

ALGORITHMES DE BANDITS POUR LE PILOTAGE DE LA CONSOMMATION ÉLECTRIQUE

Margaux Brégère ¹, Pierre Gaillard ², Yannig Goude ³, Gilles Stoltz ⁴

¹*EDF R&D, 7 Boulevard Gaspard Monge, 91 120 Palaiseau, margaux.bregere@u-psud.fr*

²*Inria, 2 Rue Simone IFF, 75012 Paris, pierre.gaillard@inria.fr*

³*EDF R&D, 7 Boulevard Gaspard Monge, 91 120 Palaiseau, yannig.goude@edf.fr*

⁴*Université Paris-Sud, 91405 Orsay, gilles.stoltz@u-psud.fr*

Résumé. L'électricité se stockant difficilement, l'équilibre entre la production et la consommation doit en permanence être maintenu. Nous proposons d'appliquer la théorie des bandits contextuels pour piloter, à l'aide d'incitations tarifaires, la demande électrique. Plus précisément, une consommation moyenne cible est fixée à chaque instant et la consommation moyenne est modélisée comme une fonction des prix envoyés et de variables contextuelles (température, heure, jour etc.). La performance des stratégies est mesurée en pertes quadratiques à travers un critère de regret. Inspiré de stratégies standards pour les bandits contextuels (LinUCB - [5, 2]), notre algorithme permet de borner ce regret en $T^{2/3}$ (aux termes poly-logarithmiques près). Des simulations sur des données publiques de UK Power Networks, dans lesquels des incitations tarifaires ont été proposées, montrent que notre stratégie influence efficacement la consommation électrique des usagers.

Mots-clés. Théorie de l'apprentissage statistique, Modèles semi- et non-paramétriques.

Abstract. We propose a contextual-bandit approach for demand side management by offering price incentives. More precisely, a target mean consumption is set at each round and the mean consumption is modeled as a complex function of the distribution of prices sent and of some contextual variables (temperature, hour, day etc.). The performance of our strategies is measured in quadratic losses through a regret criterion. We offer $T^{2/3}$ upper bounds on this regret (up to poly-logarithmic terms), for strategie inspired by standard strategies for contextual bandits (LinUCB - [5, 2]). Simulations on data gathered by UK Power Networks, in which price incentives were offered, show that our strategie is effective and may indeed manage demand response by suitably picking the price levels.

Keywords. Statistical Learning Theory, Semi and Non-Parametric Models.

1 Introduction

L'équilibre entre la consommation et la production d'énergie doit être continuellement maintenu. Actuellement, EDF prévoit la consommation de ses clients et actionne en conséquence ses différents moyens de production. Avec le développement des énergies renouvelables et intermittentes sujettes aux changements météorologiques, ajuster la production pour répondre au mieux à la demande deviendra de plus en plus complexe. Le

déploiement de nouveaux compteurs capables de collecter les données de consommation quasi-instantanément mais aussi d’envoyer des signaux aux clients permet d’envisager le pilotage de charge. L’enjeu est de choisir dynamiquement les bons signaux à envoyer aux clients afin d’influencer leur consommation et que cette dernière s’ajuste à une consommation cible. Les algorithmes devront en parallèle apprendre la réaction des clients aux différents signaux tout en optimisant l’envoi de ces derniers. Nous proposons d’adapter des algorithmes de bandit contextuel. Notre contribution consiste à modéliser ce système de gestion (Section 2). Les principaux résultats sont énoncés en Section 3 et des simulations effectuées sur des données de consommation sont présentées en Section 4.

2 Modélisation

Nous considérons une population homogène de clients dont les consommations à chaque instant t dépendent de variables exogènes (température, jour de la semaine, heure etc.) qui seront rassemblées dans un vecteur $x_t \in \mathcal{X}$ ainsi que du tarif appliqué j . La consommation individuelle est modélisée comme suit : $\psi(x_t, j)$ + bruit blanc. Nous supposons que le fournisseur d’électricité, disposant de $K \geq 2$ tarifs peut partitionner la population et envoyer le tarif j à une proportion $p_{t,j}$ de clients. Nous noterons pas la suite $p_t = (p_{t,1}, \dots, p_{t,K})^T$ le vecteur des proportions. Dès lors, la consommation moyenne de ses clients est

$$Y_{t,p_t} = \sum_{j=1}^K p_{t,j} \psi(x_t, j) + \text{bruit.}$$

Les modèles additifs généralisés (se référer au monographe [7] pour plus de détails) sont largement utilisés pour la modélisation de la consommation électrique (voir [4, 3]) qui est vue comme la somme des effets des variables explicatives. Ces effets sont généralement modélisés par des splines (fonctions \mathcal{C}^2 polynomiales par morceaux) pour les variables continues. Quand aux variables discrètes, un effet par modalité est considéré. Dans les simulations de la Section 4, les variables contextuelles considérées à un instant t sont la température τ_t , la position dans l’année y_t , l’instant de la journée h_t , le jour de la semaine w_t . De fait, pour un tarif j le modèle consommation adopté sera de la forme :

$$\psi(x_t, j) = \sum_{h=1}^{48} \left[s_h^\tau(\tau_t) + s_h^y(y_t) + \eta_h \right] \mathbf{1}_{h_t=h} + \sum_{w=1}^7 \zeta \mathbf{1}_{w=w_t} + \xi_j, \quad (1)$$

où les fonctions $(s_h^\tau)_h$ et $(s_h^y)_h$ sont des splines. Pour chaque spline, choisir le nombre de nœuds (points de rupture des polynômes) ainsi que leur position permet de déterminer une base linéaire dans laquelle la spline s’exprime. Il existe alors un paramètre inconnu θ de dimension d et une fonction de transfert connue ϕ tels que pour tout j , $\psi(x_t, j) = \phi(x_t, j)^T \theta$ (se référer à l’article [1] pour plus de détails). En étendant ϕ linéairement à sa seconde

composante, à un vecteur de prix p , la consommation à un instant t est

$$Y_{t,p} = \phi(x_t, p) + p^T \varepsilon_t, \quad (2)$$

où les vecteurs $\varepsilon_1, \varepsilon_2, \dots$ sont supposés ρ -sous gaussiens, i.i.d avec $\mathbb{E}[\varepsilon_1] = (0, \dots, 0)^T$. On notera $\Gamma = \text{Var}(\varepsilon_1)$ la matrice de covariance.

La modélisation proposée du pilotage de charge est résumée dans le protocole 1. À un instant t , le fournisseur d'électricité reçoit une consommation moyenne cible c_t . Il choisit alors des niveaux des prix $p_t \in \mathcal{P}$. L'ensemble \mathcal{P} est inclus dans le simplexe de dimension K et pourra être en pratique réduit (au gré de contraintes métier). Après avoir choisi une répartition des niveaux de prix p_t , il n'observe que Y_{t,p_t} (supposée bornée entre 0 et C , où C est connue) et fait donc face à un problème de bandits contextuels (en raison des x_t).

Protocole 1 Bandits contextuels pour le suivi de cible

- 1: **Entrées** : Borne sur les consommations C , Fonction de transfert $\phi : \mathcal{X} \times \mathcal{P} \rightarrow \mathbb{R}^d$
 - 2: **Inconnus** : Paramètre de transfert $\theta \in \mathbb{R}^d$, Matrice de covariance $\Gamma \in \mathcal{M}_K(\mathbb{R})$
 - 3: **Pour** $t = 1, 2, \dots$ **faire**
 - 4: Observer un contexte $x_t \in \mathcal{X}$ et une cible $c_t \in (0, C)$
 - 5: Choisir les proportions de prix $p_t \in \mathcal{P}$
 - 6: Observer une consommation résultante de ces choix $Y_{t,p_t} = \phi(x_t, p_t)^T \theta + p_t^T \varepsilon_t$
 - 7: Subir une perte $(Y_{t,p_t} - c_t)^2$
 - 8: **Fin**
 - 9: **Objectif** :
 - 10: Minimiser la perte cumulée $L_T = \sum_{t=1}^T (Y_{t,p_t} - c_t)^2$
-

3 Algorithme et borne sur le regret

Notons tout d'abord que la perte instantanée pour p fixé vérifie

$$\ell_{t,p} \stackrel{\text{def}}{=} \mathbb{E}[(Y_{t,p} - c_t)^2] = (\phi(x_t, p)^T \theta - c_t)^2 + p^T \Gamma p.$$

Dès lors, nous introduisons le regret \bar{R}_T , que nous chercherons à borner :

$$\bar{R}_T = \sum_{t=1}^T (\phi(x_t, p_t)^T \theta - c_t)^2 + p_t^T \Gamma p_t - \sum_{t=1}^T \min_{p \in \mathcal{P}} \left\{ (\phi(x_t, p)^T \theta - c_t)^2 + p^T \Gamma p \right\}.$$

Afin de borner ce regret, les hypothèses suivantes sont faites sur les paramètres : pour tout vecteur de contexte x_t et pour tout vecteur $p \in \mathcal{P}$,

$$\|\phi(x_t, p)\|_\infty \leq 1, \quad \|\theta\|_\infty \leq C, \quad \phi^T(x_t, p)\theta \in (0, C) \quad \text{et} \quad p^T \Gamma p \leq G.$$

L'algorithme proposé s'inspire largement de LinUCB (voir [5, 2]). Pour estimer les pertes instantanées, il faut avant tout estimer Γ et θ . Ce dernier est estimé par régression λ -Ridge :

$$\hat{\theta}_t \stackrel{\text{def}}{=} V_t^{-1} \sum_{s=1}^t Y_{s,p_s} \phi(x_t, p_x), \quad \text{où} \quad V_t \stackrel{\text{def}}{=} \lambda I_d + \sum_{s=1}^t \phi(x_s, p_s) \phi(x_s, p_s)^T.$$

En fixant un risque $\delta \in (0, 1)$, pour $t \geq 1$, peu importe le choix des $(p_s)_{1 \leq s \leq t}$, et avec probabilité au point $1 - \delta$, l'estimateur $\hat{\theta}_t$ vérifie (Lemme 1 de l'article [1])

$$\left\| V_t^{1/2} (\hat{\theta}_t - \theta) \right\| \leq B_t(\delta), \quad \text{où} \quad B_t(\delta) \stackrel{\text{def}}{=} \sqrt{\lambda d C} + \rho \sqrt{2 \ln \frac{1}{\delta} + d \ln \left(1 + \frac{t}{\lambda}\right)}.$$

La matrice de covariance Γ est estimée sur les n premières itérations (se référer à l'article [1] pour de plus amples détails) et avec probabilité $1 - \delta/2$, l'estimateur $\hat{\Gamma}_n$ vérifie $\sup_{p \in \mathcal{P}} |p^T (\hat{\Gamma}_n - \Gamma) p| \leq \gamma$. Pour chaque $p \in \mathcal{P}$, nous définissons alors un estimateur de la perte instantanée

$$\hat{\ell}_{t,p} \stackrel{\text{def}}{=} \left([\phi(x_t, p)^T \hat{\theta}_{t-1}]_C - c_t \right)^2 + p^T \hat{\Gamma}_n p,$$

ainsi qu'un terme de déviation/exploration $\alpha_{t,p}$:

$$\alpha_{t,p} \stackrel{\text{def}}{=} \gamma + \min \left\{ (C^2 + G), 2C B_{t-1}(\delta t^{-2}) \|V_{t-1}^{-1/2} \phi(x_t, p)\| \right\}.$$

L'algorithme optimiste, assurant un regret en $\mathcal{O}(T^{2/3})$ (Théorème 1 démontré dans [1]) choisit alors, pour $t \geq n + 1$,

$$p_t \in \arg \min_{p \in \mathcal{P}} \{ \hat{\ell}_{t,p} - \alpha_{t,p} \} \tag{3}$$

Théorème 1. *En fixant un niveau de risque $\delta \in (0, 1)$ et $T \geq 1$. Sous les hypothèses de bornitude définies ci-dessus et avec une phase d'exploration de $n = \mathcal{O}(T^{2/3})$ itérations, l'algorithme (3) assure, avec probabilité au moins $1 - \delta$,*

$$\bar{R}_T = \mathcal{O} \left(T^{2/3} \ln^2 \left(\frac{T}{\delta} \right) \sqrt{\ln \frac{1}{\delta}} \right).$$

4 Simulations

Les simulations s'appuient sur le jeu de données “*SmartMeter Energy Consumption Data in London Households*”¹ publiée par *UK Power Networks*. Elles regroupent la consommation sur l'année 2013, par demi-heure (en kWh par demi-heure), d'un millier de foyers

1. <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>

londoniens soumis à des variations de prix de l'électricité. Un tarif unique (*Low*, *Normal* ou *High*), communiqué un jour à l'avance, était envoyé à l'ensemble de la population à chaque demi-heure (se référer au rapport [6] pour plus de détails sur cette expérimentation). Les températures disponibles en ligne (NOAA²) ont été ajoutées de sorte que le modèle (1) soit considéré. Pour tester l'algorithme, un simulateur de données a été créé et permet, pour un contexte x_t et un vecteur de prix p_t de générer une consommation réaliste Y_{t,p_t} selon le modèle (2) (se référer à l'article [1] pour description complète des simulations). Les consommations cibles considérées sont atteignables, c'est-à-dire comprises entre la consommation moyenne associée au tarif le plus haut et celle associée au tarif le plus bas. L'ensemble des proportions admissibles \mathcal{P} est restreint à une grille, et seulement deux tarifs peuvent être envoyés en même temps : *Normal* et *Low* ou *High*. Nous supposons aussi que le fournisseur dispose d'un historique d'un an de données de consommation pour lesquelles seul le tarif *Normal* a été appliqué. Il commence ensuite à explorer la possibilité d'envoyer plusieurs tarifs : c'est la période de test, d'une durée de 30 jours, de l'algorithme 3.

L'algorithme a été exécuté 200 fois et les regrets subis représentés en Figure 1 sont comparés à l'ordre de grandeur théorique de la borne. Le bas de la figure 2 représente le vecteur p_t au cours des itérations pour une exécution. Au premier jour de la période d'essai, l'algorithme explore l'effet des tarifs en envoyant le même tarif à tous les clients (les vecteurs p_t sont des masses de Dirac), tandis qu'à la fin de la période d'essai, ils exploitent intelligemment la possibilité de scinder la population en deux groupes de tarifs. En haut de la figure, une approximation de la consommation moyenne attendue (moyenne des 200 consommations observées) est tracée (ligne noire continue) et comparée à la consommation cible (ligne rouge en pointillés). L'algorithme doit choisir les tarifs permettant le meilleur compromis biais-variance (la perte attendue comporte un terme de variance). Il en résulte un estimateur légèrement biaisé de la consommation moyenne.

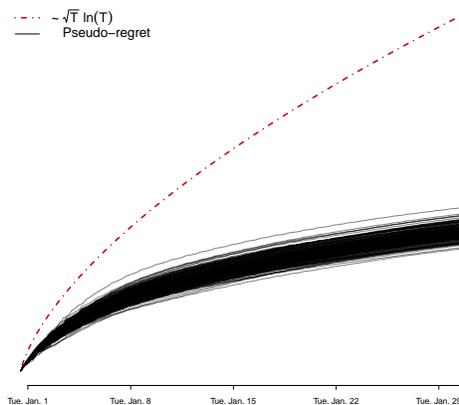


FIGURE 1 – Courbe des regrets des 200 exécutions et borne $c\sqrt{T} \ln T$ pour $c > 0$ bien choisie.

2. <https://www.noaa.gov/>

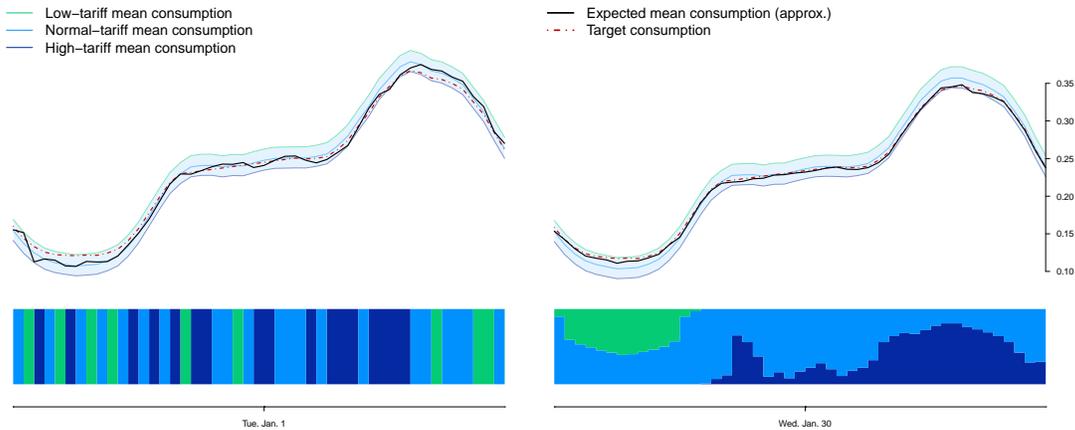


FIGURE 2 – À gauche : 1er Janvier (premier jour de la période de test). À droite : 30 Janvier (dernier jour de la période de test). En haut : 200 exécutions sont considérées. La courbe noire représentent la consommation moyennée sur les 200 exécutions, la consommation cible est en pointillé rouge et les consommations moyennes associées à chaque tarif *Low*, *Normal*, *High*, sont respectivement en vert, bleu et marine. L’enveloppe des cibles atteignables est en bleu pastel. En bas : Une seule exécution est considérée et les proportions p_t sont représentées au cours du temps.

Références

- [1] Margaux Brégère, Pierre Gaillard, Yannig Goude, and Gilles Stoltz. Target tracking for contextual bandits : Application to demand side management. 2019.
- [2] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS’11)*, pages 208–214, 2011.
- [3] Pierre Gaillard, Yannig Goude, and Raphaël Nedellec. Additive models and robust aggregation for gecom2014 probabilistic electric load and electricity price forecasting. *International Journal of forecasting*, 32(3) :1038–1050, 2016.
- [4] Yannig Goude, Raphael Nedellec, and Nicolas Kong. Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, 5(1) :440–446, 2014.
- [5] Lihong Li, Wei Chu, John Langford, and Robert Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW’10)*, pages 661–670, 2010.
- [6] J Schofield, R Carmichael, S Tindemans, M Woolf, M Bilton, and G Strbac. Residential consumer responsiveness to time-varying pricing. 2014.
- [7] Simon Wood. *Generalized Additive Models : An Introduction with R*. CRC Press, 2006.