_{dis} = \sum_g \hat{X}_g(t)$.

Le résultat principal que nous obtenons fait l'objet de la Figure 2. Pour l'année 2010, la méthode KWF donne un MAPE (Mean Absolute Percent Error) d'environ 2.1% qui par construction ne dépend pas du nombre de classes. L'approche concurrente qui consiste à obtenir des classes de clients, de les prévoir et d'agréger la somme pour anticiper la synchrone est systématiquement plus performante (MAPE moins élevée), et cela vaut pour les deux notions de distance utilisées pour les classes : celle liée à la distance euclidienne sur les descripteurs ou à la distance WER.

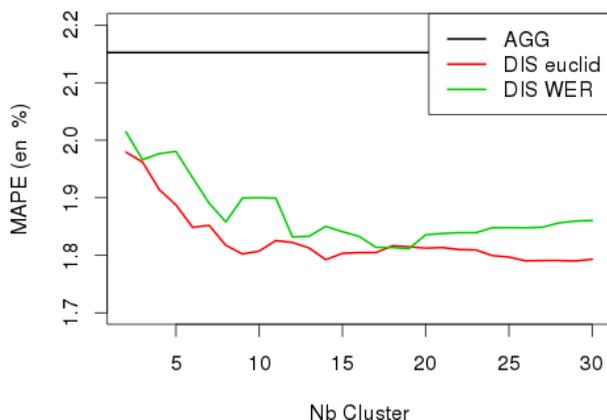


Figure 2: MAPE moyen annuel par nombre de classes selon le classification avec la distance euclidienne sur les descripteurs et avec la distance WER.

Le gain le plus faible (environ 8%) que l'on obtient est observé avec deux classes. A l'autre extrême on observe des gains allant jusqu'à 18% pour les partitions entre 15 et 20 classes de super-consommateurs.

Les gains sont considérables. Nous pouvons alors essayer de comprendre leur origine. Pour ce faire, nous représentons dans la Figure 3 des trajectoires de consommation obtenues par agrégation des consommateurs en deux et trois classes.

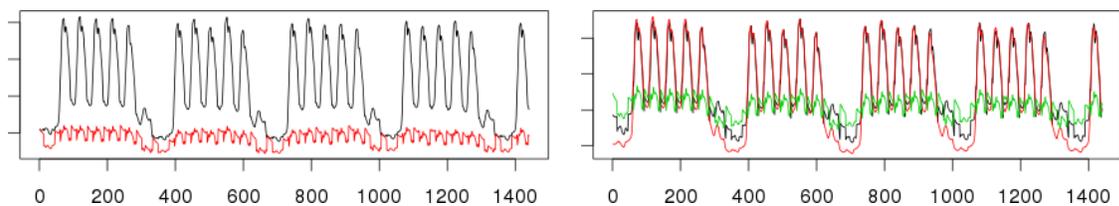


Figure 3: Agrégats de consommateurs individuels en 2 et 3 classes selon la distance RC.

Nous pouvons alors observer que les trajectoires obtenues conservent des caractéristiques de forme qui sont à la fois différenciées par classes et régulières à l'intérieur de chacune des classes. La méthode de prévision KWF s'appuyant sur la forme des courbes, elle est alors capable d'exploiter cette similarité intra-classe pour chercher des situations de consommation du passé qui sont similaires au contexte actuel même en présence de peu de données.

Bibliographie

- [1] C. Alzate and M. Sinn (2013) Improved electricity load forecasting via kernel spectral clustering of smartmeter. In *International Conference on Data Mining*, 948:943 – 948.
- [2] A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi (2012) Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité. *Journal de la Société Française de Statistique*, 153(2) :52 – 78.
- [3] A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi (2013) Functional Clustering using Wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(1).
- [4] A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi (2014) Une approche fonctionnelle pour la prévision non-paramétrique de la consommation d'électricité. *Journal de la Société Française de Statistique*, 155(2) :202 – 219
- [5] Y. Goude, R. Nedellec, and N. Kong (2013) Local Short and Middle term Electricity Load Forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, 5(1) :440 – 446.
- [6] M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi (2010) Optimized Clusters for Disaggregated Electricity Load Forecasting. *REVSTAT Statistical Journal*, 8(2):105-124.
- [7] L. R. Kaufman and P. Rousseeuw (1990) *Finding groups in data: An introduction to cluster analysis*. Hoboken NJ John Wiley & Sons Inc.
- [8] A. Pierrot and Y. Goude (2011) Short-Term Electricity Load Forecasting with Generalized Additive Models. *Proceedings of ISAP power*, pages 593 – 600.
- [9] D. Steinley and M. J. Brusco (2008). Selection of variables in cluster analysis: An empirical comparison of eight procedures. *Psychometrika*, 73(1), 125-144.
- [10] K. Zhou, S. Yang, and C. Shen (2013) A review of electric load classification in smart grid environment. *Renewable and Sustainable Energy Reviews*, 24:103 – 110.