Stochastic bandit algorithms for demand side management

Margaux Brégère

université PARIS-SACLAY

Stochastic Bandit Algorithms for Demand Side Management

Thèse de doctorat de l'université Paris-Saclay

École Doctorale n°574, Mathématique Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées
Unité de recherche : Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France.
Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue en visioconférence totale le 10 décembre 2020, par

Margaux BRÉGÈRE

Composition du Jury

Christophe GIRAUD Professeur, Université Paris-Saclay	Président
Rob J HYNDMAN Professeur, Université Monash	Rapporteur & Examinateur
Odalric-Ambrym MAILLARD Chargé de recherche, HDR, Inria Lille	Rapporteur & Examinateur
Émilie KAUFMANN Chargée de recherche, Inria Lille	Examinatrice
Nadia OUDJANE Ingénieure-chercheuse senior, HDR, EDF R&D	Examinatrice
Gilles Stoltz Directeur de recherche, CNRS/Université Paris-Saclay	Directeur de thèse
Pierre Gaillard Chercheur, Inria Grenoble-Rhône-Alpes	Co-Encadrant de thèse
Yannig Goude Ingénieur-chercheur senior, EDF R&D	Co-Encadrant de thèse

Thèse de doctorat

NNT : 2020UPASM022


eDF


MATHEMATIQUES
ORSAY


Inria
INVENTEURS DU MONDE NUMÉRIQUE

Table of Contents

Remerciements

Paralysée devant la page blanche, terrorisée par la possibilité d'un oubli et incapable de trouver les mots justes ; chers collègues, famille et amis qui lisez ces quelques lignes, pardonnez mes maladresses et la banalité de ces remerciements, ils ne sont qu'une piètre esquisse de toute la gratitude que j'aurais souhaité exprimer.

Mes premiers mots sont évidemment pour Gilles, Yannig et Pierre qui ont initié cette belle aventure. Pour le sujet passionnant que vous avez proposé, pour la confiance que vous m'avez accordée et pour votre complémentarité, merci. Arriver à jongler entre théorie et pratique tout en conciliant trois lieux, trois équipes et trois avis différents a tout d'abord suscité quelqu'appréhension. Cette dernière s'est instantanément évaporée à vos côtés, de par votre encadrement, savant mélange de liberté et de conseils avisés. Je n'aurais pu rêver d'un environnement doctoral plus stimulant et je sais que vous y êtes pour beaucoup. Gilles, bien naïve j'ai été de me croire rigoureuse ! Très vite forcée de constater que mes exigences étaient bien approximatives, je te remercie pour ton enseignement scientifique. Ta plume et la qualité de ta rédaction légendaire (tout du moins au sein de la communauté de tes doctorants) n'ont de cesse de susciter mon admiration. Note tout de même que satisfaire l'exigence qui en découle est loin d'être de tout repos ! Merci pour ton implication constante ; je n'ai eu vent de directeurs de thèse plus dévoués à leurs étudiants que tu ne l'as été : un rendez-vous par semaine, quitte à se retrouver dans un café à Massy, en terrasse place Stanislas à Nancy, ou encore sur le chemin entre Ulm et Montparnasse – pour des conversations plus ou moins scientifiques… Mais au-delà de ta rigueur mathématique, de ton gout de l'esthétisme et de ton sens de l'organisation, j'ai été particulièrement touchée par l'attention humaine que tu portes à tes doctorants. Enfin, ta considération pour la cause des femmes en sciences ainsi que ta vision des interactions entre recherche en entreprise et monde académique ont su créer une atmosphère bienveillante, inspirante et propice au travail. Yannig, merci d'avoir apporté la dimension appliquée indispensable à ma thèse. Je reste impressionnée par l'ampleur de tes connaissances aussi bien scientifiques qu'industrielles, et surtout par le recul que tu portes sur les problématiques énergétiques et mathématiques. Tes visions d'expert sur les évolutions du système électrique et de l'apprentissage statistique m'ont éclairée et ont conforté mon souhait de poursuivre dans cette voie. Je te remercie par ailleurs pour le précieux soutien moral et humain que tu m'as apporté et qui a largement facilité mes choix professionnels. Admirative de ton parcours et curieuse des plaisanteries que tu me réserves, je suis ravie de poursuivre aux côtés de ta bonne humeur et de ton chauvinisme (mais comment ne pas l'être lorsque l'on vient d'une si belle région ?). Pierre, enfin, le dernier ingrédient, et non des moindres, du trio d'exception, merci pour tes idées brillantes et les après-midis à gribouiller sur le tableau blanc de l'Inria – j'en sortais éreintée, et impressionnée par ton esprit mathématique aiguisé. Peut-être parce qu'il n'y a pas si longtemps, toi aussi, tu étais doctorant à EDF sous la direction de Gilles et Yannig, tu as su m'apporter toute la relativisation dont j'avais besoin pour mener à bien mes travaux. Je te remercie chaleureusement pour tes conseils amicaux, tes histoires des plus alambiquées et pour l'exemple que tu as été ; j'ai toutefois longtemps redouté de reproduire ton expérience de joyeux rêveur (même un peu tête-en-l'air ?) qui consistait à oublier son ordinateur dans le métro parisien trois jours avant sa soutenance – la pandémie mondiale aura le seul mérite d'avoir calmé cette crainte.

Quelques mots en anglais puisque c'est par des aller-retours en Australie que cette thèse a terminé son voyage – chose plutôt insolite en 2020. Rob Hyndmann, thank you for accepting to review this manuscript, for your careful reading and your relevant remarks. I appreciated your deep knowledge and understanding of forecasting and your brilliant ideas to improve the thesis experimental results. I am honored to count you in my jury. Je suis aussi extrêmement reconnaissante envers Odalric-Ambrym Maillard d'avoir accepté de rapporter mon travail et honorée de sa présence dans mon jury. Merci pour votre relecture assidue d'expert en apprentissage par renforcement et pour vos suggestions pertinentes quant à l'exploration d'approches « non-bandits ». Un grand merci à Nadia Oudjane qui a accepté, avec tout l'enthousiasme qu'on lui connait, de faire partie de mon jury. Depuis quelques semaines, nous cherchons à mélanger contrôle stochastique et apprentissage statistique ; impatiente que tu me transmettes une partie de tes connaissances en optimisation, je me réjouis de cette collaboration. Merci évidement à Émilie Kaufmann : je suis ravie de te retrouver dans mon jury et curieuse de ton avis d'experte en bandits. Enfin, Christophe, je te réserve quelques lignes plus loin, mais merci, déjà, d'avoir accepté de présider mon jury.

Un grand merci à Malo pour une collaboration joyeuse et amicale. Merci de m'avoir initiée aux joies des codes écrits proprement (même si je peine encore à mettre en œuvre tes enseignements) et pour ta patience face à mes *push* incessants de .pdf, *notebook* Jupyter déjà exécutés et autres fichiers indésirables sur Git. Muito obrigada a Ricardo Bessa for giving me the opportunity to visit you in Porto, for your welcome at INESC, for your ideas and for your support. Family and friends suddenly became very interested in deep learning : it was necessary to visit me. Thank you also for sharing lovely restaurant addresses with me. I am honored to have worked with you and hope that this collaboration was only the start of a great venture.

Par ordre d'apparition, j'adresse mes pensées les plus reconnaissantes aux professeurs inoubliables qui m'ont menée aux mathématiques : Mesdames Astier et Thuillier ainsi que Messieurs Germain et Cognet. Je tiens aussi à remercier la FMJH pour m'avoir soutenue lors de mon année de master ; une année décisive. C'est là, au détour du cours génial de statistiques en grande dimension, que j'ai découvert une toute nouvelle branche des mathématiques. Merci Christophe pour ta direction parfaite du master, pour l'énergie que tu y as mise et pour les intuitions mathématiques que tu nous as transmises. Tu sembles toujours savoir ce qui est idéal pour tes élèves, mais tu leur laisses le plaisir de se découvrir – tout en disposant quelques précieux indices sur leur chemin… Alors que je ne savais me décider entre déterminisme et probabilisme, tu m'as sagement conseillée sur mon choix de stage, et il y a quatre ans, tu m'as donné le contact de Yannig lorsque je t'ai parlé de mon envie grandissante de thèse en statistiques. C'est avec un plaisir immense que je te retrouve aujourd'hui à la place de président du jury.

Contrairement aux idées reçues, la thèse est loin d'être un travail solitaire, et il est temps, cette fois dans un ordre plus ou moins aléatoire, de remercier mes collègues et amis qui m'ont soutenue et épaulée. Commençons par l'équipe « prévision de consommation électrique à moyen et court termes » à la R&D d'EDF. Un immense merci à Gilles, avec qui je ne cesse de refaire le monde, à Raphaël pour son humour et ses encouragements et à Audrey pour être un merveilleux modèle. Merci aussi à Manel, Bérénice, Audrey, Sandra, Marc, Carl, Joseph, David et Hui, avec qui j'ai la joie d'encadrer la pimpante Lucile sur

un stage plutôt pointu de calibration d'hyper-paramètres avec des techniques de bandits. Ce fut un plaisir d'être membre de la communauté « Doct'Osiris » et je tiens à remercier particulièrement les anciens : Paulin, Cécile et Rodolphe, pour leurs conseils avisés. Mes pensées chaleureuses aux équipes Sierra et Willow de l'Inria Paris et un grand merci à Loucas, Raphaël, Ulysse, Alex, Yana, Gül, Hadrien, Francis, Alessandro, Adrien et Grégoire pour les Friday Beers au Ground Control et les soirées sponsorisées par les GAFAM et autres entreprises richissimes à ICML. Enfin, je suis extrêmement reconnaissante envers les membres de l'équipe probabilités et statistiques du LMO pour leur accueil et leur bienveillance, et envers Stéphane Nonnenmacher pour le suivi de ma thèse et l'organisation compliquée d'une soutenance *covid-free*. Un grand merci à mes frères de thèse Hédi et Malo que j'étais ravie de retrouver les jeudis après-midi, et à Armand pour nos discussions interminables, mais pour le moins passionnantes. Merci aussi à Pierre, Pierre et Mélanie pour l'aventure Math.en.Jeans et à toute la communauté des doctorants d'Orsay avec qui il est si facile d'échanger une heure de TD, un paquet de copies ou, bien plus souvent, deux ou trois cafés. I am very grateful to the Power and Energy team at INESCTEC for its warm welcome in Porto, and especially to Kamalanathan for his very clear explanations on electrical demand elasticity and to Ricardo for his kind support with Tensorflow. Une mention spéciale pour les doctorants qui ont eu la mauvaise surprise d'atterrir dans un de mes bureaux ; merci Guilhem, Camille, Raphaël et Augustin pour votre soutien et votre indulgence, et, évidemment, merci Julia, pour ta force tranquille. Il semblerait que lorsqu'il est dispensé dans des contrées plus ou moins lointaines (Porquerolles, Stockholm, Los Angeles, pour n'en citer que quelques-unes), un exposé statistique gagne en intérêt. Aussi, la thèse a été l'occasion de participer à de nombreuses conférences ; j'y ai découvert une communauté soudée et passionnée et je n'ose mentionner tous ceux qui m'ont, de près ou de loin, soutenue dans mes recherches. Je tiens à remercier tout de même les doctorants des 8ème Rencontres des Jeunes Statisticiens pour une semaine productive au soleil et, pour leur bienveillance, tous les membres de l'équipe statistique d'AgroParistech, que j'ai eu le plaisir de retrouver aux JdS et qui seront bientôt nos voisins à Saclay (à leur plus grand damne).

Ces trois années n'auraient pas été les mêmes sans ceux qui, au détour d'un verre, d'une soirée ou d'un week-end plus ou moins prolongé, ont su m'éloigner des bornes supérieures de regret. Merci à Fériel, pour ta présence constante ; à Ariane, pour les pinceaux et pigments qui ont coloré cette thèse, pour nos projets entrepreneuriaux et pour avoir longtemps cru en ma capacité à prévoir les chiffres du loto ; à Olivier, qui a eu l'idée surprenante de se lancer dans une thèse sur le photovoltaïque (en physique !), pour nos retrouvailles à EDF ; à Naïs, pour nos discussions d'écologistes féministes en devenir et nos aventures artistiques futures et à Geneviève, pour ta relecture objective de mon introduction et pour nos diners bien trop espacés. Je remercie aussi du fond du cœur mes amis d'école et trouverai un moyen de me faire pardonner les oublis : je ne pouvais citer toute la promo (ni celle d'avant, ni celle d'après…). Merci à toute la clique des parisiens et notamment à Matthias, William, Fabien, Alexandre, Pierre et Jean-Malo ; mais aussi à Zaïd, pour m'avoir fait découvrir mes abdominaux pendant le premier confinement – j'en profite pour remercier la fine équipe « confi'tness » qui a égayé l'écriture de mon manuscrit par ses cardio-apéros Zoom et pour faire un petit clin d'œil à Alice qui nous rejoignait de Sydney. Merci aux pêcheurs à Cannes, les Nicolas et Barthélémy, mais aussi aux toulousains et notamment à Lucia, Jean-Baptiste, Guillemette, Julien, Marc, Léonard, Marie, Thibault, Thomas, Bertrand, Damien et Marion pour les nuits blanches et les brunchs interminables en votre

compagnie. Enfin, une pensée pour Alexis et nos questions existentielles, et pour Camille, Thomas et les week-ends en bivouac ou en char à voile ponctués de « d'accord ».

Ces remerciements sont aussi l'occasion d'une pensée pour mon arrière-grand-mère, professeure de mathématiques au lycée à une époque où de nombreuses écoles d'ingénieurs fermaient leurs portes aux jeunes femmes et où les honneurs scientifiques restaient essentiellement masculins. J'espère qu'elle serait heureuse du chemin parcouru par le monde académique ces dernières décennies. Et si certaines difficultés demeurent, la bienveillance générale à mon égard des trois dernières années me laisse espérer qu'être une femme est bien loin d'être un problème en mathématiques !

Quoiqu'insensible à l'esthétisme des mathématiques et quelque peu déconcertée (à l'exception de Cécile) par la perspective d'une thèse en statistiques, ma famille m'a offert son soutien indéfectible ; je lui en suis éternellement reconnaissante. Merci à mes petits frères, mes premiers élèves (et non des moindres…) : Louis, pour s'être demandé comment il était possible de soigner une équation et Julien dont la présence parisienne fut mon plus beau réconfort. Une pensée émue pour mes grands-parents qui ont accepté avec plaisir de participer au supplice d'une présentation mathématique en anglais et, de surcroit, en visioconférence ! Et puis, Papa et Maman, merci de croire autant en moi, c'est évidemment à vous que je dédie ce travail !

Ambroise, enfin, tu m'as soutenue et supportée dès le début de cette grande aventure (et de façon encore plus certaine ces deux dernières années), merci pour ta gentillesse, pour tes instigations et, surtout, pour toutes les raisons qui ne regardent que nous.

Contributions and thesis outline

Introduction and modeling

Chapter 1 is an introduction (in French) to the industrial problem we tried to model: how to adapt bandits theory to the sequential learning problem of demand side management. It also gives an overview of the thesis results.

Chapter 2 briefly introduces the multi-armed bandit model and the Upper Confidence Bound (UCB) algorithm, initially studied by Auer et al. [2002a]. Then, we propose an elementary bandit approach for demand side management by offering price incentives. We focus on the main differences between our framework and classical bandit theory. Finally, we define a pseudo-regret criteria and, by adapting the UCB algorithm, we offer $\sqrt{T \ln T}$ upper bound on it.

Chapter 3 gives a non-exhaustive review of electricity demand forecasting methods and presents an open data set used in the thesis experiments. Next, the focus is on generalized additive models, a powerful and efficient semi-parametric approach to model electricity consumption.

Bandit algorithms for demand side management

Chapter 4 proposes a contextual-bandit approach for demand side management of an homogeneous population by offering price incentives. The electrical demand is modeled using methods presented in Chapter 3. We propose an algorithm inspired by LinUCB (see Li et al., 2010) and offer $T^{2/3}$ upper bounds on this regret (up to poly-logarithmic terms). Simulations show the efficiency of our strategies.

Chapter 4 generalizes this approach.

Towards application

Chapter 6 proposes an approach for clustering customers according to their consumption behavior, with a view to dropping the previous homogeneous population assumption. It also looks at hierarchical forecasting in the context of energy demand and proposes an approach combining an aggregation algorithm with a projection onto a coherent subspace.

Chapter 7 proposes a method to generate individual power consumption profiles using conditional variational auto-encoders. We built this data generator for an ex-ante assessment of our demand side management policies. A large set of consumers is clustered according to their consumption behaviour and price responsiveness, then consumption profiles are simulated for each cluster.

Synthesis

Chapter 8 generalizes the theoretical results of Chapter 4 to provide a contextual-bandit approach for personalized demand side management. The previous assumption of a homogeneous population is dropped and, by clustering of non-homogenous population into several homogenous groups (by the method of Chapter 6), a protocol for personalized demand side management. Experiments, using the data simulator provided in Chapter 7 to test the proposed strategies, conclude the chapter.

Publications

Chapter 4 and Chapter 7 are based on two articles published during the thesis:

- ⋆ Margaux Brégère, Pierre Gaillard, Yannig Goude and Gilles Stoltz, Target Tracking for Contextual Bandits : Application to Demand Side Management, *Proceedings of the 36th International Conference on Machine Learning*, PMLR 97:754-763, 2019.

- ⋆ Margaux Brégère and Ricardo J. Bessa, Simulating Tariff Impact in Electrical Energy Consumption Profiles With Conditional Variational Autoencoders," *IEEE Access*, vol. 8, pp. 131949-131966, 2020.

Chapter 6 is based on a submitted article:

- ⋆ Margaux Brégère and Malo Huard, Online Hierarchical Forecasting for Power Consumption Data, *arXiv:2003.00585 [stat.ML]*, 2020.

Notation

Without further indications, $\|x\|$ denotes the Euclidean norm of a vector x. Other norms are indexed by a subscript: e.g., the supremum norm of x is is denoted by $\|x\|_\infty$. The estimation of any variable X is denoted by $\widehat{X}$; alternative estimations (provided after smoothing, aggregation, projection etc.) are denoted by $\widetilde{X}$. Most of the time $\bar{X}$ denotes some average value. We tried to use notation as consistent as possible throughout the manuscript; they are all redefined within the chapters. The notation of the variables that appear in several chapters (sometimes in various forms) are summarized here.

α, β	Exploration terms
B	Confidence bound related to power consumption estimation
c	Power consumption target
C	Boundness constants on power consumption
δ	Risk level
ε, E, e	Noise
φ	Mapping function to model the expected power consumption
Γ	Boundness constants on the infinity norm of the covariance matrix relative to power consumption
g	Sub-set of households
$\mathcal{G}$	Set of household sub-sets
G	Number of household clusters
h	Half-hour
H	Number of half-hours in a day
i	Household or cluster of households
$\mathcal{I}$	Set of all households
κ	Constants appearing in deviation inequalities
K	Number of arms (Chapter 2) or incentive signals (tariffs)
λ	Regularization parameter for ridge regression
ℓ	Loss
L	Cumulative loss
μ	Average the power consumption
υ	Humidity (Chapter 6)
π	Position in the year
p	Distribution of the incentive signal sent
$\mathcal{P}$	Set of of legible distributions of the incentive signal sent
ρ	Sub-Gaussian constant
r	Instantaneous regret
R	Cumulative regret
σ^2	Variance of the power consumption
Σ	Covariance matrix relative to power consumption
s, t	Time steps (half-hour for Chapters 3 – 6; days for Chapters 7 and 8)
τ	Temperature
θ	Parameter modeling the power consumption to estimate
T, T_0	Time horizons
V	Matrix modeling knowledge acquired in past iterations
ξ	Effect of incentive signal (tariff) on power consumption
$\mathcal{X}$	Parametric space of exogenous variables
x, X	Vector of exogenous variables (expect for Chapter 6: power consumption)
y, Y	Power consumption

Note for non-French speakers:

"Bandit manchot" is the French translation for "one-armed bandit";
however, a word-to-word translation would be "crook penguin".

With this pun I tried to illustrate each chapter of the manuscript.


MB

1

Introduction et vue d'ensemble des résultats

L'objectif de la thèse est de développer des méthodes d'apprentissage séquentiel, et plus précisément des algorithmes de bandits, pour le pilotage la consommation électrique. Après une brève présentation des enjeux industriels et du cadre théorique, les principales contributions de la thèse sont exposées ; les perspectives envisagées pour la suite des travaux concluent cette introduction.

1 Pilotage de la consommation électrique

Les recherches se sont concentrées sur le pilotage de la consommation électrique (c'est-à-dire sur sa maîtrise), un enjeu de taille pour la fiabilité du système électrique, qui dans un contexte de transition énergétique et numérique, n'a de cesse de se complexifier. Ce système électrique comprend l'ensemble des activités nécessaires à la mise à disposition de l'énergie électrique : sa production, son transport (*via* les lignes haute tension), sa distribution (*via* les lignes moyenne et basse tensions), et sa fourniture (contrat, offre tarifaire...).

1.1 Équilibre entre production et consommation électriques

Pour l'heure, l'électricité ne se stocke à grande échelle qu'à des coûts prohibitifs et *via* des dispositifs peu performants. Sous peine d'effondrement du système électrique, l'équilibre entre la production et la consommation doit donc être rigoureusement maintenu à chaque instant. La gestion de cet équilibre est complexe et nécessite avant tout de s'attarder quelque peu aux spécificités de la demande électrique d'une part, et à celles du "mix de production" d'autre part.

La consommation électrique française

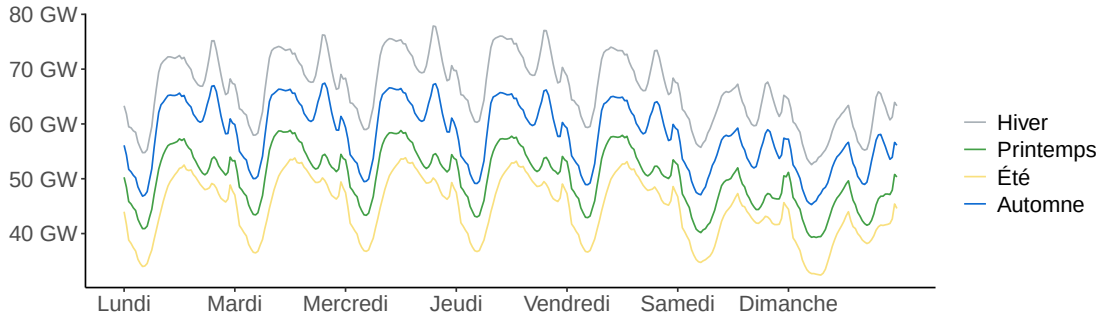


Figure 1.1 – Puissance électrique instantanée hebdomadaire (moyennée sur l'année 2016) consommée en France (en GW) selon les saisons.

Les besoins électriques des entreprises comme des particuliers (qui représentent environ un tiers de l'électricité consommée) sont divers et varient au cours de la journée, de la semaine, des conditions météorologiques ou encore d'évènements exceptionnels. De fait, la demande est généralement bien plus importante au cours de la journée que la nuit et son profil journalier se déforme selon que le jour est ouvrable ou non. Thermosensible, la consommation française varie aussi au gré des saisons, l'utilisation des chauffages électriques entraînant une hausse de la demande en hiver. Les profils hebdomadaires de la consommation française tracés en figure 1.1 pour les quatre saisons de l'année 2016 illustrent ces dépendances. Cette même année la consommation d'électricité en France métropolitaine, qui représente environ un quart de l'énergie finale consommée, s'est élevée à 483,2 TWh, ce qui correspond à environ 7 000 kWh consommés par an et par habitant – contre plus de 50 000 kWh/habitant en Islande et une moyenne mondiale proche de 2 700 kWh/habitant. Notons que les données utilisées pour tracer ces courbes sont les relevés de

consommation, disponibles en *open data*[1], fournis par le Réseau de Transport Électrique (RTE). Enfin, de nombreux évènements, prévisibles ou non, impactent significativement la consommation. Par exemple, en mars 2020, les mesures pour lutter contre la pandémie de Covid-19 telles que la fermeture des commerces non essentiels et le ralentissement de l'activité du secteur industriel ont entraîné une chute de 15% de la demande électrique française ; des baisses de la consommation plus réjouissantes sont aussi observées lors des vacances scolaires ou des jours fériés.

La production électrique française

Afin de répondre à cette demande, les centres de production sont répartis aux quatre coins de la France. À titre d'exemple, les différents moyens de production français mis en œuvre le 3 octobre 2017 (date officielle du début de la thèse) sont représentés en figure 1.2. Dans les centrales électriques, les générateurs sont entraînés par des machines thermiques alimentées par combustion d'une énergie primaire. Lorsque ce combustible est fossile (charbon, gaz naturel ou pétrole), la centrale est dite "thermique" tandis que le terme "bioénergies" est utilisé pour des combustibles organiques (biomasse, déchets). Ces énergies représentent respectivement 7,9 et 1,8% de l'électricité produite sur l'année 2019. Dans les centrales nucléaires (70,6% de la production), c'est la fission de noyaux d'atomes lourds qui dégage de la chaleur. Les barrages hydro-électriques (11,2%) et les fermes éoliennes (6,3%) utilisent l'énergie de l'eau et du vent, transformée en énergie mécanique *via* une turbine hydraulique ou d'une hélice, pour entraîner un générateur électrique. En cas de sur-production (ou de prix de l'électricité très faibles), dans certains barrages, il est possible de remontrer l'eau par pompage, et ainsi stocker l'électricité sous forme d'énergie potentielle. Enfin, les panneaux photovoltaïques (2.2%) convertissent une partie de l'énergie du rayonnement solaire en courant continu.

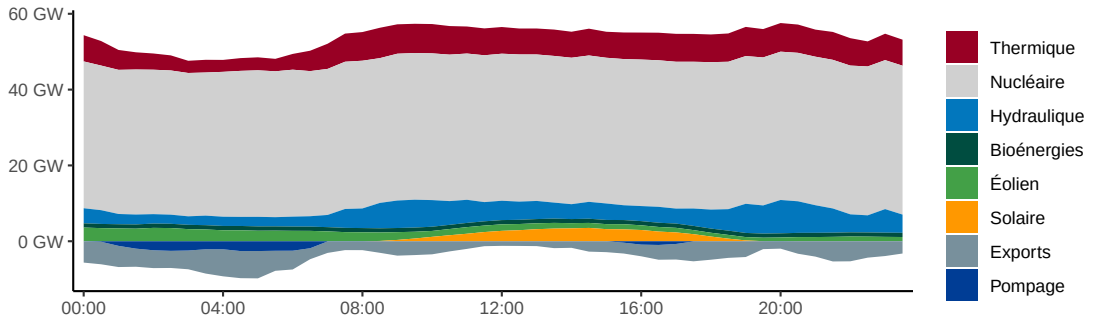


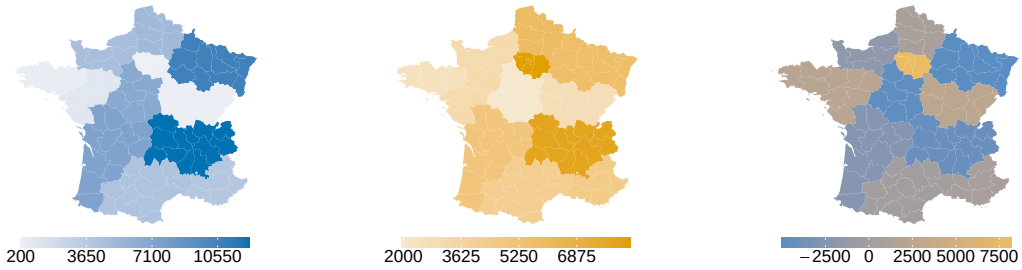
Figure 1.2 – Puissance électrique instantanée (en GW) générée en France le 3 octobre 2017.

La gestion de l'équilibre

RTE transporte l'électricité, *via* les lignes à haute tension, des centrales aux postes sources ; elle est ensuite distribuée *via* le réseau de distribution (Enedis) aux particuliers et aux entreprises (les plus énergivores sont directement rattachées au réseau de transport). Cet acheminement est complexe du fait de la multiplicité des points de génération de l'électri-

[1] `https://www.rte-france.fr/eco2mix`

cité et des nombreux échanges entre régions et avec les pays frontaliers. La figure 1.3 présente la production, la consommation et les échanges électriques entre les régions françaises le 3 octobre 2017 à midi. RTE est en charge de l'équilibre global entre la consommation et la production d'électricité mais impose à chaque fournisseur (tel EDF) de gérer son propre équilibre, qui doit prouver à chaque instant qu'il injecte physiquement la consommation de son portefeuille de clients. Les écarts compensés par RTE sont pénalisés (voir la fin du chapitre 5 pour plus de détails).



(a) Production **(b)** Consommation **(c)** Échanges

Figure 1.3 – Puissance produite (à gauche), consommée (au centre) et échangée (à droite) sur le réseau électrique en France le 3 octobre 2017 à 12:00 (en MW).

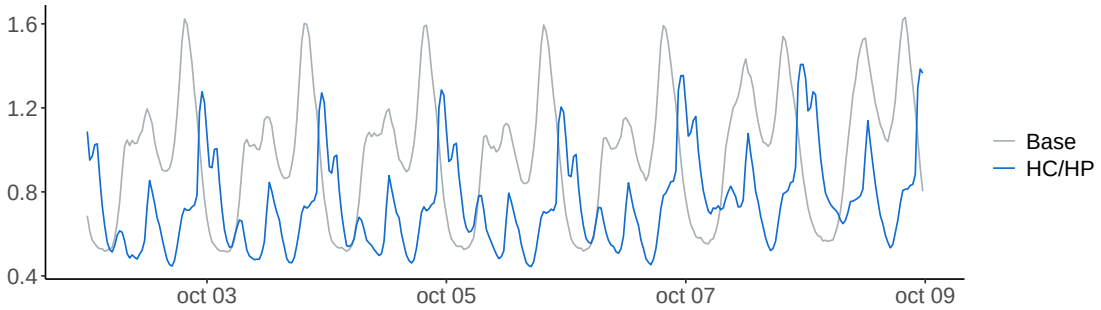


Figure 1.4 – Profils de consommation normés (échelle unitaire) utilisés par Enedis la semaine du 2 au 8 octobre 2017 pour les clients en contrat de base (tarification standard en gris) et les clients en contrat heures creuses/heures pleines (tarification HC/HP en bleu).

Les principales sources de production électrique présentent des caractéristiques bien différentes. Les énergies éolienne et solaire sont intermittentes : elles dépendent des vents et de l'ensoleillement, quand les autres moyens de production sont plus "programmables" (notons toutefois que la production d'électricité hydraulique nécessite que les barrages soient remplis, que l'utilisation des stocks hydrauliques est soumise à des contraintes écologiques ou touristiques et que certaines centrales hydroélectriques fonctionnent sans retenue d'eau – ou au fil de l'eau – et donc sans possibilité de stockage de celle-ci). Le volant principal de la gestion de la demande repose sur la flexibilité des centrales programmables, dont l'activité est déterminée par anticipation de la demande. Les prévisions de cette dernière doivent donc être aussi précises et fiables que possible, pour permettre d'actionner au mieux les différents moyens de production. Elles sont réalisées à différents horizons : les

prévisions à moyen terme (de quelques mois à quelques semaines) permettent de planifier les opérations de maintenance des centrales et les prévisions à court et très court termes (d'une quinzaine de jours à quelques heures) de déterminer leur activité. Prévoir la consommation électrique est donc une activité essentielle pour les producteurs, fournisseurs et distributeurs d'électricité ; elle est plus largement détaillée au chapitre 3. Notons que les centrales nucléaires sont historiquement peu flexibles (bien qu'elles soient plus "manœuvrables" aujourd'hui) et demeurent par ailleurs bien moins ajustables à la consommation que les centrales thermiques, de sorte qu'en réponse à des pointes de consommation, les producteurs ont généralement recours à des combustibles fossiles. Pour les exploitants des centrales nucléaires, une courbe de consommation lisse serait idéale. L'électricité française étant principalement nucléaire, EDF a été pionnier en proposant dès 1965 des tarifs heures creuses/heures pleines qui incitent les clients à consommer hors des pointes de consommation (lorsque l'électricité est moins chère). Les profils de consommation normés (*i.e.* divisés par les consommations moyennes des foyers) de clients à un tarif de base et à un tarif heures creuses/heures pleines sont représentés sur la figure 1.4 pour la semaine du 2 au 8 octobre 2017 et attestent de l'efficacité de tels contrats (les données sont disponibles en *open data* sur le site d'Enedis[2]). En effet, pour le profil HC/HP, les pointes de consommation sont observées plutôt la nuit, en décalé par rapport au profil de base. De telles stratégies dites d'effacement, qui visent à réduire provisoirement de la consommation des clients (en échange d'avantages financiers), se sont ensuite développées et ce notamment avec les acteurs industriels énergivores.

1.2 Évolution du système électrique

Un système au cœur de la transition écologique...

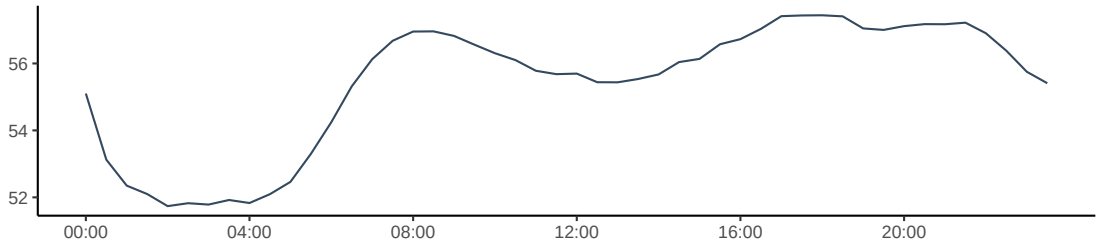


Figure 1.5 – Profil journalier de l'impact carbone du kWh (en gCO_2/kWh), moyenné sur l'année 2016 et au pas de temps demi-horaire (données disponibles sur le site web de RTE).

La volonté d'intégrer des énergies renouvelables, sujettes aux changements météorologiques, au "mix" de production complexifie la gestion de l'équilibre : il devient de plus en plus difficile de planifier la production en fonction de la demande. Par exemple, le 3 octobre 2017, les panneaux photovoltaïques n'ont produit de l'électricité qu'entre 8 et 20 heures et l'électricité éolienne était essentiellement générée la nuit (cf. figure 1.2). Ces plages horaires sont évidemment amenées à évoluer au gré des conditions météorologiques et plus la part du renouvelable intermittent augmentera, plus les fournisseurs chercheront à inciter les consommateurs à s'adapter à la production électrique, c'est-à-dire déplacer

[2] `https://www.enedis.fr/coefficients-des-profils`

leurs usages électriques lorsque de l'électricité est disponible. Notons aussi que les fermes éoliennes et solaires génèrent, en comparaison aux centrales classiques, de petites quantités d'électricité ; elles sont nombreuses et viennent se greffer un peu partout sur le réseau, modifiant ainsi sa structure jusqu'alors plutôt "centralisée". Toujours dans une perspective écologique, le profil journalier moyen (sur l'année 2016) de l'impact carbone du kWh est tracé en figure 1.5. Il est clair que l'électricité consommée en journée rejette plus de CO_2 que la même quantité d'électricité consommée la nuit. Ceci s'explique par le recours aux centrales thermiques pour répondre aux pointes de consommation. C'est donc au moment où l'énergie est produite en grande quantité qu'elle a l'impact carbone le plus fort. Dans un souci de diminution du CO_2 rejeté par la production d'électricité, l'idéal serait donc de lisser au maximum la demande électrique.

... et de la transition numérique

Le système électrique évolue aussi du côté des consommateurs : de nouveaux usages apparaissent tels que les véhicules électriques et certains foyers produisent eux-mêmes une partie de leur électricité, grâce à des panneaux photovoltaïques installés sur les toits des maisons (auto-consommation). L'installation de compteurs intelligents permet d'envisager le pilotage de certains appareils électriques dont les usages peuvent être décalés dans le temps, par exemple la recharge des voitures électriques, le chauffage des ballons d'eau chaude ou encore la mise en fonctionnement des pompes de piscine. Ces compteurs pourraient aussi assurer un accès aux données de consommation rapide et précis ainsi qu'une communication quasi-instantanée entre les acteurs du système électrique. Des solutions de pilotage de la demande électrique peuvent ainsi être imaginées : les producteurs d'électricité et les gestionnaires des réseaux pourraient communiquer directement avec les consommateurs en leur offrant des incitations à consommer lorsque la production est importante, ou inversement, à réduire leur consommation lorsque la situation est plus critique. Des réseaux intelligents (plutôt connus sous la dénomination de *smart grids*), dans lesquels l'information circule en temps réel, émergent déjà, à l'échelle de villes ou de quartiers ; par exemple à Carros et dans le Grand Lyon, où Enedis pilote les projets *Nice Grid* et *Smart Electric Lyon*. Enfin, notons que le développement des véhicules électriques permet d'envisager un stockage décentralisé de l'électricité, *via* l'utilisation des batteries, qui peuvent être mises à disposition du réseau et ainsi se charger ou se décharger en fonction de la demande et de la production électrique (technologie *vehicle-to-grid* ou V2G).

Dans ce nouveau système électrique (illustré en figure 1.6), décentralisé, bas-carbone et connecté, le pilotage dynamique de la consommation devient un enjeu de taille pour la gestion de l'équilibre production-consommation électrique : il permettrait de lisser la consommation, évitant ainsi les pointes de consommation et le recours aux centrales thermiques d'une part, et de mieux intégrer les énergies renouvelables intermittentes d'autre part.

1.3 Stratégies de gestion de l'énergie

Plus concrètement, le pilotage de la consommation électrique pourrait se matérialiser par l'envoi de signaux – tels que des changements des prix de l'électricité – qui permettrait d'inciter les usagers à la consommation en cas de production d'électricité importante (*via* un prix réduit) ou, au contraire, à sa réduction (*via* un prix plus élevé) lors de chutes de production (en l'absence de vent et de soleil, par exemple). Ils pourront alors ajuster

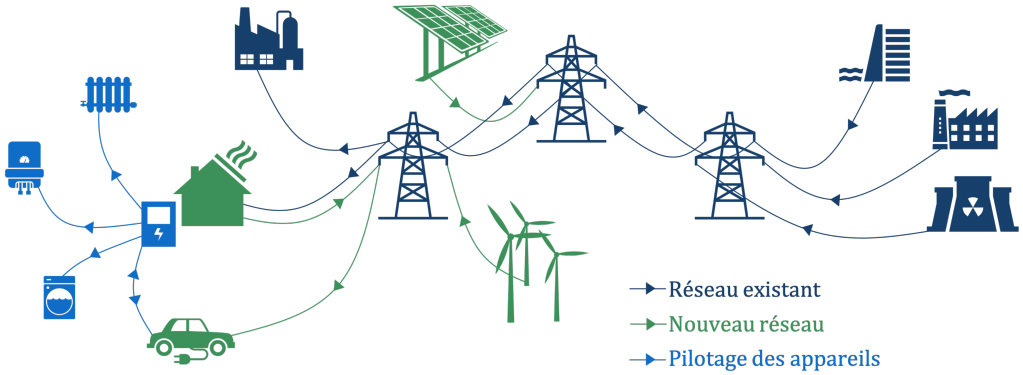


Figure 1.6 – Illustration d'un système électrique en transition écologique et numérique.

la part variable de leur consommation, et décaler dans le temps certains usages, éventuellement *via* des systèmes de contrôle des appareils électriques (*smart home*). Notons qu'EDF met aussi en place des systèmes de pilotage à distance directement au niveau des appareils électriques (sans passer par l'envoi de signaux d'incitation tarifaire) ; par exemple au cours de l'expérimentation "Une Bretagne d'avance"[3], des interruptions des chauffages sont commandées à distance afin de réduire la consommation au moment des pointes (les durées des coupures sont calculées pour ne pas altérer le confort de vie des consommateurs, qui peuvent, à tout moment reprendre le contrôle de leur chauffage). Ces possibilités de maîtrise de la consommation électrique (*demand side management* ou encore *demand response* en anglais) sont actuellement largement étudiées (voir Palensky and Dietrich, 2011 pour un aperçu plus complet des solutions développées).

Positionnement de la thèse

Afin de piloter de la demande électrique au niveau des consommateurs, un enjeu crucial est de choisir, de façon dynamique, les bons signaux à envoyer aux usagers, qui moduleront leur consommation en conséquence afin qu'elle s'ajuste au mieux à la production d'électricité. Les algorithmes utilisés pour choisir ces signaux devront apprendre la réaction des consommateurs face aux envois tout en optimisant ces derniers ; nous parlerons de compromis exploration – exploitation.

L'objectif de cette thèse est la conception de systèmes automatiques d'envoi de signaux d'incitation tarifaire et ce à l'échelle des consommateurs individuels. Remarquons qu'une fois les choix tarifaires effectués et les signaux envoyés aux usagers, seules les consommations électriques associées à ces choix sont observées. Le problème s'inscrit ainsi dans un cadre dit d'information partielle (ou "*bandit information*" en anglais) – en opposition au cadre d'information complète ("*full information*"), pour lequel il serait possible d'observer les consommations électriques associées à chaque choix tarifaire. Ainsi, nous chercherons à piloter la consommation électrique, par envoi d'incitations tarifaires, en utilisant les outils de la théorie des bandits stochastiques (que nous présentons dans la section suivante).

[3] La commission de régulation de l'énergie (CRE) documente les projets de type *smart grids* sur le site `http://www.smartgrids-cre.fr`

L'enjeu principal réside dans une modélisation pertinente de ce problème d'apprentissage par renforcement (elle est détaillée en section 3), viennent ensuite les questions de l'adaptation des algorithmes de bandits aux cas du pilotage de la consommation puis de leurs applications.

2 Théorie des bandits

Cette partie présente brièvement les fondements de la théorie des bandits à plusieurs bras sur laquelle repose l'essentiel des contributions théorique de la thèse. Après avoir décrit le cadre général, nous nous attarderons sur deux algorithmes fondamentaux : l'algorithme *Upper Confidence Bound* (UCB) et sa version "bandits linéaires" LinUCB, qui ont très largement inspiré les stratégies proposées par la suite. Ce cadre mathématiques est plus formellement introduit dans le chapitre 2, où sont aussi rappelées les preuves des résultats théoriques énoncés ci-dessous. Enfin, nous décrirons la modélisation adoptée pour le pilotage de la demande électrique et comment nous envisageons d'adapter la théorie des bandits.

2.1 Modèle de bandits stochastiques à plusieurs bras

Si le modèle de bandits à plusieurs bras a initialement été introduit par Thompson, 1933 dans l'objectif d'améliorer certains procédés de tests cliniques, il tire son nom de la manière imagée dont il est souvent formulé : un statisticien passe les portes d'un casino et fait face à une rangée de machines à sous – aussi appelées bandits manchots – dont les probabilités de gains lui sont inconnues et variables d'une machine à l'autre. À chaque fois qu'il joue sur une machine, il reçoit une récompense, régie par la loi de probabilité associée à la machine choisie. Ne connaissant pas ces lois, pour maximiser ses gains, il doit tester les différentes machines à sous tout en jouant le plus possible sur celle qui semble maximiser ses récompenses. Nous parlerons de compromis exploration-exploitation. Quelles machines faut-il jouer ? Quand et comment ? Comment être sûr d'avoir repéré la meilleure machine ? La théorie mathématique des bandits explore différentes stratégies de maximisation des gains.

Nous considérons K machines à sous numérotées de 1 à K, et au tour t, nous notons I_t la machine à sous choisie par le statisticien. Comment choisir I_t sachant que précédemment, aux tours $1, 2, \dots, t-1$, les machines $I_1, I_2 \dots, I_{t-1}$ ont été jouées et que les gains $Y_1, Y_2, \dots, Y_{t-1}$ ont été observés ? Plus formellement, une distribution ν_k est associée à chaque machine k et lorsque le joueur la choisit, c'est-à-dire lorsque $I_t = k$, il reçoit une récompense

$$Y_t \,|\, I_t = k \;\sim \nu_k \,.$$

Le modèle des bandits stochastiques à plusieurs bras repose sur l'hypothèse que les récompenses associées à une même machine sont indépendantes et identiquement distribuées. Un algorithme de bandits se définit par une stratégie, c'est-à-dire par des règles qui permettent de sélectionner, à chaque tour, la machine à jouer.

L'espérance $E(\nu_k)$ des gains de la machine k (c'est-à-dire le gain moyen de la machine k) sera notée μ_k, de sorte que, si le joueur joue un très grand nombre T de fois sur cette machine k, son gain total sera proche de $T\mu_k$. Si ces espérances étaient connues, la meilleure stratégie consisterait à ne jouer que sur une seule et même machine : la machine $k^\star$ ayant

l'espérance la plus élevée. Afin d'évaluer la performance d'une stratégie, un critère de pseudo-regret est généralement introduit ; il compare, en espérance, les gains associés à la stratégie adoptée par le statisticien et ceux de la meilleure stratégie possible. Dans le cas du problème de bandits à plusieurs bras, en jouant T tours, un joueur omniscient remporterait un gain en moyenne égal à $T\mu_{k^\star}$. Remarquons qu'à un tour t, conditionnellement à la machine choisie I_t, le gain espéré est μ_{I_t} – puisqu'il vérifie $\mathbb{E}[Y_t|I_t] = \mu_{I_t}$ (c'est l'hypothèse fondamentale du modèle de bandits stochastiques à plusieurs bras). Aussi, le pseudo-regret se définit par :

$$\bar{R}_T = T\mu_{k^\star} - \sum_{t=1}^{T} \mu_{I_t} \,.$$

Remarque 1. Idéalement, il faudrait comparer $T\mu_{k^\star}$ avec le gain cumulé $\sum_{t=1}^{T} Y_t$ et non avec $\sum_{t=1}^{T} \mu_{I_t}$. Mais ces deux quantités sont liées (elles sont égales en espérance) et grâce à des inégalités de déviation (telle que l'inégalité d'Azuma-Hoeffding) il est possible de montrer qu'elles sont proches et donc qu'en cherchant à maximiser $\sum_{t=1}^{T} \mu_{I_t}$, le statisticien maximise aussi son gain réel $\sum_{t=1}^{T} Y_t$. Tout ceci est plus largement discuté au chapitre 2, dans lequel la notion de pseudo-regret est plus formellement introduite.

Lorsque les gains sont bornés, disons entre 0 et 1, le pseudo-regret est toujours compris entre 0 et T (puisque les espérances $\mu_1, \ldots, \mu_K$ sont toutes dans l'intervalle $[0, 1]$) ; il est donc, au pire, linéaire en T. Notre objectif sera de montrer que, lorsque l'on joue un grand nombre de fois, le pseudo-regret moyen de nos stratégies converge vers 0, autrement dit que

$$\frac{\bar{R}_T}{T} \overset{T\to\infty}{\to} 0 \,.$$

Obtenir des bornes sous-linéaires en T pour le pseudo-regret permet ainsi de s'assurer du bien-fondé d'une stratégie et c'est ce que démontreront les résultats théoriques de la thèse.

2.2 Algorithme *Upper Confidence Bound*

L'algorithme étudié par Auer et al. [2002a] se fonde sur l'estimation d'espérances μ_1, μ_2, ..., μ_K. Le statisticien qui adopte la stratégie *Upper Confidence Bound* (UCB) commence par jouer sur chaque machine une fois. Puis, à chaque tour t et pour chaque machine k, il considère une estimation $\widehat{\mu}_{t-1,k}$ de μ_k. Cette estimation est calculée à partir des gains observés lorsque la machine k a été jouée : elle est la moyenne empirique

$$\widehat{\mu}_{t-1,k} = \frac{1}{N_{t-1,k}} \sum_{s=1}^{t-1} Y_s \mathbb{1}_{\{I_s=k\}} \quad \text{où} \quad N_{t-1,k} = \sum_{s=1}^{t-1} \mathbb{1}_{\{I_s=k\}} \,,$$

et où $\mathbb{1}_{\{I_t=k\}}$ est égal à 1 si $I_t = k$ et 0 sinon. L'entier $N_{t-1,k}$ correspond donc au nombre de coups joués sur la machine k avant le tour t. À chacune des machines, le statisticien associe un niveau de confiance $\alpha_{k,t}$, qui quantifie l'incertitude qu'il a sur l'estimation $\widehat{\mu}_{t-1,k}$. Ses calculs lui permettent d'affirmer que si jamais il joue la machine k au tour t, il aura de très fortes chances d'obtenir un gain dont l'espérance est comprise dans l'intervalle

$$\left[\, \widehat{\mu}_{t-1,k} - \alpha_{t,k} \,,\, \widehat{\mu}_{t-1,k} + \alpha_{t,k} \right] .$$

Très optimiste face à ses incertitudes, le statisticien décide alors de jouer comme si les machines allaient lui rapporter des gains aussi grands que plausiblement possible, c'est-à-dire comme si chaque machine k allait lui rapporter le gain $\widehat{\mu}_{t-1,k} + \alpha_{t,k}$. Ainsi, au tour t,

il joue la machine associée au gain le plus important selon ses suppositions et choisit

$$I_t \in \underset{k\in\{1,\dots,K\}}{\operatorname{argmax}} \left\{ \widehat{\mu}_{t-1,k} + \alpha_{t,k} \right\}.$$

Les caractéristiques principales des niveaux de confiance $\alpha_{k,t}$ sont qu'ils diminuent lorsque la machine k est jouée : observer un nouveau gain associé à la machine k permet de mettre à jour $\widehat{\mu}_{t-1,k}$ et d'être ainsi plus confiant sur cette estimation. De plus, ils augmentent lentement avec t, de sorte que le statisticien finira toujours par retourner jouer sur une machine qui rapporte peu (pour s'assurer qu'elle est bien sous-optimale). Ces niveaux de confiance permettent ainsi de trouver un bon compromis entre exploration (lorsque $\alpha_{k,t}$ est grand, le statisticien n'est pas confiant sur son estimation $\widehat{\mu}_{t-1,k}$ et s'il choisit $I_t = k$, c'est pour améliorer sa connaissance de la machine k) et exploitation (lorsque $\alpha_{k,t}$ est petit, le statisticien est très confiant sur son estimation $\widehat{\mu}_{t-1,k}$ et s'il choisit $I_t = k$, c'est pour de bonnes raisons : la machine k lui a déjà rapporté beaucoup).

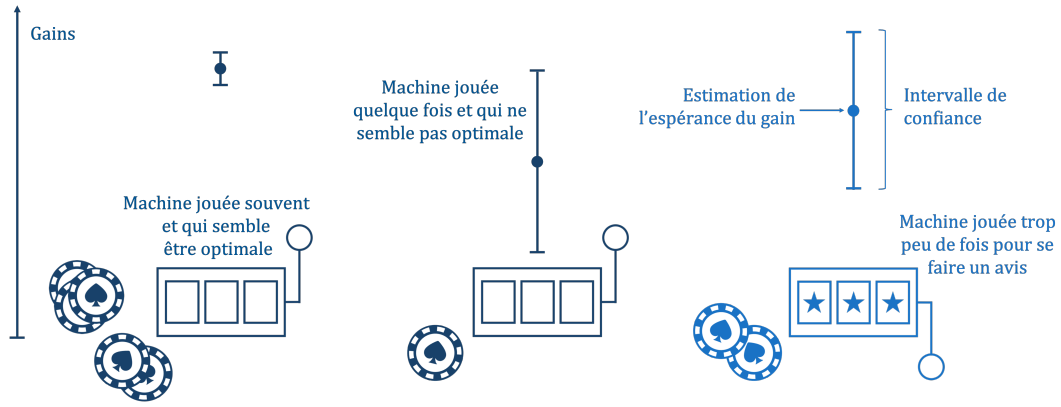


Figure 1.7 – Illustration de l'algorithme *Upper Confidence Bound* (UCB). La machine de gauche a été jouée souvent (l'intervalle de confiance $[\,\widehat{\mu}_{t-1,k} - \alpha_{t,k}\,,\,\widehat{\mu}_{t-1,k} + \alpha_{t,k}]$ est petit et l'estimation du gain est importante : cette machine a déjà rapporté beaucoup). À l'inverse, l'estimation de l'espérance du gain de la machine du milieu est faible et l'intervalle de confiance est grand : le peu de fois où elle a été jouée, elle n'a pas rapporté beaucoup. Son intervalle de confiance n'est toutefois pas assez large pour exiger un nouveau tour d'exploration. Enfin la machine de droite a le niveau $\widehat{\mu}_{t-1,k} + \alpha_{t,k}$ le plus élevé, elle est donc jouée ; le gain observé va permettre de mettre à jour le gain moyen (qui semble pour l'heure n'être ni bon ni mauvais) et de diminuer la taille de l'intervalle de confiance.

Remarque 2. Pour l'algorithme UCB, les niveaux de confiance sont fondés sur l'inégalité de Azuma-Hoeffding (cf. chapitre 2 pour plus de détails) et définis, pour $t \geqslant K$, par

$$\alpha_{t,k} = \sqrt{\frac{2\ln t}{N_{t-1,k}}}\,.$$

L'algorithme UCB est illustré en figure 1.7. De telles stratégies ont été largement étudiées et permettent d'obtenir de bons résultats tant sur le plan théorique que pratique ; en choisissant correctement les niveaux de confiance (ceci est largement détaillé au chapitre 2), il est possible de montrer que l'espérance du pseudo-regret est sous-linéaire, et plus précisément d'ordre

$$\mathcal{O}\big(\sqrt{T\ln T}\,\big)\,.$$

2.3 Algorithme LinUCB

L'algorithme LinUCB, introduit par Li et al. [2010], est une généralisation de l'algorithme UCB au cas où il existe une dépendance linéaire entre l'espérance des gains et certaines variables contextuelles.

Dans ce cadre étendu du modèle de bandits à plusieurs bras, l'espace d'action du statisticien n'est plus forcément discret, il ne choisit alors plus une machine $k \in \{1, \dots, K\}$, mais une action que nous noterons p_t et qui appartient à un espace $\mathcal{P}$. À chaque tour t, il observe un vecteur de variables contextuelles noté x_t qui oriente son choix $p_t \in \mathcal{P}$; il reçoit ensuite le gain Y_t. L'hypothèse majeure du modèle de bandits contextuels avec dépendance linéaire est que, conditionnellement au contexte x_t et à l'action p_t, l'espérance du gain satisfait

$$\mathbb{E}\big[Y_t \,|\, x_t, p_t\big] = \varphi(x_t, p_t)^{\mathrm{T}}\theta \,,$$

où φ est une fonction connue dite de transfert (ou de "*mapping*") et θ est un vecteur paramètres inconnus. L'intérêt d'une telle hypothèse se comprend sûrement mieux avec un exemple : supposons, comme dans l'application originale des modèles de bandits aux test cliniques, que nous cherchions à comparer l'efficacité de deux médicaments. Un tour t correspond à l'arrivée d'un patient à qui il faut attribuer un traitement p_t. Le gain modélise la guérison ou non du malade et le contexte x_t pourrait alors rassembler diverses informations sur le patient (son âge, sa taille, etc.). En supposant que les taux de guérison décroissent linéairement avec l'âge des patients mais qu'ils varient d'un traitement à l'autre, le modèle précédent peut tout à fait convenir.

Notons qu'en omettant les variables contextuelles et en restreignant l'espace $\mathcal{P}$ aux vecteurs $\mathrm{e}_1, \dots, \mathrm{e}_K$ de la base canonique de $\mathbb{R}^K$ (c'est-à-dire les K-vecteurs $(1, 0, 0, \dots)$, $(0, 1, 0, \dots)$, etc.), nous retrouvons le modèle de bandits à plusieurs bras.

Dorénavant, l'action optimale dépend du contexte (un traitement peut être plus efficace chez les enfants que chez les adultes, par exemple) ; et pour le joueur omniscient, qui connaît θ, la meilleure action à prendre au tour t est

$$p_t^\star \in \operatorname*{argmax}_{p \in \mathcal{P}} \varphi(x_t, p)^{\mathrm{T}}\theta \,.$$

Exactement de la même façon que dans le cas du modèle de bandits à plusieurs bras, il est alors possible de définir le pseudo-regret, qui compare, en espérance, la meilleure stratégie à celle adoptée :

$$\bar{R}_T = \sum_{t=1}^{T} \varphi(x_t, p_t^\star)^{\mathrm{T}}\theta - \sum_{t=1}^{T} \varphi(x_t, p_t)^{\mathrm{T}}\theta \,.$$

L'algorithme LinUCB repose sur une estimation du paramètre θ (tout comme l'algorithme UCB reposait sur l'estimation des espérances $\mu_1, \mu_2, \dots, \mu_K$) obtenue à l'aide des observations passées. Au tour t, nous noterons $\widehat{\theta}_{t-1}$ cette estimation, calculée à partir des variables $x_1, p_1, Y_1, \dots, x_{t-1}, p_{t-1}, Y_{t-1}$ (par régression Ridge). Pour chaque action p, un niveau de confiance $\alpha_{t,p}$, qui quantifie le niveau d'exploration de l'action, est aussi défini. Dès lors, après un premier tour d'exploration (p_1 choisit aléatoirement, par exemple), à un tour $t \geqslant 2$, l'algorithme LinUCB choisit l'action p_t de façon optimiste :

$$p_t \in \operatorname*{argmax}_{p \in \mathcal{P}} \big\{ \varphi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1} + \alpha_{t,p} \big\} \,,$$

ce qui assure un regret sous-linéaire d'ordre

$$\mathcal{O}\big(\sqrt{T \ln T}\,\big)\,.$$

Tout ceci est parfaitement documenté dans chapitre 19 de la monographie Lattimore and Szepesvári [2020].

2.4 Modélisation adoptée pour le pilotage de la demande électrique

Le pilotage de la consommation électrique peut être vu comme un problème d'apprentissage séquentiel. À chaque tour t, l'algorithme choisit un signal incitatif à envoyer aux usagers, qui modulent en conséquence leur consommation électrique. Cette dernière est ensuite observée par l'algorithme qui apprend ainsi les réactions des consommateurs face aux changements tarifaires. Nous rappelons que l'algorithme ne peut pas observer ce qu'il se serait produit avec un signal différent, nous sommes bien dans un cadre d'information partielle et il est ainsi naturel de considérer un modèle de bandits.

Au tour $t+1$, l'algorithme décide du nouveau signal à envoyer, et ainsi de suite. Afin de maintenir l'équilibre entre la production et la consommation électrique, il est impératif que les usagers consomment exactement l'électricité produite. Aussi, le choix des signaux à envoyer dépendra d'une consommation cible qui sera donnée à l'algorithme, cette cible correspond par exemple à la production électrique. Comme certaines variables météorologiques et calendaires influent significativement sur la consommation électrique, il est essentiel qu'elles soient elles aussi observées par l'algorithme. Au tour $t+1$, ce dernier décidera donc du signal à envoyer en fonction de la production électrique et des variables contextuelles, et il le fera en utilisant toutes les informations (cibles, variables contextuelles, signaux tarifaires et consommations électriques) recueillies aux tours précédents. La figure 1.8 illustre ce problème d'apprentissage séquentiel ; il sera formalisé plus loin par le protocole 1.

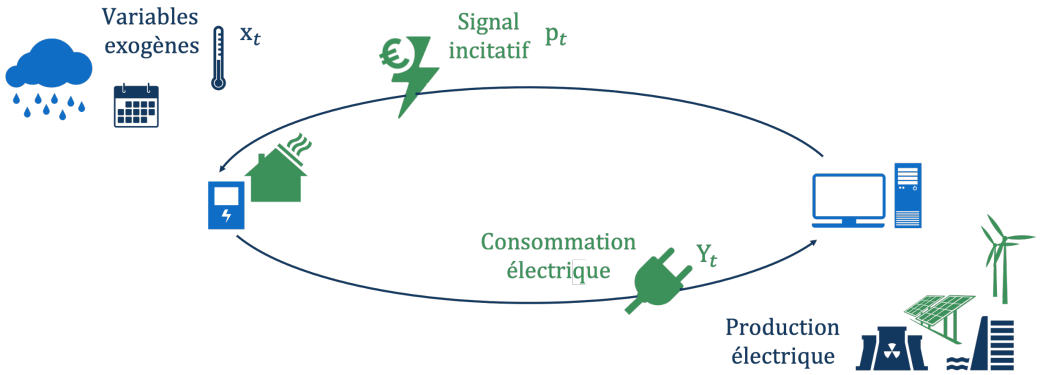


Figure 1.8 – Illustration d'un protocole d'apprentissage séquentiel pour le pilotage de la demande électrique.

Les algorithmes de bandits sont largement utilisés pour modéliser les problèmes d'apprentissage séquentiel puisque la machine à sous du tour t est choisie en fonction des machines jouées et des gains remportés aux tours précédents. Dans le cas du pilotage de charge, les signaux à envoyer aux clients joueront le rôle des machines à sous ou des

actions à prendre p_t et la consommation, celui des gains Y_t. Le contexte x_t regroupe les variables météorologiques et calendaires qui influent significativement sur la consommation (en plus du prix de l'électricité). Les modèles diffèreront cependant du cadre classique : l'objectif n'est plus de maximiser les gains mais d'approcher au mieux une cible connue (la production d'électricité).

3 Algorithme optimiste pour le pilotage de la consommation d'une population homogène – cf. chapitres 4 et 5

Afin de concevoir des systèmes automatiques efficaces pour répondre aux enjeux industriels énoncés, la modélisation de la problématique

Appliquer la théorie mathématique des bandits au problème d'apprentissage séquentiel du pilotage de la consommation électrique

était déterminante. Les littératures sur les modèles de bandits d'une part, et sur les modélisations de la consommation électrique d'autre part, étant très riches, il était primordial de formaliser convenablement le problème. Après avoir fixé une modélisation de la consommation électrique (en nous autorisant certaines hypothèses de départ), un protocole pour le pilotage de la demande a été établi. Les recherches se sont alors concentrées sur l'élaboration d'un algorithme de bandits destiné à piloter la consommation électrique d'une population homogène d'usagers. Il permet d'atteindre, en grande probabilité, un regret sous-linéaire et ses performances théoriques ont été illustrées par des expériences réalisées sur un jeu de données de consommation de foyers londoniens soumis à des changements dynamiques du prix de l'électricité.

Avant d'aborder formellement la question du pilotage, nous tenions à présenter succinctement le jeu de données qui a été utilisé tout au long de la thèse et qui a largement inspiré notre modélisation du problème.

Données issues du projet *Low Carbon London*

Les données utilisées pour tester les algorithmes de bandits sont disponibles en *open data*[4], ce qui permet de s'inscrire dans une démarche de reproductibilité des résultats par la communauté. Elles regroupent les relevés de consommation d'un millier de foyers londoniens ayant participé au projet *Low Carbon London* mené par *UK Power Networks*, au pas de temps demi-horaire et tout au long de l'année 2013. Ces foyers ont été soumis à des prix de l'électricité dynamiques : trois tarifs (bas, standard ou élevé) pouvaient être appliqués. L'ensemble des foyers recevait le même signal de prix. Les tarifs ainsi que leur plage horaire étaient communiqués un jour à l'avance *via* les compteurs électriques et ont été conçus pour être représentatifs des types de signaux qui pourraient être utilisés pour piloter la consommation des usagers. Ces données précieuses, présentées en détail au chapitre 3, ont orienté une partie des travaux.

Nous nous sommes en effet appuyés sur le projet *Low Carbon London* pour modéliser les signaux d'incitation tarifaire : nous nous focaliserons sur un nombre fini de tarifs à

[4] `https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households`

envoyer aux consommateurs (tels que bas, standard ou élevé – une autre approche aurait pu considérer des prix de l'électricité continus, comme c'est le cas sur le "marché de gros" de l'électricité).

De plus, une analyse descriptive poussée de ces données (cf. chapitre 3) a permis de mieux comprendre l'impact d'un tarif sur la consommation et de pouvoir, par la suite, proposer une modélisation cohérente. À titre d'exemple, la figure 1.9 représente l'effet de l'envoi d'un signal tarifaire sur la consommation moyenne de l'ensemble de la population. Il est clair que le tarif élevé induit une baisse de la consommation et le tarif bas, un pic. Mais notons que ces effets sont très dépendants des heures de la journée pour lesquels ils sont appliqués : les baisses ou les hausses de consommation engendrées par des changements de prix de l'électricité sont plus significatives en journée que la nuit.

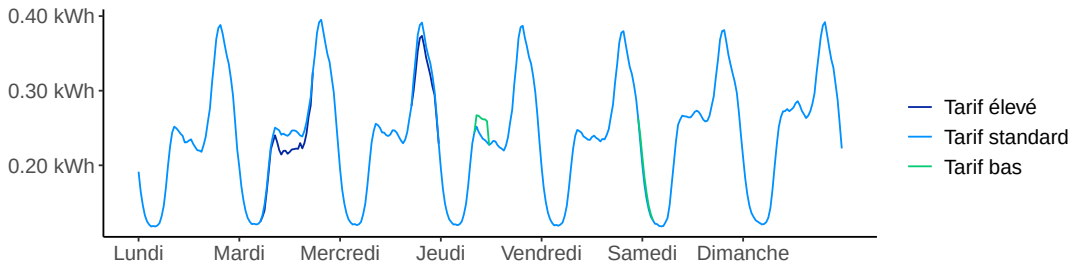


Figure 1.9 – Impact des variations du prix de l'électricité (élevé en marine, standard en bleu et bas en vert) sur la consommation électrique des foyers du projet *Low Carbon London* sur une semaine de l'année 2013.

3.1 Modélisation de la consommation et protocole d'apprentissage séquentiel

Notre première modélisation considère une population homogène de consommateurs ; ces derniers répondent donc tous, en moyenne, de la même manière à un changement des prix de l'électricité. Nous supposons que le fournisseur d'électricité dispose de K tarifs. À chaque demi-heure, selon qu'il est maître de sa production ou non (dans le cas d'un fournisseur d'énergies intermittentes, par exemple), il fixe ou reçoit une consommation cible (qui sera une entrée de l'algorithme de bandits) et observe des variables contextuelles (température, jour de la semaine, etc.). Il peut scinder la population en plusieurs parties et envoyer à chaque sous-population un tarif différent. En choisissant correctement la taille de chaque sous-population ainsi que les tarifs à envoyer, il espère alors approcher au mieux la consommation cible, c'est-à-dire la production d'électricité. Le schéma de la figure 1.10 illustre ce processus, qui est répété toutes les demi-heures.

Plus formellement, au tour t, l'algorithme reçoit une consommation cible c_t ainsi qu'un vecteur de variables contextuelles x_t. Il choisit alors les proportions $p_{t,1}, \dots, p_{t,K}$, regroupées dans un vecteur p_t, de sorte qu'une proportion $p_{t,1}$ de la population reçoit le tarif 1, une proportion $p_{t,2}$ la tarif 2 et ainsi de suite. La population étant supposée homogène, la façon dont elle est partitionnée n'a pour l'heure aucune importance. Après avoir envoyé les tarifs, l'algorithme observe la consommation Y_t. Afin de quantifier la justesse de son choix p_t, il subit alors une perte

$$\ell_t = \big(Y_t - c_t\big)^2 ,$$

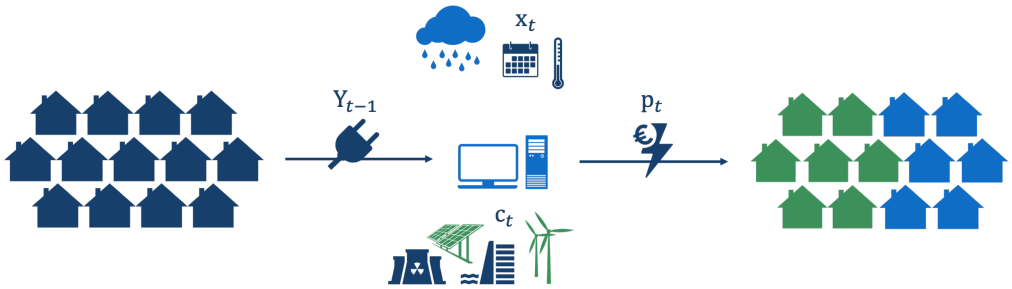


Figure 1.10 – Illustration d'une itération du protocole de pilotage de la demande électrique d'une population homogène. En entrée, l'algorithme reçoit les variables contextuelles x_t et une consommation cible c_t (la production d'électricité) et a observé la consommation au tour précédant Y_{t-1}. Il a aussi accès à l'historique (tours 1 à $t-1$) des consommations, des variables contextuelles et des tarifs choisis. Il choisit alors les proportions de la population à qui il envoie les différents tarifs. Sur cet exemple, il scinde la population en deux et envoie le tarif bleu à 6/13 de la population et le tarif vert au 7/13 restant.

c'est-à-dire la différence au carré entre la cible et la consommation observée. Cette perte peut être vue comme l'opposé des gains des machines à sous du problème classique de bandits.

Détaillons à présent la modélisation de la consommation électrique ; elle est la clé qui nous permettra de proposer un algorithme en réponse à ce protocole de pilotage résumé plus loin, au protocole 1.

Modélisation de la consommation avec un modèle additif généralisé

Tout au long de la thèse, la consommation électrique sera majoritairement modélisée à l'aide de modèles additifs généralisés, très largement utilisés à EDF et présentés en détail au chapitre 3. La consommation électrique est alors vue comme la somme des effets des différentes variables contextuelles (température, jour de la semaine, etc.). À chaque tour t, la consommation suit donc un modèle de la forme :

$$Y_t = f_1(p_t) + f_2(\text{température}) + f_3(\text{jours de la semaine}) + \cdots + \text{bruit}\,,$$

où les fonctions f_1, f_2, etc. sont des fonctions lisses (généralement obtenues par projection sur des bases de *splines* – cf. chapitre 3). Lorsqu'un tel modèle est supposé, il est possible de définir une fonction de transfert φ et de se ramener au modèle linéaire, de sorte que, à un tour t, conditionnellement à x_t et p_t, l'espérance de la consommation Y_t est :

$$\mathbb{E}\big[Y_t \,|x_t, p_t\big] = \varphi(x_t, p_t)^{\mathrm{T}}\theta\,,$$

où θ est un vecteur de paramètres inconnu. L'établissement, à partir du modèle additif généralisé, du modèle linéaire ci-dessus est expliqué au chapitre 3.

Nous verrons un peu plus loin qu'il était aussi primordial de modéliser le terme de bruit. Comme nous avons observé sur les données que les consommations résultantes de l'application de tarifs spéciaux pouvaient être très variables (car ces derniers engendrent une modification des comportements de clients) ; le bruit devra déprendre des signaux p_t

choisis. Le modèle adopté est le suivant : au tour t, lorsque le fournisseur choisi le vecteur de proportion p_t, la consommation électrique Y_t est égale à

$$Y_t = \varphi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t \,.$$

Les vecteurs aléatoires $\varepsilon_1, \varepsilon_2, \ldots$, sont sous-gaussiens[5], indépendants et identiquement distribués, de moyenne nulle et de matrice de covariance Σ. Cette dernière permet de modéliser la variance de la consommation électrique, qui diffère d'un signal tarifaire envoyé à l'autre et n'est pas connue du fournisseur d'électricité. Notre modélisation est plus amplement discutée au chapitre 4 (un second modèle, pour lequel le bruit est indépendant des vecteurs p_t choisis, y est aussi introduit). Ainsi, l'effet des incitations tarifaires impacte la consommation électrique au niveau de son espérance (*via* la fonction f_1 et donc le paramètre θ) d'une part, et de sa variance (*via* la matrice de covariance Σ), d'autre part.

Protocole 1 Pilotage de la consommation électrique

Entrée
Fonction de transfert φ
Paramètres inconnu
Vecteur θ
Matrice de covariance Σ
Pour $t = 1, 2, \ldots$
Observation d'un vecteur de contexte x_t et d'une cible c_t
Choix des proportions de population p_t
Observation de la consommation résultante $Y_t = \varphi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t$
Perte subie $(Y_t - c_t)^2$
Objectif
Minimiser la perte cumulée $L_T = \sum_{t=1}^{T} (Y_t - c_t)^2$

3.2 Minimisation du regret

Au vu du protocole, un fournisseur d'électricité omniscient (c'est-à-dire connaissant le vecteur θ et la matrice Σ) observant le vecteur x_t et la cible c_t choisira le vecteur de proportion qui minimise l'espérance de sa perte :

$$p_t^\star = \operatorname*{argmin}_p \big\{\ell_{t,p}\big\}$$

$$\text{où} \quad \ell_{t,p} = \mathbb{E}\big[(\varphi(x_t, p)^{\mathrm{T}}\theta + p^{\mathrm{T}}\varepsilon_t - c_t)^2 \,|\, x_t, p\big] = \Big(\varphi(x_t, p)^{\mathrm{T}}\theta - c_t\Big)^2 + p^{\mathrm{T}}\Sigma p \,.$$

Ces calculs, détaillés au chapitre 4, montrent que le meilleur choix à prendre à un tour t dépend du vecteur θ et de la matrice Σ. En effet, il est possible que les meilleures proportions à choisir ne soient pas celles qui sont associées à des consommations en moyenne très proches de la cible. La variance (*via* la matrice Σ) de la demande électrique joue un rôle majeur dans notre modélisation : les tarifs spéciaux étant généralement associés à une

[5] Pour $\rho > 0$, un vecteur aléatoire ε de dimension d est ρ–sous-gaussien si, $\forall\, \nu \in \mathbb{R}^d$, $\mathbb{E}\big[\mathrm{e}^{\nu^{\mathrm{T}}\varepsilon}\big] \leqslant \mathrm{e}^{\rho^2\|\nu\|^2/2}$.

variabilité importante, il est parfois préférable de rester à un tarif standard et être sûr de perdre un peu plutôt que d'appliquer un tarif spécial et risquer de perdre beaucoup. Notre algorithme présenté ci-après va permettre de gérer ce compromis biais-variance.

Notons que p_t correspond à l'action prise par le statisticien dans le cadre des bandits linéaires et $p_t^\star$ à la meilleure action a jouer au tour t. Exactement comme dans ce cadre classique présenté en seconde partie, nous introduisons alors le pseudo-regret

$$\bar{R}_T = \sum_{t=1}^{T} \ell_{t,p_t} - \sum_{t=1}^{T} \ell_{t,p_t^\star} \, .$$

Nous avons ensuite proposé un premier algorithme de bandits (cf. section 3 du chapitre 4) qui s'inspire très largement de l'algorithme LinUCB. L'idée est d'estimer, à chaque tour t, pour chaque vecteur de proportions p, l'espérance de la perte $\ell_{t,p}$ qui sera subie ainsi que de calculer un intervalle de confiance autour de cette estimation. Cette dernière est rendue possible grâce à la modélisation, sous forme linéaire, de la consommation électrique. Elle repose sur l'estimation préalable du paramètre θ (exactement comme dans l'algorithme LinUCB) ainsi que sur l'estimation de la matrice Σ (grâce à des tours d'exploration déterministe – cf. chapitre4).

Pour t et p donnés, l'estimation et le niveau de confiance sont respectivement notés $\widehat{\ell}_{t,p}$ et $\alpha_{t,p}$. Notre algorithme, lui aussi optimiste, agit comme si les pertes allaient être aussi petites que plausiblement possible et choisit les proportions p_t selon le critère :

$$p_t \in \operatorname*{argmin}_p \left\{ \widehat{\ell}_{t,p} - \alpha_{t,p} \right\} .$$

Remarque 3. Comme nous regardons des pertes et non des gains, les signes + deviennent des signes − et au lieu de choisir la machine associée au gain plausiblement possible le plus grand, nous choisissons les proportions associées à la perte plausiblement possible la plus petite (de même pour le regret qui a été défini en changeant les signes).

La modélisation posée a permis de démontrer que le pseudo-regret de notre algorithme était bien sous-linéaire et nous énonçons ici une version très simplifiée de notre résultat – qui correspond au Théorème 4 du chapitre 4).

Théorème. Pour un risque $\delta \in]0,1[$, en choisissant correctement les niveaux de confiance $\alpha_{t,p}$, avec probabilité supérieure à $1-\delta$, le pseudo-regret satisfait

$$\bar{R}_T = \sum_{t=1}^{T} \ell_{t,p_t} - \sum_{t=1}^{T} \min_p \ell_{t,p} = \mathcal{O}\big(T^{2/3} \ln^2 \big(T/\delta\big) \sqrt{\ln(1/\delta)} \big) \, .$$

Dans certains cas très favorables (qui supposent notamment que la variance de la consommation est indépendante des tarifs envoyés), nous avons montré que le pseudo-regret était de l'ordre de $\ln^2 T$ (cf. Théorème 5).

Perspective : optimalité théorique de l'algorithme

D'un point de vue purement théorique, il reste à regarder l'optimalité de l'algorithme proposé. En effet, nos résultats montrent que l'algorithme optimiste permet d'apprendre

les bons signaux incitatifs à envoyer, mais pas qu'il est optimal. Pour ce faire, il faudrait s'intéresser aux *lower bounds* de notre modèle de bandits, pour obtenir un résultat du type : "peu importe la stratégie adoptée, le pseudo-regret sera toujours supérieur à ...". Cela permettrait de quantifier l'écart entre le regret de l'algorithme développé et le regret du meilleur algorithme possible, et donc d'évaluer les possibilités d'amélioration.

3.3 Gestion des effets rebond et de bord

La possibilité d'envoyer des tarifs demi-heure par demi-heure n'étant absolument pas réaliste en pratique, un aspect important fut de généraliser les travaux précédents afin d'intégrer ces contraintes opérationnelles. Le fournisseur choisira désormais un profil journalier de tarif qu'il enverra à ses clients la veille ; et il observera un profil de consommation journalier en fin de journée. Cette extension permet aussi de modéliser les effets rebond et de bords observés lors de l'envoi d'un tarif spécial. Plus précisément, lorsqu'un tarif élevé est envoyé, l'effet de ce tarif dure généralement plus longtemps que la plage horaire sur laquelle il est effectivement appliqué (les consommateurs veulent être sûrs de ne pas consommer quand les prix sont élevés et éteignent leurs appareils avant l'application effective du tarif). À l'inverse, pour un tarif bas, l'effet dure généralement moins longtemps (les usagers ne consomment que lorsque les prix sont effectivement faibles). C'est ce que nous désignerons par "effet de bord". De plus lorsqu'un tarif haut est appliqué sur une certaine plage horaire, une baisse de la consommation est observée, mais cela génère généralement une hausse de la consommation à un autre moment de la journée car les usages électriques ne peuvent pas être décalés indéfiniment (à l'inverse l'application d'un tarif bas peut engendrer une baisse de consommation à un autre moment de la journée). Ce phénomène est connu sous le nom "d'effet rebond".

3.4 Extension aux pertes quelconques

Ces premiers résultats théoriques ont été étendus à des fonctions de pertes plus générales. Lorsque le fournisseur d'électricité choisit le vecteur de proportions p_t, il subit désormais une perte

$$\ell_t = f(Y_{t,p_t}, c_t)\,,$$

où f est une fonction vérifiant certaines hypothèses explicitées au chapitre 5. Cette fonction peut par ailleurs varier d'une demi-heure (ou d'un jour) à l'autre.

3.5 Premières applications

Une fois ces résultats théoriques établis, l'objectif final était leur mise en application. Tester un algorithme de bandits est par ailleurs loin d'être évident. En effet, seuls les relevés de consommation électrique associés aux choix tarifaires fixés par les fournisseurs d'électricité sont disponibles. Un jeu de données en "information complète" – comportant, à chaque instant, les relevés associés à tous les choix tarifaires possibles – est cependant nécessaire à l'évaluation des algorithmes. Ces relevés indisponibles sont alors simulés. Ils aspirent à être les plus réalistes possibles afin d'attester des performances réelles des algorithmes et envisager une mise en œuvre opérationnelle.

Lors des premières expériences menées sur le jeu de données *Low Carbon Data*, les relevés de consommation ont été simulés selon le modèle de consommation pré-établi. Bien que peu réalistes sur le plan opérationnel (l'envoi se fait demi-heure par demi-heure, sans

aucune restriction sur les choix – un tarif élevé pendant plusieurs jours est possible, par exemple), elles ont appuyé la théorie.

Notre première méthodologie d'expérimentation soulève toutefois une question de fond importante : est-il raisonnable de tester les algorithmes avec des données générées selon le modèle présupposé par l'algorithme ? Évidemment, ces premiers tests sont nécessaires pour illustrer la performance des stratégies proposées, mais si une mise en œuvre opérationnelle est envisagée, la robustesse des algorithmes doit être évaluée : que se passe-t-il si les données de consommation se suivent pas exactement la modélisation considérée ? Nous verrons en seconde partie de la section suivante comment nous avons tenté d'aborder cette problématique.

4 Segmentation des foyers et simulation de données : vers l'application des résultats théoriques

La mise en application des résultats théoriques nécessitait une connaissance approfondie des problématiques de prévision de consommation électrique, qui sont extrêmement liées aux recherches, et ce à deux niveaux : les modèles de prévision sont, d'une part, intégrés aux algorithmes de bandits – qui doivent estimer correctement la consommation pour optimiser le choix des signaux – et permettent, d'autre part, de simuler des données – et ainsi de tester les algorithmes. Aussi, des travaux sur la prévision de la consommation d'agrégats de foyers, segmentés selon leur profil, et reliés par des contraintes dites hiérarchiques (telle que : la consommation totale est la somme des consommations des différents agrégats) ont ensuite été menés. Ils ont permis de mieux appréhender l'hétérogénéité des profils de consommation, et ont ouvert la voie vers le pilotage personnalisé (les résultats théoriques suggéraient par ailleurs qu'une extension levant l'hypothèse d'homogénéité de la population était possible). Les recherches se sont ensuite concentrées sur l'élaboration d'un générateur de données de consommation électrique construit à partir d'auto-encodeurs variationnels conditionnels, qui permettra de tester la robustesse des algorithmes à des données qui ne suivent pas exactement les modèles supposés par ces derniers.

4.1 Prévision de consommation électrique de groupes de foyers reliés hiérarchiquement – cf. chapitre 6

L'hypothèse précédente d'homogénéité de la population n'est pas vraiment réaliste, d'une part car au sein d'un même réseau électrique, les conditions météorologiques peuvent différer et d'autre part, car les habitudes de consommation sont propres à chaque foyer. Ces derniers sont toutefois généralement catégorisés selon la taille de leur logement, le nombre de personnes qui l'occupent, le type de chauffage, les appareils électriques utilisés etc., ou grâce à un historique de relevés de consommation. Dans un second temps de la thèse, nous nous sommes focalisés sur la prévision de la consommation électrique d'agrégats de foyers, préalablement regroupés selon leurs habitudes de consommation et sur l'utilisation de liens hiérarchiques entre ces agrégats pour améliorer les prévisions. L'objectif est de prévoir la consommation à différents niveaux d'agrégation. En effet, si la prévision de la consommation globale est essentielle pour assurer l'équilibre sur tout le réseau électrique, des prévisions à des échelles plus petites permettent d'envisager un pilotage plus local (si les prévisions sont faites région par région, par exemple) et plus personnalisé (si les pré-

visions sont réalisées par groupe comportemental). Notons qu'en pratique, la somme des consommations locales est égale à la consommation globale, mais que ce n'est forcément pas le cas pour les prévisions (qui sont souvent réalisées indépendamment les unes des autres). En outre, plus le niveau d'agrégation est élevé, plus la courbe de consommation est lisse (les comportements erratiques des individus se moyennent) et plus cette dernière est facile à prévoir. Il ne semble donc pas inutile de prévoir les consommations aux différents niveaux d'agrégation, indépendamment, puis de les combiner, en vue de les améliorer.

Le jeu de données utilisé pour les expériences est décrit au chapitre 6, il rassemble les relevés de consommation d'environ 1 500 ménages, répartis sur plusieurs régions de Grande-Bretagne. Il n'est plus question ici de mesurer l'impact d'un tarif dans le but d'implémenter des stratégies de pilotage de la consommation (les foyers de ce jeu de données n'étant pas soumis à des prix variables de l'électricité), mais plutôt de développer des méthodes de segmentation des usagers et de prévision de leur consommation afin de s'affranchir de l'hypothèse d'homogénéité de la population, et de pouvoir ainsi concevoir des algorithmes de pilotage personnalisé.

L'objectif de ces travaux est de proposer des méthodes de prévision de la consommation électrique de groupes de foyers, et ce à différents niveaux d'agrégation. Nous considérons, par exemple, que la population est segmentée en différentes régions, chaque région ayant sa propre météo, ainsi qu'en différents groupes, chaque groupe ayant un profil de consommation différent. Nous souhaitons obtenir des prévisions de consommation pour chaque sous-groupe d'une même région, pour chaque groupe, pour chaque région et aussi pour la population totale. Cette hiérarchie à trois niveaux est illustrée en figure 1.11.

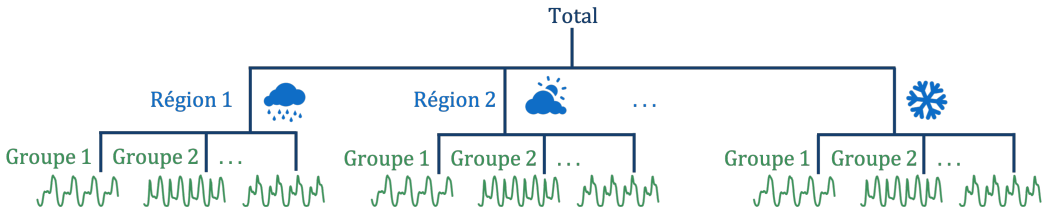


Figure 1.11 – Illustration de la prévision sous contrainte hiérarchique.

Segmentation des foyers

Avant de prévoir les consommations électriques, il a fallu segmenter les foyers. Différentes méthodes ont été testées : des segmentations fondées sur les caractéristiques des foyers (critères socio-démographiques, type de chauffage) ainsi qu'une segmentation utilisant les historiques de consommation. Pour cette seconde méthode, l'historique de relevés de consommation de chaque foyer est d'abord résumé en quelques variables (grâce à une technique de réduction de dimension), qui sont ensuite données à des algorithmes de segmentation classiques (de type *k-means*, par exemple). Ces derniers répartissent alors les foyers dans les différents groupes. Les profils de consommation électrique hebdomadaire moyens sont tracés sur la figure 1.12 et illustrent des habitudes de consommation significativement différentes d'un groupe à l'autre.

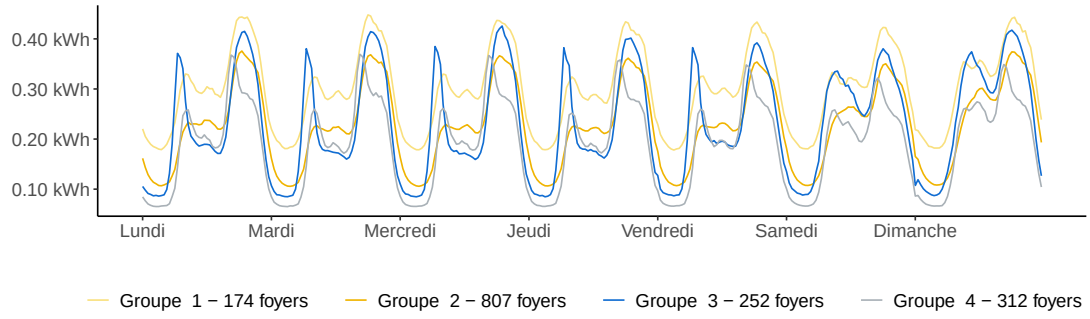
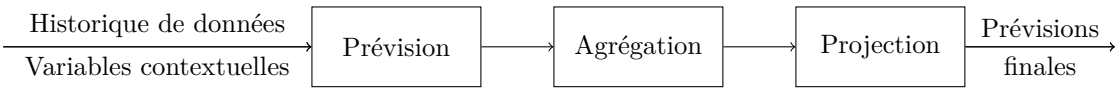


Figure 1.12 – Segmentation des foyers en fonction de relevés de consommation.

Méthodologie : une prévision en trois étapes

Une fois ces groupes définis, une hiérarchie similaire à celle de la figure 1.11 est alors obtenue. La méthode utilisée pour obtenir les prévisions de consommation est la suivante. Pour chaque nœud de la hiérarchie, une prévision est réalisée. Différentes méthodes (fondées sur les modèles linéaires, les modèles additifs généralisés ou les modèles de forêts aléatoires), détaillées au chapitre 6, ont par ailleurs été testées et comparées. Par la suite, pour chaque nœud, l'ensemble des prévisions est utilisé pour prévoir de nouveau la consommation du même nœud. Pour ce faire, nous avons utilisé des algorithmes d'agrégation. Ces algorithmes reçoivent en entrée plusieurs prévisions et renvoient une combinaison de ces prévisions, c'est-à-dire une moyenne pondérée des prévisions. Les poids de cette moyenne sont appris par l'algorithme au fur à mesure des itérations. De tels algorithmes ont fait leur preuve à EDF et sont généralement utilisés pour mélanger les prévisions d'un même agrégat de foyers mais issues de méthodes de prévisions différentes (cf. Goude [2008] et Gaillard [2015] pour plus de détails). L'idée d'utiliser ici les algorithmes d'agrégation était de pouvoir corriger automatiquement les mauvaises prévisions : si la consommation d'un des nœuds est mal prévue alors que celles des nœuds d'à côté le sont correctement, l'algorithme est capable d'utiliser les bonnes prévisions pour corriger la prévision du nœud qui pose problème. Tout cela fonctionne car les prévisions sont liées par les comportements des usagers et les variables contextuelles d'une part, mais aussi par les contraintes hiérarchiques (la somme des consommations de sous-groupes comportementaux d'une région est égale à la consommation de la région etc.) d'autre part. Par ailleurs, comme rien jusqu'alors n'assure que ces contraintes soient vérifiées, une dernière étape, dite de projection, permet d'améliorer encore les prévisions tout en s'assurant qu'elles sont cohérentes (c'est-à-dire que les contraintes hiérarchiques sont vérifiées). La démarche est résumée par le schéma ci-dessous.



Résultats

Nous avons démontré que, sous certaines hypothèses prises sur l'algorithme d'agrégation utilisé, la méthode proposée permettait d'améliorer les prévisions dans leur ensemble ; c'est à dire que la somme des erreurs sur toutes les prévisions diminue (cf. Théorème 9). Les résultats numériques obtenus sur les données de Grande-Bretagne, en utilisant trois

algorithmes d'agrégation vérifiant les hypothèses du théorème, ont été encore plus convaincants que la théorie. En effet, les prévisions ont été améliorées à la fois au niveau des petits agrégats, mais aussi au niveau global ; c'est à dire que les erreurs sur les prévisions locales d'une part, et l'erreur sur la prévision globale d'autre part, ont diminué.

4.2 Simulations de données de consommation électrique – cf. chapitre 7

Nous rappelons que pour tester efficacement un algorithme de bandits pour le pilotage de la demande électrique, il est nécessaire de disposer des relevés de consommation associés à chaque choix possible de l'algorithme. Comme seule la consommation résultante des prix effectivement appliqués est observée, un tel jeu de données n'existe pas. Les relevés indisponibles seront donc simulés et devront être les plus réalistes possibles. Idéalement, les données de consommation simulées doivent reproduire les effets rebond et de bords induits par des changements tarifaires. Pour ce faire, nous générerons des profils de consommation journaliers : le simulateur prendra en entrée des variables contextuelles telle que la température, le jour de la semaine etc. ainsi qu'un profil tarifaire journalier (constitué des tarifs appliqués pour chaque demi-heure de la journée) et renverra un profil de consommation réaliste (mais aléatoire – pour tenir compte de la variabilité de la consommation électrique). Afin de tester la robustesse des algorithmes proposés, nous avons opté pour une approche "boîte noire" orientée données qui ne suppose aucune connaissance préalable sur la consommation électrique et ne nécessite aucune paramétrisation de modèle. Elle repose sur l'utilisation d'auto-encodeurs variationnels conditionnels, des systèmes constitués de réseaux de neurones introduits par Kingma et al. [2014] et présentés au chapitre 7. Afin d'attester du bien-fondé d'une telle méthode, les données simulées ont été comparées à celles générées par des méthodes s'appuyant sur les modèles additifs généralisés, qui requièrent une expertise métier.

Méthodologie

Notre simulateur tente de reproduire des profils journaliers de consommation électrique semblables à ceux du jeu de données *Low Carbon Data*. Les foyers ont tout d'abord été répartis, en fonction de leurs habitudes de consommation et de leur réaction face aux changements des prix de l'électricité, en plusieurs groupes (grâce à une méthode similaire à celle présentée dans le paragraphe précédent). Pour chaque groupe, un auto-encodeur variationnel conditionnel, prenant en entrée les profils journaliers de consommation du groupe ainsi que les profils tarifaires et les variables contextuelles, a ensuite été entraîné. Il était alors possible de simuler de nouveaux relevés de consommation pour chaque groupe, pour n'importe quel profil tarifaire et n'importe quelles conditions météorologiques. Un jeu de données en information complète a ainsi pu être obtenu.

Résultats

Les résultats ont montré que le simulateur capturait correctement l'effet des variables contextuelles ainsi que l'effet d'un profil tarifaire journalier : il est capable de reproduire les effets de bord et de rebond. De plus, pour les mêmes variables contextuelles et les mêmes prix en entrée du simulateur, les échantillons générés différaient d'un groupe de consommateurs à l'autre ; ce qui montre que l'approche de segmentation proposée répartit correctement les foyers en fonction de leur réactivité à un profil tarifaire. Sur la figure 1.13, des profils de consommation simulés sont tracés sur les trois premiers jours de l'année 2013.

Ils correspondent tous à la consommation d'un même groupe de foyers, soumis à différents profils tarifaires : un tarif standard en noir, un tarif bas en début de la journée en jaune et un tarif élevé le soir en bleu.

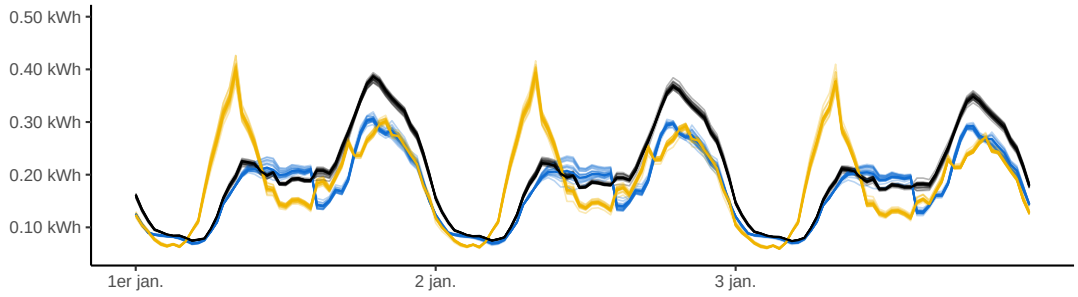


Figure 1.13 – Simulation de profils de consommation pour différents profils tarifaires.

Enfin, nous avons remarqué que le simulateur de profils de consommation peinait à générer des données cohérentes lorsqu'un nouveau profil tarifaire (c'est-à-dire non observé dans les données d'apprentissage) était donné en entrée. Cela paraît assez intuitif : n'ayant pas d'exemple à disposition, le simulateur n'a aucune idée du profil de consommation à associer à ces nouveaux tarifs. Afin d'améliorer ce dernier, nous envisageons d'ajouter de nouvelles données d'apprentissage, issues d'expériences différentes de celle menée sur les foyers du projet *Low Carbon London*, et ce grâce à des méthodes dites de *transfer learning*.

5 Synthèse et perspectives : vers un pilotage personnalisé – cf. chapitre 8

Les derniers travaux tentent de concilier les contributions précédentes et visent à proposer un algorithme de pilotage de la consommation personnalisé (les signaux incitatifs envoyés à un usager dépendent de son profil de consommation) prenant en compte des contraintes opérationnelles (un tarif élevé ne peut par exemple pas être appliqué trop longtemps).

5.1 Modélisation d'une population non-homogène et algorithme optimiste

La contribution finale de la thèse (cf. chapitre 8) considère une population qui n'est plus homogène mais segmentée en G groupes de foyers ayant les mêmes habitudes de consommation électrique. Chaque groupe est vu comme une sous-population homogène, soumise à des conditions météorologiques qui lui sont propres et sa consommation électrique est modélisée de la même manière que l'était la population homogène dans la première contribution.

À chaque tour t, le fournisseur d'électricité souhaite maintenir l'équilibre global entre la production et la consommation et fixe (ou reçoit) donc une consommation cible c_t^{TOT}. Afin de pouvoir favoriser l'intégration des énergies intermittentes à un niveau plus local, par exemple en incitant les foyers proches d'une ferme éolienne à consommer en période de vent fort, nous supposons que le fournisseur peut aussi fixer des cibles locales c_t^{g}, où g

est un sous-ensemble de groupes de foyers. En fonction des conditions météorologiques de chaque groupe, ainsi que des cibles fixées, l'algorithme choisit alors, pour chaque groupe $i = 1, \dots, G$, un vecteur de proportions p_t^i (la proportions $p_{t,1}^i$ du groupe i reçoit le tarif 1, la proportion $p_{t,2}^i$ le tarif 2, etc.) et observe ensuite les G consommations $Y_{t,p_t^1}^1, \dots, Y_{t,p_t^G}^G$. Il subit alors la perte

$$\ell_t = \Big(\sum_{i=1}^{G} Y_{t,p_t^i}^i - c_t^{\mathrm{TOT}} \Big)^2 + \Big(\sum_{i \in \mathrm{g}} Y_{t,p_t^i}^i - c_t^{\mathrm{g}} \Big)^2 + \dots ,$$

c'est à dire la somme des pertes associées à chacune des cibles. La figure 1.14 schématise ce procédé, répété à chaque tour t.

Figure 1.14 – Illustration d'une itération du protocole de pilotage de la demande électrique pour une population non-homogène. En entrée, l'algorithme reçoit des variables contextuelles de chaque groupe, une consommation cible globale (la production d'électricité), ainsi que certaines cibles locales (induites par exemple par la production d'énergie solaire ou éolienne). Il a accès à l'historique de consommation de la population (avec les variables contextuelles et les tarifs choisis sur les itérations passées). Pour chaque groupe, il choisit alors les tarifs à envoyer ainsi que leurs proportions respectives. Sur cet exemple, il ne scinde que les groupe du haut, à qui il envoie le tarif bleu au 4/7 du groupe et le tarif vert au 3/7 restant, le groupe du milieu reçoit le tarif vert et celui du bas le bleu.

En généralisant les travaux présentés au chapitre 4, un algorithme de bandits optimiste est proposé et permet de répondre à ce problème de pilotage de charge personnalisé et contraint. Le Théorème 10 du chapitre 8 démontre que cet algorithme assure un regret sous-linéaire, puisqu'avec grande probabilité, il vérifie

$$\bar{R}_T = \mathcal{O}\big(T^{2/3}\big) .$$

Notons tout de même que cette borne sur le regret dépend du nombre de groupes de foyers G, *via* un facteur $2^{G-1}G$.

5.2 Résultats expérimentaux et améliorations envisagées

Des expériences ont été menées, dans des cas simplifiés, sur les données *Low Carbon London*. Elles utilisent le simulateur de données construit à l'aide des auto-encodeurs variationnels et ont montré que l'algorithme apprenait rapidement à envoyer des profils tarifaires permettant d'approcher les cibles à la fois locales et globales. Mais les algorithmes proposés se sont avérés difficiles à implémenter et nous avons pris certaines libertés vis-à-vis de

la théorie dans les expériences présentées. Ces résultats expérimentaux ont donc vocation à être améliorés.

Tout d'abord, nous rappelons que l'algorithme partage les sous-populations en choisissant les proportions

$$p_t \in \left\{ \operatorname*{argmin}_p \widehat{\ell}_{t,p} - \alpha_{t,p} \right\}.$$

À chaque tour t, il doit ainsi résoudre un problème de minimisation. Ce dernier peut avoir plusieurs solutions et est compliqué à résoudre (car il n'est pas convexe). En outre, l'ensemble des choix possibles de l'algorithme peut être immense. À titre d'exemple, lorsque le choix des proportions p_t est restreint et que tous les foyers d'un même groupe reçoivent le même tarif, dans le cas où trois tarifs sont disponibles et où l'algorithme choisit chaque jour, pour quatre groupes différents, un profil tarifaire demi-heure par demi-heure, $(3^{48})^4 \approx 4.10^{91}$ possibilités (c'est plus que le nombre d'atomes dans l'univers) s'offrent à lui. Heureusement, certaines de ces possibilités peuvent d'ores et déjà être éliminées : d'un point de vue opérationnel, il n'est par exemple pas envisageable d'envoyer un profil tarifaire passant d'un tarif élevé à un tarif bas toutes les demi-heures. En pratique nous avons jusqu'alors restreint les choix des vecteurs p_t à un ensemble fini relativement petit, s'affranchissant ainsi de la question de la minimisation d'une fonction non convexe sur un espace de grande dimension – qui méritait pourtant d'être approfondie.

Notons aussi que les niveaux de confiance $\alpha_{t,p}$ stipulés par la théorie sont généralement très grands, ce qui conduit à des algorithmes qui explorent énormément (et exploitent alors peu les connaissances acquises aux cours des itérations). Trouver le bon niveau d'exploration est un second enjeu expérimental important.

5.3 Perspectives

Au cours de la thèse, les premières hypothèses, notamment celle concernant l'homogénéité de la population, ont été abandonnées et des algorithmes de pilotage personnalisés et prenant en compte certaines contraintes opérationnelles ont alors pu être proposés. Il reste encore toutefois de nombreuses pistes à explorer pour envisager la mise en place effective d'algorithmes de bandits pour le pilotage de la demande électrique.

Notons que les résultats, à la fois théoriques et appliqués, obtenus sur l'amélioration des prévisions de consommation d'agrégats de foyers reliés par des contraintes hiérarchiques n'ont pas directement impacté les recherches sur le pilotage de la consommation. Ils soulèvent cependant de nouvelles problématiques. Serait-il possible d'intégrer ces travaux à l'algorithme de pilotage personnalisé ? À chaque itération, ce dernier estime les consommations électriques associées à chaque groupe de foyers (avant d'estimer les pertes puis les niveaux de confiance). L'idée serait d'améliorer ces prévisions à l'aide d'étapes d'agrégation et de projection. Nous espérons qu'alors, les algorithmes seront capables d'apprendre encore plus vite les bons signaux de prix à envoyer aux usagers.

Nous tenions aussi à noter que pour nos modèles, les groupes de foyers sont fixés à l'avance et restent constant au cours du temps. D'un point de vue opérationnel, il serait intéressant de pouvoir faire varier ces groupes. Un foyer dont les habitudes de consommation changent (lorsque les enfants quittent la maison pour aller faire leurs études, ou suite à l'achat d'une voiture électrique, par exemple) pourrait ainsi être déplacé d'un groupe

à l'autre. Aussi, il semblerait intéressant de coupler les algorithmes de bandits avec des algorithmes de segmentation séquentiels. Dès lors, de nouveaux problèmes apparaissent : si un groupe change significativement de comportement, l'algorithme de bandits devra "oublier" ce qu'il a appris et ré-apprendre les nouvelles habitudes de consommation du groupe.

Ces deux perspectives principales soulèvent une nouvelle problématique visant à

Intégrer des méthodes séquentielles de prévisions et de segmentation des foyers aux algorithmes de pilotage personnalisé de la consommation électrique

et qui pour l'heure, demeure ouverte. La modélisation et la formalisation du problème, qui constituent un enjeu préalable de taille, permettront sans doute d'envisager plus sereinement la conception de ces nouvelles approches.

6 Résumé de la démarche scientifique et plan du manuscrit

La structure du manuscrit et les liens entre les chapitres sont schématisés en figure 1.15 (en anglais, comme le reste du manuscrit). Le chapitre 2 introduit mathématiquement les modèles de bandits à plusieurs bras ainsi que l'algorithme UCB. Un algorithme de bandits pour le pilotage dans un cas basique y est aussi proposé et une borne sur son pseudo-regret démontrée. Le chapitre 3 présente le jeu de données *Low Carbon London* ainsi qu'un bref historique des méthodes de prévision de la consommation électrique. Il s'attarde plus longuement sur les modèles additifs généralisés, largement utilisés à EDF et à l'origine de notre modélisation de la consommation électrique. Les chapitres 4 et 5 introduisent les

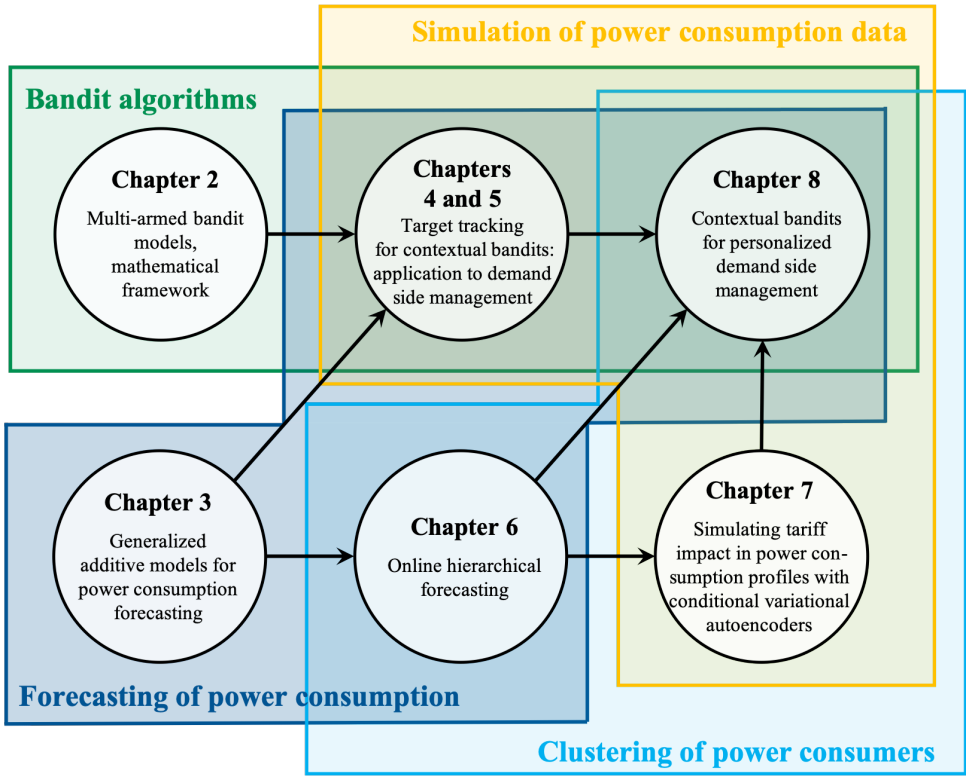


Figure 1.15 – Contents and organization of the thesis manuscript

premiers algorithmes de bandits pour le pilotage de la consommation électrique d'une population homogène. Le chapitre 6 présente les travaux sur la segmentation des foyers ainsi que sur la prévision d'agrégats de foyers reliés par des contraintes hiérarchiques et le chapitre 7, le simulateur de données. Enfin, l'algorithme de bandits pour un pilotage personnalisé est proposé au chapitre 8.

La figure 1.16 retrace la chronologie de la thèse. Le sujet posé en début de thèse, ainsi que toutes les questions qui se sont soulevées au fil des recherches et auxquelles nous avons tenté de répondre sont en bleu clair. Les contributions sont brièvement explicitées en bleu marine et les différents verrous qu'il reste à lever ainsi que les perspectives pour les travaux futurs sont mentionnés en vert.

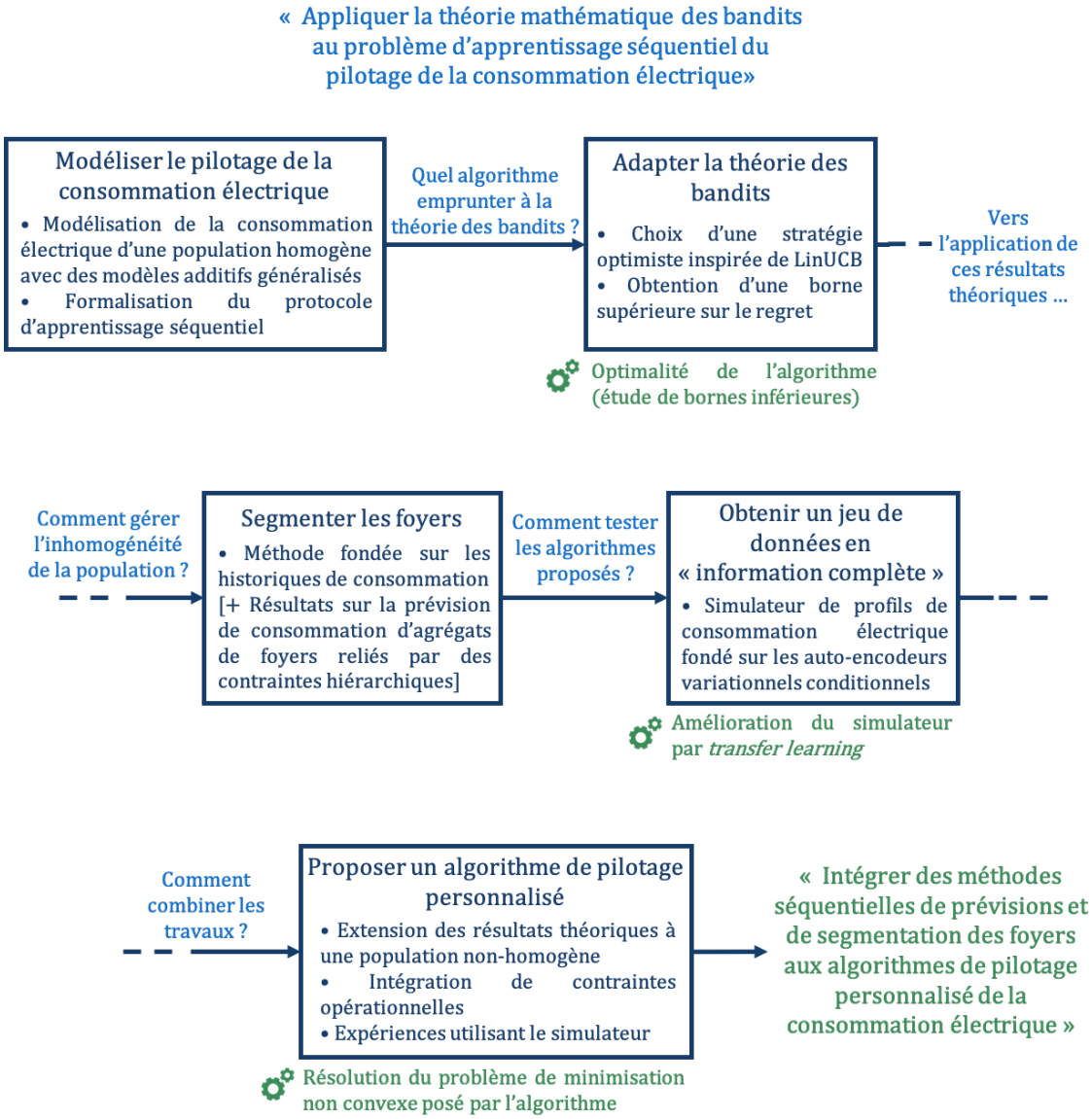


Figure 1.16 – Chronologie et démarche scientifique de la thèse. En bleu clair : les questions soulevées au cours de la thèse, en marine : les contributions apportées en réponse et en vert : les perspectives.


MB

2

Multi-armed bandit models, mathematical framework

This chapter briefly introduces the multi-armed bandit model and the Upper Confidence Bound (UCB) algorithm, initially studied by Auer et al. [2002a]. Both distribution-dependent and distribution-free regret bounds are recalled. Then, we propose an elementary bandit approach for demand side management by offering price incentives. We focus on the main differences between our framework and classical bandit theory. Finally, we define a pseudo-regret criteria and, by adapting the UCB algorithm, we offer $\sqrt{T \ln T}$ upper bound on it.

1 Introduction

William R. Thompson originally introduced the multi-armed bandit problem for a medical application (see Thompson, 1933). Traditionally, clinical trials identify the best of two treatments with A/B testing: by blindly splitting a large population in two sub-groups, they compare subjects' responses to variant A against variant B and determine which of the two variants is more effective. Thompson's idea was to consider a sequential approach: patients volunteering to take part in the trial arrive one by one and an agent decides which of the two treatments is assigned to them using the responses of the previous patients. Mathematically, this trial is model by two slot machines, also called "one-armed bandits" characterized by two unknown probabilistic distributions. A gambler (or an agent) enters in the casino and starts playing. At each new round, she has to decide which machine to play with: should she test the one that seems inferior but hasn't been played much yet or continue playing the one that looks best currently? She thus faces an exploration-exploitation dilemma.

Bandits are classic reinforcement learning problems generally divided in two frameworks: stochastic bandits, for which the feedback (response to treatment or slot machine reward) is sampled from a probability distribution specific to the chosen arm (treatment or slot machine) and adversarial bandits, for which there is no assumption on how the feedbacks are generated. We refer to the exhaustive survey of Bubeck and Cesa-Bianchi [2012] and to the pedagogic book of Lattimore and Szepesvári [2020] which present the different frameworks and algorithms that have been proposed over the last few decades. The modelings considered in this thesis fall under the umbrella of stochastic bandits and among the different existing algorithms we will focus on Upper Confidence Bound (UCB) strategies.

As far as we know, the bandits have not yet been set up in clinical trials, although it is still a high-potential research topic and the design of these solutions is ongoing (see, e.g. Bastani and Bayati, 2020 and Aziz et al., 2020). They are already use for many applications dealing with sequential decision-making under uncertainty. For example, some solutions have been developed for configuring web interfaces, by including recommendation, dynamic pricing and ad placement (see, among others Mary et al., 2015 and Vernade et al., 2017) and for financial portfolio design (see, e.g. Shen et al., 2015). Throughout this thesis, we propose some adaptation of the bandit theory for demand side management. This chapter is dedicated to the presentation of the well-known multi-armed bandit framework and to the introduction of an elementary method for demand side management with bandits.

Section 2 below recalls the mathematical modeling of the multi-armed bandit problem and the UCB algorithm. In Section 3, we propose a basic approach of demand side management and Section 5 establishes some theoretical results for an optimistic algorithm which chooses tariffs to influence the power consumption of a population of consumers.

2 Stochastic multi-armed bandits

2.1 The multi-armed stochastic bandit model

In a multi-armed stochastic bandit model, a gambler faces a row of K slot machines (sometimes known as "one-armed bandits") and has to decide which one to play with to maximize her rewards (see Figure 2.1). A collection of K probability distributions denoted by $\nu_1, \dots, \nu_K$ over $\mathbb{R}$ defines a multi-armed stochastic bandit problem: with each arm $k \in \{1, \dots, K\}$ is associated the probability distribution ν_k with an expectation $\mu_k = \mathrm{E}(\nu_k)$. In what follows, we denote by $\mu^\star = \max_{k=1,\dots,K} \mu_k$ the expectation associated with the best slot machine(s) to play. At each round $t = 1, 2, \dots$, the gambler picks an arm I_t and gets a reward Y_t, drawn at random according to ν_{I_t}: this is the only feedback she has access to. Therefore, conditionally to the chosen arm, the reward is independent from the past and we have $Y_t | I_t \sim \nu_{I_t}$. Protocol 2 below sums up this online procedure.

Figure 2.1 – A row of K slot machines.

Protocol 2 Simplest Case of Multi-Armed Bandits

Input
Number of arms K
Unknown parameters
K probability distributions $\nu_1, \dots, \nu_K$ over $\mathbb{R}$
for $t = 1, 2, \dots$ **do**
Choose an arm $I_t \in \{1, \dots, K\}$
Get and observe the reward $Y_t \,|\, I_t \sim \nu_{I_t}$
end for
Aim
Minimize in high probability or in expectation the pseudo-regret $\bar{R}_T = T\,\mu^\star - \sum_{t=1}^T \mu_{I_t}$

If the player knew the distributions $\nu_1, \dots, \nu_K$ then her best strategy would be: play an optimal arm $k^\star$ such that $\mu_{k^\star} = \mu^\star$ at each round. But the gambler starts from scratch, with no information on the probability distributions: she picks the arm I_t according to the past actions $I_1, \dots I_{t-1}$ and the past observations $Y_1, \dots, Y_{t-1}$. A bandit strategy maps the past experience of the player $I_1, Y_1, \dots, I_{t-1}, Y_{t-1}$ to her next choice I_t. By denoting by $X_{t,k}$ the reward of arm k at round t, to evaluate the performance of a strategy, we consider the regret R_T, first introduced by Lai and Robbins [1985],

$$R_T \triangleq \max_{k \in \{1,\dots,K\}} \sum_{t=1}^T X_{t,k} - \sum_{t=1}^T Y_t\,, \quad \text{with} \quad Y_t = X_{t,I_t}\,.$$

We emphasize that, as rewards are random, the best arm to play at a round t is not necessary an optimal arm and that $\max_{k \in \{1,\dots,K\}} \sum_{t=1}^T X_{t,k}$ may be reached for a sub-

optimal arm. But in bandit framework and contrary to a "full-information" setting where all $X_{t,1}, \dots, X_{t,K}$ are observed, the only feedback we have is the reward Y_t associated with the picked arm I_t. This is why, in stochastic multi-armed bandits, the strategy is generally compared to the best reward in expectation $\mu^\star$. Moreover, we notice that, with the filtration $\mathcal{F}_t = \sigma(I_1, Y_1, \dots, I_t, Y_t)$, at a round t, the arm picked I_t is $\mathcal{F}_{t-1}$-measurable variable such that $\mathbb{E}[Y_t | cF_{t-1}] = \mathbb{E}[Y_t | I_t] = \mu_{I_t}$. By summing over the round $t = 1, \dots, T$, the tower rule gives

$$\mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] = \sum_{t=1}^{T} \mathbb{E}\Big[\mathbb{E}\big[Y_t \,|\, \mathcal{F}_{t-1}\big]\Big] = \mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t}\right].$$

This expression suggests to consider the so-called pseudo-regret $\bar{R}_T$, the random variable which satisfies $\mathbb{E}[T\mu^\star - \sum_{t=1}^{T} Y_t] = \mathbb{E}[\bar{R}_T]$, and is defined as the difference between the cumulative expected reward of the best possible strategy and the one of the strategy associated with the choices $(I_t)_{t \geqslant 1}$

$$\bar{R}_T \triangleq T\mu^\star - \sum_{t=1}^{T} \mu_{I_t}\,.$$

We highlight that we have $\mathbb{E}[\bar{R}_T] \leqslant \mathbb{E}[R_T]$. For any $k = 1, \dots, K$, we denote the gap of the arm k by $\Delta_k \triangleq \mu^\star - \mu_k$. Therefore, a sub-optimal arm k is characterized by a positive gap $\Delta_k > 0$ whereas, for an optimal arm $k^\star$, for which $\mu_{k^\star} = \mu^\star$, the gap is null. Notice that to bound the pseudo-regret, it suffices to bound, for each sub-optimal arm k, the random integer $N_{T,k}$, which counts the number of times arm k has been picked between rounds 1 and T. Indeed, with these notations, the pseudo-regret can we rewritten

$$\bar{R}_T = \sum_{t=1}^{T} \Big(\mu^\star - \mu_{I_t}\Big) = \sum_{t=1}^{T} \Delta_{I_t} = \sum_{k \,|\, \Delta_k > 0} \Delta_k N_{T,k}\,, \quad \text{with} \quad N_{T,k} = \sum_{s=1}^{T} \mathbb{1}_{\{I_s = k\}}\,.$$

In the sequel, we focus on controlling the pseudo-regret in high probability or in expectation. We highlight that under suitable assumption, such results on $\bar{R}_t$ ensure also a control of the regret R_T. For example, by assuming that the rewards are bounded between 0 and 1, applying the Azuma-Hoeffding inequality to the sequence of the $\mathcal{F}_t$-adapted random variables $(Y_t - \mu_{I_t})_{t \geqslant 1}$ gives that, for any $\delta \in (0, 1)$, with probability at least $1 - \delta$,

$$\left| \sum_{t=1}^{T} Y_t - \sum_{t=1}^{T} \mu_{I_t} \right| \leqslant \sqrt{2T \ln \frac{2}{\delta}}\,.$$

Therefore, we will pay a bound of order $\sqrt{T}$ in addition to the bound on the pseudo-regret. To maximize the rewards, bandit algorithms should pick optimal arms as often as possible. But to be sure that these arms are indeed optimal, they must test all the different possibilities. They thus face an exploration-exploitation trade-off. We highlight that the pseudo-regret is always smaller than

$$\bar{R}_T \leqslant T\mu^\star - \min_{k \in \{1,\dots,K\}} \sum_{t=1}^{T} \mu_k = T \max_{k \in \{1,\dots,K\}} \Delta_k\,,$$

and is therefore naturally linear in T. We aim to do better by providing bandit algorithms with sub-linear pseudo-regrets.

From now on, we focus on rewards that are bounded between 0 and 1, so the probability distributions $\nu_1, \dots, \nu_K$ are defined over $[0, 1]$.

2.2 Upper confidence bound strategies

The present section recalls the Upper Confidence Bound algorithm studied by Auer et al. [2002a]. Generally, the theoretical results on this algorithm provide some upper bounds on the pseudo-regret of the form:

$$\mathbb{E}\left[\,\bar{R}_T\,\right] \leqslant \mathcal{O}\Big(\sqrt{T\ln T}\Big)\,.$$

In Chapters 4, 5 and 8, we will establish some regret bounds in high probability for the contextual bandit models we will consider. For the best of our knowledge, there is no high-probability logarithmic bound on the pseudo-regret for the UCB algorithm, which is an any time strategy (there is no need to know the time horizon T in advance). This is why, after presenting this well-known algorithm and the results associated with it, we introduce a δ-risk level version of UCB that must be folklore knowledge. The regret bound will state with probability at least $1-\delta$. From this result, with the time horizon information, we may choose the risk δ to deduce a regret bound in expectation. We will also remark that the UCB algorithm in a moving risk-level version of our δ-UCB algorithm.

This section is cut into four parts: the three first subsections recall the UCB algorithm and the two classical bounds obtained on its pseudo-regret (namely, on the expectation of its regret), while the last subsection introduces a δ-risk level UCB-version. Theorem 2 and Corollary 2 establish some regret bounds for this new algorithm with high probability and in expectation, respectively.

2.2.1 Principle of optimism and upper confidence bound algorithm

Based on the principle of optimism in the face of uncertainty, upper confidence bound strategies choose arms to play "optimistically" depending on the past observations. More precisely, at any round t, for each arm $k = 1, \ldots, K$, they compute an empirical mean reward $\widehat{\mu}_{t-1,k}$ from the previous rewards associated with arm k. They also introduce a confidence level $\alpha_{t,k}$ on it: with high probability, the true mean reward μ_k satisfies

$$\mu_k \in \Big[\widehat{\mu}_{t-1,k} - \alpha_{t,k},\, \widehat{\mu}_{t-1,k} + \alpha_{t,k}\Big]\,.$$

Generally, confidence levels $\alpha_{t,k}$ increase with t to promote exploration and decrease whenever arm k is chosen. Indeed, a new reward Y_t, drawn from ν_k leads to an update of the empirical mean $\widehat{\mu}_{t,k}$, on which we are thus more confident. Then, algorithms act as if the environment is as favorable as plausibly possible and as if the reward for arm k at t was $\widehat{\mu}_{t,k} + \alpha_{t,k}$. Therefore they choose the one which maximizes this quantity.

Algorithm 1 was studied by Auer et al. [2002a] and is illustrated in Figure 2.1. It starts by K rounds of deterministic exploration (one per arm), then at each round $t \geqslant K+1$ and for each arm k, it computes the estimations $\widehat{\mu}_{t-1,k}$ of the means μ_k:

$$\widehat{\mu}_{t-1,k} \triangleq \frac{1}{N_{t-1,k}} \sum_{s=1}^{t-1} Y_s \mathbb{1}_{\{I_s=k\}}, \quad \text{with} \quad N_{t-1,k} = \sum_{s=1}^{t-1} \mathbb{1}_{\{I_s=k\}},$$

and sets the confidence levels

$$\alpha_{t,k} = \sqrt{\frac{2\ln t}{N_{t-1,k}}}$$

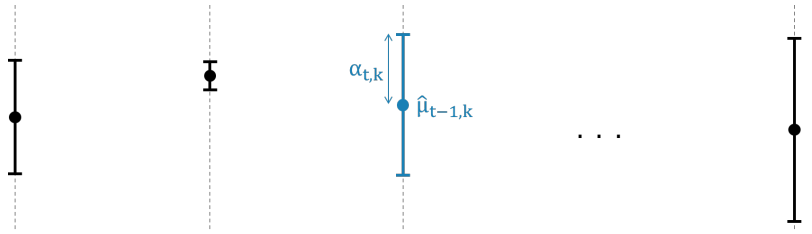


Figure 2.2 – Illustration of the principle of optimism. The upper bounds of the segments correspond to the terms $\widehat{\mu}_{t-1,k} + \alpha_{t,k}$. The second arm is well-estimated ($\alpha_{t,2}$ is small), while the last arm has been picked only a few times ($\alpha_{t,K}$ is large). Even if the empirical mean of the blue arm is smaller than the one of the second arm, the upper confidence bound algorithm chooses it: a large uncertainty on $\widehat{\mu}_{t-1,k}$ forces exploration.

on it. Then, it chooses the next arm optimistically:

$$I_t \in \underset{k\in\{1,\dots,K\}}{\operatorname{argmax}} \left\{ \widehat{\mu}_{t-1,k} + \alpha_{t,k} \right\}.$$

Algorithm 1 UCB (Upper Confidence Bound) Algorithm [Auer et al., 2002a]

Unknown parameters
K probability distributions $\nu_1, \dots, \nu_K \in [0,1]$
Initialization
for each arm the counter $N_{0,k} = 0$ and the empirical mean $\widehat{\mu}_{0,k} = 0$
for $t = 1, \dots, T$ **do**
  if $t \leqslant K$ **then**
    $I_t = t$
  else
    Choose optimistically the next arm $I_t \in \underset{k\in\{1,\dots,K\}}{\operatorname{argmax}} \ \widehat{\mu}_{t-1,k} + \sqrt{\dfrac{2\ln t}{N_{t-1,k}}}$
  end if
  Observe the reward $Y_t \in [0,1]$
  Update for each arm the counter $N_{t,k} = N_{t-1,k} + \mathbb{1}_{\{I_t=k\}}$
  Update for each arm the empirical mean $\widehat{\mu}_{t,k} = \dfrac{1}{N_{t,k}}\left(\widehat{\mu}_{t-1,k} N_{t-1,k} + Y_t \mathbb{1}_{\{I_t=k\}}\right)$
end for

The idea of using confidence bounds came from the work of Lai and Robbins [1985] and a first version of UCB algorithm can be found in Lai [1987]. The one stated in Algorithm 1 and the analysis provided below is from Auer et al. [2002a]. The following proofs of regret bound can also be found, under different versions, in the survey Bubeck and Cesa-Bianchi [2012] or in the book Lattimore and Szepesvári [2020].

2.2.2 Statement of the regret bound

Theorem 1 below states a regret bound for the UCB algorithm. The bound is called distribution-dependent because it depends on the probability distributions $\nu_1, \dots, \nu_K$ through the gaps $\Delta_1, \dots, \Delta_K$.

Theorem 1. *If the distributions $\nu_1, \dots, \nu_K$ have supports all included in $[0,1]$, then, for*

any $T \geqslant 1$, the pseudo-regret of Algorithm 1 satisfies

$$\mathbb{E}\big[\,\bar{R}_T\,\big] \leqslant \sum_{k\,|\,\Delta_k>0} \left(2 + \frac{8\ln T}{\Delta_k}\right).$$

To prove this result, we will use the adaption of Azuma-Hoeffing inequality stated in Lemma 1 below and proved in Appendix, which is based on Hoeffding's lemma (see Lemma 2). It holds because all the rewards lie in $[0,1]$ and mostly because for an arm k, the rewards Y_t for the rounds t such that $I_t = k$ are independent and identically distributed. This can been proved using Doob's optional skipping (see Doob, 1953, Chapter III, Theorem 5.2). With the re-indexation

$$\tau_{k,n} = \min\{t \geqslant 1,\, N_{t,k} = n\}\,,$$

this trick ensures that the variables $\big(Y_{\tau_{k,n}}\big)_n$ are independent and identically distributed according to ν_k.

Lemma 1 (Azuma-Hoeffing inequality with a random number of summands)**.** *For any $k \in \{1,\dots,K\}$, for any $t \geqslant k$ (so that $N_{t,k} \geqslant 1$), for any $\delta \in (0,1)$,*

$$\mathbb{P}\left(\mu_k > \widehat{\mu}_{t,k} + \sqrt{\frac{\ln(1/\delta)}{2N_{t,k}}}\right) \leqslant t\delta$$

and by symmetry,

$$\mathbb{P}\left(\mu_k < \widehat{\mu}_{t,k} - \sqrt{\frac{\ln(1/\delta)}{2N_{t,k}}}\right) \leqslant t\delta\,.$$

Proof of Theorem 1. Foremost, we recall that to bound the expectation of the pseudo-regret, it suffices to bound the expectation of the random integer $N_{t,k}$, for each sub-optimal arm k. Indeed, by linearity of the expectation, the expectation of the pseudo-regret satisfies

$$\mathbb{E}\big[\bar{R}_T\big] = \sum_{k\,|\,\Delta_k>0} \Delta_k \mathbb{E}\big[N_{T,k}\big]\,.$$

The proof breaks down into three steps: Step 1 states the causes leading the optimistic algorithm to play a sub-optimal arm; then, in Step 2, the expectations $\mathbb{E}\big[N_{T,k}\big]$ are upper-bounded; finally, Step 3 concludes the proof.

⋆ *Step 1: Reasons to play a sub-optimal arm.* Let $k^\star$ be an optimal arm, so $\Delta_{k^\star} = 0$. First of all, at a round $t \geqslant K+1$, a sub-optimal arm $k\,|\,\Delta_k > 0$ is picked only if at least one of the following events happens:

(i)	$\mu_k < \widehat{\mu}_{t-1,k} - \alpha_{t,k}$	(μ_k over-estimated)
(ii)	$\mu^\star > \widehat{\mu}_{t-1,k^\star} + \alpha_{t,k^\star}$	($\mu^\star$ under-estimated)
(iii)	$\Delta_k \leqslant 2\alpha_{t,k}$	(arm k not picked often enough)

We will see that the first two events (i) and (ii) hold rarely while the last event ensures the exploration-exploitation trade-off. It occurs if $\Delta_k \leqslant 2\alpha_{t,k}$, which is equivalent, by replacing the confidence level by its definition to $N_{t-1,k} \leqslant {}^{8\ln t}/_{\Delta_k^2}$. These expressions suggest

that the confidence level $\alpha_{t,k}$ is too big compared to the gap Δ_k because the sub-optimal arm k has not been picked often enough: $N_{t-1,k}$ is too small to get a sharp enough estimation of μ_k good enough to distinguish arm k from an optimal arm.

By definition of Algorithm 1, if arm k is picked, in particular, we have

$$\widehat{\mu}_{t-1,k^\star} + \alpha_{t,k^\star} \leqslant \widehat{\mu}_{t-1,k} + \alpha_{t,k}\,. \tag{2.1}$$

So, if inequalities (i) and (ii) do not hold, the gap Δ_k is bounded by:

$$\Delta_k = \mu^\star - \mu_k \overset{\bar{\text{(i)}} \text{ and } \bar{\text{(ii)}}}{\leqslant} \widehat{\mu}_{t-1,k^\star} + \alpha_{t,k^\star} - \widehat{\mu}_{t-1,k} + \alpha_{t,k} \overset{(2.1)}{\leqslant} 2\alpha_{t,k} = 2\sqrt{\frac{2\ln t}{N_{t-1,k}}}\,.$$

$\star$ *Step 2: Bounds on the expectation of the number of times each sub-optimal arm has been played.* For the first K rounds, the algorithm play deterministically by picking $I_t = t$, so $N_{T,k} = 1 + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k\}}$. Then, for any $T \geqslant K+1$ and any sub-optimal arm $k \,|\, \Delta_k > 0$, by decomposing it depending on whether the events (i), (ii) or (iii) defined below occur, the number of times arm k has already been played after round T satisfies

$$\begin{aligned} N_{T,k} &= \sum_{t=1}^{T} \mathbb{1}_{\{I_t=k\}} \\ &\leqslant 1 + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (i)}\}} + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (ii)}\}} + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (iii)}\}}\,. \end{aligned} \tag{2.2}$$

Therefore, the expectation the number of times sub-optimal arm k has been played is bounded by

$$\mathbb{E}[N_{T,k}] \leqslant 1 + \sum_{t=K+1}^{T} \Big(\mathbb{P}\big(I_t = k \text{ and (i)}\big) + \mathbb{P}\big(I_t = k \text{ and (ii)}\big)\Big) + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (iii)}\}}\right].$$

A straightforward application of Lemma 1 with $\delta = t^{-4}$ leads to

$$\begin{aligned} \mathbb{P}\big(I_t = k \text{ and (i)}\big) \leqslant \mathbb{P}(\text{i}) = \mathbb{P}\Big(\mu_k < \widehat{\mu}_{t-1,k} - \alpha_{t,k}\Big) &= \mathbb{P}\left(\mu_k < \widehat{\mu}_{t-1,k} - \sqrt{\frac{2\ln t}{N_{t-1,k}}}\right) \\ &\leqslant (t-1)t^{-4} \leqslant t^{-3}\,, \end{aligned}$$

and symmetrically, $P\big(I_t = k \text{ and (ii)}\big) \leqslant t^{-3}$. It remains to bound the last term of Equation (2.2):

$$\begin{aligned} \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (iii)}\}} &= \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and } N_{t-1,k} \leqslant 8\ln t/\Delta_k^2\}} \\ &\overset{(t\leqslant T)}{\leqslant} \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and } N_{t-1,k} \leqslant 8\ln T/\Delta_k^2\}}\,. \end{aligned}$$

For any t, the indicator $\mathbb{1}_{\{I_t=k \text{ and } N_{t-1,k} \leqslant 8\ln T/\Delta_k^2\}}$ can equal 1 only if $N_{t,k} \leqslant 8\ln T/\Delta_k^2 + 1$; so the sum $\sum_{s=1}^{t} \mathbb{1}_{\{I_s=k\}} = N_{t,k}$ is controlled by this number and we get the deterministic upper-bound:

$$\sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (iii)}\}} \leqslant \left(\frac{8\ln T}{\Delta_k^2} + 1\right) - 1 = \frac{8\ln T}{\Delta_k^2}\,,$$

where the -1 is because $I_k = k$ is not included in the sum over $t = K+1, \dots, T$. Finally, by combining the two bounds above, the expectation of $N_{T,k}$ is upper-bounded by

$$\begin{aligned}\mathbb{E}[N_{T,k}] &\leqslant 1 + \sum_{t=K+1}^{T} \Big(\mathbb{P}\big(I_t = k \text{ and (i)}\big) + \mathbb{P}\big(I_t = k \text{ and (ii)}\big) \Big) + \mathbb{E}\Big[\sum_{t=K+1}^{T} \mathbb{1}_{\{I_t = k \text{ and (iii)}\}} \Big] \\ &\leqslant 1 + 2 \sum_{t=K+1}^{T} t^{-3} + \frac{8 \ln T}{\Delta_k^2} \\ &\leqslant 1 + \int_K^{+\infty} 2t^{-3} \mathrm{d}t + \frac{8 \ln T}{\Delta_k^2} \leqslant 2 + \frac{8 \ln T}{\Delta_k^2} \,.\end{aligned}$$

⋆ *Step 3: Bound in the expectation of the pseudo-regret.* By summing the bounds on $\mathbb{E}[N_{T,k}]$ over the sub-optimal arms and using $\Delta_k \leqslant 1$, we control the expectation of the pseudo-regret:

$$\mathbb{E}\big[\bar{R}_t\big] = \sum_{k \,|\, \Delta_k > 0} \Delta_k \mathbb{E}\big[N_{T,k}\big] \leqslant \sum_{k \,|\, \Delta_k > 0} \left(2 + \frac{8 \ln T}{\Delta_k} \right).$$

□

2.2.3 Distribution-free regret bound

The regret bound stated in the previous section depends on the distributions $\nu_1, \dots, \nu_K$ through the gaps $\Delta_1, \dots, \Delta_K$, for $k = 1, \dots, K$. This dependence is not always desirable: the smaller these gaps, the higher the regret bound. For example, for a fixed horizon time T, if the magnitude of gap associated with a sub-optimal arm is equal to $1/T$, the established regret bound is linear. Therefore, there is some bandit problem $\nu_1, \dots, \nu_K$ for which the previous regret bound has no interest. However, from the previous analysis, a so-called distribution-free regret bound, namely which does not depend on the gaps Δ_k, can be obtained: this is Corollary 1 below.

Corollary 1 (Distribution-free regret bound in expectation)**.** *If the distributions $\nu_1, \dots, \nu_K$ have supports all included in $[0,1]$, then, for any $T \geqslant K$, the pseudo-regret of Algorithm 1 satisfies*

$$\mathbb{E}[\bar{R}_T] \leqslant \sqrt{KT(2 + 8 \ln T)} \,.$$

Proof of Corollary 1. We deduce this distribution-free regret bound from the one of Theorem 1 using the Cauchy–Schwarz inequality. We recall that $\mathbb{E}\big[\bar{R}_T\big] = \sum_{k \,|\, \Delta_k > 0} \Delta_k \mathbb{E}[N_{T,k}]$, so applying the mentioned inequality leads to

$$\begin{aligned}\mathbb{E}\big[\bar{R}_T\big] &= \sum_{k \,|\, \Delta_k > 0} \Big(\Delta_k \sqrt{\mathbb{E}\big[N_{T,k}\big]} \Big) \Big(\sqrt{\mathbb{E}\big[N_{T,k}\big]} \Big) \\ &\leqslant \sqrt{\sum_{k \,|\, \Delta_k > 0} \Delta_k^2 \mathbb{E}\big[N_{T,k}\big]} \sqrt{\mathbb{E}\Big[\sum_{k \,|\, \Delta_k > 0} N_{T,k} \Big]}.\end{aligned}$$

As the sum of $N_{T,k}$ over all the arms is equal to T, the second factor of the above inequality equals $\sqrt{T}$. Moreover, we recall (see Step 2 of the proof of Theorem 1) that the expectation

of $N_{t,k}$, for a sub-optimal arm k, is bounded by

$$\mathbb{E}\big[N_{t,k}\big] \leqslant 2 + \frac{8\ln T}{\Delta_k^2}.$$

Therefore, as gaps Δ_k lie in $[0,1]$, the expectation of the pseudo-regret satisfies

$$\begin{aligned}\mathbb{E}\big[\bar{R}_T\big] &\leqslant \sqrt{\sum_{k\,|\,\Delta_k>0} \Delta_k^2\Big(2 + \frac{8\ln T}{\Delta_k^2}\Big)}\sqrt{T} \leqslant \sqrt{\sum_{k\,|\,\Delta_k>0} \Big(2 + 8\ln T\Big)}\sqrt{T} \\ &\leqslant \sqrt{KT\big(2+8\ln T\big)}\,.\end{aligned}$$

□

Remark 1. *Note that, with $\alpha > 2$, the α-UCB version – see, for example, Bubeck and Cesa-Bianchi [2012] – is defined with the deviation levels $\frac{\alpha \ln t}{2N_{t,k}}$. In this case, one may obtain the bound in expectation*

$$\mathbb{E}[\bar{R}_T] \leqslant \sum_{k\,|\,\Delta_k>0} \left(\frac{\alpha-2}{\alpha} + \frac{2\alpha\ln T}{\Delta_k}\right).$$

2.2.4 An optimistic algorithm with a δ-risk level

Algorithm 2 stated below is a δ-risk level version of UCB for which the only difference with Algorithm 1 (see lines 9 and 10) is the confidence term $\alpha_{t,k}$, which now equals

$$\alpha_{t,k} = \sqrt{\frac{\ln(t^3/\delta)}{2N_{t-1,k}}}\,.$$

We detail in the analysis of the pseudo-regret below how to establish, from these confidence levels, a regret bound with probability at least $1-\delta$, for $\delta \in (0,1)$. Exactly as in the previous sections, after the first K rounds of deterministic exploration (one per arm), Algorithm 2 chooses the next arm optimistically:

$$I_t \in \operatorname*{argmax}_{k\in\{1,\dots,K\}} \widehat{\mu}_{t-1,k} + \alpha_{t,k}\,.$$

For this new algorithm, a distribution-dependent and a distribution-free regret bounds are obtained with high probability (see paragraphs 2.2.4.1 and 2.2.4.2, respectively). We also show that, when the time horizon T is known, it is possible to deduce a bound on the expectation of the pseudo-regret from this high-probability regret bound.

2.2.4.1 High-probability regret bound

Theorem 2. *If the distributions $\nu_1,\dots,\nu_K$ have supports all included in $[0,1]$, then, for any $T \geqslant K$, the pseudo-regret of Algorithm 2 satisfies*

$$\bar{R}_T \leqslant \sum_{k\,|\,\Delta_k>0} \left(1 + \frac{2\ln(T^3/\delta)}{\Delta_k}\right),$$

with probability at least $1-\delta$.

Algorithm 2 δ-Upper Confidence Bound Algorithm

Unknown parameters K probability distributions $\nu_1, \dots, \nu_K \in [0,1]$
Input
risk level $\delta \in (0,1)$
Initialization
for each arm the counter $N_{0,k} = 0$ and the empirical mean $\widehat{\mu}_{0,k} = 0$
for $t = 1, \dots, T$ **do**
  if $t \leqslant K$ **then**
    $I_t = t$
  else
    Choose optimistically the next arm $I_t \in \underset{k \in \{1,\dots,K\}}{\operatorname{argmax}} \ \widehat{\mu}_{t-1,k} + \sqrt{\dfrac{\ln(t^3/\delta)}{2N_{t-1,k}}}$
  end if
  Observe the reward $Y_t \in [0,1]$
  Update for each arm the counter $N_{t,k} = N_{t-1,k} + \mathbb{1}_{\{I_s = k\}}$
  Update for each arm the empirical mean $\widehat{\mu}_{t,k} = \dfrac{1}{N_{t,k}}\big(\widehat{\mu}_{t-1,k} N_{t-1,k} + Y_t \mathbb{1}_{\{I_t = k\}}\big)$
end for

Proof of Theorem 2. The proof of this theorem is similar to the one of Theorem 1. The only difference is in Step 2: instead of bounding the expectation of $N_{T,k}$ for the sub-optimal arms, we will bound the random integers $N_{T,k}$ with probability at least $1-\delta$. By replacing the confidence levels by their new expressions $\alpha_{t,k} = \sqrt{\ln(t^3/\delta)/2N_{t-1,k}}$ at a round $t \geqslant K+1$, a sub-optimal arm $k \,|\, \Delta_k > 0$ is still picked only if at least one of the events (i), (ii) or (iii) defined in the proof of Theorem 1, holds (see Step 1). We now consider the event $\mathcal{E}$ below, which occurs with probability $1-\delta$ and which gathers, for all $t \geqslant K$, and for all sub-optimal arms, the events (i) and (ii):

$$\mathcal{E} \triangleq \Big\{ \forall\, t \geqslant K+1,\ \forall\, k \,|\, \Delta_k > 0 : \widehat{\mu}_{t-1,k} \leqslant \mu_k + \alpha_{t,k} \\ \text{and, for } k^\star : \widehat{\mu}_{t-1,k^\star} \geqslant \mu^\star - \alpha_{t,k^\star} \Big\}.$$

Therefore, over $\mathcal{E}$, at each round $t \geqslant K+1$, for all sub-optimal arms k (such that $\Delta_k > 0$), the estimations $\widehat{\mu}_{t-1,k}$ do not over-estimate the means μ_k and for the optimal arm $k^\star$ (such that $\Delta_k^\star = 0$) for which the estimation $\widehat{\mu}_{t-1,k^\star}$ does not under-estimate the best mean $\mu^\star$. For each round $t \geqslant K+1$, there are at most K inequalities in $\mathcal{E}$ (if the optimal arm is unique, there are exactly K) and each of them occurs with probability at least $1-(t-1)\delta/t^3$ (by applying Lemma 1 with a risk level δ/t^3). Using Lemma 1 and a union bound, we get that the probability that $\mathcal{E}$ does not occur is bounded by

$$\mathbb{P}[\bar{\mathcal{E}}] \leqslant K \sum_{t=K+1}^{+\infty} \times \frac{\delta}{t^2} \leqslant K\delta \int_K^{+\infty} \frac{1}{t^2}\,\mathrm{d}t = K\delta \left[-\frac{1}{t} \right]_K^{+\infty} = \delta\,.$$

If the event $\mathcal{E}$ occurs then, for any $T \geqslant K+1$, for any sub-optimal arm $k \,|\, \Delta_k > 0$,

$$N_{T,k} \leqslant 1 + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (i)}\}} + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (ii)}\}} + \sum_{t=K+1}^{T} \mathbb{1}_{\{I_t=k \text{ and (iii)}\}}$$
$$\overset{\text{over } \mathcal{E}}{\leqslant} 1 + \sum_{t=K+1}^{T} \mathbb{1}_{\left\{I_t=k \text{ and } N_{t-1,k} < 2\ln(t^3/\delta)/\Delta_k^2\right\}}.$$

We showed in Step 2 of the proof of Theorem 1 that $\sum_{t=K+1}^{T} \mathbb{1}_{\left\{I_t=k \text{ and } N_{t-1,k} < 8\ln t/\Delta_k^2\right\}}$ was deterministically bounded by $8\ln T/\Delta_k^2$. In the same way, we upper-bound the second term of the inequality above:

$$\sum_{t=K+1}^{T} \mathbb{1}_{\left\{I_t=k \text{ and } N_{t-1,k} < 2\ln(t^3/\delta)/\Delta_k^2\right\}} \leqslant \frac{2\ln(T^3/\delta)}{\Delta_k^2}.$$

As in Step 3, it remains to sum the bounds on $N_{T,k}$ over the sub-optimal arms to control the pseudo-regret with probability $1-\delta$ and conclude the proof. Indeed, when $\mathcal{E}$ occurs, we have

$$\bar{R}_t = \sum_{k \,|\, \Delta_k > 0} \Delta_k N_{T,k} \leqslant \sum_{k \,|\, \Delta_k > 0} \left(1 + \frac{2\ln(T^3/\delta)}{\Delta_k}\right).$$

□

From this high-probability regret bound, by choosing a risk-level of order $1/T$ (or even smaller), it is possible to get a bound on the expectation of the pseudo-regret. We highlight that this δ-risk level algorithm is not an any time strategy: to obtain a regret bound in expectation, we use the time horizon information. By taking a moving risk level $\delta_t = 1/t$ at each time step, we recover the classical Upper Confidence Bound algorithm (see Algorithm 1). In this case, as the algorithm does not depend anymore on a risk level, and as far as we know, there is no bound in high probability. However, it guarantees the bound on the pseudo-regret expectation stated in Theorem 1. Corollary 2 below states the result in expectation, when the time horizon T is known.

Corollary 2 (regret bound in expectation)**.** *If the distributions $\nu_1, \ldots, \nu_K$ have supports all included in $[0,1]$, then, for any $T \geqslant K$, the expectation of the pseudo-regret of Algorithm 2 run with $\delta = 1/T$, satisfies*

$$\mathbb{E}[\bar{R}_T] \leqslant 1 + \sum_{k \,|\, \Delta_k > 0} \left(1 + \frac{8\ln T}{\Delta_k}\right).$$

Proof of Corollary 2. As all Δ_k are bounded between 0 and 1, the pseudo-regret $\bar{R}_T$ is always smaller than T. For any u, we have

$$\mathbb{E}[\bar{R}_T] \leqslant T\,\mathbb{P}(\bar{R}_T > u) + u\,\mathbb{P}(\bar{R}_T \leqslant u).$$

By choosing $u = \sum_{k \,|\, \Delta_k > 0} \left(1 + 2\ln(T^3/\delta)/\Delta_k\right)$, we decompose the expectation of $\bar{R}_T$ depending on whether it is controlled by the bound established in Theorem 2 (that occurs with probability $1-\delta$) or not (in which case we bound it by T). Therefore, we get

$$\mathbb{E}[\bar{R}_T] \leqslant T\delta + \sum_{k \,|\, \Delta_k > 0} \left(1 + \frac{2\ln(T^3/\delta)}{\Delta_k}\right).$$

By choosing $\delta = 1/T$, we conclude the proof. □

2.2.4.2 Distribution-free regret bound

Corollary 3 (Distribution-free regret bound in high probability)**.** *If the distributions $\nu_1, \dots, \nu_K$ have supports all included in $[0,1]$, then, for any $T \geqslant K$, the pseudo-regret of Algorithm 2 satisfies*

$$\bar{R}_T \leqslant \sqrt{KT(\ln T^3/\delta + 1)}\,,$$

with probability at least $1-\delta$. Moreover, by choosing $\delta = 1/T$, the expectation of the pseudo-regret satisfies

$$\mathbb{E}[\bar{R}_T] \leqslant 1 + \sqrt{KT(8\ln T + 1)}\,.$$

Proof of Corollary 3. We deduce this distribution-free regret bound from the one of Theorem 2 using the Cauchy–Schwarz inequality. Indeed, applying the inequality leads to

$$\begin{aligned}\bar{R}_T &= \sum_{k\,|\,\Delta_k>0} \left(\Delta_k \sqrt{N_{T,k}}\right)\left(\sqrt{N_{T,k}}\right) \\ &\leqslant \sqrt{\sum_{k\,|\,\Delta_k>0} \Delta_k^2 N_{T,k}} \sqrt{\sum_{k\,|\,\Delta_k>0} N_{T,k}}.\end{aligned}$$

As the sum of $N_{T,k}$ over all the arms is equal to T and as, with probability at least $1-\delta$, for all sub-optimal arms k, we have $N_{t,k} \leqslant 1 + \ln(T^3/\delta)/\Delta_k^2$; the pseudo-regret is bounded by:

$$\begin{aligned}\bar{R}_T &\leqslant \sqrt{\sum_{k\,|\,\Delta_k>0} \Big(1 + \ln(T^3/\delta)\Big)}\sqrt{T} \\ &\leqslant \sqrt{KT\Big(1 + \ln(T^3/\delta)\Big)}\,,\end{aligned}$$

with probability at least $1-\delta$. To get the bound in expectation, we do exactly as in the proof of Corollary 2. □

3 Stochastic multi-armed bandits for demand side management

Previous results and algorithms are well-known and we now present the main objectives of the thesis and how it will be possible to use bandit theory to perform demand side management, that is, to influence the power consumption in order to maintain the balance between the production and the consumption of electricity. The simplest modeling is introduced: it relies on the multi-armed bandit problem introduced above and will become more complex through Chapters 4, 5 and 8. We also propose an algorithm for this simple bandit approach and provide a $\mathcal{O}(T \ln T)$ bound on its regret.

3.1 Framework, objectives and examples

3.1.1 Consumption modeling

We consider the power consumption of an homogeneous population of households which, at each round t depends, among others, on some exogenous factors (temperature, wind, season, day of the week, etc.), forming a context vector $x_t \in \mathcal{X}$, where $\mathcal{X}$ is some parametric space. To manage demand, the electricity provider sends some incentive signals to its customers (for example, it changes electricity price by making it more expensive to reduce consumption or less expensive to encourage customers to consume more now rather than in some hours). We assume that $K \geqslant 2$ price levels (tariffs) are available. The consumption of the population getting tariff $k \in \{1, \dots, K\}$ is assumed to be of the form $\phi(x_t) + Y_t$, where the random variable $Y_t \sim \nu_k$ models the effect of tariff k (which we assume to be independent of the context in this chapter). We make no assumption on the distributions $(\nu_k)_{1,\dots,K}$.

Therefore, at a round t, the electricity provider picks the tariff I_t which is sent to all households, and observes the resulting mean power consumption:

$$\phi(x_t) + Y_t, \quad \text{with} \quad Y_t \,|\, I_t \sim \nu_{I_t}\,.$$

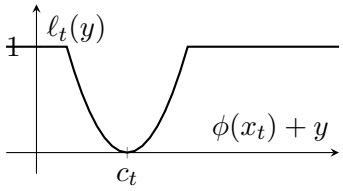
The mean is obtained by averaging the power consumption of households over the whole population. In this basic modeling, only the distributions $(\nu_k)_{1,\dots,K}$ are unknown, that is, the electricity provider has a complete knowledge on how the consumption depends on contextual variables x_t but it has no idea on how the population reacts to a tariff change. Thus, we assume that the function ϕ is known and also that its support is included in $[0, C]$. In this new framework, the tariffs play the role of the slot machines and the effect of a tariff the one of the reward. If we wanted to maximize the power consumption, we could directly apply the algorithms presented in Section 2.2. But our aim is different: we do not want to maximize consumption at all cost but rather fit it as well as possible with electricity production. We explain below how to model this new objective.

3.1.2 Target, loss function and online protocol

To manage demand response, the electricity provider may set a power consumption target c_t to reach at each round t. This target is assumed to be bounded by the constant C. When it chooses a tariff k, it observes the resulting tariff effect Y_t, drawn from the distribution ν_k. To measure the relevance of its choice, it fixes a loss function $\ell_t : \mathbb{R} \to [0, 1]$ and suffers the loss $\ell_t(Y_t) \in [0, 1]$. This loss function varies over time: in particular, it depends on the target, but may also change with the electrical market, the costs of electricity production, the meteorological conditions, etc. We will discuss some realistic modeling of the true losses suffered by the electricity provider in Chapter 5 and provide below two basic examples of loss functions.

Example 1: *Quadratic clipped losses.* Without paying attention to the financial costs caused by poor electricity demand management, we consider here the quadratic difference between the power consumption and its target, which is a smooth function. It is the loss function considered in Chapter 4 and 8. In the present chapter, we only deal with bounded losses, and this is why we clip the quadratic loss as explained below.

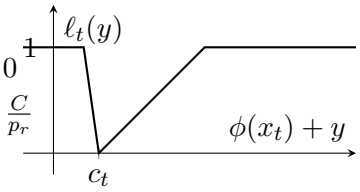
$$\ell_t(y) = \begin{cases} \frac{1}{C^2}\big(\phi(x_t) + y - c_t\big)^2 & \text{if} \quad |f(x_t) + y - c_t| \leqslant C \\ 1 & \text{else.} \end{cases}$$



As we assumed that targets c_t and consumptions with no tariff effects $\phi(x_t)$ were bounded between 0 and C, we always get $(\phi(x_t) - c_t)^2 \leqslant C^2$. So if the effect Y_t of the tariff is not too important, we will generally have that the quadratic difference between the power consumption and its target $(\phi(x_t) + Y_t - c_t)^2$ is equal to $C^2\ell_t(Y_t)$. So most of the time, $C^2\ell_t(y)$ will be equal to $(\phi(x_t) + y - c_t)^2$ and the clipping is merely considered to make sure that the losses lie in $[0, 1]$.

Example 2: *Asymmetric absolute clipped losses.* For the electricity provider, it can be sometimes costlier to under-produce than to over-produce (or vice versa). To model this phenomenon, we could introduce some weights $p_{t,l}$ and $p_{t,r}$ to ponder the losses depending on whether the consumption is above or below the target. We can then define an asymmetric absolute loss function by

$$\ell_t(y) = \begin{cases} p_{t,l}\big(c_t - \phi(x_t) - y\big) & \text{if} \quad -\frac{C}{p_l} \leqslant c_t - \phi(x_t) - y \leqslant 0 \\ p_{t,r}\big(\phi(x_t) + y - c_t\big) & \text{if} \quad 0 \leqslant \phi(x_t) + y - c_t \leqslant \frac{C}{p_r} \\ 1 & \text{else.} \end{cases}$$



We will also consider asymmetric absolute losses in Chapter 5.

We point out that, at each round t, the loss function ℓ_t is known beforehand, it encompasses all dependencies on the context x_t through the consumption level $\phi(x_t)$ but also the target consumption c_t. The only unknowns are the distributions $\nu_1, \ldots, \nu_K$ that model the effect of tariffs and the electricity provider aims to choose tariffs to minimize the cumulative losses

$$L_T \triangleq \sum_{t=1}^{T} \ell_t(Y_t)\,.$$

3.2 Introduction of a pseudo-regret, differences with classical bandit theory and literature discussion

There exists many variants of the multi-armed stochastic bandits problem presented in Section 2.2 and, after presenting it, we point out that our "target-tracking" framework differs from the existent ones.

3.2.1 Bias-variance trade-off and pseudo-regret

Our setting differs from the classical bandits because even if feedbacks are also of the form $Y_t \,|\, I_t = k \sim \nu_k$, our aim is different: we do not want to maximize Y_t but to be as close as possible to a given target, so to minimize $\ell_t(Y_t)$. The best tariff to choose is then

$$k_t^\star \in \operatorname*{argmin}_{k \in \{1,\ldots,K\}} \mathbb{E}_{Y\sim\nu_k}\big[\ell_t(Y)\big]\,.$$

Generally, the expected value of a function of a random variable X is not the value of the function applied to $\mathbb{E}[X]$, namely

$$\mathbb{E}_{Y\sim\nu_k}\big[\ell_t(Y)\big] \quad \neq \quad \ell_t\big(\mathrm{E}(\nu_k)\big)\,.$$

Therefore, it is no longer enough to focus on the expectations of the distributions $\nu_1 \ldots, \nu_K$. Moreover, estimating $\mathbb{E}_{Y\sim\nu_k}\big[\ell_t(Y)\big]$ is expensive since ℓ_t can be complicated and changes over time.

For example, we consider, for any $x \in \mathcal{X}$ and $c \in [0, C]$ the two arms ν_1 and ν_2 associated with the normal distribution $\mathcal{N}(\varphi(x)-c, \text{variance})$ and the Dirac $\delta_{\varphi(x)-c+\text{bias}}$, respectively. If for a round t, we observe $x_t = x$ and $c_t = c$, we get

$$E_{Y\sim\nu_1}\Big[\big(\varphi(x)+Y-c\big)^2\Big] = \text{variance} \quad \text{and} \quad E_{Y\sim\nu_2}\Big[\big(\varphi(x)+Y-c\big)^2\Big] = \text{bias}\,.$$

Thus, without taking into account the clipping between 0 and 1, the quadratic loss function of Example 1 leads to a bias-variance trade-off. Indeed, the best tariff to play is not necessary the one which is the closest in expectation to the target (*i.e.* arm 1) and it may be better to chose a tariff more stable (*i.e.*, arm 2).

From now on, we introduce the expected losses at round t associated with tariff k:

$$\ell_{t,k} \triangleq \mathbb{E}_{Y\sim\nu_k}\big[\,\ell_t\big(Y\big)\,\big]\,.$$

With this notation, at a round t, the best tariff to pick is $k_t^\star \in \operatorname{argmin}_{k\in\{1,\ldots K\}} \ell_{t,k}$ and the expected loss suffered is ℓ_{t,I_t} . We recall that in the classical multi-armed bandit problem, the best arm to play was always $k^\star \in \operatorname{argmax}_{k\in\{1,\ldots,K\}}\mu_k$ and the expected reward was μ_{I_t}. Moreover, we focused on the pseudo-regret

$$\max_{k\in\{1,\ldots K\}} \sum_{t=1}^{T} \mu_k - \sum_{t=1}^{T} \mu_{I_t} = T\mu^\star - \sum_{t=1}^{T} \mu_{I_t}\,.$$

Similarly, we can now define a pseudo-regret as the difference between the cumulative expected loss associated with the chosen strategy and the one for the best possible strategy (namely, the strategy adopted if the distributions $\nu_1, \ldots, \nu_K$ are known):

$$\bar{R}_T = \sum_{t=1}^{T} \ell_{t,I_t} - \sum_{t=1}^{T} \min_{k\in\{1,\ldots K\}} \ell_{t,k}\,.$$

Here, contrary to the classical setting, the minimum over the set of arms $\{1, \ldots K\}$ is inside the sum. We recall that the pseudo-regret is a random variable because the chosen tariffs $I_1, \ldots, I_T$ can be random. Protocol 3 summarizes the online setting described above.

3.2.2 Adversarial bandits

Because of the dependence on ℓ_t, for a chosen tariff k, the losses $\ell_t(Y_t)$ for the instants such that $I_t = k$ (namely, the variables $\big(\ell_{\tau_{k,n}}(Y_{\tau_{k,n}})\big)_n$ with $\tau_{k,n} = \min\{t \geqslant 1,\ N_{t,k} = n\}$) are not i.i.d anymore. As a consequence, we could have considered adversarial bandits: in such a setting, there is no assumption on how the rewards are generated and the environment

Protocol 3 Target Tracking for Multi-Armed Bandits

Input
Number K of available tariffs
Unknown parameters
K probability distributions $\nu_1, \dots, \nu_K$ on $\mathbb{R}$
for $t = 1, 2, \dots$ **do**
Observe a loss function $\ell_t : \mathbb{R} \to [0, 1]$, which does not depend on the past
Choose a tariff $I_t \in \{1, \dots, K\}$
Observe a resulting effect $Y_t \sim \nu_{I_t}$
Suffer a loss $\ell_t(Y_t)$
end for
Aim
Minimize in high probability or in expectation the pseudo-regret

$$\bar{R}_T = \sum_{t=1}^{T} \ell_{t,I_t} - \sum_{t=1}^{T} \min_{k \in \{1,\dots K\}} \ell_{t,k}$$

is called the "adversary". More precisely, at a round t, for each tariff k, the adversary chooses the losses $\ell_{t,k}^{\text{adv}}$ and the gambler suffers the one associated with the tariff she picks: $\ell_{t,I_t}^{\text{adv}}$. The pseudo-regret is then defined as the difference between the expected loss of the gambler strategy and the expected loss of the best constant strategy (namely, a strategy which consists in picking always the same arm):

$$\bar{R}_T^{\text{adv}} = \mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}^{\text{adv}}\right] - \min_{k \in \{1,\dots,K\}} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,k}^{\text{adv}}\right].$$

Note that, for adversarial bandits, the pseudo-regret is not random. Algorithm Exp3 (Exponential weights for Exploration and Exploitation, see Auer et al., 2002b) run for losses that are in $[0, 1]$ ensures a regret bound of $\sqrt{2TK \ln K}$. In our framework, the notations and assumptions introduced above, the adversary would have chosen the losses this way:

$$\ell_{t,I_t}^{\text{adv}} = \ell_t(Y_t) \quad \text{with} \quad Y_t \,|\, I_t \sim \nu_{I_t}\,.$$

and the pseudo-regret would have been

$$\bar{R}_T^{\text{adv}} = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t\big(Y_t\big)\right] - \min_{k \in \{1,\dots,K\}} \mathbb{E}\left[\sum_{t=1}^{T} \ell_t\big(X_{t,k}\big)\right] \quad \text{with} \quad X_{1,k}, \dots, X_{T,k} \sim \nu_k\,.$$

Then, the tower rule would have led to

$$\begin{aligned}\bar{R}_T^{\text{adv}} &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\big[\ell_t(Y_t) \,|\, \mathcal{F}_{t-1}\big]\right] - \min_{k \in \{1,\dots,K\}} \sum_{t=1}^{T} \mathbb{E}_{X \sim \nu_k}\big[\ell_t(X)\big] \\ &= \mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \min_{k \in \{1,\dots,K\}} \sum_{t=1}^{T} \ell_{t,k}\,,\end{aligned}$$

where we recall that $\ell_{t,k} = \mathbb{E}_{Y \sim \nu_k}[\ell_t(Y)]$. In our adaptation of the stochastic bandit framework, the evaluated strategies are compared to non-constant strategies and the random

pseudo-regret is

$$\bar{R}_T = \sum_{t=1}^{T} \ell_{t,I_t} - \sum_{t=1}^{T} \min_{k\in\{1,\dots,K\}} \ell_{t,k}\,.$$

Therefore, the expectation of pseudo-regret for stochastic bandits satisfies

$$\mathbb{E}\big[\bar{R}_T\big] \geqslant \bar{R}_T^{\text{adv}}\,,$$

and the two pseudo-regrets are equal in expectation only if the best tariff to pick does not vary over time (namely if the best strategy is constant). To bound $\bar{R}_T$ ensures a better control of losses and our framework is thus better adapted to stochastic bandit than to adversarial bandits. Indeed, in the next section, under the assumption on how the effects Y_t are generated and by knowing the loss functions ℓ_t, we see how to obtain some bound in high probability and in expectation on the pseudo-regret $\bar{R}_T$ of order $\sqrt{T\ln T}$.

3.2.3 Contextual bandits

We highlight that in stochastic contextual bandit theory (we further discuss these bandit settings in Chapter 4), the best arm also varies over time. Here, we recall that it depends on the loss function ℓ_t (so on the target, the contextual variables etc.). But, to the best of our knowledge, the aim is always to maximize the cumulative reward or to find the best arm to play for a given time budget (which is a different problem called best arm identification, see among others Audibert and Bubeck [2010]).

The framework closest to ours would therefore be that of the contextual bandits, in which the gambler aims to maximize its reward X_t drawn from distributions that may vary with time. More formally, a contextual bandit protocol generally follows the following procedure (see, among others Perchet and Rigollet, 2013). At a round $t \geqslant 1$, the gambler observes a context $z_t \in \mathcal{C}$, where $\mathcal{C}$ is some context space. Then, she chooses an arm $I_t \in \{1,\dots,K\}$ and gets the reward

$$X_t \,|\, \{I_t = k\} \sim \nu_k(z_t)\,.$$

For each arm k, the distribution ν_k now depends on the context z_t, so, given z_t, the best arm to play is:

$$k^\star(z_t) \in \operatorname*{argmax}_{k\in\{1,\dots,K\}} \mathrm{E}\left(\nu_k(z_t)\right).$$

Depending on how the distributions evolve, so under suitable conditions on $\mathrm{E}(\nu_k(z_t))$, some regret bounds have already been obtained. For example, linear contextual bandits assume that, for each arm k, there is an unknown vector θ_k such that

$$\mathrm{E}\left(\nu_k(z_t)\right) \triangleq \mu_k(z_t) = z_t^{\mathrm{T}}\theta_k\,.$$

Under such assumptions, the LinUCB algorithm proposed by Li et al. [2010] offered $\sqrt{T}$ upper-bounds on its pseudo-regret (up to polylogarithmic terms), this framework is further discuss in Chapter 4. Bandit problems under less restrictive assumptions on $\mu_k(z_t)$ have also been studied, see among others Valko et al. [2013] and Foster et al. [2018].

An optimal strategy maps each context z_t to an optimal arm $k^\star(z_t)$. To learn one, we will therefore have to estimate for each z_t, and each $k \in \{1, \dots, K\}$, the expectation $\mathrm{E}\left(\nu_k(z_t)\right) = \mu_k(z_t)$. For a finite number of contexts $\mathcal{C} = \{z^1, \dots, z^{|\mathcal{C}|}\}$, by playing the UCB algorithm (or another bandit algorithm) per context, we may obtain some regret bounds of $\sqrt{T}$ order with a multiplicative constant which depends on the number of contexts $|\mathcal{C}|$. Indeed, if we assume that for a fixed context z^i, for any $T \geqslant 1$, the pseudo-regret associated this context z^i satisfies

$$\bar{R}_T^i = \sum_{t=1}^{T} \mu_{I_t}(z_i) - \min_{k \in \{1,\dots,K\}} T\mu_k(z^i) \leqslant \square \sqrt{KT \ln T}\,,$$

where $\square$ is a constant; by denoting by $T(i)$ the number of times the context z^i have been picked by the environment, we get that

$$\begin{aligned}\bar{R}_T &= \sum_{t=1}^{T} \max_{k \in \{1,\dots,K\}} \mu_k(z_t) - \sum_{t=1}^{T} \mu_{I_t}(z_t) = \sum_{i \in \mathcal{C}} \left(\max_{k \in \{1,\dots,K\}} \sum_{t=1}^{T} \mu_k(z^i) - \sum_{t=1}^{T} \mu_{I_t}(z^i) \right) \mathbb{1}_{\{z_t = z^i\}} \\ &= \sum_{i \in \mathcal{C}} \left(\max_{k \in \{1,\dots,K\}} T(i)\mu_k(z^i) - \sum_{t \,|\, z_t = z^i} \mu_{I_t}(z^i) \right) = \sum_{i \in \mathcal{C}} \bar{R}_{T_i}^i \leqslant \square \sum_{i \in \mathcal{C}} \sqrt{KT(i) \ln T(i)}\,,\end{aligned}$$

with $\sum_{i \in \mathcal{C}} T(i) = T$. By using Jensen's inequality on the squared root function, which is concave, we then obtain

$$\bar{R}_T \leqslant \square \sqrt{K|\mathcal{C}|T \ln T}\,.$$

Imposing a finite number of context is quite restrictive and the environment should generally be able to pick a context z_t in a larger space. Another possibility is the discretization of set of contexts. Under some assumption on the regularity (Lipschitz, Hölder etc.) of the functions $z \to \mu_k(z)$, a grid $\{z^1, \dots, z^{|\mathcal{C}|}\}$ of $|\mathcal{C}|$ contexts may be considered. It has to be fine enough to ensure that, no matter how the experiment picks the context z_t, there exists $i \in \{1, \dots, |\mathcal{C}|\}$, such that, for any arm k,

$$|\mu_k(z^i) - \mu_k(z_t)\,| \leqslant \varepsilon\,,$$

where ε is an approximation error. Then, by summing over t and using the bound obtained of a finite number of contexts, we get

$$\bar{R}_T = \varepsilon T + \sum_{i \in \mathcal{C}} \bar{R}_T^i \leqslant \varepsilon T + \square \sqrt{K|\mathcal{C}|T \ln T}\,.$$

To obtain a sub-linear bound, the approximation error ε has to be of order $1/T^\alpha$, with $\alpha > 0$, which requires a very fine grid and therefore a large number of contexts $|\mathcal{C}|$ (generally of order $1/\varepsilon^\beta \sim T^{\alpha\beta}$, with $\beta > 0$).

In our framework, we recall that, when a tariff k is picked, we observe the tariff effect Y_t, with $Y_t \,|\, I_t = k \sim \nu_k$ and suffer the loss $\ell_t(Y_t) \in [0, 1]$, or, in other works, we observe the reward

$$X_t = 1 - \ell_t(Y_t), \quad \text{with} \quad 1 - \ell_t(Y_t) \,|\, \ell_t,\, I_t = k \;\sim\; \nu_k(\ell_t)\,.$$

Therefore, in a contextual bandit setting, the context (previously denoted by z_t) refers to the loss function ℓ_t. With Example 1, the two variables x_t and c_t are enough to define ℓ_t and we could consider $z_t = (x_t, c_t)$. More generally, z_t may contain exogenous context

variables x_t, the target consumption c_t and some additional information on the type of loss (quadratic, absolute, etc.) considered at round t. With our notation, the loss function ℓ_t gathers all this information. We emphasize that an optimal strategy maps each loss function ℓ_t to an optimal arm $k^\star(\ell_t)$ and with the notations introduced above,

$$\ell_{t,k} = \mathbb{E}_{Y\sim\nu_k}[\ell_t(Y)] = \mathrm{E}\big(\nu_k(\ell_t) = \big)1 - \mu_k(z_t)\,, \quad \text{where} \quad z_t = \ell_t.$$

Therefore, for a finite number of contexts $\mathcal{L} = \{\ell^1, \ldots, \ell^{|\mathcal{L}|}\}$, with for example the loss functions $\ell^i = \max\{(\varphi(x^i) + y - c^i)/C^2, 1\}$, we could get that

$$\bar{R}_T \leqslant \square\sqrt{K|\mathcal{L}|T\ln T}\,.$$

Imposing a finite number of context is too restrictive for our framework and the discretization of set of contexts, namely of the set of loss functions requires a very fine grid to get a sub-linear bound. It is likely that the order of the obtained regret bound is worst than $\sqrt{T}$ and computationally, the discretization is not efficient (we recall that the algorithm requires K exploration rounds in each context, namely for each loss function). In such a framework, it seems therefore unreasonable to consider an infinite number of types of loss (like asymmetric absolute with weights changing at each round, see Example 2). Such a solution is thus not conceivable, both theoretically and practically.

Another possibility would be to make some strong assumptions (like a linear dependence) on the way the distributions ν_k evolve with the loss functions ℓ_t, but this would completely restrict our framework.

Remark 2. *This contextual point of view can be adapted to the framework of the adversarial bandits by considering $|\mathcal{C}|$ Exp3 algorithms (one per context) or the Exp4 algorithm introduced by Auer et al. [2002b], see Section 4.2 of Bubeck and Cesa-Bianchi [2012] for further details. The obtained bound will be of order $\sqrt{T|\mathcal{C}|K\ln K}$, as for the stochastic framework with a finite number of contexts (or of loss functions $\ell^1, \ldots, \ell^{|\mathcal{L}|}$). As in the stochastic version of contextual bandits, we could deal with an infinite number of contexts using discretization approaches. Quickly, these solutions would become computationally inefficient and would provide regret bounds of order high than $\sqrt{T}$.*

In what follows, we show how to use both knowledge of ℓ_t and stochastic assumption $Y_t \,|\, I_t = k \sim \nu_k$ (which will be $Y_t \,|\, I_t = k \sim \nu_{k,t}$ in Chapter 4) in a way to provide some regret bounds of the form $\square\sqrt{T}$, where $\square$ depends on K and $\ln T$ and is a small compared to the bound obtain for contextual bandit solutions. The considered algorithm is easily implemented and has a computational cost much lower than those related to contextual bandits.

3.2.4 Multi-arm bandits to electricity demand management: literature discussion

Recently, even at the same time as the thesis work, bandit approaches have been developed for demand side management. We discuss here the work we are aware of and point out the differences with our frameworks.

Wang et al. [2014] consider a system that consists of K electric loads (or arms) that can be deployed by an aggregator. The dynamics of a load is described by a pair of two-state Markov chains, one characterizing state transitions when the load is deployed and one characterizing state transitions when the load is not deployed. In each case, the load may be in one of two states, depending on whether it is available for consumption curtailment or not. Therefore, when a load is deployed (or chosen) and if it is available, it curtails the electrical consumption. At each time step t, the aggregator chooses up to N loads to participate in demand response, where N is the "budget". The objective of the aggregator is to maximize the expected consumption curtailment, subjected to the contraint that the number of chosen loads at t equals N. They propose an ε-greedy algorithm (an algorithm that explores with probalility ε and exploits with probability $1-\varepsilon$) and show numerically that its regret is of logarithmic order in time, and outperforms a UCB-inspired algorithm. Our approach differs from that of Wang et al. [2014]. Indeed, the arm of their bandit problem (which correspond to electric loads) are defined by Markov chains with unknown transition probabilities whereas they correspond to tariffs and are defined by unknown probability distribution in our framework. But the main difference is that we do not aim at minimizing feedbacks (namely, reducing consumption) but at minimizing losses (which are a function of feedbacks and a target).

While we were developing our approach, Moradipari et al. [2018] proposed a framework quite similar to ours: At each time step, they observe a target and choose tariffs in order to minimize a function of the electrical demand and the target); but they propose a Thompson Sampling based algorithm to minimize the regret. They provide a discussion on regret bounds for their algorithm. We point out that their approach does not model the dependencies between power consumption and some contextual variables, such as temperature (although this will be the case in our framework presented in Chapter 4.

Very recently, Li et al. [2020] modeled the demand reduction of customers with a Bernoulli distributions. They thus consider a simple customer behavior model, where each customer i may either respond to a demand respons event by reducing one unit of power consumption with probability $0 \leqslant p_i \leqslant 1$, or not respond with probability $1-p_i$. At each time t, there is a demand response event with a nonnegative demand reduction target determined by the power system. The aggregator aims to select a subset of customers, such that the total demand reduction is as close to the target as possible. They consider quadratic losses and to learn and select the right users, they propose a bandit approach and a UCB-inspired algorithm: Combinatorial Upper Confidence Bound-Average. They consider both a fixed time-invariant target and time-varying targets, and show that their algorithm achieves $\mathcal{O}(\ln T)$ and $\mathcal{O}(\sqrt{T}\ln T)$ regrets respectively. Finally, Chen et al. [2020] complicate the customer behavior model by including the influences of environmental factors (temperatures, changes in lifestyles, etc.) and propose and algorithm based on Thompson Sampling. Numerical simulations are performed to demonstrate its learning effectiveness.

4 Upper confidence bound algorithm for target tracking

We propose an adaptation of the UCB algorithm recalled above to track some moving targets (defined throughout the loss functions ℓ_t) and minimize the pseudo-regret introduced in Protocol 3. We consider a δ-risk level optimistic algorithm in the next section and we

provide a pseudo-regret analysis based on the proofs of Section 2.2. Finally, we present a moving risk level version of this algorithm, which defines an anytime strategy and ensures a bound on the expectation of the pseudo-regret.

5 Upper confidence bound algorithm for target tracking

We propose an adaptation of the UCB algorithm recalled above to track some moving targets (defined throughout the loss functions ℓ_t) and minimize the pseudo-regret introduced in Protocol 3. We consider a δ-risk level optimistic algorithm in the next section and we provide a pseudo-regret analysis based on the proofs of Section 2.2. Finally, we present a moving risk level version of this algorithm, which defines an anytime strategy and ensures a bound on the expectation of the pseudo-regret.

5.1 Optimistic algorithm

We emphasize that, in general,

$$\mathbb{E}_{Y\sim\nu_k}\big[\ell_t\big(Y\big)\big] \neq \ell_t\big(\mathbb{E}_{Y\sim\nu_k}[Y]\big)\,,$$

so it is not useful to estimate the means of the laws $(\nu_k)_{k=1,\dots,K}$ as it was done in the UCB algorithm for the multi-armed bandit problem. Instead, to hope to choose the best tariff at round t, we have to compute, for each tariff k, an estimation of the expected loss $\ell_{t,k}$ (and not an estimation of the expected effect). In the same way that we estimated the mean $\mu_k = \mathrm{E}(\nu_k)$ with

$$\widehat{\mu}_{t-1,k} = \frac{1}{N_{t-1,k}} \sum_{s=1}^{t-1} Y_s \mathbb{1}_{\{I_s=k\}}\,,$$

for a fixed function ℓ, we can estimate the expectation $\ell_k = \mathbb{E}_{X\sim\nu_k}[\ell(X)]$ with

$$\widehat{\ell}_k = \frac{1}{N_{t-1,k}} \sum_{s=1}^{t-1} \ell(Y_s) \mathbb{1}_{\{I_s=k\}}\,.$$

In particular, for a round t, by taking $\ell = \ell_t$, depending on which tariff I_s have been picked, we average the losses $\ell_t(Y_s)$, where the indexes s and t may be different. Therefore, for $k \in \{1,\dots,K\}$, we compute the empirical mean of the loss associated with tariff k (the increase in the algorithm complexity is discussed in the remark below):

$$\widehat{\ell}_{t,k} \triangleq \frac{1}{N_{t-1,k}} \sum_{s=1}^{t-1} \ell_t\big(Y_s\big) \mathbb{1}_{\{I_s=k\}} \quad \text{with} \quad N_{t-1,k} = \sum_{s=1}^{t-1} \mathbb{1}_{\{I_s=k\}}\,.$$

Moreover, it should be noted, that, because there is no control on the functions ℓ_t, it would not have been possible to just use $\ell_t(Y_t)$, with the same indexes. Indeed, the functions ℓ_t may induce some dependencies between the losses. For example, the environment could choose $\ell_1 = \cdots = \ell_{\lfloor T/2\rfloor} : x \mapsto 0$ and $\ell_{\lfloor T/2\rfloor+1} = \cdots = \ell_T : x \mapsto 1$; in this case for a fixed tariff k, when it is picked, the losses $\ell_t(Y_t)$, for $I_t = k$ (namely, for $n = 1, 2, \dots,$ the variables $\ell_{\tau_{k,n}}(Y_{\tau_{k,n}})$ with $\tau_{k,n} = \min\{t \geqslant 1,\, N_{t,k} = n\}$) are not identically distributed (whereas, by using Doob's optional skipping, for $n = 1, 2, \dots,$ the variables $Y_{\tau_{k,n}}$ and $\ell_t(Y_{\tau_{k,n}})$ are i.i.d, by assuming ℓ_t deterministic).

Once the expected losses have estimated, the optimistic algorithm, stated in Algorithm 3, considers the same confidence levels as Algorithm 2 and after K first deterministic rounds $I_t = t$, it chooses at each round $t \geqslant K+1$ the tariff

$$I_t \in \operatorname*{argmin}_{k\in\{1,\dots,K\}} \widehat{\ell}_{t,k} - \alpha_{t,k} \quad \text{with} \quad \alpha_{t,k} = \frac{\ln(t^3/\delta)}{2N_{t-1,k}}\,.$$

Remark 3. *The main difference between Algorithm 3 and Algorithm 2 is in the estimation of the losses: it has to be updated at each time step, depending on the loss function ℓ_t and computationally, this has a cost. Previously, at a round t, for a new observation Y_t, only the empirical mean of the reward associated with arm I_t was updated, with*

$$\widehat{\mu}_{t,I_t} = \frac{1}{N_{t,I_t}}\big(N_{t-1,I_t}\widehat{\mu}_{t,I_t} + Y_t\big) \quad \text{and} \quad \forall\, k \neq I_t,\ \widehat{\mu}_{t,k} = \widehat{\mu}_{t-1,k}\,.$$

Therefore three basic operations were done: one multiplication, one addition, one division; and for the round $t+1$, only the $2K$ variables $\widehat{\mu}_{t,k}$ and $N_{t,k}$ were needed. While, with the new setting, at a round t, the empirical mean of the loss associated with tariff I_t is

$$\widehat{\ell}_{t+1,k} = \frac{1}{N_{t,k}}\sum_{s=1}^{t} \ell_{t+1}(Y_s)\mathbb{1}_{\{I_s=k\}}\,.$$

So all the variables $I_1, Y_1, \dots I_t, Y_t$ have to be stored (namely, $2T$ quantities) and t evaluations of the loss function ℓ_t, t additions and K divisions have to be done to provide $\widehat{\ell}_{t+1,k}$ for each tariff.

5.2 Distribution-free analysis of the pseudo-regret

Theorem 3. *For all distributions $\nu_1, \dots, \nu_K$ with supports in $\mathbb{R}$, for any $T \geqslant 1$ and any sequence of loss functions with supports in $[0,1]$, $\ell_1, \ell_2, \dots, \ell_T$, the pseudo-regret of Algorithm 3 satisfies*

$$\bar{R}_T \triangleq \sum_{t=1}^{T} \ell_{t,I_t} - \sum_{t=1}^{T} \min_{k\in\{1,\dots K\}} \ell_{t,k} \leqslant K + 2\sqrt{2KT\ln\left(T^3/\delta\right)}\,,$$

with probability at least $1-\delta$.

Remark 4. *Even if the best tariff may change at each round, we obtain a bound of the same order $\mathcal{O}(\sqrt{T\ln T})$ as the classical distribution-free one for the UCB algorithm.*

Proof of Theorem 3. The analysis of the pseudo-regret is very similar to the proofs of Theorems 1 and 2. In the classical setting, a distribution-dependent bound was firstly obtained and then, the distribution-free bound was deducted. As the best tariff to play may change at each step, such a trick is not possible here. Indeed, previously the regret bound depended on the gaps $\Delta_k = \mu_\star - \mu_k$. Now, for a round t and a tariff k, we introduce the time dependent gap

$$\Delta_k^t \triangleq \ell_{t,k} - \ell_{k_t^\star,t} \quad \text{with} \quad k_t^\star \in \operatorname*{Argmin}_k \ell_{t,k}\,,$$

so $k_t^\star$ denote a best tariff to play. Therefore, it seems complex to consider a distribution-dependent analysis. However, by using the same type of concentration inequalities, we

Algorithm 3 δ-Tracking Upper Confidence Bound Algorithm

Unknown parameters
K probability distributions $\nu_1, \dots, \nu_K \in \mathbb{R}$
Input
risk level $\delta \in (0,1)$
initialization
for each tariff the counter $N_{k,0} = 0$
for $t = 1, \dots, T$ **do**
  if $t \leqslant K$ **then**
    $I_t = t$
  else
    Observe the loss function $\ell_t : \mathbb{R} \to [0,1]$
    Compute for each tariff k the empirical loss $\widehat{\ell}_{t,k} = \dfrac{1}{N_{t-1,k}} \displaystyle\sum_{s=1}^{t-1} \ell_t\big(Y_s\big) \mathbb{1}_{\{I_s = k\}}$
    Choose optimistically the next tariff

$$I_t \in \operatorname*{Argmin}_{k \in \{1,\dots,K\}} \ \widehat{\ell}_{t,k} - \sqrt{\frac{\ln(t^3/\delta)}{2N_{t-1,k}}}$$

  end if
  Observe $Y_t \in \mathbb{R}$
  Update for each tariff the counter $N_{t,k} = N_{t-1,k} + \mathbb{1}_{\{I_t = k\}}$
end for

manage to obtain a distribution-free regret bound. Indeed, a straightforward application of Lemma 1 on the estimations of the losses, computed with the variables $\ell_t(Y_1), \dots, \ell_t(Y_{t-1})$ (see Remark 6 in Appendix for further details), gives that for all $k \in \{1, \dots, K\}$, all $t \geqslant K$ (so $N_{t,k} \geqslant 1$), and all $\delta \in (0,1)$,

$$\mathbb{P}\left(\ell_{t,k} > \widehat{\ell}_{t,k} - \sqrt{\frac{\ln \frac{1}{\delta}}{2N_{t,k}}}\right) \leqslant t\delta \quad \text{and} \quad \mathbb{P}\left(\widehat{\ell}_{t,k} < \ell_{t,k} + \sqrt{\frac{\ln \frac{1}{\delta}}{2N_{t,k}}}\right) \leqslant t\delta. \tag{2.3}$$

The proof is broken down into three steps: as in the proof of Theorems 1, Step 1 provides the necessary conditions for playing a sub-optimal tariff; then the pseudo-regret associated with each tariff is bounded in Step 2; Step 3 concludes the proof. The main difference with the previous analysis is in Step 2.

$\star$ *Step 1: Reasons to play a sub-optimal tariff.* At a round $t \geqslant K+1$, the instantaneous pseudo-regret $r_t = \ell_{t,I_t} - \ell_{t,k_t^\star}$ (bounded by 1) can then be rewritten $r_t = \Delta^t_{I_t}$. Note that if $I_t = k$ one of these three inequalities is satisfied:

(i)	$\ell_{t,k} > \widehat{\ell}_{t,k} + \alpha_{t,k}$	(the loss associated with k is underestimated)
(ii)	$\ell_{t,k_t^\star} < \widehat{\ell}_{t,k_t^\star} - \alpha_{t,k_t^\star}$	(the loss associated with $k_t^\star$ is overestimated)
(iii)	$\Delta^t_k \leqslant 2\alpha_{t,k}$	(tariff k has not been played enough)

Indeed, with $I_t = k$, Algorithm (3) ensures

$$\widehat{\ell}_{t,k} - \alpha_{t,k} \leqslant \widehat{\ell}_{k_t^\star,t} - \alpha_{t,k_t^\star} \quad \Rightarrow \quad \widehat{\ell}_{t,k} - \widehat{\ell}_{k_t^\star,t} \leqslant \alpha_{t,k} - \alpha_{t,k_t^\star}\,.$$

If (i) and (ii) are not satisfied, then

$$\Delta_k^t = \ell_{t,k} - \ell_{k_t^\star,t} \overset{\text{(i) and (ii)}}{\leqslant} \widehat{\ell}_{t,k} + \alpha_{t,k} - \widehat{\ell}_{k_t^\star,t} + \alpha_{t,k_t^\star} \overset{\text{(Algorithm 3)}}{\leqslant} 2\alpha_{t,k}\,.$$

As in the proof of Theorem 2, we introduce the following event, which gathers events (i) and (ii) for all rounds $t \geqslant K+1$:

$$\mathcal{E} \triangleq \left\{ \forall\, t \geqslant K+1,\, \forall\, k \,|\, \Delta_k^t > 0 : \widehat{\ell}_{t,k} \leqslant \ell_{t,k} + \alpha_{t,k} \quad \text{and} \quad \widehat{\ell}_{t,k_t^\star} \geqslant \ell_{t,k_t^\star} - \alpha_{t,k_t^\star} \right\}. \tag{2.4}$$

Moreover, by a union bound and by using inequalities (2.3) we get, in a similar manner, that event $\mathcal{E}$ holds with probability as least $1-\delta$. From now on, we consider event $\mathcal{E}$, so for any round t, a sub-optimal tariff k is played only if $\Delta_k^t \leqslant 2\alpha_{t,k}$.

$\star$ *Step 2: Bounds on the contribution of tariff k in the pseudo-regret under event $\mathcal{E}$.* With the notation introduced above, notice that the pseudo-regret equals

$$\bar{R}_T = \sum_{k=1}^{K} \sum_{t=1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k\}} = \sum_{k=1}^{K} \bar{R}_{T,k}\,, \quad \text{with} \quad \bar{R}_{T,k} \triangleq \sum_{t=1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k\}}\,.$$

We point out that for the classical setting, the gaps Δ_k^t did not depend on t so the contribution of arm k to the pseudo-regret was $\bar{R}_{T,k} = \Delta_k N_{T,k}$ and we just had to bound the integers $N_{T,k}$. Here, we focus, for $k \in \{1,\dots,K\}$, on $\bar{R}_{T,k}$, the contribution of tariff k to the pseudo-regret. As the strategy on the first K steps is deterministic with $I_t = t$, the tariff k is played exactly once between step 1 and step K and depending on which of the events the event (i), (ii) or (iii) occurs, we have

$$\begin{aligned}\bar{R}_{T,k} \leqslant 1 &+ \sum_{t=K+1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k \text{ and (i)}\}} \\ &+ \sum_{t=K+1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k \text{ and (ii)}\}} + \sum_{t=K+1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k \text{ and (iii)}\}}\,.\end{aligned} \tag{2.5}$$

In the classical bandit framework, we directly upper-bounded $\Delta_k \sum_{t=1}^{T} \mathbb{1}_{\{I_t=k \text{ and (iii)}\}}$ using that all $N_{t,k}$ were bounded by $8\ln T/\Delta_k^2$. Here, the gaps are not anymore constant, but by noticing that

$$\Delta_k^t\, \mathbb{1}_{\{I_t=k \text{ and (iii)}\}} = \Delta_k^t\, \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\alpha_{t,k}\}} \leqslant 2\alpha_{t,k}\, \mathbb{1}_{\{I_t=k\}}\,,$$

the last term of Equation (2.5) satisfies

$$\begin{aligned}\sum_{t=K+1}^{T} \Delta_k^t\, \mathbb{1}_{\{I_t=k \text{ and (iii)}\}} &= \sum_{t=K+1}^{T} \Delta_k^t\, \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\alpha_{t,k}\}} \\ &\leqslant \sum_{t=K+1}^{T} 2\alpha_{t,k} \mathbb{1}_{\{I_t=k\}} = \sum_{t=K+1}^{T} 2\sqrt{\frac{\ln(t^3/\delta)}{2N_{t-1,k}}} \mathbb{1}_{\{I_t=k\}} \\ &\leqslant \sqrt{2\ln(T^3/\delta)} \sum_{t=K+1}^{T} \frac{1}{\sqrt{N_{t-1,k}}} \mathbb{1}_{\{I_t=k\}}\,,\end{aligned}$$

where we simply upper-bound t by T to get the last inequality. As the integer $N_{t-1,k}$ increases by 1 each time tariff k is played, we have

$$\sum_{t=K+1}^{T} \frac{1}{\sqrt{N_{t-1,k}}} \mathbb{1}_{\{I_t=k\}} = \sum_{n=1}^{N_{T-1,k}} \frac{1}{\sqrt{n}} = 1 + \sum_{n=2}^{N_{T-1,k}} \frac{1}{\sqrt{n}} \,.$$

By using the integral test for convergence, we may upper bound the sum by an integral as follows: for any $N \geqslant 2$,

$$\sum_{n=2}^{N} 1/\sqrt{n} \leqslant \int_1^N \frac{1}{\sqrt{x}} \mathrm{d}x = \big[2\sqrt{x}\big]_1^N = 2\sqrt{N} - 2 \,,$$

therefore, we get

$$\sum_{n=1}^{N_{T-1,k}} 1/\sqrt{n} \leqslant 1 + 2\sqrt{N_{T-1,k}} - 2 \leqslant 2\sqrt{N_{T-1,k}} \,.$$

By combining all the inequalities above, under the event $\mathcal{E}$, only events (iii) hold. Thus, with the decomposition stated in Equation (2.5), for all $k = 1, \ldots, K$, the contribution of tariffs k in the pseudo-regret is bounded by

$$\bar{R}_{T,k} \leqslant 1 + 2\sqrt{2\ln(T^3/\delta)}\sqrt{N_{T-1,k}} \,.$$

⋆ *Step 3: Regret bound with high probability.* With $\sum_{k=1}^{K} N_{T-1,k} = T - 1$, by applying Jensen's inequality, we get

$$\sum_{k=1}^{K} \frac{1}{K}\sqrt{N_{T-1,k}} \leqslant \sqrt{\sum_{k=1}^{K} \frac{N_{T-1,k}}{K}} = \sqrt{\frac{T-1}{K}} \leqslant \sqrt{T/K} \,.$$

By summing the contribution of each tariff, if the event $\mathcal{E}$ holds, the pseudo-regret is bounded by:

$$\bar{R}_T = \sum_{k=1}^{K} \bar{R}_{T,k} \leqslant K + 2\sqrt{2KT\ln(T^3/\delta)} \,.$$

We recall that event $\mathcal{E}$ holds with probability at least $1 - \delta$, which concludes the proof. □

To obtain a bound on the expectation of the pseudo-regret, we may as in Corollary 2, take $\delta = 1/T$ and get

$$\mathbb{E}\big[\bar{R}_T\big] = 1 + K + 4\sqrt{2KT\ln T} \,.$$

If the time horizon T is unknown, it is still possible to bound the pseudo-regret's expectation by considering a moving risk level $\delta_t = 1/t$, at each time step t (similarly, Algorithm 1 was a $\delta_t = 1/t$-version of Algorithm 2). So, for each round $t \geqslant K + 1$, we can consider the optimistic algorithm which picks:

$$I_t \in \operatorname*{argmin}_{k\in\{1,\ldots,K\}} \widehat{\ell}_{t,k} - \alpha_{t,k} \quad \text{with} \quad \alpha_{t,k} = \sqrt{\frac{2\ln t}{N_{t-1,k}}} \,, \tag{2.6}$$

and which, for the first K rounds, chooses $I_t = t$. Corollary 4 below states a regret bound in expectation for this moving risk-level algorithm.

Corollary 4. *For all distributions $\nu_1, \dots, \nu_K$ with supports in $\mathbb{R}$, for any $T \geqslant K$ and any sequence of loss functions with supports in $[0,1]$, $\ell_1, \ell_2, \dots, \ell_T$, the expectation of the pseudo-regret of Algorithm 2.6 satisfies*

$$\mathbb{E}\big[\bar{R}_T\big] \leqslant 3K + 4\sqrt{2KT\ln T}\,.$$

Proof of Corollary 4. The proof looks like the one of Theorem 3, the only difference is that, there is no need to introduce an event $\mathcal{E}$. By replacing the confidence levels $\alpha_{t,k}$ by $\sqrt{2\ln t/N_{t-1,k}}$, for any $t \geqslant 1$, as $\Delta_k^t \leqslant 1$, using the linearity of the expectation, Equation 2.5, which still holds, leads to:

$$\mathbb{E}\big[\bar{R}_{T,k}\big] \leqslant 1 + \sum_{t=K+1}^{T} \mathbb{P}[\text{(i)}] + \mathbb{P}[\text{(ii)}] + \mathbb{E}\Big[\Delta_k^t \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\sqrt{2\ln t/N_{t-1,k}}\}}\Big]\,. \tag{2.7}$$

By applying Lemma 1 for each round t between $K+1$ and T, with $\delta = t^{-4}$, we obtain (see Step 2 of the proof of Theorem 1 for further details)

$$\sum_{t=K+1}^{T} \mathbb{P}[\text{(i)}] + \mathbb{P}[\text{(ii)}] \leqslant \sum_{t=K+1}^{T} 2t^{-3} \leqslant 2\,. \tag{2.8}$$

Then, exactly as in Step 2 of the proof of Theorem 3, by using that

$$\Delta_k^t \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\alpha_{t,k}\}} \leqslant 2\alpha_{t,k}\mathbb{1}_{\{I_t=k\}}\,,$$

with $\alpha_{t,k} = \sqrt{2\ln t/N_{t-1,k}}$, we get that

$$\begin{aligned}\sum_{t=K+1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\sqrt{2\ln t/N_{t,k}}\}} &\leqslant \sum_{t=K+1}^{T} 2\sqrt{\frac{2\ln t}{N_{t,k}}}\mathbb{1}_{\{I_t=k\}} \\ &\leqslant 2\sqrt{2\ln T}\sum_{t=K+1}^{T}\sqrt{\frac{1}{N_{t,k}}}\mathbb{1}_{\{I_t=k\}} = 2\sqrt{2\ln T}\sum_{n=1}^{N_{T,k}}\frac{1}{\sqrt{n}}\,.\end{aligned}$$

Then, the integral test for convergence gives

$$\sum_{t=K+1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\sqrt{2\ln t/N_{t-1,k}}\}} \leqslant 4\sqrt{2N_{T-1,k}\ln T}\,.$$

By summing over the tariffs, Jensen's inequality (see Step 3 of the proof of Theorem 3) leads to the deterministic bound:

$$\sum_{k=1}^{K}\sum_{t=K+1}^{T} \Delta_k^t \mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\sqrt{2\ln t/N_{t,k}}\}} \leqslant 4\sqrt{2KT\ln T}\,. \tag{2.9}$$

Finally, by injecting Equations (2.8) and (2.9) into Equation (2.7), we bound the expectation of the pseudo-regret of Algorithm (2.6) by

$$\begin{aligned}\mathbb{E}\big[\bar{R}_T\big] = \sum_{k=1}^{K}\mathbb{E}\big[\bar{R}_{T,k}\big] &\leqslant \sum_{k=1}^{K}\left(1 + 2 + \mathbb{E}\Big[\sum_{t=K+1}^{T}\Delta_k^t\mathbb{1}_{\{I_t=k \text{ and } \Delta_k^t \leqslant 2\sqrt{2\ln t/N_{t-1,k}}\}}\Big]\right) \\ &\leqslant 3K + 4\sqrt{2KT\ln T}\,,\end{aligned}$$

which concludes the proof. □

Remark 5. *These analyses also hold when the gaps are constant over t; but, within a multiplicative constant, the previous distribution-free regret bounds are better. For example, for Algorithm 1, the first and well-known analysis provided a distribution-free regret bound equal to*

$$\sqrt{KT(2+8\ln T)} \overset{T\to\infty}{\sim} 2\sqrt{2KT\ln T}\,.$$

While, with this new analysis, we obtain the bound $3K + 4\sqrt{2KT\ln T} \overset{T\to\infty}{\sim} 4\sqrt{2KT\ln T}$.

6 Perspectives

We point out that we develop this target tracking with bandit framework for demand side management. However, it may be possible to consider other applications. Indeed, in a decision process, as soon as the aim is not to maximize the reward but to get close to a target (by introducing some loss functions ℓ_t that are known and may change throughout the rounds), the propose method can be considered.

The main drawbacks of the modeling above are the assumptions made on the power consumption. Indeed, we assumed the effects of the tariff Y_t to be independent on the contextual variables, which is not really suitable. For example, a change in the electricity price during the night may have less impact than a change during a peak-hour. The temperature may also affect the tariff effects: if people are thermo-sensitive, they should be unlikely to reduce their consumption when they need to heat their households. Moreover, we made the strong assumption that we knew the function ϕ, namely the part of the consumption that did not depend on the tariff. In a more realistic approach of demand side management, the effect of every exogenous variables (temperature, days, tariffs, etc.) on the expected power consumption should to be learn by the electricity provider (namely, without any prior knowledge on how people consume). Therefore, in Chapter 4, the bandit feedback considered will be the power consumptions (and not any more the tariff effects). Furthermore, for a chose tariff $I_t = k$, they will be of the form $Y_t = \phi(x_t, k) + \text{noise}$, where the function ϕ is unknown.

Appendix

Proof of Azuma-Hoeffding inequality with a random number of summands

Here we prove Lemma 1 used in Step 2 of the proof of Theorem 1, to bound the probability that event (i) or (ii) occurs. It is based on Hoeffding's lemma, which is recalled in Lemma 2 below and proved at the end of the section.

Lemma 2 (Conditional Hoeffding's lemma)**.** *Let X be a random variable such that, almost surely, $X \in [a,b]$, then for any σ-algebra $\mathcal{G}$,*

$$\forall x \in \mathbb{R}, \quad \mathbb{E}\Big[\mathrm{e}^{x(X-\mathbb{E}[X|\mathcal{G}])}\Big|\mathcal{G}\Big] \leqslant \mathrm{e}^{\frac{x^2}{8}(b-a)^2}.$$

In the proof of Lemma 1, together with the filtration $(\mathcal{F}_t) \triangleq \big(\sigma(I_1, Y_1, \dots, I_t, Y_t)\big)_{t\geqslant 0}\big)$, we introduce the $\mathcal{F}_t$-martingale

$$Z_t \triangleq \sum_{s=1}^{t} \big(Y_s - \mu_k\big)\mathbb{1}_{\{I_s=k\}} = N_{t,k}\Big(\widehat{\mu}_{t,k} - \mu_k\Big).$$

Indeed, I_t is $\mathcal{F}_{t-1}$-measurable so, given how rewards are drawn, we have first $\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = \mathbb{E}[Y_t|I_t] = \mu_{I_t}$ and therefore,

$$\mathbb{E}\big[(Y_t - \mu_k)\mathbb{1}_{\{I_t=k\}}|\mathcal{F}_{t-1}\big] = (\mu_{I_t} - \mu_k)\mathbb{1}_{\{I_t=k\}} = 0\,.$$

Therefore, Z_t is $\mathcal{F}_t$-adapted and its conditional expectation equals 0. As each increment $(Y_s - \mu_k)\mathbb{1}_{\{I_s=k\}}$ is bounded between $-\mu_k$ and $1-\mu_k$, by applying the classical Azuma-Hoeffding inequality to the $\mathcal{F}_t$-martingale Z_t, we would have obtained, for any $\varepsilon > 0$,

$$\mathbb{P}[Z_t \geqslant \varepsilon] \leqslant \mathrm{e}^{2\varepsilon^2/t},$$

and in particular, with $\varepsilon = \sqrt{t\ln(1/t\delta)/2}$,

$$\mathbb{P}\left(\widehat{\mu}_{t,k} - \mu_k \geqslant \sqrt{\frac{t}{N_{t,k}}}\sqrt{\frac{\ln(1/t\delta)}{2N_{t,k}}}\right) \leqslant t\delta.$$

To avoid the factor $\sqrt{t/N_{t,k}}$ in the deviation bound, we use, in the proof, the fact that the estimation $\widehat{\mu}_{t,k}$ is computed on random number of observations, and we thus reach a better deviation bound than the direct application of the classical inequality.

Proof of Lemma 1. The proof is based on the fact that $(M_t)_{t\geqslant 0}$ is a super-martingale with respect to the filtration $(\mathcal{F}_t)$, with for any $x > 0$,

$$M_t \triangleq \exp\left(xZ_t - \frac{x^2}{8}N_{tk}\right) = \exp\Big(\sum_{s=1}^{t}\big(x(Y_s-\mu_k) - x^2/8\big)\mathbb{1}_{\{I_s=k\}}\Big).$$

Indeed, as Z_t and $N_{t,k}$ are $\mathcal{F}_t$-adapted, M_t is too. To prove that, for any $t \geqslant 1$, $\mathbb{E}[M_t|\mathcal{F}_{t-1}] \leqslant M_{t-1}$, we recall that the random variable $\big(Y_t - \mu_k\big)\mathbb{1}_{\{I_t=k\}}$ is almost surely bounded between $-\mu_k$ and $1-\mu_k$ and that its expectation is null:

$$\mathbb{E}\Big[\big(Y_t - \mu_k\big)\mathbb{1}_{\{I_t=k\}}\Big] = \mathbb{E}\Big[\mathbb{E}\big[(Y_t - \mu_k)\mathbb{1}_{\{I_t=k\}}|\mathcal{F}_{t-1}\big]\Big] = 0\,.$$

Therefore, by using Lemma 2, we obtain that

$$x \in \mathbb{R}, \quad \mathbb{E}\Big[\exp\Big(x(Y_t-\mu_k)\mathbb{1}_{\{I_t=k\}}\Big)|\mathcal{F}_{t-1}\Big] \leqslant \mathrm{e}^{\frac{x^2}{8}}\,. \tag{$\star$}$$

But, we can reach an even better bound by decomposing the conditional expectation according to the value of the $\mathcal{F}_{t-1}$-measurable random variable $\mathbb{1}_{\{I_t=k\}}$:

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(x(Y_t-\mu_k)\mathbb{1}_{\{I_t=k\}}\Big)|\mathcal{F}_{t-1}\Big] &= \mathbb{E}\Big[\exp\Big(x(Y_t-\mu_k)\mathbb{1}_{\{I_t=k\}}\Big)|\mathcal{F}_{t-1}\Big]\mathbb{1}_{\{I_t=k\}} + \mathrm{e}^0\mathbb{1}_{\{I_t\neq k\}} \\ &\overset{(\star)}{\leqslant} \mathrm{e}^{\frac{x^2}{8}}\mathbb{1}_{\{I_t=k\}} + \mathbb{1}_{\{I_t\neq k\}} = \exp\Big(\frac{x^2}{8}\mathbb{1}_{\{I_t=k\}}\Big)\,.\end{aligned}$$

So, $(M_t)_{t\geqslant 0}$ is a super-martingale:

$$\mathbb{E}\big[M_t\big|\mathcal{F}_{t-1}\big] = M_{t-1}\mathbb{E}\Big[\exp\big((x(Y_t-\mu_k)-x^2/8)\mathbb{1}_{\{I_t=k\}}\big)\Big|\mathcal{F}_{t-1}\Big] \leqslant M_{t-1}\,,$$

and, for any $x>0$, we obtain that

$$\mathbb{E}\Big[\exp\Big(xZ_t - \frac{x^2}{8}N_{t,k}\Big)\Big] = \mathbb{E}[M_t] \leqslant \mathbb{E}[M_0] = 1\,. \tag{$\star\star$}$$

For each deterministic possible value $s=1,\dots,t$ of $N_{t,k}$, for any $x>0$, and any $\varepsilon>0$, we use a Markov-Chernoff bound to obtain

$$\begin{aligned}\mathbb{P}\Big(Z_t \geqslant \varepsilon \text{ and } N_{t,k}=s\Big) &\leqslant \mathrm{e}^{-x\varepsilon}\mathbb{E}\Big[\mathrm{e}^{xZ_t}\mathbb{1}_{\{N_{t,k}=s\}}\Big] = \mathrm{e}^{-x\varepsilon+\frac{x^2}{8}s}\,\mathbb{E}\big[\mathrm{e}^{xZ_t-\frac{x^2}{8}s}\mathbb{1}_{\{N_{t,k}=s\}}\big] \\ &\leqslant \mathrm{e}^{-x\varepsilon+\frac{x^2}{8}s}\mathbb{E}\Big[\exp\Big(xZ_t-\frac{x^2}{8}N_{tk}\Big)\Big] \overset{(\star\star)}{\leqslant} \mathrm{e}^{-x\varepsilon+\frac{x^2}{8}s}\,.\end{aligned}$$

And, by choosing $x=4\varepsilon/s$, we get that,

$$\forall\,\varepsilon>0,\,\forall\,s\geqslant 1, \quad \mathbb{P}\Big(Z_t\geqslant\varepsilon \text{ and } N_{t,k}=s\Big) \leqslant \mathrm{e}^{-2\varepsilon^2/s} \tag{$\star\star\star$}$$

We recall that we aim to bound

$$\mathbb{P}\left(\widehat{\mu}_{t,k}-\mu_k \leqslant \sqrt{\frac{\ln 1/\delta}{2N_{t,k}}}\right) = \mathbb{P}\left(Z_t \leqslant \sqrt{\frac{N_{t,k}\ln 1/\delta}{2}}\right).$$

so it only remains to sum the inequality above over s for each $s=1,\dots t$, with $\varepsilon=\sqrt{\frac{s\ln 1/\delta}{2}}$, to conclude the proof:

$$\begin{aligned}\mathbb{P}\left(Z_t \leqslant \sqrt{\frac{N_{t,k}\ln 1/\delta}{2}}\right) &\leqslant \sum_{s=1}^{t}\mathbb{P}\Big(Z_t\leqslant\sqrt{\frac{s\ln 1/\delta}{2}} \text{ and } N_{t,k}=s\Big) \\ &\overset{(\star\star\star)}{\leqslant} \sum_{s=1}^{t}\exp\Big(-\frac{2s\ln 1/\delta}{2s}\Big) = t\delta\,.\end{aligned}$$

□

We now prove the conditional version of Hoeffding's lemma stated and used above.

Proof of Lemma 2. Let us fix $x \in \mathbb{R}$ and denote by $Y = X - \mathbb{E}[X|\mathcal{G}]$ the random variable that is almost surely bounded between $A = a - \mathbb{E}[X|\mathcal{G}] \leqslant 0$ and $B = b - \mathbb{E}[X|\mathcal{G}] \geqslant 0$ (which are $\mathcal{G}$-measurable random variables). By using the convexity of $y \mapsto \mathrm{e}^{xy}$ on

$$Y = \frac{B-Y}{B-A}A + \frac{Y-A}{B-A}B\,, \quad \text{we get} \quad \mathrm{e}^{xY} \leqslant \frac{B-Y}{B-A}\mathrm{e}^{xA} + \frac{Y-A}{B-A}\mathrm{e}^{xB}\,.$$

By taking the expectation conditionally to $\mathcal{G}$ and using that A and B are $\mathcal{G}$-measurable, and keeping in mind that $\mathbb{E}[Y|\mathcal{G}] = 0$, we get

$$\mathbb{E}[\mathrm{e}^{xY}|\mathcal{G}] \leqslant \frac{B}{B-A}\mathrm{e}^{xA} + \frac{-A}{B-A}\mathrm{e}^{xB}\,.$$

It only remains to apply Lemma 3 below with $p = -A/(B-A)$ (which is in $[0,1]$ because $A \leqslant 0$ and $B \geqslant 0$) and $u = x(B-A) = u(b-a)$ to conclude the proof. □

Lemma 3 (Non-conditional Hoeffding's lemma for Bernoulli distributions)**.** *For any $p \in [0,1]$ and any $u \in \mathbb{R}$,*

$$(1-p)\mathrm{e}^{-pu} + p\mathrm{e}^{(1-p)u} \leqslant \mathrm{e}^{u^2/8}\,.$$

Proof of Lemma 3. We consider the random variable X which follows a Bernoulli distribution of parameter $p \in [0,1]$, so $\mathbb{E}[X] = p$ and $0 \leqslant X \leqslant 1$ almost surely. The inequality of the lemma is a straightforward application of (non-conditional) Hoeffding's lemma (see Hoeffding, 1994) which states that, for all $u \in \mathbb{R}$,

$$\mathbb{E}\big[\mathrm{e}^{uX}\big] = p\mathrm{e}^u + 1 - p \leqslant \exp\left(u\mathbb{E}[X] + \frac{u^2}{8}\right) = \mathrm{e}^{up}\mathrm{e}^{u^2/8}\,.$$

By dividing the inequality above by e^{-pu}, we conclude the proof. □

Remark 6. *We point out that Lemma 1 also holds for the expected losses $\ell_{t,k}$ and their estimations $\widehat{\ell}_{t,k}$ (computed with the variables $\ell_t(Y_1), \ldots, \ell_t(Y_{t-1})$). We recall that, for a round t and a tariff k, these quantities are defined by:*

$$\ell_{t,k} = \mathbb{E}_{Y\sim\nu_k}\big[\ell_t(Y)\big] \quad \text{and} \quad \widehat{\ell}_{t,k} = \frac{1}{N_{t-1,k}}\sum_{s=1}^{t}\ell_t(Y_s)\mathbb{1}_{\{I_s=k\}}\,.$$

In the proof above, it suffices to replace the variables μ_k and $\widehat{\mu}_{t,k}$ by $\ell_{t+1,k}$ and $\widehat{\ell}_{t+1,k}$, respectively, and to update the $\mathcal{F}_t$-martingale $Z_t = \sum_{s=1}^{t}\big(Y_s - \mu_k\big)\mathbb{1}_{\{Is=k\}}$ with

$$Z_t = \sum_{s=1}^{t}\big(\ell_{t+1}(Y_s) - \ell_{t+1,k}\big)\mathbb{1}_{\{Is=k\}}$$

to get the result. Therefore, for all $k \in \{1,\ldots,K\}$, for all $t \geqslant K$ (so $N_{t,k} \geqslant 1$), and for all $\delta \in (0,1)$, we have

$$\mathbb{P}\left(\ell_{t,k} > \widehat{\ell}_{t,k} - \sqrt{\frac{\ln\frac{1}{\delta}}{2N_{t,k}}}\right) \leqslant t\delta \quad \text{and} \quad \mathbb{P}\left(\widehat{\ell}_{t,k} < \ell_{t,k} + \sqrt{\frac{\ln\frac{1}{\delta}}{2N_{t,k}}}\right) \leqslant t\delta\,.$$


MB

3

Forecasting of power consumption

After a brief introduction on the industrial challenges of forecasting electricity consumption and a non-exhaustive review of commonly used methods, the chapter presents the open data set "Low Carbon London" created and published by UK Power Networks. It contains electricity consumption records of households subjected to dynamic electricity tariffs. A succinct descriptive analysis is also provided. Next, the focus is on generalized additive models, a powerful and efficient semi-parametric approach to model electricity consumption. An application of these methods to the Low Carbon London data set concludes the chapter.

1 Introduction

1.1 Motivations and industrial challenges

Electricity is still difficult to store, except at prohibitive costs, and this is why the balance between production and consumption needs to be constantly maintained. Indeed, the security of the system and the power grid must be ensured to avoid blackouts. This management is becoming increasingly difficult since many intermittent means of production, such as wind power or photo-voltaic, are connected to the grid. The flexible means of electricity production (water dams, nuclear plants, coal and gas plants) are currently adapted according to the energy consumption forecasts. A slight improvement in electricity demand forecasting can bring significant benefits by reducing production costs, limiting the financial penalties imposed by system operators in case of mismanagement of the production/consumption balance and increasing the trading advantages, especially during the peak periods (see, e.g. Bunn and Farmer [1985])

In the short-term (from a few hours to two weeks) or in the middle-term (from two weeks to five years), electricity consumption forecasts are therefore essential for scheduling and optimizing the use of power plants. In addition, in the long-term (from five to fifty years), electricity consumption forecasts provide prospects for the evolution of the customer portfolio. They are therefore useful both for adapting commercial offers accordingly and for defining an investment strategy.

Electricity consumption forecasts are required at different levels of aggregation. Foremost, forecasts of global consumption (e.g. for an entire country), on the one hand, and of consumption at the interfaces between the transport network (high-voltage lines) and the distribution network (medium- and low-voltage lines), on the other hand, are essential both for network operators, which must dispatch electricity, and for electricity providers, which must produce the quantity of electricity corresponding to that consumed. However, with the integration of decentralized means of production such as wind and solar farms, as well as the development of auto-consumption, it also becomes essential to forecast consumption at the lowest aggregate levels. Indeed, due to the increase in the uptake of distributed generation and storage systems, dis-aggregated load forecasting becomes essential. Electricity grids are becoming "smart" and there is no doubt that forecasting of power consumption plays a key role in their proper management. In the same time, smart meters, which are being massively deployed, will be a valuable new source of information and may also offer new services to consumers.

Forecasting electricity consumption over different time horizons and geographical scales is therefore crucial for electricity suppliers and grid operators. In the current context of energy and digital transition, models must be as efficient as possible and will have to adapt to the evolution of power systems and to the new challenges that accompany it.

1.2 History of forecasting methods and literature discussion

EDF is active in many power generation technologies as nuclear, hydro, wind, solar, biomass, geothermal, fossil-fired and marine energies. To response to the electricity demand by managing these production units, it developed over the last decades accurate consumption forecasting models. The company collected electricity production and con-

sumption data history over many years and gained sound knowledge of power consumption forecasting.

1.2.1 Exogenous variables useful for forecasting power consumption

Electricity consumption varies according to many variables, which are primarily meteorological, calendar and related to electricity pricing options. First of all, because of the large demand of electrical heating in cold weather, temperature is essential to obtain relevant forecasts. Among others, Engle et al. [1986] and Taylor and Buizza [2002] investigated the non-linear relationship between temperature and electricity consumption; they used a semiparametric regression and artificial neural networks, respectively. Moreover, Taylor [2003] showed that some exponential smoothings of the temperature, which model buildings inertia or consumers reaction delay, are likely to improve the predictions. Other weather variables like wind, humidity, precipitation or cloud cover may also influence the electricity consumption, see Taylor and Buizza [2003]. Calendar variables are also taken into account in the vast majority of the power consumption models (see, for example Haida and Muto [1994]). Finally, many demand-response strategies (which influence consumption in one direction or the other) rely on changes in the price of electricity, which therefore also have an impact on electricity consumption; see, among others, Lescoeur and Galland [1987] which presents the French experience of using marginal cost pricing for good demand side management, and Kostková et al. [2013] which gives an overview of the load management methods, techniques and programs theoretically described or practically used in several countries. Therefore, to model and forecast electricity consumption, the proposed approaches provide predictions based on exogenous weather, calendar and price information. They may also consider past load data, using traditional time series techniques.

1.2.2 Parametric and semi-parametric methods

Among the parametric approaches, Ramanathan et al. [1997] proposed a multiple regression model (one for each hour of the day) with a dynamic error structure, based on auto-regressive models. Moreover, univariate methods based on exponential smoothing and SARIMA (seasonal auto-regressive integrated moving average) models can be found in Hyndman et al. [2002] and Abraham and Nath [2001], respectively. Before 2006, the electricity consumption approaches used by the operational EDF entities consisted in regression methods coupled with classical times series models such as (S)ARIMA models (see, e.g., Ernoult et al., 1983).

More complex relationships between the electricity consumption and its covariates have also been studied. For example, in Engle et al. [1986], the non-linear effect of the temperature is estimated using an extension of smoothing splines and Harvey and Koopman [1993] captured the seasonal patterns and the temperature effect using a time-varying spline model. Antoniadis et al. [2006] proposed an approach based on functional kernel non-parametric regression estimation techniques. With Generalized Additive Models (GAM), studied by Fan and Hyndman [2012] and Pierrot and Goude [2011], the expected power consumption is modeled as a sum of independent exogenous variable effects. These effects are approached with smooth functions which can capture nonlinearities. This semi-

parametric approach is currently used in an operational manner in many EDF entities in order to provide forecasts at different time horizons and at various levels of aggregation.

The main drawback of a such modeling is its weak ability to adapt to changes in power consumption behaviors due to the evolution of economic growth, the development of new electrical uses (for instance, electrical vehicles), the opening of new electricity markets, etc Ba et al. [2012] proposed an online learning algorithm to track the smoothing functions of additive models and therefore to adapt the electricity consumption models to an ever-changing environment. It should also be noted that Wood et al. [2015] developed generalized additive model fitting methods for large data sets, considering a practical application in electricity grid load prediction.

Finally, it must be emphasized that the previous methods focused on point forecasting (generally, the expected consumption). However, probabilistic electricity consumption forecasting gives additional information on the variability and uncertainty of the consumption and is becoming valuable for production planning or smart grid management. With regard to this objective Hyndman and Fan [2010] proposed to forecast the probability distribution of annual and weekly peak electricity demand using GAMs and later Fasiolo et al. [2020] provided a generalization of GAMs for fitting additive quantile regression models.

1.2.3 Non-parametric methods

Load forecasting has not escaped the current machine learning trend and many applications in the electrical field, based on neural networks or decision trees have been successfully carried out. For example, Ben Taieb and Hyndman [2014] and Chen et al. [2004] proposed, in load forecasting competitions, methods based on gradient boosting and support vector machine, respectively. Random forests also provided relevant power consumption forecasts (see, e.g. Dudek, 2015) as well as artificial neural networks (see among others, Asar and Mcdonald, 1994 and Ringwood et al., 2001 and more lately Kong et al., 2017). Since several years, black-box models have also been examined in EDF Research and Development department. This recent research gives promising results even if these models suffer, compared to GAM, of a lack of interpretability: this is one of the reasons they are not yet used by operational EDF entities.

1.2.4 Online expert aggregation

To conclude this review of the methods used at EDF for forecasting electricity consumption, it should be noted that it is sometimes useful to combine forecasts obtained from different models. Some of the predictions may be relevant in specific conditions (peak hours, winter, etc.) while others are better at other times; this is why it can be interesting to combine them. Goude [2008] and Gaillard [2015] proposed and implemented algorithms for online predictors aggregation which significantly improve the electricity consumption forecasts. These black-box algorithms take as input several predictors (several forecasting models) and output a linear combination of the forecasts provided by the predictors. The weights in the linear combination are constantly adjusted according to the most recent past prediction errors. This is further presented and discussed in Chapter 6. The `R`-Package

`Opera`[1], see Goude and Gaillard [2016], gathers these algorithms, which are used in certain EDF entities.

As the demand side management strategies implemented in this thesis will target aggregates of a hundred to a thousand households, the focus is on short and medium term forecasting of both global and local energy consumption (at an aggregation level of around 100 households). It is essential, before tackling the implementation of bandit algorithms, to study in more detail the specificities of electricity consumption at these levels and terms and to propose a realistic model of the latter. Section 2 presents the power consumption data set used throughout the manuscript, called "Low Carbon London data set" in the sequel. Most of consumption models we will use will be based on generalized additive models. In Section 3, GAM theory is briefly introduced. All the presented results can be found in the exhaustive monograph *Generalized Additive Models: An Introduction with R* by Wood [2006]. Finally, an application of GAMs to the Low Carbon London data is provided in Section 4.

2 Low Carbon London data

In Chapters 4, 7 and 8, we consider an open data set created and published by UK Power Networks and containing electricity consumption (in kWh per half-hour) of around 5,000 households throughout 2013[2]. Since weather has a strong impact on energy consumption, open source data points of London air temperature were added to the data set. These records were gathered by the U.S.A. scientific agency NOAA (National Oceanic and Atmospheric Administration) and are available online[3]. The next subsection presents the data in more detail and is followed by a descriptive analysis.

2.1 Underlying dataset

Between November 2011 and February 2014, 5,567 London households took part in the UK Power Networks led Low Carbon London project. These households were recruited as a balanced sample representative of the Greater London population and an CACI Acorn group[4] was assigned to each of them. Among these households, a sub-group of approximately 1,100 was subjected to a dynamic Time of Use (ToU) tariff throughout the year 2013. The tariff values were among High (67.20 p/kWh), Low (3.99 p/kWh), or Normal (11.76 p/kWh); and the (half-hourly) intervals where these prices are applied, were announced one-day-ahead via the smart meter or text message. According to UK Power Networks, the signals sent were designed to be representative of those that could be used in the future, whether to manage the integration of renewables into the electricity generation mix or to test the potential of using a high price to reduce stress on grids during periods of over-consumption. All ToU households received the same tariffs and non-ToU households were on a flat rate tariff of 14.228 p/kWh; we refer to them as Standard (Std) customers. The report Schofield et al. [2014] provides a full description of this experimentation and an exhaustive analysis of the results.

[1] `https://CRAN.R-project.org/package=opera`
[2] `https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households`
[3] `www.noaa.gov`
[4] `https://acorn.caci.co.uk/`

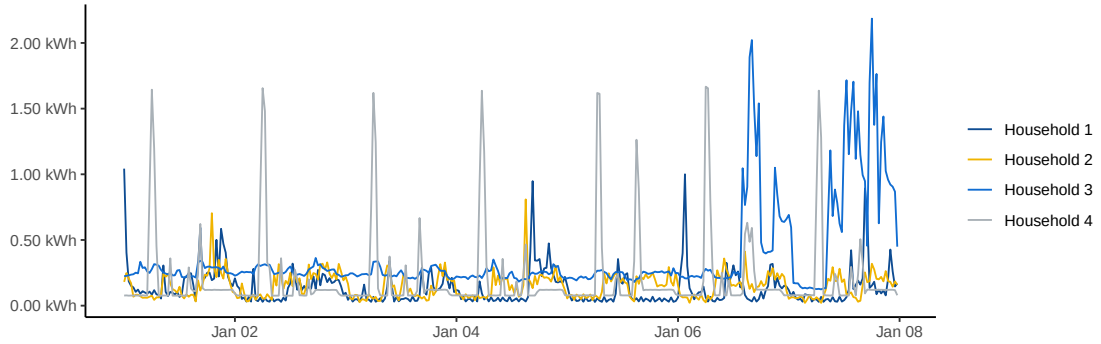


Figure 3.1 – Four households electricity consumption (in kWh) per half-hour, over seven days.

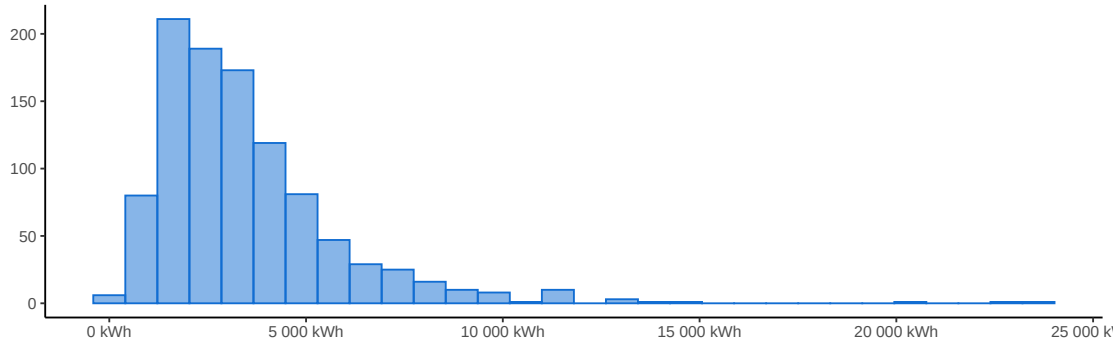


Figure 3.2 – Distribution of the annual electricity consumptions (in kWh) of 1,007 ToU households.

2.1.1 Electricity consumption and tariff data

The data set contains energy consumption readings in kWh for each customer at half-hourly intervals. Time slots and price signal schedules are also available. Since the dissertation focuses on demand management strategies, we are interested in household behaviors in response to changes in electricity prices. Therefore, the experiments will be based on ToU households consumption readings. Only households with more than 95% of data available (1,007 ToU households) are kept and the same number of Std households are sampled to build a control group. The missing values in the time series were filled by linear interpolation, using the previous and next interval records for small gaps. For longer periods of missing data, records were missing over longer time intervals and we then considered the energy consumption half an hour by half an hour. By selecting the records associated with a specific half-hour, we imputed the missing records by linear interpolation, i.e. from the consumption of the days preceding and following at the same half-hour. Energy consumption readings for the first seven days of 2013 are plotted for four randomly selected ToU households in Figure 3.1. Households 1 and 2 present rather similar consumption curves, with periods of low and high consumption for each day. The times of these drops or peaks differ from household to household. The curve for Household 3 is almost flat and at a very low level during the first five days, it becomes much more significant, with many consumption peaks during the last two days. It would seem that the members of this household went on holiday (probably for Christmas and New Year), leaving a few

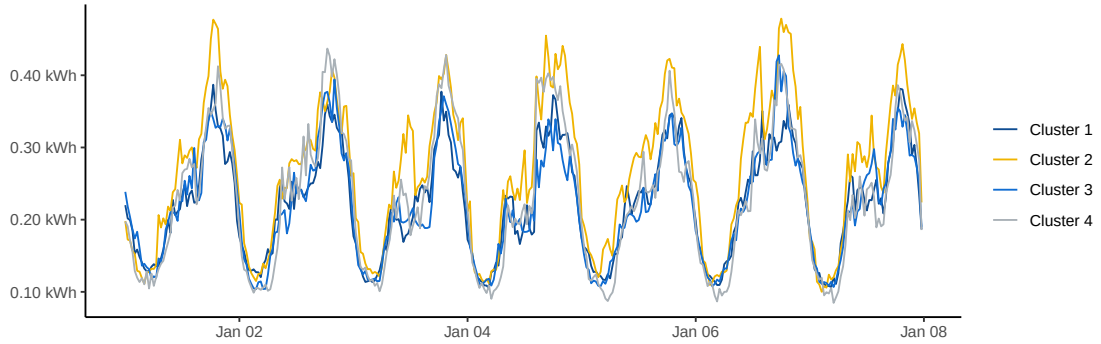


Figure 3.3 – Electricity consumption (in kWh) per half-hour, over seven days of four clusters of 100 households.

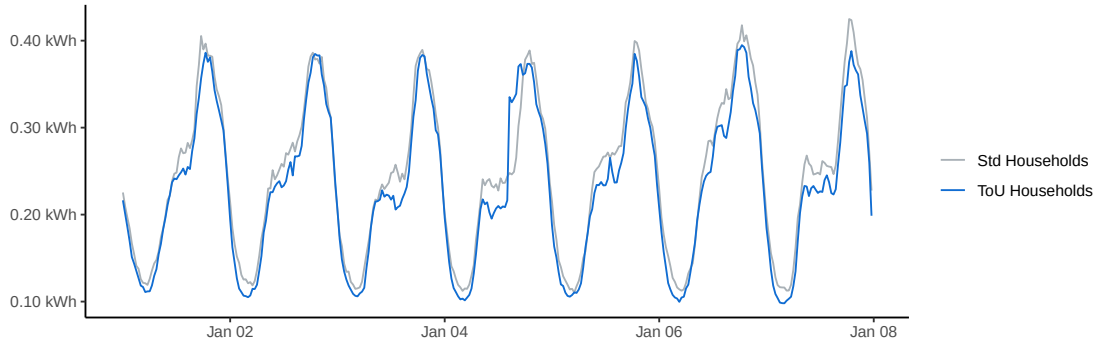


Figure 3.4 – Electricity consumption (in kWh) per half-hour, over seven days, averaged over 1,007 ToU households (in blue) and 1,007 Std households (in grey).

appliances on standby, and that on their return, the household started consuming again. The consumption curve for Household 4 shows a peak each morning, which could be due to the connection of an energy intensive appliance. Individual electricity consumption is thus extremely erratic, making it difficult to model and predict. The histogram of Figure 3.2 represents the annual consumption of the 1,007 households. The majority of households consumes between 2,000 and 5,000 kWh per year, although a few are very energy intensive.

A level of aggregation of at least a hundred households enables to smooth consumption, which will then be more easily modeled and predictable (whereas individual consumption is erratic and unpredictable). Figure 3.3 represents, still for the first seven days of 2013, the average electricity consumption of four clusters of one hundred households. And the higher aggregation level, the smoother the electricity consumption. Indeed, the average electricity consumption of 1,007 ToU and 1,007 Std households is shown in Figure 3.4, these curves are much smoother than those for individual consumptions. Even if the ToU households received the Normal tariff for this period almost every time (from January, 1st to 7.), the average electricity consumption of the ToU and Std households differ. Indeed, the consumption Tou is always lower than the consumption Std and this difference is even more significant between 8 a.m. and 6 p.m. From now on, we denote by $Y_1, Y_2, \ldots$ the time series of the average power consumption of the ToU household at half hourly intervals.

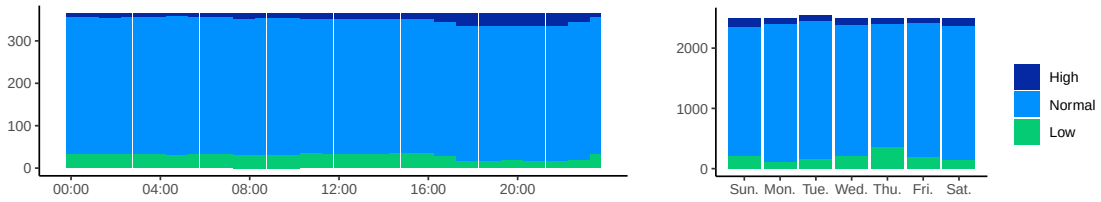


Figure 3.5 – Distributions of tariffs Low (in green), Normal (in blue) and High (in navy) according to the half-hour of the day (on the left) and to the days of the week (on the right).

2.1.2 Electricity tariffs data

For the ToU households, electricity price varies among three tariffs: Low, Normal and High and are denoted by $j_1, j_2, \ldots$. Throughout the Low Carbon London project, these tariffs have been carefully chosen and their distributions are not uniform at all. Therefore, we must be aware of possible biases that exist in the data set and we should be careful with the conclusions drawn from our experiments. Figure 3.5 shows the distribution of the three tariffs according to the time of day and the days of the week. It should be noted that the high tariff is more often given at the evening peak of consumption (between 6 and 10 p.m.) and during the weekends. This is because we want to have a better understanding of the possible drop in consumption caused by a high price in this particular time slot. It should also be noted that, during the same day, a special rate (Low or High) applies only for a few hours; the rest of the day is under the Normal tariff. There are thus operational constraints, which were considered during the Low Carbon London project experience. This is an important component to take into account: when testing strategies, it will be necessary to make sure that the tariffs are not sent over too long a time slot (we will neglect these constraints in our first models, though).

2.1.3 Meteorological and calendar data

As mentioned above, since weather has a strong impact on energy consumption, we added half-hourly data points of air temperature in London $\tau_1, \tau_2, \ldots$ obtained from hourly public observations after linear interpolation. It is common in the electricity forecasting literature (e.g., Fan and Hyndman, 2012) to use nonlinear distributed lag models (namely to τ_{t-1}, τ_{t-2} etc. to predict Y_t). Here, we use an exponentially weighted temperature – a "smoothed temperature" $\bar{\tau}_t$, that models the thermal inertia of buildings and is defined as follows: for any round $t \geqslant 1$, it is defined by

$$\bar{\tau}_t = \begin{cases} \tau_1^1 & \text{if } t = 1 \\ (1-a)\tau_t + a\bar{\tau}_{t-1} = (1-a)\sum_{k=0}^{t-2} a^k \tau_{t-k} + a^{t-1}\tau_1 & \text{else,} \end{cases}$$

where the smoothing parameter a is in $[0, 1]$. Using this smoothed temperature is more parsimonious than taking all the lag temperatures τ_{t-1}, $\tau_{t-2}, \ldots$ into account and it is likely to improve forecasts (see among others, Taylor, 2003 and Goude et al., 2014). Note that between these two approaches Carroll et al. [1997] also proposed to use linear combinations of past temperatures of the form $\tilde{\tau}_t = \sum_{k=0}^{t} \beta_k \tau_{t-k}$ to improve predictions. To tune a, we performed an exhaustive grid search (by testing many values on the prediction models described in Section 4) and set $a = 0.998$. This value is consistent with the one generally used in electricity consumption forecasting models developed at EDF R&D. Figure 3.6 represents both realized and smoothed temperatures. As desired, the exponentially

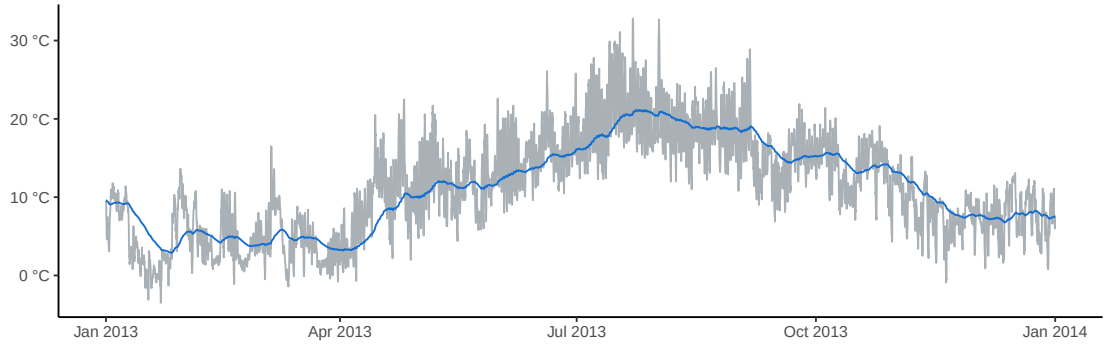


Figure 3.6 – Half-hourly temperature (in grey) and exponentially smoothed temperature (in blue).

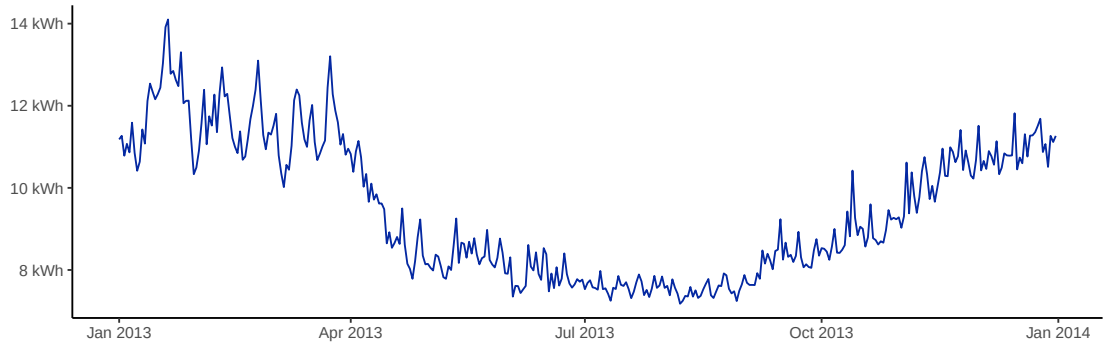


Figure 3.7 – Average daily ToU household electricity consumption for the year 2013.

smoothed temperature is not at all erratic and there is also a certain delay, compared to the realized temperature, in its variations.

Finally, energy consumption also depends on calendar variables such as the half-hour of the day, the day, the season etc. Thus, for a round t, three additional variables were created: h_t, the half-hour of the record, the categorical variable "type of day" w_t that takes values 0 on Sunday, 1 on Monday and so on, and the position in the year κ_t, a continuous variable which increases linearly from 0 (on January, 1st) to 1 (on December, 31st).

The power consumption records provided by the Low Carbon London project, the temperatures archived by NOAA, the exponentially smoothed temperatures and the calendar variables created are summarized in Table 3.1.

2.2 Descriptive analysis

Electricity consumption is characterized by several seasonalities. In Figure 3.7, the average daily consumption is plotted over the year 2013. It varies over the months: in winter, due to the termo-sensitivity of households, electricity consumption is high, whereas it decreases in summer. In July and August, consumption is even lower than in the other hot months: many households probably went on summer holidays, leaving the buildings empty.

Variable	Notation
Average (over 1,007 ToU households) half-hour energy consumption	Y_t
Half-hour tariff among Low, Normal and High	j_t
Half-hour London air temperature	τ_t
Half-hour London air smooth temperature	$\bar{\tau}_t$
Type of day (0 for Sunday, 1 for Monday etc.)	w_t
Position in the year (0 on January, 1st at 00:00 and 1 on December, 31st at 23:30)	π_t
Half-hour index (0 at 00:00 and 47 at 23:30)	h_t

Table 3.1 – Summary of the variables provided and created.

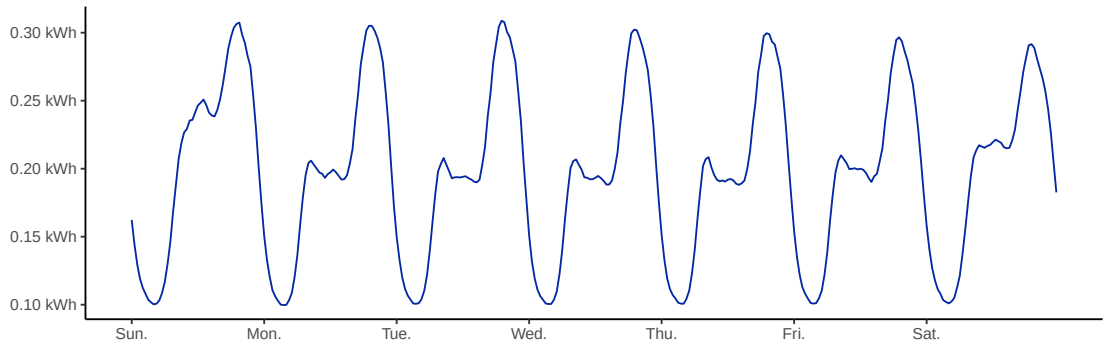


Figure 3.8 – Average ToU household electricity consumption at half-hourly intervals, averaged over the 52 weeks of the year 2013.

2.2.1 Seasonalities and calendar variables effect

Figure 3.8 shows the average weekly consumption at half-hourly intervals. It should be noted that there is a daily seasonality: for each day, the consumption curve follows the same evolution. It is low at night, with a first peak around 8 a.m., when consumers wake up. Power consumption then stagnates until the evening, when there is a second, even more important peak: household members go home, cook dinner, turn on the TV, etc. and it collapses for the night. It is therefore highly correlated with the activities of household members. Furthermore, household electricity consumption is highest at weekends, when individuals are not working and potentially staying at home. In particular, the curves significantly differ between 10 a.m. and 4 p.m. for Saturday and Sunday, on the one hand, and for working days, on the other hand; in this time slot, people can stay at home on weekends while they have to go to work on the other days of the week. Moreover, we notice a delayed morning peak: individuals probably get up later (see Figure 3.9). Therefore, electricity consumption also presents a weekly seasonality.

2.2.2 Temperature effect

In Figure 3.10, we plotted the average daily power consumption and the average daily temperatures (realized and smoothed) over the year 2013. Visually, it is clear that temperatures and consumption are strongly correlated. More precisely, in winter, there is a strong negative correlation between temperature and energy consumption (see left of Figure 3.11). This is due to the temperature sensitivity of households that use electric

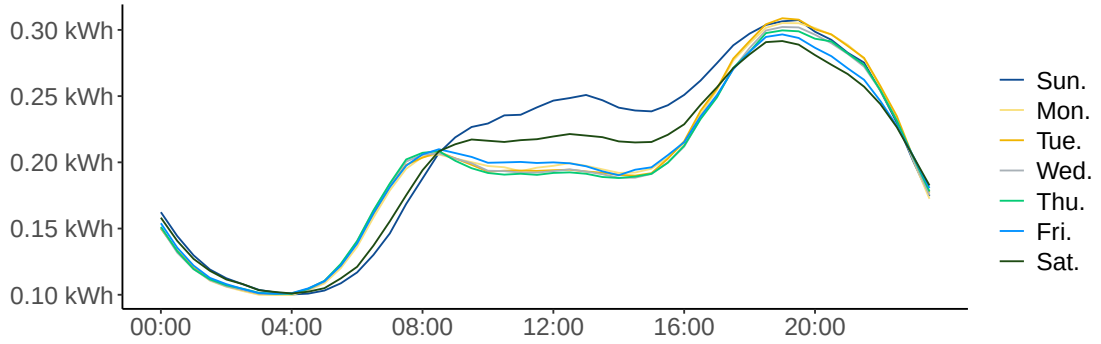


Figure 3.9 – Average ToU household electricity consumption at half-hourly intervals, averaged over days of the year 2013, depending on the day of the week (Sun., Mon. etc.).

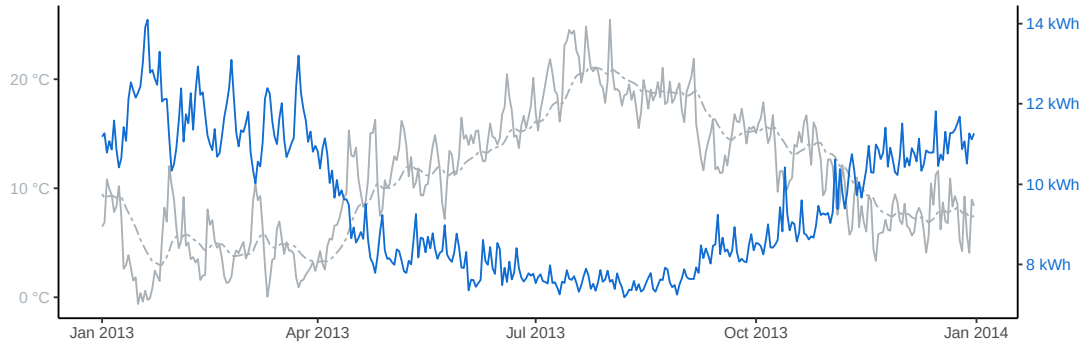


Figure 3.10 – Average daily ToU household electricity consumption in khW (in blue) and average daily realized (solid line) and smoothed (dashed line) temperatures in °C (in grey) for the year 2013.

heating. In summer, the relationship between temperature and consumption is less visible. It is slightly positive because of the use of air conditioning (see right of Figure 3.11).

2.2.3 Tariff effect

Finally, we investigate the effect of tariff changes on the electricity consumption. Figure 3.12 displays boxplots of the average consumption depending on both the tariff and the hour of the day. Except between 5 p.m. and 11 p.m., the higher the price, the lower the electricity consumption. We explain in the following paragraph the reason why the effect of a tariff change is counter-intuitive in the evening. An important point to note here is that the effect of a tariff change, i.e. the decrease or increase in load, seems to depend strongly on the time of day when the tariff change is applied: for example, a change during the night has practically no effect. As load adjustment to the electricity price is not yet automated, during the night, people are probably unlikely to wake up to turn on their electrical devices (although some, such as washing machines, can be programmed). On the opposite, a change in the price of electricity during the day has a significant impact. It should also be noted that the load variance is generally greater for Low tariffs than for High or Normal tariff. This seems quite logical, since reacting to a change in tariff requires a change in consumption habits, which is not always possible. This variability in reactions

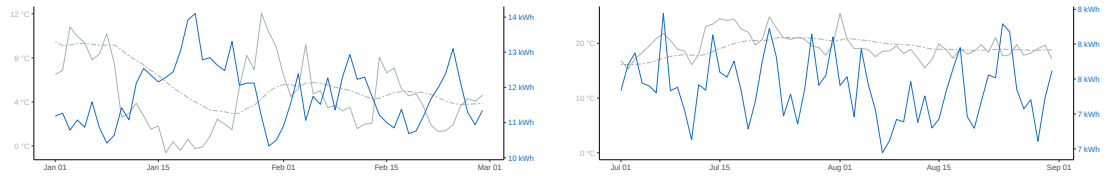


Figure 3.11 – Average daily ToU household electricity consumption in khW (in blue) and average daily realized (solid line) and smoothed (dashed line) temperatures in °C (in grey) in January and February (to the left) and in July and August 2013 (to the right).

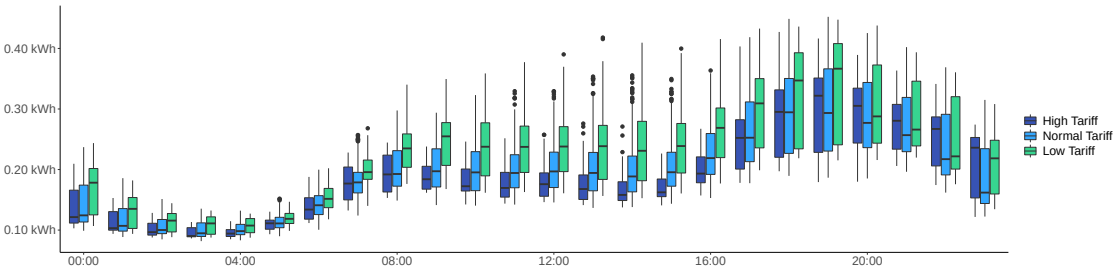


Figure 3.12 – Boxplot of electricity consumption for ToU households (in green for Low tariff, blue for Normal tariff and navy for High tariff) depending on the hour of the day.

to tariff changes is much higher for Low tariff than for High tariff: there is only a loss of profit if one does not react to Low tariff (whereas there is potentially more to lose if one continues to consume when the tariff is High). On the opposite, a High tariff provokes generally the same reaction: limiting the power consumption as much as possible to pay less. Therefore the variance is most of the time lower than for Normal and Low tariffs. We note, however, that the boxplots associated with Normal tariff are calculated with many more observations than those associated with Low or High tariffs (see Figure 3.5), and that these results must thus be considered with some caution.

Surprisingly, for a few hours during the evening consumption peak, the median load value for High tariff is higher than that of Normal tariff. Remembering the purposes of the Low Carbon London project, this actually makes sense: tariff were designed to test the potential to use high price signals to reduce stress on local distribution grids during periods of stress. Therefore, there is some bias in the data set and we should compare the load of ToU households with that of Std households to properly measure the impact of a tariff change. In Figure 3.13, the average daily electricity consumptions associated with the three tariffs are plotted separately for ToU and Std sub-populations. We emphasize that for the Std households, a change in tariff should not have any impact on their consumption: they were always on a constant tariff. Nevertheless, looking at the graphs to the left of Figure 3.13, the load of Std households is higher during the evening consumption peak for High tariff than for Normal or Low tariff. Therefore, High tariff was sent during periods of stress and it is logical to observe a less significant effect on the electricity consumption of ToU households. Finally, Figure 3.14 shows the electricity consumption over one week for both populations. Compared to the control group, namely Std households, ToU households react significantly to a change of the tariff: High tariff generally leads to a decrease in consumption, and conversely, Low tariff to an increase in consumption; however, these effects are less visible at night.

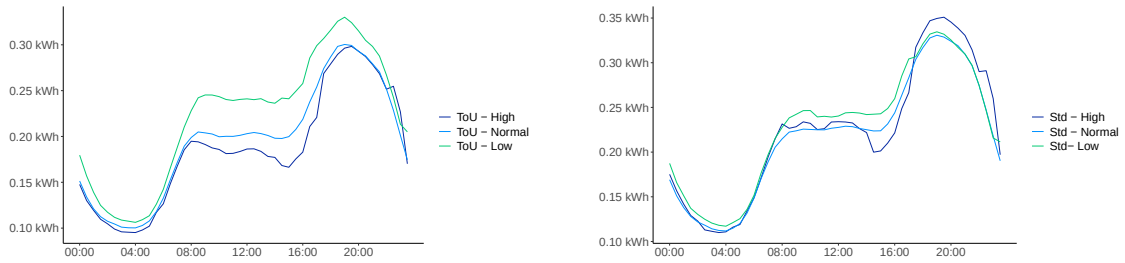


Figure 3.13 – Daily average electricity consumption for ToU households (on the left) and Std Households (on the right) depending on the tariff received only by ToU households (in green for Low tariff, blue for Normal tariff and navy for High tariff).

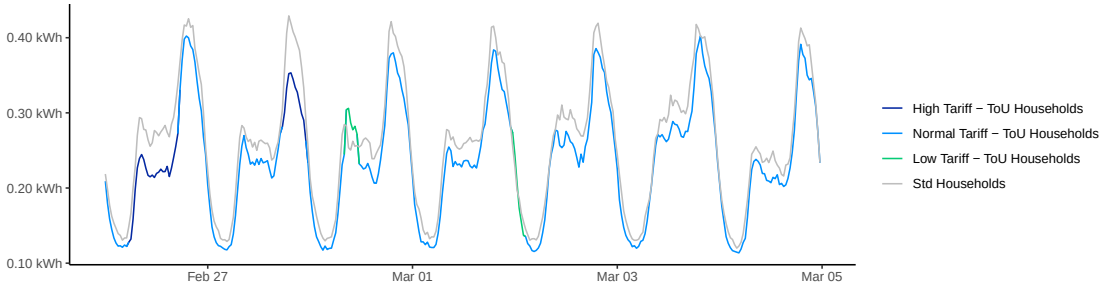


Figure 3.14 – Average electricity consumption of ToU (in green for Low tariff, blue for Normal tariff and navy for High tariff) and Std (in grey) households from Sunday February 26. to Saturday March 4.

This descriptive analysis points out the correlation between the average electricity consumption Y_t and the exogenous variables introduced below. To model and predict the load, we will therefore consider the realized and smoothed temperatures, τ_t and $\bar{\tau}_t$, respectively, the position in the year π_t, the day of the week w_t, the half-hour index h_t and the tariff j_t. All these variables are gathered in a contextual variable vector denoted by x_t. Throughout the thesis, we will generally consider some semi-parametric modeling of the load in the form:

$$Y_t = f(x_t) + \text{noise}\,,$$

where the function f has to be estimated. As previously mentioned, GAMs form a powerful and efficient semi-parametric approach used by EDF to model electricity consumption. We will illustrate this in Section 4 on the Low Carbon London data set. But above all, we will briefly present generalized additive models (GAM) in the next section. We will also show how a GAM can be expressed as an over-parameterized linear model. This writing leads to the assumptions made in the following chapters: they will concern the modeling of energy consumption and will be fundamental to obtain theoretical results on the convergence of the proposed bandit algorithms.

3 Generalized additive models

Generalized additive models were originally introduced by Hastie and Tibshirani [1986] to blend properties of generalized linear models with additive models. The model is defined

by an equation linking a response variable Y to some predictor variables, x. The variable of interest Y is assumed to follow an exponential family distribution (e.g. normal, binomial or Poisson distributions). This distribution is specified along with a link function g (such as the identity or logarithmic function) relating the expected value of Y to the predictor variables via a structure such that

$$g\big(\mathbb{E}[Y]\big) = f_1(x_1) + f_2(x_2) + f_{34}(x_3, x_4) + \dots ,$$

where the function f_i are smooth functions of the covariates. In order to define these "smooth functions", some spaces in which they can be represented will be introduced. The specification of the dependencies between the response variable and its covariates is therefore quite flexible, it can be non linear and non parametric, so that many different effects can be modeled (as opposed to parametric modeling where the relationships are very detailed). Generally, these smooth functions are estimated by penalized regression splines (which are functions defined piece-wise by polynomials), and their degrees of smoothness are chosen from data using a generalized cross-validation criteria.

For convenience, in the next subsection, we focus on an univariate cubic spline regression and how to set its degree of smoothness. Then, we will generalize this penalized regression method to any generalized additive model. We recall that all the results below are extracted from the monograph by Wood [2006].

3.1 Uni-variate semi-parametric regression: an example with cubic splines

We consider the basic model, where a random response variable Y depends on a single variable x,

$$Y = f(x) + \varepsilon\,, \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)\,.$$

With no loss of generality,we assume that the covariate x lies between 0 and 1. To estimate the smooth function f, we first choose an appropriate basis of functions $b_1, \dots b_q$, defining the space of functions to which f (or a close approximation thereof) belongs. Therefore, we assume that there exists a representation of f of the form

$$f(x) = \sum_{\ell=1}^{q} b_\ell(x)\theta_\ell\,,$$

where the coefficients θ_ℓ are unknown and have to be estimated. This transformation yields a linear model and it is now possible to estimate $\theta_1, \dots, \theta_q$ with classical linear regression methods. Indeed, denoting by $\theta = (\theta_1, \dots, \theta_q)$ the vector of coefficients, we get the linear model

$$Y = f(x) + \varepsilon = \big(b_1(x), \dots, b_q(x)\big)^{\mathrm{T}}\theta + \varepsilon\,,$$

so it is enough to estimate the vector θ to get an approximation of f. We point out that there exist various options to approximate the function f based on, for example, kernels, Fourier transformations, wavelets, etc.

Given a set of n observations (y_t, x_t), the n equations $y_t = f(x_t) + \varepsilon_i$ can be stacked together and by using the expression of f in the function basis b_1, $b_2, \dots, b_q$, they can be

written in matrix notation as

$$Y = X\theta + E \quad \text{where} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} b_1(x_1) & \dots & b_q(x_1) \\ b_1(x_2) & \dots & b_q(x_2) \\ \vdots & \vdots & \vdots \\ b_1(x_n) & \dots & b_q(x_n) \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

The line t of the design matrix X is made of the images of x_t by the functions $b_1, \dots, b_q$. We assume that the noises $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed random variables (with a centered Gaussian distribution of variance σ^2) and we are therefore in the classical linear regression framework.

A cubic spline basis. By assuming that f can be correctly approximated by a cubic spline, namely a continuous up to second derivative piecewise-cubic polynomial function, we may define a finite basis of functions to represent it. The points where the pieces of cubic polynomials join are called the knots of the spline. Their number q and locations are specified by the statistician. They are generally chosen evenly spaced across the range of observed values of x or at quantiles of the distribution of x values. Here, we denote the locations of the knots by $x_0^\star, \dots, x_{q-1}^\star$ (as we assumed that x lies in the interval $[0,1]$, we set $x_0^\star = 0$ and $x_{q-1}^\star = 1$) and once they are defined, we may consider a classical basis (proposed in the monographs by Wahba, 1990, Gu, 2002 and Wood, 2006): for $x \in [0,1]$, the functions b_1, $b_2, \dots b_q$ are defined by

$$b_1(x) = 1, \quad b_2(x) = x, \quad b_{i+2} = R(x, x_i^\star), \quad \text{for} \quad i = 1, \dots, q-2 \tag{3.1}$$

$$\text{with} \quad R(x, x^\star) = \frac{1}{4}\Big(\big(x^\star - 1/2\big)^2 - 1/12\Big)\Big(\big(x - 1/2\big)^2 - 1/12\Big) - \frac{1}{24}\Big(\big(|x - x^\star| - 1/\big)^4 - 1/2\big(|x - x^\star| - 1/2\big)^2 + 7/240\Big).$$

In Figure 3.15, these functions, scaled between 0 and 1, are plotted for $q = 7$ and for knots evenly spaced, that is, with $x_i^\star = i/6$, for $i = 0, \dots, 6$.

We emphasize that f may be represented in an other bases of functions, like in a polynomial basis ($b_1(x) = 1$, $b_2(x) = x$, $b_3(x) = x^2$, and so on) or in a truncated power function basis (e.g. $b_1(x) = 1$, $b_2(x) = x$, $b_3(x) = |x - x_1^\star|_+$, $b_4(x) = |x - x_2^\star|_+$ etc., with $|x|_+ = \max\{x, 0\}$).

Control of the smoothness of f. The smaller q, the smoother the function f. Indeed, for the cubic spline basis defined above, for $q = 2$, the estimated function will be linear. For splines, q is generally the number of knots and this corresponds to $q - 1$ portions of polynomials. When it increases, f becomes less and less smooth and too many knots lead to over-fitting of the data. In order to choose q properly, we could test different values. As models with $q - 1$ knots (whether uniformly spaced or at quantiles of the distribution of x values) are not nested within q knots models, such an approach would be problematic. Indeed, it would not be enough to add a function to the $q - 1$-dimension base to get the q-dimension base, it would actually be necessary to redefine all the functions again, and it would also be ineffective from a computational point of view. It could also be possible to start from a large value of q and thus a fine grid of knot and to drop them one by one, successively. At the end of this process, the knots may be unevenly spaced, leading to a poor model performance; especially since for spline regression models, the estimated

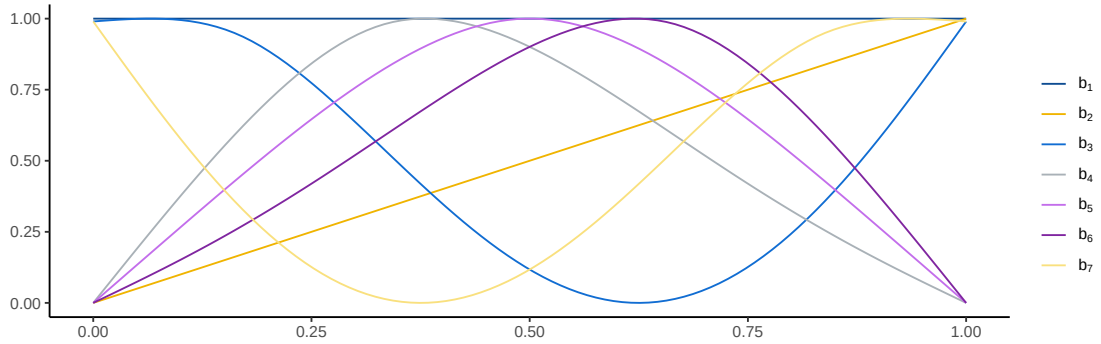


Figure 3.15 – A rank-7 cubic spline basis, with knot locations uniformly distributed between 0 and 1. Functions, defined in Equation (3.1), were scaled between 0 and 1.

models depend quite strongly on the locations chosen for the knots. A good alternative to control the degree of smoothness of f is to set the size of the basis q to a large number (slightly larger than assumed), and to add a regularity term in the regression minimization problem. Therefore, the vector θ is estimated by minimizing

$$\| Y - X\theta \|^2 + \lambda \int_0^1 f''(x)^2 \mathrm{d}x \,,$$

where $\lambda > 0$ is an hyper-parameter controlling the trade-off between model fit and model smoothness (we further detail how to choose it below). A very erratic function, with a boom and bust curve, generally leads to strong variations in the second derivative, which, therefore, has an important squared norm. Such a regularization term involving the second derivative thus allows to smooth the function f: the greater this regularization term, the more the function will tend to be linear (these are the functions with zero second derivative). We point out that adding such regularization terms is common for linear regression: among others, Ridge regression (see, e.g. Hoerl et al., 1962) is used to deal with the problem of multi-collinearity by adding a L^2-norm regularization term $\lambda\|\theta\|_2$ and Lasso regression (see Tibshirani, 1996) performs both θ estimation and variable selection with a L^1-norm regularization term $\lambda\|\theta\|_1$. Because of the representation of f in the function basis $b_1, \ldots, b_q$, the regularization term can be written as a quadratic form of θ:

$$\int_0^1 f''(x)^2 dx = \theta^{\mathrm{T}} S \theta \,,$$

where the positive definite matrix S is known (see, e.g. Lancaster and Šalkauskas, 1986, Gu, 2002 or Wood, 2006). Therefore, the function to minimize becomes $\|Y - X\theta\| + \lambda\theta^{\mathrm{T}} S\theta$. It is continuously differentiable and convex in θ; by canceling its gradient we obtain the penalized least-square estimator

$$\widehat{\theta} = (X^{\mathrm{T}}X + \lambda S)^{-1} X^{\mathrm{T}} Y \,.$$

It only remains to set the hyper-parameter λ. This can be done by using the following generalized cross validation criteria proposed in the monograph by Wood [2006] and implemented in the `R`-package `mgcv`, see Wood [2020]. Let us denote by $\widehat{f}_\lambda^{-t}$, the estimation of f associated with the hyper-parameter λ computed without using the observation (y_t, x_t),

that is, for any $x \in [0,1]$, we have

$$\widehat{f}_\lambda^{-t}(x) = \sum_{\ell=1}^{q} b_\ell(x)\widehat{\theta}_\ell^{-t}(\lambda) \quad \text{with} \quad \widehat{\theta}^{-t}(\lambda) = (X^{-t\mathrm{T}}X^{-t} + \lambda S)^{-1}X^{-t\mathrm{T}}Y^{-t}\,,$$

where we denote by X^{-t} and Y^{-t} the design matrix and the response vector associated with the observations (y_s, x_s), for $s \neq t$, respectively. Then, for any $\lambda > 0$, we consider the leave-one-out cross-validation criteria:

$$\mathrm{CV}(\lambda) = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{f}_\lambda^{-t}(x_i) - y_t\right)^2.$$

The computation of this score requires to fit n models for each λ and is therefore completely inefficient in practice. However, one can show that each term of the above sum may be fairly well approximated by $(\widehat{f}_\lambda(x_t) - y_t)^2/(1 - A_{tt})^2$, where $\widehat{f}_\lambda$ is computed using all the observation (y_t, x_t) and A is the influence matrix $A = X(X^{\mathrm{T}}X + \lambda S)^{-1}X^{\mathrm{T}}$. By approximating each weight $1 - A_{tt}$ by the mean weight $\mathrm{Tr}(I_n - A)/n = (n - \mathrm{Tr}(A))/n$ (with I_n the n-identity matrix), the generalized cross validation score is obtained:

$$\mathrm{GCV}(\lambda) = \frac{n}{\big(n - \mathrm{Tr}(A)\big)^2}\sum_{t=1}^{n}\left(\widehat{f}_\lambda(x_t) - y_t\right)^2,$$

and hyper-parameter λ is chosen by minimizing it.

3.2 Additive models

In this section, we extend the uni-variate model defined above to a multivariate model. By considering several explanatory covariates $x_1, x_2, \ldots$, we now assume that the response variable Y satisfies:

$$Y = f_1(x_1) + f_2(x_2) + \ldots + \text{noise} = \sum_i f_i(x_i) + \varepsilon\,, \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)\,. \tag{3.2}$$

We specify a function basis for each effect. For categorical effects (i.e., if some x_i is a categorical variable taking values $1, 2, \ldots, q_i$), the function is assumed to be a sum of indicators:

$$f_i(x_i) = \sum_{\ell=1}^{q_i} \mathbb{1}_{\{x_i=\ell\}}\theta_\ell^i\,.$$

Otherwise, for continuous variables, the functions f_i may be represented with cubic splines (others functions can be considered like polynomial, B-splines, thin plate splines etc., see Wood, 2006 for further details), that is

$$f_i(x_i) = \sum_{\ell=1}^{q_i} b_\ell^i(x_i)\theta_\ell^i\,.$$

We highlight that the model now contains several functions, which introduces an identifiability problem. Indeed, any constant could be simultaneously added to f_1 and subtracted from f_2 and so on, without changing the resulting model. Before fitting the model, some identifiability constraints have therefore to be imposed. For example, if we consider a set of n observations $(y_t, x_{1,t}, x_{2,t}, \ldots)_{1\leqslant t\leqslant n}$ we may require that for all $i \geqslant 2$, $\sum_{t=1}^{n} f_i(x_{i,t}) = 0$.

For each variable x_i, let us denote by X^i, θ^i and S^i the design matrix, the parameter vector and the regularization matrix associated with the estimation of the effect f_i. We emphasize that for categorical variables, X^i is obtained by one-hot encoding the variable x_i, that is for observation t, the t^{th} line of X^i is

$$\left(\mathbb{1}_{\{x_{i,t}=1\}}, \dots, \mathbb{1}_{\{x_{i,t}=q_i\}}\right),$$

and that S^i is null. For continuous variables, as detailed the previous subsection, the design matrix is made of the lines $(b_1^i(x_{i,t}), \dots, b_{q_i}^i(x_{i,t}))$ and the regularization matrix depends on the regularization term imposed on the smoothness of f_i. For all variables, the parameter vector θ^i is made of the coefficients $\theta_1^i, \dots, \theta_{q_i}^i$. Therefore, to estimate the smooth functions f_i, it is enough to find the vectors θ^i which minimize

$$\left\| Y - \sum_i X^i \theta^i \right\|^2 + \sum_i \lambda_i {\theta^i}^{\mathrm{T}} S^i \theta^i .$$

We now introduce the design matrix X and the regularization matrix S of the additive model:

$$X = \left[X^1 \,|\, X^2 \dots\right] \quad \text{and} \quad S_\lambda = \sum_i \lambda_i \mathbf{S}^i \quad \text{where} \quad \mathbf{S}^i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & S^i & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

With θ, the parameter vector obtained by aggregating the vectors $\theta^1, \theta^2, \dots$, Equation (3.2) yields the linear model

$$Y = X\theta + \varepsilon ,$$

and by minimizing $\|Y - X\theta\|^2 + \sum_i \lambda_i \theta^{\mathrm{T}} \mathbf{S}^i \theta$, we obtain the estimator

$$\widehat{\theta} = \left(X^{\mathrm{T}} X + S_\lambda\right)^{-1} X^{\mathrm{T}} Y .$$

Exactly as in the previous section, the vector $\lambda = (\lambda_1, \lambda_2, \dots)$ is chosen with a generalized cross validation criteria.

We emphasize that, eventually, we may consider cross-effects, namely functions that depend on two (or more) variables x_i and $x_{i'}$. The smooth function $f_{ii'}$ is then approximated by introducing a basis per variable

$$f_{ii'}(x_i, x_{i'}) = \sum_{\ell=1}^{q_i} \sum_{\ell'=1}^{q_i'} b_\ell^i(x_i) b_{\ell'}^{i'}(x_{i'}) \theta_{\ell\ell'}^{ii'} .$$

Exactly as for univariate effects, we obtain a linear expression of the model (see Chapter 4 of Wood, 2006 for further details).

3.3 Generalized additive models [GAM]

A GAM may model even more complex relationships between the response variable Y and its covariates x. Indeed, we recall that it relates Y, a random variable with a specified distribution from the exponential family, to a sum of smooth functions of the covariates x via a link function g:

$$g\big(\mathbb{E}[Y]\big) = f_1(x_1) + f_2(x_2) + f_{34}(x_3, x_4) + \dots, \quad \text{with} \quad Y \sim \text{exponential family distribution.}$$

Generalized linear models (GLMs) derive from linear models, and, similarly, GAMs derive from additive models. While linear models are fitted by least squares, GLMs are estimated with a maximum likelihood estimator that can be found using an iteratively reweighted least squares algorithm (IRLS), see Nelder and Wedderburn [1972] for further details. In the previous section, we showed how additive models were estimated by penalized least squares, and in the same way, GAMs are generally fitted by penalized likelihood maximization. The penalized iteratively re-weighted least squares algorithm (P-IRSL) described and implemented in the perfectly well-documented `R`-package `mgcv` can be used to fit GAMs. As we will mainly use additive models, we do not give more details on this algorithm but we refer to the monograph by Wood [2006] for an exhaustive presentation. However, we emphasize that the P-IRLS approach will be used, through the `gam` function of the `mgcv`-package, in many of our experiments. To use it, we need to choose the distribution of the response variable (binomial, Gaussian, Gamma, Poisson distribution etc.). We highlight this interesting extension: for the Gaussian distribution, it is possible to consider a variance which also depends on the covariates; a second model, for the standard deviation, must then be specified (see `gaulss` function for further details)

4 Application to the Low Carbon London data set

In this last section, thanks to the `mgcv`-package, we fit two generalized additive models on the data set presented in Section 2. The first one is a toy mode while the second one is more complex and will be used in the last section to measure the impact of a tariff change.

4.1 Estimations and predictions of power consumption

Experiment design and assessment of the forecasts. To evaluate the predictions, we introduce the well-known root mean square error (RMSE) and mean absolute percentage of error (MAPE). For n observations $Y_1, \dots Y_n$ and respective forecasts $\widehat{Y}_1, \dots \widehat{Y}_n$, these errors are defined by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2} \quad \text{and} \quad \text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \widehat{Y}_t}{Y_t}\right|.$$

The generalized additive models will be fitted on a sub-sample of the Low Carbon London data, and we will use the remaining observations as a test set (i.e. to evaluate the predictions and to be sure that the models are not over-fitting the training data). More precisely, we split the $T = 48 \times 365 = 17,520$ observations (corresponding to average energy consumption of ToU households, at half-hour intervals, throughout 2013) into a train set and a test set . The test set is composed of 12 weeks, evenly space over 2013, it thus represents approximately a quarter of the observations. The train set consists of the rest of the observations. After fitting a model, for each half-hourly time step $t \in \{1, \dots, T\}$, given the covariates x_t, we can then predict the power consumption denoted by $\widehat{Y}_t$. For both sets, we can then compute the RMSE and the MAPE. These in-sample errors and out-of-sample errors, computed on the train set and the test respectively, assess the quality of the forecast. We recall that for each time step $t \in \{1, \dots, T\}$, the response variable Y_t is the average electricity consumption of the ToU households for the half-hour considered and that the covariates are weather and calendar variables (see Table 3.1).

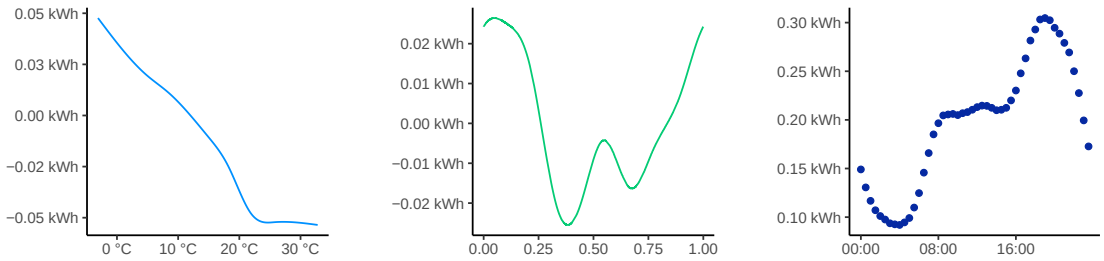


Figure 3.16 – Effects on the energy consumption $\widehat{s}_\tau$ and $\widehat{s}_\pi$ of the temperature (on the left), and of the position in the year (in the middle), respectively; and of the half-hour of the day (on the right), $(\widehat{\alpha}_h)_{0\leqslant h\leqslant 47}$, as estimated under the generalized additive model of Equation (3.3).

A basic model. Firs, the following elementary model is fitted on the data: it only takes into account the effect of the realized temperature τ_t and of the position in the year π_t, which are modeled by splines, and of the half-hour h_t, which is a categorical variable. Therefore, it can be written as

$$Y_t = s_\tau(\tau_t) + s_\pi(\pi_t) + \sum_{h=0}^{47} \alpha_h \mathbb{1}_{\{h_t=h\}} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim \mathcal{N}\big(0, \sigma^2\big)\,, \tag{3.3}$$

where the functions s_τ and s_π are cubic splines and cyclic cubic splines (the function has the same value at its upper and lower boundaries), respectively. The coordinates of s_τ and s_π in their spline bases and the coefficients α_h are estimated on the train set using the function `gam` of the `mgcv` package. Once the model fitted, the resulting estimators are denoted by $\widehat{s}_\tau$, $\widehat{s}_\pi$ and $\widehat{\alpha}_h$. In Figure 3.16, we plotted the coefficients $\widehat{\alpha}_h$ are plotted in function of h, as well as both functions $\widehat{s}_\tau$ and $\widehat{s}_\pi$ (on the range of the values, observed on the train set, of τ_t and π_t, respectively). These effects may be quite well interpreted. Indeed, on the right-most figure, , we can clearly recognize a daily profile of electricity consumption; note that the half-hour effect is significant and varies between approximately 0.1 and 0.3 kWh. In the center of the figure, the plot of the effect of temperature leads to the conclusions drawn from the descriptive analysis, i.e., the lower the temperature, the higher the consumption (due to the termo-sensitivity of the households). From around 22°C, the temperature can continue to rise without any effect on consumption. It should however be noted that in other countries, where the use of air conditioners (AC) is widespread, electricity consumption increases with high temperatures (due to the use of AC in summer). Finally, the position in the year has a smaller effect (between ± 0.02 kWh) on the power consumption; it is negative around May and August, and positive in winter.

We emphasize that in some of the electricity demand literature, the log demand is modeled instead of demand because this allows the covariates to have multiplicative (not additive) effects, see among others Fan and Hyndman [2012]. Here, we tried both modelings and obtained better results with the additive model (3.3).

Once these effects have been estimated, for any half-hourly time step $t = 1, \ldots, T$ of train or test set, the model provides the forecasts

$$\widehat{Y}_t = \widehat{s}_\tau(\tau_t) + \widehat{s}_\pi(\pi_t) + \sum_{h=0}^{47} \widehat{\alpha}_h \mathbb{1}_{\{h_t=h\}}\,.$$

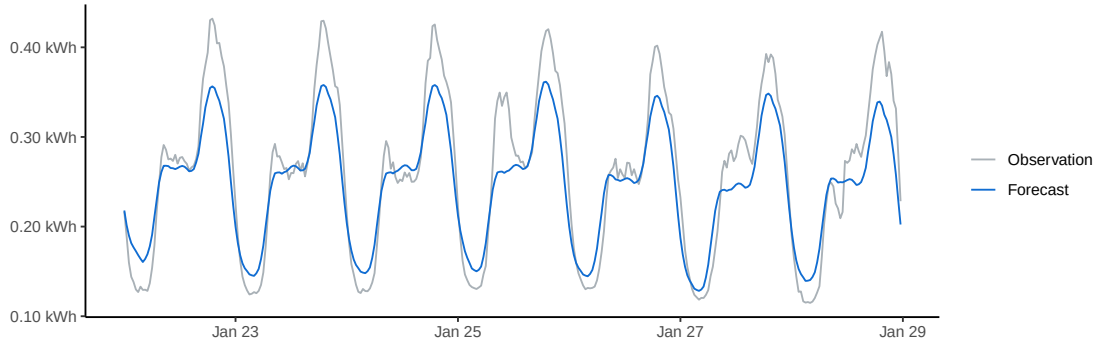


Figure 3.17 – Observations (in grey) and forecast (in blue) provided by the generalized additive model of Equation (3.3) of the average electricity consumption of ToU households from Sunday January 22. to Saturday January 28.

	RMSE (in Wh)	MAPE (in %)
Estimation	25.2	10.91
Prediction	25.9	10.93

Table 3.2 – Root mean squared error (RMSE) and mean absolute percentage of error (MAPE) for the estimation (in-sample forecasts) and for the prediction (out-of-sample forecasts) based on the GAM of Equation (3.3).

For the first week of the test set, both forecasts $\widehat{Y}_t$ (in blue) and observations Y_t (in grey) are plotted in Figure 3.17. While the model correctly catches the daily seasonality, it poorly estimates the drops and peaks around this mean daily consumption and forecasts from a day to an other are quite similar (which is consistent with the fact that we did not consider any "type of day" information). Table 3.2 provides the RMSE and MAPE computed on both train and test sets from the forecasts. Results are clearly improvable, which is why we consider the more complex model below.

A second model. Since the half-hour h_t has a strong impact on the electricity power consumption, it may be more efficient to consider a model per half-hour (see e.g. Fan and Hyndman, 2012 and Goude et al., 2014). Therefore, with the realized and smoothed temperatures τ_t and $\bar{\tau}_t$, the type of day (among Sunday–0, Monday–1 etc.) w_t, the position in the year π_t and the tariff (which is Low, Normal or High) j_t, the additive model to fit on the train set is

$$Y_t = \sum_{h=0}^{47} \Bigg(\xi^h_{\text{Low}} \mathbb{1}_{\{j_t=\text{Low}\}} + \xi^h_{\text{Normal}} \mathbb{1}_{\{j_t=\text{Normal}\}} + \xi^h_{\text{High}} \mathbb{1}_{\{j_t=\text{High}\}} + \sum_{w=0}^{6} \zeta^h_w \mathbb{1}_{\{w_t=w\}} + s^h_\tau(\tau_t) + s^h_{\bar{\tau}}(\bar{\tau}_t) + s^h_\pi(\pi_t) + \varepsilon^h_t \Bigg) \mathbb{1}_{\{h_t=h\}}, \quad (3.4)$$

with $\varepsilon^h_t \sim \mathcal{N}\big(0, \sigma^2(h)\big)$. We therefore consider 48 independent models, related to 48 noise variances $\sigma^2(h)$. The RMSE and MAPE scores are much better (see Table 3.3), which confirms the value of using calendar variables and smoothed temperatures, on one hand, and of considering a mode per half-hour, on the other hand, to forecast the load. Looking at both observations and forecasts plotted in Figure 3.18 for the first week of the test set, there is no doubt that this new model is much better than the previous one. We highlight

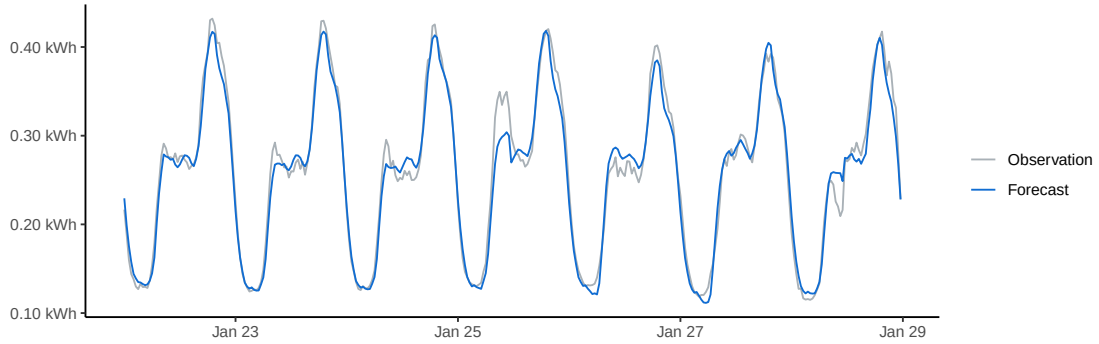


Figure 3.18 – Observations (in grey) and forecasts (in blue) provided by the generalized additive model of Equation (3.4) of the average electricity consumption for ToU households from Sunday January 22. to Saturday January 28.

	RMSE (in Wh)	MAPE (in %)
Estimation	12.6	4.47
Prediction	12.9	4.69

Table 3.3 – Root mean squared error (RMSE) and mean absolute percentage of error (MAPE) for the estimation (in-sample forecasts) and for the prediction (out-of-sample forecasts) based on the GAM of Equation (3.4).

that it is not uncommon that EDF obtains MAPE less than 2% for the forecasting of its entire customer portfolio load and that these forecasts are therefore extremely valuable for the management of the electricity production.

Remark 7. *In Chapter 4, we will assume that, for a time step t, a vector x_t composed of the exogenous variables τ_t, $\bar{\tau}_t$, π_t, w_t and h_t, and a tariff j, the power consumption Y_t follows the linear model*

$$Y_t = \varphi(x_t, j)^{\mathrm{T}}\theta + \text{noise}\,,$$

where the mapping function φ results from an underlying additive model. For example, it may be of the form

$$\varphi(x_t, j_t) = \left(b_1^1(x_1), b_2^1(x_1), \ldots, \mathbb{1}_{\{j_t=\text{Low}\}}, \mathbb{1}_{\{j_t=\text{Normal}\}}, \mathbb{1}_{\{j_t=\text{High}\}}\right).$$

4.2 Measurement of the tariff impact on power consumption

From the model fitted in the previous section, it is possible to measure (in expectation) the impact of a tariff change on the electricity consumption. Indeed, for any covariates h_t, j_t, w_t, π_t, τ_t and $\bar{\tau}_t$, the additive model provides an estimator of the expected consumption:

$$\widehat{Y}_t = \widehat{\xi}_{j_t}^{h_t} + \widehat{\zeta}_{w_t}^{h_t} + s_\tau^{h_t}(\tau_t) + s_{\bar{\tau}}^{h_t}(\bar{\tau}_t) + s_\pi^{h_t}(\pi_t)\,.$$

By replacing, in the above equation, j_t by Low, Normal and High, the model outputs, for any time step t, three estimations, which we denote by $\widehat{Y}_t^{\text{Low}}$, $\widehat{Y}_t^{\text{Normal}}$ and $\widehat{Y}_t^{\text{High}}$ respectively. Figure 3.19 shows the forecasts for a week for which several tariff changes occur. The curve color depends on the tariff (green for Low, blue for Normal and navy for High). We superimpose the forecasts $\widehat{Y}_t^{\text{Normal}}$; we then get an estimated measure of the expected

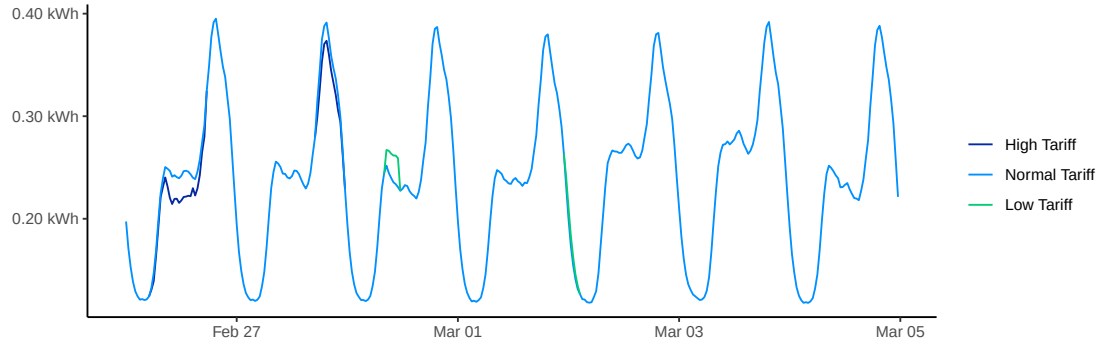


Figure 3.19 – Forecasts (given by the generalized additive model of Equation (3.3) of the average electricity consumption of ToU (in green for Low tariff, blue for Normal tariff and navy for High tariff) households from Sunday February 26. to Saturday March 4.

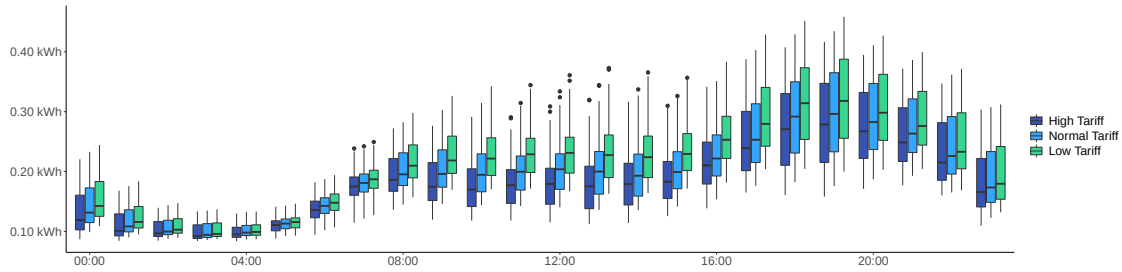


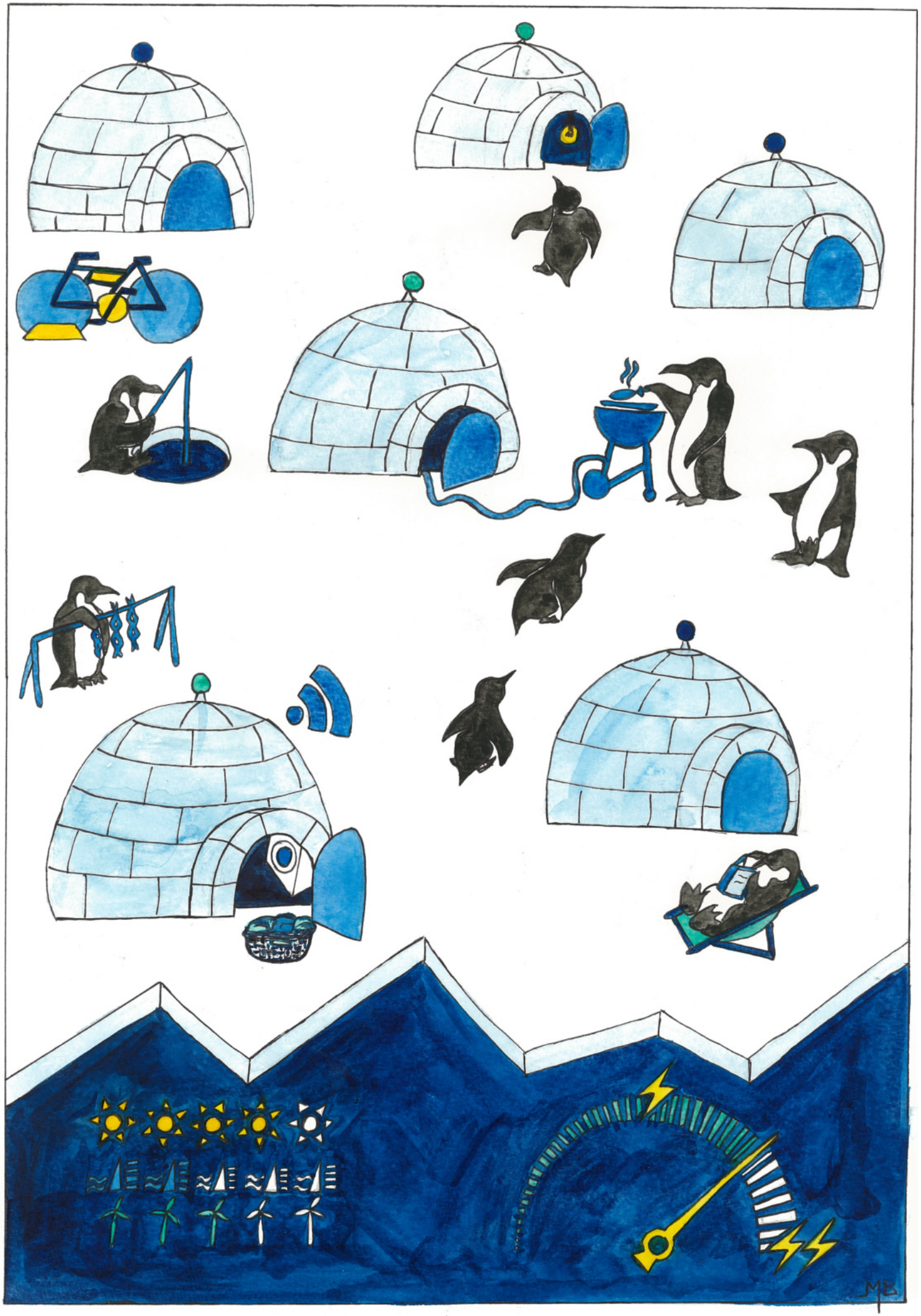
Figure 3.20 – Boxplot of the estimated electricity consumption of ToU households (in green for Low tariff, blue for Normal tariff and navy for High tariff) depending on the hour of the day. Estimations come from the generalized additive model of Equation (3.3).

amountof electricity gained or lost with the tariff change. As expected (and as seems logical), High and Low tariffs lead to lower and higher consumption, respectively. Moreover, it seems that the application of Low tariff during the night of March 1st to 2nd had no effect.

Figure 3.20 provides box-plots of the estimations of the expected average power consumptions $\widehat{Y}_t^{\text{Low}}$, $\widehat{Y}_t^{\text{Normal}}$ and $\widehat{Y}_t^{\text{High}}$ depending on both the tariff and the half-hour. Compared to the ones of Figure 3.12, which were computed with the observed average power consumptions, these box-plots are based on the same number of observations (365, one for each day of 2013). But what is really interesting about this graph is that it frees us from the bias induced by the choice of tariff application times. We recall that the report of the Low Carbon London project (see Schofield et al., 2014) mentioned that High tariff were especially chosen during stress period (for example, when the electricity provider forecasts that the consumption will be very high because of a cold temperature). Consequently, we initially obtained surprising results: during the evening load peak, the electricity consumption associated with High tariff was equal or even higher than that associated with Normal tariff. Now, we can see that for each half-hour of the day, the consumption is lower for High tariff than for Normal tariff. Therefore, our modeling reveals that the application of the tariff has indeed led to a decrease in consumption (as we saw when comparing ToU households to Std households).

We highlight that box-plot associated with tariffs Low and High are actually the trans-

lation of the tariff Normal box-plot by ξ^h_{Low} and ξ^h_{High}, respectively. By including more complex relationships between the tariff and the other covariates, or even by considering a model per tariff, we could obtain sharper results on the distribution of the expected daily power consumption associated with each tariff. Moreover, by using the option `gaulss`-family in the `gam`-function, we could also have estimated the variance of the noise, and especially its dependence on the tariff, within the framework of an additive model. Then, we could have estimated the probability distribution associated with each tariff, for each half-hour of the day. Unfortunately, the number of observations in the data set is not sufficient to implement such solutions. Finally, we point out that even if we had such probability distributions, we would still have to look at the intra-day correlations between the half-hourly power consumptions and the tariffs of the day. Indeed, for a given half-hour, the tariff profile of the whole day influences the associated electricity consumption, and not just the tariff for the half-hour in question. We will further detail the modeling of these dependencies in the last section of the next chapter.


MB

4

Target tracking for contextual bandits: application to demand side management

This chapter proposes 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 such as the temperature, weather, and so on. 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)—and even faster rates under stronger assumptions—for strategies inspired by standard strategies for contextual bandits (like LinUCB, see Li et al., 2010). Simulations on a real data set gathered by UK Power Networks, in which price incentives were offered, show that our strategies are effective and may indeed manage demand response by suitably picking the price levels.

The first five sections of this chapter have been written, published and presented in collaboration with Pierre Gaillard, Yannig Goude and Gilles Stoltz at ICML 2019 (International Conference on Machine Learning).

1 Introduction

Electricity management is classically performed by anticipating demand and adjusting accordingly production. The development of smart grids, and in particular the installation of smart meters (see Yan et al., 2012, Mallet et al., 2014), come with new opportunities: getting new sources of information, offering new services. For example, demand-side management (also called demand-side response; see Albadi and El-Saadany, 2007, Siano, 2014 for an overview) consists of reducing or increasing consumption of electricity users when needed, typically reducing at peak times and encouraging consumption of off-peak times. This is good to adjust to intermittency of renewable energies and is made possible by the development of energy storage devices such as batteries or even electric vehicles (see Fischer et al., 2015, Kikusato et al., 2018); the storages at hand can take place at a convenient moment for the electricity provider. We will consider such a demand-side management system, based on price incentives sent to users via their smart meters. We propose here to adapt contextual bandit algorithms to that end, which are already used in online advertising. Other such systems were based on different heuristics (see Shareef et al., 2018, Wang et al., 2015).

The structure of this chapter is to first provide a modeling of this management system, in Section 2. It relies on making the mean consumption as close as possible to a moving target by sequentially picking price allocations. The literature discussion of the main ingredient of our algorithms, contextual bandit theory, is postponed till Section 2.3. Then, our main results are stated and discussed in Section 3: we control our cumulative loss through a $T^{2/3}$ regret bound with respect to the best constant price allocation. A refinement as far as convergence rates are concerned is offered in Section 4. A section with simulations based on a real data set concludes the chapter: Section 5. For the sake of length, most of the proofs are provided in the supplementary material.

2 Setting and models

Our setting consists of a modeling of electricity consumption and of an aim – tracking a target consumption. Both rely on price levels sent out to the customers.

2.1 Modeling of the electricity consumption

We consider a large population of customers of some electricity provider and assume it homogeneous, which is not an uncommon assumption, see Mei et al. [2017]. The consumption of each customer at each round t depends, among others, on some exogenous factors (temperature, wind, season, day of the week, etc.), which will form a context vector $x_t \in \mathcal{X}$, where $\mathcal{X}$ is some parametric space. The electricity provider aims to manage demand response: it sets a target mean consumption c_t for each time instance. To achieve it, it changes electricity prices accordingly (by making it more expensive to reduce consumption or less expensive to encourage customers to consume more now rather than in some hours). We assume that $K \geqslant 2$ price levels (tariffs) are available. The individual consumption of a given customer getting tariff $j \in \{1, \ldots, K\}$ is assumed to be of the form $\phi(x_t, j) +$ white noise, where the white noise models the variability due to the customers, and where ϕ is some function associating with a context x_t and a tariff j an expected consumption $\phi(x_t, j)$. Details on and examples of ϕ are provided below. At round t, the electricity provider sends tariff j to a share $p_{t,j}$ of the customers; we denote by p_t the convex vector $(p_{t,1}, \ldots, p_{t,K})$. As the population is rather homogeneous, it is unimportant to know to which specific customer a given signal was sent; only the global proportions $p_{t,j}$ matter. The mean consumption observed equals

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

The noise term is to be further discussed below; we first focus on the ϕ function by means of examples.

Example 3: *Linear model.* The simplest approach consists in considering a linear model per price level, i.e., parameters $\theta_1, \ldots, \theta_K \in \mathbb{R}^{\dim(\mathcal{X})}$ with $\phi(x_t, j) = \theta_j^{\mathrm{T}} x_t$. We denote $\theta = (\theta_j)_{1 \leqslant j \leqslant K}$ the vector formed by aggregating all vectors θ_j. This approach can be generalized by replacing x_t by a vector-valued function $b(x_t)$. This corresponds to the case where it is assumed that the $\phi(\,\cdot\,, j)$ belong to some set $\mathcal{H}$ of functions $h : \mathcal{X} \to \mathbb{R}$, with a basis composed of $b_1, \ldots, b_q$. Then, $b = (b_1, \ldots, b_q)$. For instance, $\mathcal{H}$ can be given by histograms on a given grid of $\mathcal{X}$.

Example 4: *Generalized additive models.* Generalized additive models form a powerful and efficient semi-parametric approach to model electricity consumption as a sum of independent exogenous variable effects (see Chapter 3 for further details). In our simulations, see (7.8), we will consider a mean expected consumption of the form $\phi(x_t, j) = \phi(x_t, 0) + \xi_j$, that is, the tariff will have a linear impact on the mean consumption, independently of the contexts. The baseline mean consumption $\phi(x_t, 0)$ will be modeled as a sum of simple $\mathbb{R} \to \mathbb{R}$ functions, each taking as input a single component of the context vector:

$$\phi(x_t, 0) = \sum_{i=1}^{Q} f^{(i)}(x_{t,h(i)})\,,$$

where $Q \geqslant 1$ and where each $h(i) \in \big\{1, \ldots, \dim(\mathcal{X})\big\}$. Some components $h(i)$ may be used several times. When the considered component $x_{t,h(i)}$ takes continuous values, these

functions $f^{(i)}$ are so-called cubic splines: $\mathcal{C}^2$–smooth functions made up of sections of cubic polynomials joined together at points of a grid (the knots). Choosing the number q_i of knots (points at which the sections join) and their locations is sufficient to determine (in closed form) a linear basis $\left(b_1^{(i)}, \ldots, b_{q_i}^{(i)}\right)$ of size q_i, see Chapter 3 for details. The function $f^{(i)}$ can then be represented on this basis by a vector of length q_i, denoted by $\theta^{(i)}$:

$$f^{(i)} = \sum_{j=1}^{q_i} \theta_j^{(i)} b_j^{(i)} .$$

When the considered component $x_{t,h(i)}$ takes finitely many values, we write $f^{(i)}$ as a sum of indicator functions:

$$f^{(i)} = \sum_{j=1}^{q_i} \theta_j^{(i)} \mathbb{1}_{\{v_j^{(i)}\}} ,$$

where the $v_j^{(i)}$ are the q_i modalities for the component $h(i)$. All in all, $\phi(x_t, j)$ can be represented by a vector of dimension $K + q_1 + \ldots + q_Q$ obtained by aggregating the ξ_j and the vectors $\theta^{(i)}$ into a single vector.

Both examples above show that it is reasonable to assume that there exists some unknown $\theta \in \mathbb{R}^d$ and some known transfer function φ such that $\phi(x_t, j) = \varphi(x_t, j)^{\mathrm{T}}\theta$. By linearly extending φ in its second component, we get

$$Y_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta + \text{noise} .$$

We will actually not use in the sequel that $\varphi(x, p)$ is linear in p: the dependency of $\varphi(x, p)$ in p could be arbitrary.

We consider the two following models that only differ by the noise term. In Model 1, we make the assumption that the variance of the noise varies from a tariff to another: non-standard pricing often leads to higher variability – power consumption data confirm this see 5.1). The second modeling assume that the global noise term does not depend on the picked tariffs p_t.

2.1.1 Modeling of the consumption with a tariff-dependent noise

We first recall that we assumed that our population is rather homogeneous, which is a natural feature as soon as it is large enough. Therefore, we may assume that the variabilities within the group of customers getting the same tariff j can be combined into a single random variable $\varepsilon_{t,j}$. We denote by ε_t the vector $(\varepsilon_{t,1}, \ldots, \varepsilon_{t,K})$. All in all, we will mainly consider the following model.

Model 1: *Tariff-dependent noise. When the electricity provider picks the convex vector p, the mean consumption obtained at round t equals*

$$Y_{t,p} = \varphi(x_t, p)^{\mathrm{T}}\theta + p^{\mathrm{T}}\varepsilon_t .$$

The noise vectors $\varepsilon_1, \varepsilon_2, \ldots$ are ρ–sub-Gaussian[1] *i.i.d. random variables with $\mathbb{E}[\varepsilon_1] = (0, \ldots, 0)^{\mathrm{T}}$. We denote by $\Sigma = \mathrm{Var}(\varepsilon_1)$ their covariance matrix.*

[1] For $\rho > 0$, a d–dimensional random vector ε is ρ–sub-Gaussian if for all $\nu \in \mathbb{R}^d$, $\mathbb{E}\left[\mathrm{e}^{\nu^{\mathrm{T}}\varepsilon}\right] \leqslant \mathrm{e}^{\rho^2\|\nu\|^2/2}$.

No assumption is made on Σ in the model above (real data confirms that Σ typically has no special form, see 5.1). However, when it is proportional to the $K \times K$ matrix [1], the noises associated with each group can be combined into a global noise, leading to the following model. It is less realistic in practice, but we discuss it because regret bounds may be improved in the presence of a global noise.

2.1.2 Modeling of the consumption with a global noise

Model 2: *Global noise. When the electricity provider picks the convex vector p, the mean consumption obtained at time instance t equals*

$$Y_{t,p} = \varphi(x_t, p)^{\mathrm{T}}\theta + \mathrm{e}_t \,.$$

The scalar noises $\mathrm{e}_1, \mathrm{e}_2, \ldots$ are ρ–sub-Gaussian i.i.d. random variables, with $\mathbb{E}[\mathrm{e}_1] = 0$. We denote by $\sigma^2 = \mathrm{Var}(\mathrm{e}_1)$ the variance of the random noises e_t.

2.2 Tracking a Target Consumption

We now move on to the aim of the electricity provider. At each time instance t, it picks an allocation of price levels p_t and wants the observed mean consumption Y_{t,p_t} to be as close as possible to some target mean consumption c_t. This target is set in advance by another branch of the provider and p_t is to be picked based on this target: our algorithms will explain how to pick p_t given c_t but will not discuss the choice of the latter. In this article we will measure the discrepancy between the observed Y_{t,p_t} and the target c_t via a quadratic loss: $(Y_{t,p_t} - c_t)^2$. We may set some restrictions on the convex combinations p that can be picked: we denote by $\mathcal{P}$ the set of legible allocations of price levels. This models some operational or marketing constraints that the electricity provider may encounter. We will see that whether $\mathcal{P}$ is a strict subset of all convex vectors or whether it is given by the set of all convex vectors plays no role in our theoretical analysis.

As explained in Section 3.1, we will follow a standard path in online learning theory: to minimize the cumulative loss suffered we will minimize some regret.

After picking an allocation of price levels p_t, the electricity provider only observes Y_{t,p_t}: it thus faces a bandit monitoring. Because of the contexts x_t, the problem considered falls under the umbrella of contextual bandits. No stochastic assumptions are made on the sequences x_t and c_t: the contexts x_t and c_t will be considered as picked by the environment. Finally, mean consumptions are assumed to be bounded between 0 and C, where C is some known maximal value. The online protocol described in Sections 2.1 and 2.2 is stated in Protocol 4. We see that the choices x_t, c_t and p_t need to be $\mathcal{F}_{t-1}$–measurable, where $\mathcal{F}_{t-1} \triangleq \sigma(\varepsilon_1, \ldots, \varepsilon_{t-1})$.

2.3 Literature Discussion: Contextual Bandits

In many bandit problems the learner has access to additional information at the beginning of each round. Several settings for this side information may be considered. The adversarial case was introduced in Auer et al. [2002b, Section 7, algorithm Exp4]: and subsequent improvements were suggested in Beygelzimer et al. [2011] and McMahan and Streeter [2009]. The case of i.i.d. contexts with rewards depending on contexts through an

Protocol 4 Target Tracking for Contextual Bandits

Input
Parametric context set $\mathcal{X}$
Set of legible convex weights $\mathcal{P}$
Bound on mean consumptions C
Transfer function $\varphi : \mathcal{X} \times \mathcal{P} \to \mathbb{R}^d$
Unknown parameters
Transfer parameter $\theta \in \mathbb{R}^d$
Covariance matrix Σ of size $K \times K$ (Model 1)
Variance σ^2 (Model 2)
for $t = 1, 2, \ldots$ **do**
Observe a context $x_t \in \mathcal{X}$ and a target $c_t \in (0, C)$
Choose an allocation of price levels $p_t \in \mathcal{P}$
Observe a resulting mean consumption

$$Y_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t \quad \text{(Model 1)}$$
$$Y_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta + \mathrm{e}_t \quad \text{(Model 2)}$$

Suffer a loss $(Y_{t,p_t} - c_t)^2$
end for
Aim
Minimize the cumulative loss $L_T = \sum_{t=1}^{T} (Y_{t,p_t} - c_t)^2$

unknown parametric model was introduced by Wang et al. [2005b] and generalized to the non-i.i.d. setting in Wang et al. [2005a], then to the multivariate and nonparametric case in Perchet and Rigollet [2013]. Hybrid versions (adversarial contexts but stochastic dependencies of the rewards on the contexts, usually in a linear fashion) are the most popular ones. They were introduced by Abe and Long [1999] and further studied in Auer [2002]. A key technical ingredient to deal with them is confidence ellipsoids on the linear parameter; see Dani et al. [2008], Rusmevichientong and Tsitsiklis [2010] and Abbasi-Yadkori et al. [2011]. The celebrated UCB algorithm of Lai and Robbins [1985] was generalized in this hybrid setting as the LinUCB algorithm, by Li et al. [2010] and Chu et al. [2011]. Later, Filippi et al. [2010] extended it to a setting with generalized additive models and Valko et al. [2013] proposed a kernelized version of UCB. Other approaches, not relying on confidence ellipsoids, consider sampling strategies (see Gopalan et al., 2014) and are currently extended to bandit problems with complicated dependency in contextual variables [Mannor, 2018]. Our model falls under the umbrella of hybrid versions considering stochastic linear bandit problems given a context. The main difference of our setting lies in how we measure performance: not directly with the rewards or their analogous quantities Y_{t,p_t} in our setting, but through how far away they are from the targets c_t.

3 A regret bound with tariff-dependent noise modeling

This section considers Model 1. We take inspiration from LinUCB (Li et al., 2010, Chu et al., 2011): given the form of the observed mean consumption, the key is to estimate the parameter θ. Denoting by I_d the $d \times d$ identity matrix and picking $\lambda > 0$, we classically

do so according to

$$\widehat{\theta}_t \triangleq V_t^{-1} \sum_{s=1}^{t} Y_{s,p_s} \varphi(x_s, p_s) \quad \text{where} \quad V_t \triangleq \lambda \mathrm{I}_d + \sum_{s=1}^{t} \varphi(x_s, p_s) \varphi(x_s, p_s)^{\mathrm{T}} . \tag{4.1}$$

3.1 Regret as a proxy for minimizing losses

We are interested in the cumulative sum of the losses, but under suitable assumptions (e.g., bounded noise) the latter is close to the sum of the conditionally expected losses (e.g., through the Hoeffding–Azuma inequality). Typical statements are of the form: for all strategies of the provider and of the environment, with probability at least $1 - \delta$,

$$L_T = \textstyle\sum_{t=1}^{T} (Y_{t,p_t} - c_t)^2 \leqslant \sum_{t=1}^{T} \mathbb{E}\big[(Y_{t,p_t} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big] + \mathcal{O}\big(\sqrt{T \ln(1/\delta)}\big).$$

All regret bounds in the sequel will involve the sum of conditionally expected losses $\overline{L}_T$ above but up to adding a deviation term to all these regret bounds, we get from them a bound on the true cumulative loss L_T. Now, the choices x_t, c_t and p_t are $\mathcal{F}_{t-1}$–measurable, where $\mathcal{F}_{t-1} = \sigma(\varepsilon_1, \dots, \varepsilon_{t-1})$. Therefore, under Model 1,

$$\begin{aligned}
\mathbb{E}\big[(Y_{t,p_t} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big] &= \mathbb{E}\Big[\big(\varphi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t - c_t\big)^2 \,\Big|\, \mathcal{F}_{t-1}\Big] \\
&= \big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + \mathbb{E}\big[(p_t^{\mathrm{T}}\varepsilon_t)^2 \,\big|\, \mathcal{F}_{t-1}\big] + \mathbb{E}\Big[2\big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big) p_t^{\mathrm{T}}\varepsilon_t \,\Big|\, \mathcal{F}_{t-1}\Big] \\
&= \big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}}\Sigma p_t \,,
\end{aligned} \tag{4.2}$$

that is, after summing,

$$\overline{L}_T = \sum_{t=1}^{T} \big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}}\Sigma p_t \,.$$

We therefore introduce the (conditional) regret

$$\overline{R}_T = \sum_{t=1}^{T} \big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}}\Sigma p_t - \sum_{t=1}^{T} \min_{p \in \mathcal{P}} \Big\{ \big(\varphi(x_t, p)^{\mathrm{T}}\theta - c_t\big)^2 + p^{\mathrm{T}}\Sigma p \Big\} .$$

This will be the quantity of interest in the sequel[2].

3.2 Optimistic algorithm

3.2.1 All but the estimation of the covariance matrix

We assume that in the first τ rounds an estimator $\widehat{\Sigma}_\tau$ of the covariance matrix Σ was obtained; details are provided in the next subsection. We explain here how the algorithm plays for rounds $t \geqslant \tau + 1$. We assumed that the transfer function φ and the bound $C > 0$ on the target mean consumptions were known. We use the notation $[x]_C = \min\big\{\max\{x, 0\}, C\big\}$ for the clipped part of a real number x (clipping between 0 and C). We then estimate the instantaneous losses (4.2)

$$\ell_{t,p} \triangleq \mathbb{E}\big[(Y_{t,p} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big] = \big(\varphi(x_t, p)^{\mathrm{T}}\theta - c_t\big)^2 + p^{\mathrm{T}}\Sigma p$$

[2] With the definition in Chapter 2, this quantity is a "pseudo-regret"; for convenience, let call it "regret".

associated with each choice $p \in \mathcal{P}$ by:

$$\widehat{\ell}_{t,p} = \Big(\big[\varphi(x_t, p)^{\mathrm{T}} \widehat{\theta}_{t-1} \big]_C - c_t \Big)^2 + p^{\mathrm{T}} \widehat{\Sigma}_\tau p \,.$$

We also denote by $\alpha_{t,p}$ deviation bounds, to be set by the analysis. The optimistic algorithm picks, for $t \geqslant \tau + 1$:

$$p_t \in \operatorname*{argmin}_{p \in \mathcal{P}} \big\{ \widehat{\ell}_{t,p} - \alpha_{t,p} \big\} \,. \tag{4.3}$$

Remark 8. *In linear contextual bandits, rewards are linear in θ and to maximize global gain, LinUCB Li et al. [2010] picks a vector p which maximizes a sum of the form $\varphi(x_t, p)^{\mathrm{T}} \widehat{\theta}_{t-1} + \widetilde{\alpha}_{t,p}$. Here, as we want to track the target, we slightly change this expression by substituting the target c_t and taking a quadratic loss. But the spirit is similar.*

3.2.2 Optimistic algorithm: estimation of the covariance matrix

The estimation of the covariance matrix Σ is hard to perform (on the fly and simultaneously) as the algorithm is running. We leave this problem for future research and devote here the first τ rounds to this estimation. We created from scratch the estimation of Σ proposed below and studied in Lemma 5, as we could find no suitable result in the literature.

For each pair

$$(i,j) \in E \triangleq \big\{ (i,j) \in \{1, \dots, K\}^2 : 1 \leqslant i \leqslant j \leqslant K \big\}$$

we define the weight vector $p^{(i,j)}$ as: for $k \in \{1, \dots, K\}$,

$$p_k^{(i,j)} = \begin{cases} 1 & \text{if } k = i = j, \\ 1/2 & \text{if } k \in \{i,j\} \text{ and } i \neq j, \\ 0 & \text{if } k \notin \{i,j\}. \end{cases}$$

These correspond to all weights vectors that either assign all the mass to a single component, like the $p^{(i,i)}$, or share the mass equally between two components, like the $p^{(i,j)}$ for $i \neq j$. There are $K(K+1)/2$ different weight vectors considered. We order these weight vectors, e.g., in lexicographic order, and use them one after the other, in order. This implies that in the initial exploration phase of length τ, each vector indexed by E is selected at least

$$\tau_0 \triangleq \left\lfloor \tfrac{2\tau}{K(K+1)} \right\rfloor$$

times. At the end of the exploration period, we define $\widehat{\theta}_\tau$ as in (4.1) and the estimator

$$\widehat{\Sigma}_\tau \in \operatorname*{argmin}_{\widehat{\Sigma} \in \mathcal{M}_K(\mathbb{R})} \sum_{t=1}^{\tau} \big(\widehat{Z}_t^2 - p_t^{\mathrm{T}} \widehat{\Sigma} p_t \big)^2 \,, \quad \text{where} \quad \widehat{Z}_t \triangleq Y_{t,p_t} - \big[\varphi(x_t, p_t)^{\mathrm{T}} \widehat{\theta}_\tau \big]_C . \tag{4.4}$$

Note that $\widehat{\Sigma}_\tau$ can be computed efficiently by solving a linear system as soon as K is small enough.

Remark 9. *We implicitly assume that $\mathcal{P}$ contains all the $p_k^{(i,j)}$.*

3.3 Statement of the regret bound

Assumption 1 – *Boundedness assumptions.* They are all linked to the knowledge that the mean consumption lies in $(0, C)$ and indicate some normalization of the modeling:

$$\|\varphi\|_\infty \leqslant 1\,, \qquad \|\theta\|_\infty \leqslant C\,, \qquad \varphi^{\mathrm{T}}\theta \in [0, C]\,.$$

As a consequence of these boundedness assumptions, $\|\theta\| \leqslant \sqrt{d}\,C$ and all eigenvalues of V_t lie in $[\lambda, \lambda + t]$, thus

$$\ln\bigl(\det(V_t)\bigr) \in \bigl[d\ln\lambda, d\ln(\lambda + t)\bigr]\,.$$

Finally, we also assume that a bound Γ is known, such that $\forall p \in \mathcal{P}$, $p^{\mathrm{T}}\Sigma p \leqslant \Gamma$. A last consequence of all these boundedness assumptions is that $L \triangleq C^2 + \Gamma$ upper bounds the (conditionally) expected losses $\ell_{t,p} = \bigl(\varphi(x_t, p)^{\mathrm{T}}\theta - c_t\bigr)^2 + p^{\mathrm{T}}\Sigma p$.

Theorem 4. *Fix a risk level $\delta \in (0,1)$ and a time horizon $T \geqslant 1$. Assume that Assumption 1 holds. The optimistic algorithm* (4.3) *with an initial exploration of length $\tau = \mathcal{O}(T^{2/3})$ rounds satisfies*

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

with probability at least $1 - \delta$.

When the covariance matrix Σ is known, no initial exploration is required and the regret bound improves to $\mathcal{O}(\sqrt{T}\ln T)$ as far as the orders of magnitude in T are concerned. These improved rates might be achievable even if Σ is unknown, through a more efficient, simultaneous, estimation of Σ and θ (an issue we leave for future research, as already mentioned at the beginning of Section 3.2.2).

We emphasize that the expected losses depend on both Σ and θ, which must therefore be estimated correctly. The regret bound derives from the two deviation inequalities on the estimators $\widehat{\theta}_t$ and $\widehat{\Sigma}_\tau$ (defined in Equations (4.1) and (4.4)), which are proved in Sections 3.3.1 and 3.3.2, respectively. The proof of the theorem is then stated in Section 3.3.3.

3.3.1 Deviation inequality on the parameter vector estimation

A straightforward adaptation of earlier results (see Theorem 2 of Abbasi-Yadkori et al., 2011 or Theorem 20.2 in the monograph by Lattimore and Szepesvári, 2020) yields the following deviation inequality.

Lemma 4. *No matter how the provider picks the p_t, we have, for all $t \geqslant 1$ and all $\delta \in (0,1)$,*

$$\sqrt{\bigl(\widehat{\theta}_t - \theta\bigr)^{\mathrm{T}} V_t \bigl(\widehat{\theta}_t - \theta\bigr)} \triangleq \bigl\|V_t^{1/2}\bigl(\widehat{\theta}_t - \theta\bigr)\bigr\| \leqslant \sqrt{\lambda}\|\theta\| + \rho\sqrt{2\ln\frac{1}{\delta} + d\ln\frac{1}{\lambda} + \ln\det(V_t)}\,,$$

with probability at least $1 - \delta$.

Actually, the result above could be improved into an anytime result ("with probability $1-\delta$, for all $t \geqslant 1$, ...") with no effort, by applying a stopping argument (or, alternatively, Doob's inequality for super-martingales), as Abbasi-Yadkori et al. [2011] did. This would slightly improve the regret bounds below by logarithmic factors. The deviation bound of Lemma 4 plays a key role in the algorithm. We introduce the following upper bound on it:

$$B_t(\delta) \triangleq \sqrt{\lambda d}\, C + \rho\sqrt{2\ln\tfrac{1}{\delta} + d\ln\left(1+\tfrac{t}{\lambda}\right)}\,. \tag{4.5}$$

Proof of Lemma 4. The proof below relies on Laplace's method on super-martingales, which is a standard argument to provide confidence bounds on a self-normalized sum of conditionally centered random vectors. See Theorem 2 of Abbasi-Yadkori et al. [2011] or Theorem 20.2 in the monograph by Lattimore and Szepesvári [2020].

Under Model 1 and given the definition of V_t, we have the rewriting

$$\begin{aligned}\widehat{\theta}_t &= V_t^{-1}\sum_{s=1}^{t}\varphi(x_s,p_s)Y_{s,p_s} = V_t^{-1}\sum_{s=1}^{t}\varphi(x_s,p_s)\big(\varphi(x_s,p_s)^{\mathrm{T}}\theta + p_s^{\mathrm{T}}\varepsilon_s\big)\\ &= V_t^{-1}\big((V_t-\lambda \mathrm{I}_d)\theta + M_t\big) = \theta - \lambda V_t^{-1}\theta + V_t^{-1}M_t\,,\end{aligned}$$

where we introduced

$$M_t = \sum_{s=1}^{t}\varphi(x_s,p_s)p_s^{\mathrm{T}}\varepsilon_s\,,$$

which is a martingale with respect to $\mathcal{F}_t = \sigma(\varepsilon_1,\dots,\varepsilon_t)$. Thus, by a triangle inequality,

$$\big\|V_t^{1/2}\big(\widehat{\theta}_t-\theta\big)\big\| = \big\|-\lambda V_t^{-1/2}\theta + V_t^{-1/2}M_t\big\| \leqslant \lambda\big\|V_t^{-1/2}\theta\big\| + \big\|V_t^{-1/2}M_t\big\|\,.$$

On the one hand, given that all eigenvalues of the symmetric matrix V_t are larger than λ (given the $\lambda \mathrm{I}_d$ term in its definition), all eigenvalues of $V_t^{-1/2}$ are smaller than $1/\sqrt{\lambda}$ and thus,

$$\lambda\big\|V_t^{-1/2}\theta\big\| \leqslant \lambda\,\frac{1}{\sqrt{\lambda}}\|\theta\| = \sqrt{\lambda}\|\theta\|\,.$$

We now prove, on the other hand, that with probability at least $1-\delta$,

$$\big\|V_t^{-1/2}M_t\big\| \leqslant \rho\sqrt{2\ln\frac{1}{\delta} + d\ln\frac{1}{\lambda} + \ln\det(V_t)}\,,$$

which will conclude the proof of the lemma.

⋆ *Introducing super-martingales.* For all $\nu \in \mathbb{R}^d$, we consider

$$S_{t,\nu} = \exp\left(\nu^{\mathrm{T}}M_t - \frac{\rho^2}{2}\nu^{\mathrm{T}}V_t\nu\right)$$

and now show that it is an $\mathcal{F}_t$–super-martingale. First, note that since the common distribution of the $\varepsilon_1,\varepsilon_2,\dots$ is ρ–sub-Gaussian, then for all $\mathcal{F}_{t-1}$–measurable random vectors ν_{t-1},

$$\mathbb{E}\left[\mathrm{e}^{\nu_{t-1}^{\mathrm{T}}\varepsilon_t}\,\middle|\,\mathcal{F}_{t-1}\right] \leqslant \mathrm{e}^{\rho^2\|\nu_{t-1}\|^2/2}\,. \tag{4.6}$$

Now,

$$S_{t,\nu} = S_{t-1,\nu}\,\exp\left(\nu^{\mathrm{T}}\varphi(x_t,p_t)p_t^{\mathrm{T}}\varepsilon_t - \frac{\rho^2}{2}\nu^{\mathrm{T}}\varphi(x_t,p_t)\varphi(x_t,p_t)^{\mathrm{T}}\nu\right)$$

where, by using the sub-Gaussian assumption (4.6) and the fact that $\sum_j p_{j,t}^2 \leqslant 1$ for all convex weight vectors p_t,

$$\mathbb{E}\Big[\exp\big(\nu^{\mathrm{T}}\varphi(x_t,p_t)p_t^{\mathrm{T}}\varepsilon_t\,\Big|\,\mathcal{F}_{t-1}\Big] \leqslant \exp\left(\frac{\rho^2}{2}\nu^{\mathrm{T}}\varphi(x_t,p_t)\underbrace{p_t^{\mathrm{T}}p_t}_{\leqslant 1}\varphi(x_t,p_t)^{\mathrm{T}}\nu\right).$$

This implies $\mathbb{E}\big[S_{t,\nu}\big|\mathcal{F}_{t-1}\big] \leqslant S_{t-1,\nu}$.

Note that the rewriting of $S_{t,\nu}$ in its vertex form is, with $m = V_t^{-1}M_t/\rho^2$:

$$\begin{aligned} S_{t,\nu} &= \exp\left(\frac{1}{2}(\nu-m)^{\mathrm{T}}\,\rho^2 V_t\,(\nu-m) + \frac{1}{2}m^{\mathrm{T}}\rho^2 V_t\,m\right) \\ &= \exp\left(\frac{1}{2}(\nu-m)^{\mathrm{T}}\,\rho^2 V_t\,(\nu-m)\right) \times \exp\left(\frac{1}{2\rho^2}\big\|V_t^{-1/2}M_t\big\|^2\right). \end{aligned}$$

⋆ *Laplace's method—integrating $S_{t,\nu}$ over $\nu \in \mathbb{R}^d$.* The basic observation behind this method is that (given the vertex form) $S_{t,\nu}$ is maximal at $\nu = m = V_t^{-1}M_t/\rho^2$ and then equals $\exp\big(\big\|V_t^{-1/2}M_t\big\|^2/(2\rho^2)\big)$, which is (a transformation of) the quantity to control. Now, because the exp function quickly vanishes, the integral over $\nu \in \mathbb{R}^d$ is close to this maximum. We therefore consider

$$\overline{S}_t = \int_{\mathbb{R}^d} S_{t,\nu}\mathrm{d}\nu\,.$$

We will make repeated uses of the fact that the Gaussian density functions,

$$\nu \longmapsto \frac{1}{\sqrt{\det(2\pi C)}}\exp\left((\nu-m)^{\mathrm{T}}C^{-1}(\nu-m)\right),$$

where $m \in \mathbb{R}^d$ and C is a (symmetric) positive-definite matrix, integrate to 1 over $\mathbb{R}^d$. This gives us first the rewriting

$$\overline{S}_t = \sqrt{\det\big(2\pi\rho^{-2}V_t^{-1}\big)}\,\exp\left(\frac{1}{2\rho^2}\big\|V_t^{-1/2}M_t\big\|^2\right).$$

Second, by the Fubini-Tonelli theorem and the super-martingale property

$$\mathbb{E}\big[S_{t,\nu}\big] \leqslant \mathbb{E}\big[S_{0,\nu}\big] = \exp\big(-\lambda\rho^2\|\nu\|^2/2\big)\,,$$

we also have

$$\mathbb{E}\big[\overline{S}_t\big] \leqslant \int_{\mathbb{R}^d}\exp\big(-\lambda\rho^2\|\nu\|^2/2\big)\mathrm{d}\nu = \sqrt{\det\big(2\pi\rho^{-2}\lambda^{-1}\mathrm{I}_d\big)}\,.$$

Combining the two statements, we proved

$$\mathbb{E}\left[\exp\left(\frac{1}{2\rho^2}\big\|V_t^{-1/2}M_t\big\|^2\right)\right] \leqslant \sqrt{\frac{\det\big(V_t\big)}{\lambda^d}}\,.$$

⋆ *Markov-Chernov bound.* For $u > 0$,

$$\begin{aligned} \mathbb{P}\Big[\big\|V_t^{-1/2}M_t\big\| > u\Big] = \mathbb{P}\left[\frac{1}{2\rho^2}\big\|V_t^{-1/2}M_t\big\|^2 > \frac{u^2}{2\rho^2}\right] &\leqslant \exp\left(-\frac{u^2}{2\rho^2}\right)\mathbb{E}\left[\exp\left(\frac{1}{2\rho^2}\big\|V_t^{-1/2}M_t\big\|^2\right)\right] \\ &\leqslant \exp\left(-\frac{u^2}{2\rho^2}+\frac{1}{2}\ln\frac{\det\big(V_t\big)}{\lambda^d}\right) = \delta \end{aligned}$$

for the claimed choice

$$u = \rho\sqrt{2\ln\frac{1}{\delta} + d\ln\frac{1}{\lambda} + \ln\det(V_t)}\,.$$

□

3.3.2 Deviation inequality on the covariance matrix estimation

Lemma 5. *For all $\delta \in (0,1)$, the estimator (4.4) satisfies: with probability at least $1-\delta$,*

$$\sup_{p\in\mathcal{P}}\left|p^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)p\right| \leqslant (K+8)\kappa_\tau\sqrt{\tau}/\tau_0 = \mathcal{O}(\kappa_\tau/\sqrt{\tau}) = \mathcal{O}\left(\frac{1}{\sqrt{\tau}}\ln^2(\tau/\delta)\sqrt{\ln(1/\delta)}\right),$$

where we recall that $\tau_0 = \lfloor 2\tau/(K(K+1))\rfloor$ and where, $\kappa_\tau = \big(C+2M_\tau\big)B_\tau(\delta/3) + M'_\tau$ with

$$M_\tau \triangleq \rho/2 + \ln(6\tau/\delta) \quad \text{and} \quad M'_\tau \triangleq M_\tau^2\sqrt{2\ln(3K^2/\delta)} + 2\sqrt{\exp(2\rho)\delta/6}.$$

For simplicity of notation, we introduce the following upper bound:

$$\upsilon_\tau(\delta) \triangleq (K+8)\kappa_\tau\sqrt{\tau}/\tau_0. \tag{4.7}$$

We derived the proof scheme of Lemma 5 from scratch as we could find no suitable result in the literature for estimating Σ in our context. We first consider the following auxiliary result.

Lemma 6. *Let $\tau \geqslant 1$. Assume that the common distribution of the $\varepsilon_1, \varepsilon_2, \ldots$ is ρ–sub-Gaussian. Then, no matter how the provider picks the p_t, we have, for all $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\left\|\sum_{t=1}^{\tau} p_t p_t^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)p_t p_t^{\mathrm{T}}\right\|_\infty \leqslant \kappa_\tau\sqrt{\tau}\,, \quad \text{with} \quad \kappa_\tau \triangleq \big(C+2M_\tau\big)B_\tau(\delta/3) + M'_\tau.$$

Proof of Lemma 6. We can show that $\widehat{\Sigma}_\tau$ defined in (4.4) satisfies

$$\sum_{t=1}^{\tau} p_t p_t^{\mathrm{T}}\widehat{\Sigma}_\tau p_t p_t^{\mathrm{T}} = \sum_{t=1}^{\tau}\widehat{Z}_t^2 p_t p_t^{\mathrm{T}}\,, \tag{4.8}$$

where we recall that $\widehat{Z}_t \triangleq Y_{t,p_t} - \big[\varphi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_\tau\big]_C$. Indeed, with,

$$\Phi\big(\widehat{\Sigma}\big) \triangleq \sum_{t=1}^{\tau}\Big(\widehat{Z}_t^2 - p_t^{\mathrm{T}}\widehat{\Sigma}p_t\Big)^2 = \sum_{t=1}^{\tau}\Big(\widehat{Z}_t^2 - \mathrm{Tr}\big(\widehat{\Sigma}p_t p_t^{\mathrm{T}}\big)\Big)^2,$$

using $\nabla_A \mathrm{Tr}(AB) = B$, we get

$$\nabla_{\widehat{\Sigma}}\Phi\big(\widehat{\Sigma}\big) = \sum_{t=1}^{\tau} 2p_t p_t^{\mathrm{T}}\Big(\widehat{Z}_t^2 - p_t^{\mathrm{T}}\widehat{\Sigma}p_t\Big),$$

which leads to (4.8) by canceling the gradient and keeping in mind that $p_t^{\mathrm{T}}\widehat{\Sigma}p_t$ is a scalar value. Let us denote, for all $t \geqslant 1$

$$Z_t \triangleq Y_{t,p_t} - \varphi(x_t,p_t)^{\mathrm{T}}\theta = p_t^{\mathrm{T}}\varepsilon_t\,.$$

To prove the lemma, we replace $\widehat{\Sigma}_\tau$ by using (4.8) and apply a triangular inequality:

$$\Big\| \sum_{t=1}^{\tau} p_t p_t^{\mathrm{T}} \big(\widehat{\Sigma}_\tau - \Sigma\big) p_t p_t^{\mathrm{T}} \Big\|_\infty \leqslant \Big\| \sum_{t=1}^{\tau} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \Big\|_\infty + \Big\| \sum_{t=1}^{\tau} Z_t^2 p_t p_t^{\mathrm{T}} - p_t p_t^{\mathrm{T}} \Sigma p_t p_t^{\mathrm{T}} \Big\|_\infty \tag{4.9}$$

We will consecutively provide bounds for each of the two terms in the right-hand side of the above inequality, each holding with probability at least $1-\delta/3$. To do so, we focus on the event defined below where all Z_t are bounded:

$$\mathcal{E}_\tau(\delta) \triangleq \big\{ \forall t = 1, \dots \tau, \quad |Z_t| \leqslant M_\tau \big\}, \tag{4.10}$$

with M_τ defined in the statement of the lemma. We will show below that $\mathcal{E}_\tau(\delta)$ takes place with probability at least $1-\delta/3$. All in all, our obtained global bound will hold with probability at least $1-\delta$, as stated in the lemma.

⋆ *Bounding the probability of the event* $\mathcal{E}_\tau(\delta)$. Recall that p_t is $\mathcal{F}_{t-1} = \sigma(\varepsilon_1, \dots, \varepsilon_{t-1})$ measurable. For $t \in \{1, \dots, \tau\}$, as ε_t is a ρ–sub-Gaussian variable independent of $\mathcal{F}_{t-1}$,

$$\mathbb{E}\Big[\exp(p_t^{\mathrm{T}} \varepsilon_t) \,\Big|\, \mathcal{F}_{t-1} \Big] \leqslant \exp\left(\frac{\rho \|p_t\|^2}{2} \right) \leqslant \exp\Big(\frac{\rho}{2}\Big) ;$$

see Footnote (1 page 241) for a reminder of the definition of a ρ–sub-Gaussian variable. Using the Markov-Chernov inequality, we obtain

$$\begin{aligned} \mathbb{P}\big(Z_t \geqslant M_\tau \,\big|\, \mathcal{F}_{t-1}\big) &\leqslant \mathbb{E}\Big[\exp(Z_t) \,\Big|\, \mathcal{F}_{t-1} \Big] \, \exp(-M_\tau) \\ &\leqslant \exp\Big(\frac{\rho}{2} - M_\tau\Big) = \frac{\delta}{6\tau}. \end{aligned} \tag{4.11}$$

Symmetrically, we get that $\mathbb{P}(Z_t \leqslant -M_\tau) \leqslant \delta/6\tau$. Combining all these bounds for $t = 1, \dots, \tau$, the event $\mathcal{E}_\tau(\delta)$ happens with probability at least $1-\delta/3$.

⋆ *Upper bound on the first term in* (4.9). By Assumption 1, we have $\varphi(x_t, p_t)^{\mathrm{T}} \theta \in [0, C]$, thus

$$|\widehat{Z}_t - Z_t| = \Big| \varphi(x_t, p_t)^{\mathrm{T}} \theta - \big[\varphi(x_t, p_t)^{\mathrm{T}} \widehat{\theta}_\tau \big]_C \Big| \leqslant C \,,$$

and therefore, on $\mathcal{E}_\tau(\delta)$,

$$|\widehat{Z}_t + Z_t| \leqslant |\widehat{Z}_t - Z_t| + |2Z_t| \leqslant C + 2M_\tau \triangleq M''_\tau \,.$$

Noting that all components of $p_t p_t^{\mathrm{T}}$ are upper bounded by 1,

$$\begin{aligned} \Big\| \sum_{t=1}^{\tau} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \Big\|_\infty &\leqslant \sum_{t=1}^{\tau} |\widehat{Z}_t^2 - Z_t^2| = \sum_{t=1}^{\tau} \big| (\widehat{Z}_t - Z_t)(\widehat{Z}_t + Z_t) \big| \\ &\leqslant M''_\tau \sqrt{\tau \sum_{t=1}^{\tau} (\widehat{Z}_t - Z_t)^2} \,, \end{aligned}$$

where the last inequality was obtained by $|\widehat{Z}_t + Z_t| \leqslant M''_\tau$ together with the Cauchy-Schwarz inequality. Using that $\big|y - [x]_C\big| \leqslant |y - x|$ when $y \in [0, C]$ and $x \in \mathbb{R}$, we note that

$$|\widehat{Z}_t - Z_t| \leqslant \Big| \varphi(x_t, p_t)^{\mathrm{T}} (\widehat{\theta}_\tau - \theta) \Big| \,.$$

All in all, we proved so far

$$\begin{aligned}\Big\| \sum_{t=1}^{\tau} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \Big\|_\infty &\leqslant M_\tau'' \sqrt{\tau (\widehat{\theta}_\tau - \theta)^{\mathrm{T}} \left(\sum_{t=1}^{\tau} \varphi(x_t, p_t) \varphi(x_t, p_t)^{\mathrm{T}} \right) (\widehat{\theta}_\tau - \theta)} \\ &= M_\tau'' \sqrt{\tau (\widehat{\theta}_\tau - \theta)^{\mathrm{T}} \left(V_\tau - \lambda I \right) (\widehat{\theta}_\tau - \theta)} \\ &\leqslant M_\tau'' \sqrt{\tau (\widehat{\theta}_\tau - \theta)^{\mathrm{T}} V_\tau (\widehat{\theta}_\tau - \theta)} = M_\tau'' \, \| V_\tau^{1/2} (\theta - \widehat{\theta}_\tau) \| \sqrt{\tau} \,,\end{aligned}$$

where $V_\tau = \lambda I + \sum_{t=1}^{\tau} \varphi(x_t, p_t) \varphi(x_t, p_t)^{\mathrm{T}}$ was used for the last steps. From Lemma 4 and the bound (4.5), we finally obtain that with probability at least $1 - \delta/3$,

$$\Big\| \sum_{t=1}^{\tau} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \Big\|_\infty \leqslant M_\tau'' \, B_\tau(\delta/3) \sqrt{\tau} = (C + 2M_\tau) \, B_\tau(\delta/3) \sqrt{\tau} \,. \tag{4.12}$$

⋆ *Upper bound on the second term in* (4.9). Recall that p_t is $\mathcal{F}_{t-1}$ measurable and that in Model 1, we defined $Z_t = Y_{t,p_t} - \varphi(x_t, p_t)^{\mathrm{T}} \theta = p_t^{\mathrm{T}} \varepsilon_t$, which is a scalar value. These two observations yield

$$\begin{aligned}\mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] &= \mathbb{E}\big[p_t Z_t^2 p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] = \mathbb{E}\big[p_t p_t^{\mathrm{T}} \varepsilon_t \varepsilon_t^{\mathrm{T}} p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] \\ &= p_t p_t^{\mathrm{T}} \, \mathbb{E}\big[\varepsilon_t \varepsilon_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] \, p_t p_t^{\mathrm{T}} = p_t p_t^{\mathrm{T}} \Sigma p_t p_t^{\mathrm{T}} \,.\end{aligned} \tag{4.13}$$

We wish to apply the Hoeffding–Azuma inequality to each component of $Z_t^2 p_t p_t^{\mathrm{T}}$, however, we need some boundedness to do so. Therefore, we consider instead $Z_t^2 \mathbb{1}_{\{|Z_t| \leqslant M_\tau\}}$. The indicated inequality, together with a union bound, entails that with probability at least $1 - \delta/3$,

$$\Big\| \sum_{t=1}^{\tau} Z_t^2 \mathbb{1}_{\{|Z_t| \leqslant M_\tau\}} p_t p_t^{\mathrm{T}} - \sum_{t=1}^{\tau} \mathbb{E}\Big[Z_t^2 \mathbb{1}_{\{|Z_t| \leqslant M_\tau\}} p_t p_t^{\mathrm{T}} \,\Big|\, \mathcal{F}_{t-1} \Big] \Big\|_\infty \leqslant M_\tau^2 \sqrt{2\tau \ln(3K^2/\delta)} \,. \tag{4.14}$$

Over $\mathcal{E}_\tau(\delta)$, using (4.13) and applying a triangular inequality, we obtain

$$\begin{aligned}\Big\| \sum_{t=1}^{\tau} Z_t^2 p_t p_t^{\mathrm{T}} - p_t p_t^{\mathrm{T}} \Sigma p_t p_t^{\mathrm{T}} \Big\|_\infty &= \Big\| \sum_{t=1}^{\tau} Z_t^2 \mathbb{1}_{\{|Z_t| \leqslant M_\tau\}} p_t p_t^{\mathrm{T}} - \sum_{t=1}^{\tau} \mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \\ &\leqslant \Big\| \sum_{t=1}^{\tau} Z_t^2 \mathbb{1}_{\{|Z_t| \leqslant M_\tau\}} p_t p_t^{\mathrm{T}} - \sum_{t=1}^{\tau} \mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \mathbb{1}_{\{|Z_t| \leqslant M_\tau\}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \\ &\qquad + \sum_{t=1}^{\tau} \Big\| \mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \mathbb{1}_{\{|Z_t| > M_\tau\}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \,.\end{aligned} \tag{4.15}$$

We just need to bound the last term of the inequality above to conclude this part. Using that $x^2 \leqslant \exp(x)$ for $x \geqslant 0$, we get

$$\mathbb{E}\Big[Z_t^2 \mathbb{1}_{\{|Z_t| > M_\tau\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \leqslant \mathbb{E}\Big[\exp\big(|Z_t|\big) \mathbb{1}_{\{|Z_t| > M_\tau\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \,.$$

Applying a conditional Cauchy-Schwarz inequality yields

$$\mathbb{E}\Big[\exp\big(|Z_t|\big) \mathbb{1}_{\{|Z_t| > M_\tau\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \leqslant \sqrt{\mathbb{E}\big[\exp\big(2|Z_t|\big) \,\big|\, \mathcal{F}_{t-1} \big] \; \mathbb{E}\big[\mathbb{1}_{\{|Z_t| > M_\tau\}} \,\big|\, \mathcal{F}_{t-1} \big]} \,.$$

Now, thanks to the sub-Gaussian property of ε_t used with $\nu = 2p_t$ and $\nu = -2p_t$, we have

$$\mathbb{E}\big[\exp\big(2|Z_t|\big) \leqslant \mathbb{E}\big[\exp(2Z_t)\,\big|\,\mathcal{F}_{t-1}\big] + \mathbb{E}\big[\exp(-2Z_t)\,\big|\,\mathcal{F}_{t-1}\big] \leqslant 2\exp(2\rho)\,.$$

The bound (4.11) and its symmetric version indicate that

$$\mathbb{P}\big(|Z_t| \geqslant M_\tau \,\big|\, \mathcal{F}_{t-1}\big) \leqslant \frac{\delta}{3\tau}\,.$$

We therefore proved

$$\mathbb{E}\Big[\exp\big(|Z_t|\big)\mathbb{1}_{\{|Z_t|>M_\tau\}}\,\Big|\,\mathcal{F}_{t-1}\Big] \leqslant \sqrt{2\exp(2\rho)\,\frac{\delta}{3\tau}}\,.$$

Thus, we have $\mathbb{E}\big[Z_t^2\mathbb{1}_{\{|Z_t|>M_\tau\}}\,\big|\,\mathcal{F}_{t-1}\big] \leqslant 2\sqrt{\exp(2\rho)\delta/(6\tau)}$ and as all components of the $p_tp_t^{\mathrm{T}}$ are in $[0,1]$,

$$\Big\|\mathbb{E}\big[Z_t^2\mathbb{1}_{\{|Z_t|>M_\tau\}}p_tp_t^{\mathrm{T}}\,\big|\,\mathcal{F}_{t-1}\big]\Big\|_\infty \leqslant 2\sqrt{\exp(2\rho)\frac{\delta}{6\tau}}. \tag{4.16}$$

Finally , combining (4.15) with (4.14) and (4.16), we get with probability $1-\delta/3$

$$\Big\|\sum_{t=1}^{\tau} Z_t^2 p_tp_t^{\mathrm{T}} - p_tp_t^{\mathrm{T}}\Sigma p_tp_t^{\mathrm{T}}\Big\|_\infty \leqslant M_\tau^2\sqrt{2\tau\ln(3K^2/\delta)} + 2\tau\sqrt{\exp(2\rho)\delta/(6\tau)} = M'_\tau\sqrt{\tau}\,,$$

where M'_τ is defined in the statement of the lemma.

$\star$ *Combining the two upper bounds into* (4.9). Combining the above upper bound with (4.9) and (4.12), we proved that with probability $1-\delta$,

$$\Big\|\sum_{t=1}^{\tau} p_tp_t^{\mathrm{T}}\Big(\widehat{\Sigma}_\tau - \Sigma\Big)p_tp_t^{\mathrm{T}}\Big\|_\infty \leqslant M'_\tau\sqrt{\tau} + M''_\tau B_\tau(\delta/3)\sqrt{\tau}\,,$$

which concludes the proof. □

By choosing specific vectors p_t, for $t = 1,\dots,\tau$ and using Lemma 6, we prove below how the estimator (4.4) satisfies the inequality of Lemma 5.

Proof of Lemma 5. Applying Lemma 6 together with

$$p_tp_t^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)p_tp_t^{\mathrm{T}} = p_t\mathrm{Tr}\Big(p_t^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)p_t\Big)p_t^{\mathrm{T}} = \mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)p_tp_t^{\mathrm{T}}\Big)p_tp_t^{\mathrm{T}}$$

we have, with probability at least $1-\delta$, that for all pairs of coordinates $(i,j) \in E$,

$$\left|\sum_{t=1}^{\tau}\mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)p_tp_t^{\mathrm{T}}\Big)\big[p_tp_t^{\mathrm{T}}\big]_{i,j}\right| \leqslant \kappa_\tau\sqrt{\tau}\,. \tag{4.17}$$

Remember that in the set E considered, we only have pairs (i,j) with $i \leqslant j$. However, for symmetry reasons, it will be convenient to also consider the vectors $p^{(i,j)}$ with $i > j$, where the latter vectors are defined in an obvious way. We note that for all $1 \leqslant i,j \leqslant K$,

$$p^{(i,j)}p^{(i,j)^{\mathrm{T}}} = p^{(j,i)}p^{(j,i)^{\mathrm{T}}}\,. \tag{4.18}$$

Now, our aim is to control

$$\left|q^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)q\right| = \left|\mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)qq^{\mathrm{T}}\Big)\right| \tag{4.19}$$

uniformly over $q \in \mathcal{P}$. The proof consists of two steps: establishing such a control for the special cases where q is one of the $p^{(i,j)}$ and then, extending the control to arbitrary vectors $q \in \mathcal{P}$, based on a decomposition of qq^{T} as a weighted sum of $p^{(i,j)}p^{(i,j)^{\mathrm{T}}}$ vectors.

⋆ *The case of the $p^{(i,j)}$ vectors.* We first consider the off-diagonal elements $1 \leqslant i < j \leqslant K$. Note that since p_t is of the form $p^{(i',j')}$ for all $1 \leqslant t \leqslant \tau$, we have

$$\big[p_t p_t^{\mathrm{T}}\big]_{i,j} = \begin{cases} 1/4 & \text{if } p_t = p^{(i,j)}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.20}$$

Using that $p_t = p^{(i,j)}$ at least for τ_0 rounds, Inequality (4.17) entails

$$\frac{\tau_0}{4}\left|\mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)p^{(i,j)}p^{(i,j)^{\mathrm{T}}}\Big)\right| \leqslant \kappa_\tau\sqrt{\tau}\,,$$

or put differently,

$$\left|\mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)p^{(i,j)}p^{(i,j)^{\mathrm{T}}}\Big)\right| \leqslant \frac{4\kappa_\tau\sqrt{\tau}}{\tau_0}\,. \tag{4.21}$$

Now, let us consider the diagonal elements. Let $1 \leqslant i \leqslant K$. We have

$$\big[p_t p_t^{\mathrm{T}}\big]_{i,i} = \begin{cases} 1 & \text{if } p_t = p^{(i,i)}, \\ 1/4 & \text{if } p_t = p^{(i,j)} \text{ for some } j > i, \\ 1/4 & \text{if } p_t = p^{(k,i)} \text{ for some } k < i, \\ 0 & \text{otherwise,} \end{cases} \tag{4.22}$$

where we recall that the p_t are necessarily of the form $p^{(k,\ell)}$ with $k \leqslant \ell$. Therefore, Inequality (4.17) yields

$$\tau_0\left|\mathrm{Tr}\left(\big(\widehat{\Sigma}_\tau - \Sigma\big)\Big(p^{(i,i)}p^{(i,i)^{\mathrm{T}}} + \frac{1}{4}\sum_{j>i} p^{(i,j)}p^{(i,j)^{\mathrm{T}}} + \frac{1}{4}\sum_{k<i} p^{(k,i)}p^{(k,i)^{\mathrm{T}}}\Big)\right)\right| \leqslant \kappa_\tau\sqrt{\tau}\,,$$

which we rewrite by symmetry – see (4.18) – as

$$\left|\mathrm{Tr}\left(\big(\widehat{\Sigma}_\tau - \Sigma\big)\Big(p^{(i,i)}p^{(i,i)^{\mathrm{T}}} + \frac{1}{4}\sum_{j\neq i} p^{(i,j)}p^{(i,j)^{\mathrm{T}}}\Big)\right)\right| \leqslant \frac{\kappa_\tau\sqrt{\tau}}{\tau_0}\,. \tag{4.23}$$

⋆ *Decomposing arbitrary vectors $q \in \mathcal{P}$.* Now, let $q \in \mathcal{P}$. We show below by means of elementary calculations that

$$qq^{\mathrm{T}} = \sum_{i=1}^{K}\sum_{j=1}^{K} u(i,j)\; p^{(i,j)}p^{(i,j)^{\mathrm{T}}}, \text{ with } u(i,j) = 2q_iq_j \text{ if } i \neq j \text{ and } u(i,i) = 2q_i^2 - q_i. \tag{4.24}$$

Indeed, by identification and by imposing $u(i,j) = u(j,i)$ for all pairs i,j, the equalities (4.20) and the symmetry property (4.18) entail, for $k \neq k'$:

$$q_kq_{k'} = \big[qq^{\mathrm{T}}\big]_{k,k'} = \sum_{i=1}^{K}\sum_{j=1}^{K} u(i,j)\;\big[p^{(i,j)}p^{(i,j)^{\mathrm{T}}}\big]_{k,k'} = \frac{u(k,k')}{4} + \frac{u(k',k)}{4} = \frac{u(k,k')}{2}\,,$$

which can be rephrased as $u(k,k') = u(k',k) = 2q_k q_{k'}$. Now, let us calculate the diagonal elements, by identification and by the equalities (4.22) as well as by the symmetry property (4.18):

$$q_k^2 = \left[qq^{\mathrm{T}}\right]_{k,k} = \sum_{i=1}^{K}\sum_{j=1}^{K} u(i,j)\ \left[p^{(i,j)}p^{(i,j)^{\mathrm{T}}}\right]_{k,k} = u(k,k) + \sum_{i\neq k}\frac{u(i,k)}{4} + \sum_{j\neq k}\frac{u(k,j)}{4}$$
$$= u(k,k) + \frac{1}{2}\sum_{i\neq k} u(i,k) = u(k,k) + \sum_{i\neq k} q_k q_i = u(k,k) + \sum_{i=1}^{K} q_k q_i - q_k^2 = u(k,k) + q_k - q_k^2\,,$$

which leads to $u(k,k) = 2q_k^2 - q_k$. We introduce the notation $P^{(i,j)} = p^{(i,j)}p^{(i,j)^{\mathrm{T}}}$ and in light of (4.21) and (4.23), we rewrite (4.24) as

$$qq^{\mathrm{T}} = \sum_{i=1}^{K} u(i,i)\left(P^{(i,i)} + \frac{1}{4}\sum_{j\neq i} P^{(i,j)}\right) + \sum_{i=1}^{K}\sum_{j\neq i}\left(u(i,j) - \frac{u(i,i)}{4}\right)P^{(i,j)}\,.$$

⋆ *Controlling arbitrary vectors* $q \in \mathcal{P}$. Therefore, substituting this decomposition of qq^{T} into the aim (4.19), and using the linearity of the trace as well as the triangle inequality for absolute values, we obtain

$$\left|q^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)q\right| = \left|\mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)qq^{\mathrm{T}}\Big)\right| \leqslant \sum_{i=1}^{K}\left|u(i,i)\right|\left|\mathrm{Tr}\left(\big(\widehat{\Sigma}_\tau - \Sigma\big)\left(P^{(i,i)} + \frac{1}{4}\sum_{j\neq i}P^{(i,j)}\right)\right)\right|$$
$$+ \sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j) - \frac{u(i,i)}{4}\right|\left|\mathrm{Tr}\Big(\big(\widehat{\Sigma}_\tau - \Sigma\big)P^{(i,j)}\Big)\right|$$

We then substitute the upper bounds (4.21) and (4.23) and get

$$\left|q^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)q\right| \leqslant \frac{\kappa_\tau\sqrt{\tau}}{\tau_0}\left(\sum_{i=1}^{K}\left|u(i,i)\right| + 4\sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j) - \frac{u(i,i)}{4}\right|\right).$$

By the triangle inequality, by the values $2q_iq_j$ of the coefficients $u(i,j)$ when $i \neq j$ and by using $|u(i,i)| \leqslant q_i$,

$$\sum_{i=1}^{K}\left|u(i,i)\right| + 4\sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j) - \frac{u(i,i)}{4}\right| \leqslant K\sum_{i=1}^{K}\left|u(i,i)\right| + 4\sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j)\right|$$
$$\leqslant K\sum_{i=1}^{K} q_i + 8\sum_{i=1}^{K}\sum_{j\neq i} q_iq_j = K + 8\sum_{i=1}^{K} q_i(1-q_i) \leqslant K + 8\,.$$

Putting all elements together, we proved $\sup_{q\in\mathcal{P}}\left|q^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)q\right| \leqslant \frac{\kappa_\tau\sqrt{\tau}}{\tau_0}(K+8) = \upsilon_\tau(\delta)$, which concludes the proof of Lemma 5. □

3.3.3 Analysis – proof of Theorem 4

The analysis exploits how well each $\widehat{\theta}_t$ estimates θ and how well $\widehat{\Sigma}_\tau$ estimates Σ. The regret bound, as is clear from Proposition 1 below, also consists of these two parts.

Proposition 1. *Fix a risk level $\delta \in (0,1)$ and an exploration budget $\tau \geqslant 2$. Assume that Assumption 1 holds. Consider the estimator $\widehat{\Sigma}_\tau$ of Σ such that $\sup_{p\in\mathcal{P}} \left|p^{\mathrm{T}}(\Sigma - \widehat{\Sigma}_\tau)p\right| \leqslant \upsilon$ with probability at least $1-\delta/2$, by taking $\upsilon = \upsilon_\tau(\delta/2) > 0$ – see Equation (4.7). Then choosing $\lambda > 0$ and*

$$\begin{aligned} a_{t,p} &= \min\Big\{L,\ 2C\, B_{t-1}(\delta t^{-2})\, \big\|V_{t-1}^{-1/2}\varphi(x_t,p)\big\|\Big\}\,, \\ \alpha_{t,p} &= \upsilon + a_{t,p}\,, \end{aligned} \tag{4.25}$$

the optimistic algorithm (4.3) ensures that with probability $1-\delta$,

$$\textstyle\sum_{t=\tau+1}^{T} \ell_{t,p_t} - \sum_{t=\tau+1}^{T} \min_{p\in\mathcal{P}} \ell_{t,p} \leqslant 2\sum_{t=\tau+1}^{T} \alpha_{t,p_t}\,.$$

Remark 10. *Li et al. [2010] pick $\alpha(t,p)$ proportional to $\big\|V_{t-1}^{-1/2}\varphi(x_t,p)\big\|$ only, but we need an additional term to account for the covariance matrix.*

Lemma 5 above studies how well $\widehat{\Sigma}_\tau$ estimates Σ and we are thus left with controlling the sum of the $a_{t,p}$ to prove Proposition 1

Proof of Proposition 1. We show below (*Step 1*) that for all $t \geqslant 2$, if

$$\big\|V_{t-1}^{1/2}\big(\widehat{\theta}_{t-1} - \theta\big)\big\| \leqslant B_{t-1}(\delta t^{-2}) \quad \text{and} \quad \|\Sigma - \widehat{\Sigma}_t\|_\infty \leqslant \upsilon\,, \tag{4.26}$$

then

$$\forall p \in \mathcal{P}, \qquad \big|\ell_{t,p} - \widehat{\ell}_{t,p}\big| \leqslant \alpha_{t,p}\,. \tag{4.27}$$

Property (5.7), for those t for which it is satisfied, entails (*Step 2*) that the corresponding instantaneous regrets are bounded by

$$r_t \triangleq \ell_{t,p_t} - \min_{p\in\mathcal{P}} \ell_{t,p} \leqslant 2\alpha_{t,p_t}\,.$$

It only remains to deal (*Step 3*) with the rounds t when (5.7) does not hold; they account for the $1-\delta$ confidence level.

⋆ *Step 1: Good estimation of the losses.* When the two events (4.26) hold, we have

$$\begin{aligned} \big|\ell_{t,p} - \widehat{\ell}_{t,p}\big| &= \Big|\big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2 + p^{\mathrm{T}}\Sigma p - \Big(\big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t\Big)^2 + p^{\mathrm{T}}\widehat{\Sigma}_t p\Big| \\ &\leqslant \big|p^{\mathrm{T}}\Sigma p - p^{\mathrm{T}}\widehat{\Sigma}_t p\big| + \Big|\big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2 - \Big(\big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t\Big)^2\Big|. \end{aligned}$$

On the one hand, $\big|p^{\mathrm{T}}\Sigma p - p^{\mathrm{T}}\widehat{\Sigma}_t p\big| \leqslant \upsilon$ while on the other hand,

$$\begin{aligned} &\Big|\big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2 - \Big(\big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t\Big)^2\Big| \\ &\qquad = \Big|\varphi(x_t,p)^{\mathrm{T}}\theta - \big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C\Big| \times \Big|\varphi(x_t,p)^{\mathrm{T}}\theta + \big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - 2c_t\Big|\,, \end{aligned}$$

where by Assumption 1, all quantities in the final inequality lie in $[0,C]$, thus

$$\Big|\varphi(x_t,p)^{\mathrm{T}}\theta + \big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - 2c_t\Big| \leqslant 2C\,.$$

Finally,

$$\begin{aligned}\left|\varphi(x_t,p)^{\mathrm{T}}\theta - \left[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\right]_C\right| &\leqslant \left|\varphi(x_t,p)^{\mathrm{T}}\theta - \varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\right| \\ &\leqslant \left\|V_{t-1}^{1/2}\big(\theta-\widehat{\theta}_{t-1}\big)\right\| \left\|V_{t-1}^{-1/2}\varphi(x_t,p)\right\|, \end{aligned} \tag{4.28}$$

where we used the Cauchy-Schwarz inequality for the second inequality, and the fact that $\big|y-[x]_C\big| \leqslant |y-x|$ when $y \in [0,C]$ and $x \in \mathbb{R}$ for the first inequality. Collecting all bounds together, we proved

$$\left|\big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2 - \Big(\left[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\right]_C - c_t\Big)^2\right| \leqslant 2C \underbrace{\left\|V_{t-1}^{1/2}\big(\theta-\widehat{\theta}_{t-1}\big)\right\|}_{\leqslant B_{t-1}(\delta t^{-2})} \left\|V_{t-1}^{-1/2}\varphi(x_t,p)\right\|,$$

but of course, this term is also bounded by the quantity L introduced at the beginning of Section 3. This concludes the proof of the claimed inequality (5.7).

⋆ *Step 2: Resulting bound on the instantaneous regrets.* We denote by

$$p_t^\star \in \operatorname*{argmin}_{p\in\mathcal{P}} \big\{\ell_{t,p}\big\} \tag{4.29}$$

an optimal convex vector to be used at round t. By definition (4.3) of the optimistic algorithm, we have that the played p_t satisfies

$$\widehat{\ell}_{t,p_t} - \alpha_{t,p_t} \leqslant \widehat{\ell}_{t,p_t^\star} - \alpha_{t,p_t^\star}, \qquad \text{that is,} \qquad \widehat{\ell}_{t,p_t} - \widehat{\ell}_{t,p_t^\star} \leqslant \alpha_{t,p_t} - \alpha_{t,p_t^\star}.$$

Now, for those t for which both events (4.26) hold, the property (5.7) also holds and yields, respectively for $p=p_t$ and $p=p_t^\star$:

$$\ell_{t,p_t} - \widehat{\ell}_{t,p_t} \leqslant \alpha_{t,p_t} \qquad \text{and} \qquad \widehat{\ell}_{t,p_t^\star} - \ell_{t,p_t^\star} \leqslant \alpha_{t,p_t^\star}.$$

Combining all these three inequalities together, we proved

$$\begin{aligned} r_t = \ell_{t,p_t} - \ell_{t,p_t^\star} &= \big(\ell_{t,p_t} - \widehat{\ell}_{t,p_t}\big) + \big(\widehat{\ell}_{t,p_t} - \widehat{\ell}_{t,p_t^\star}\big) + \big(\widehat{\ell}_{t,p_t^\star} - \ell_{t,p_t^\star}\big) \\ &\leqslant \alpha_{t,p_t} + (\alpha_{t,p_t} - \alpha_{t,p_t^\star}) + \alpha_{t,p_t^\star} = 2\alpha_{t,p_t}, \end{aligned}$$

as claimed. This yields the $2\sum \alpha_{t,p_t}$ in the regret bound, where the sum is for $t \geqslant \tau+1$.

⋆ *Step 3: Special cases.* We conclude the proof by dealing with the time steps $t \geqslant \tau+1$ when at least one of the events (4.26) does not hold. By a union bound, this happens for some $t \geqslant \tau+1$ with probability at most

$$\frac{\delta}{2} + \delta \sum_{t\geqslant\tau+1} t^{-2} \leqslant \frac{\delta}{2} + \delta \int_2^\infty \frac{1}{t^2}\mathrm{d}t = \delta,$$

where we used $\tau \geqslant 2$. These special cases thus account for the claimed $1-\delta$ confidence level. □

Remark 11. *The main difference with the regret analysis of LinUCB provided by Chu et al. [2011] or Lattimore and Szepesvári [2020] is in the first part of* Step 1*, as we need to deal with slightly more complicated quantities: not just with linear quantities of the form* $\varphi(x_t,p)^{\mathrm{T}}\theta$*. Steps 2 and 3 are easy consequences of Step 1.*

We are now left with proving the following lemma to conclude the analysis.

Lemma 7. *No matter how the environment and provider pick the x_t and p_t,*

$$\sum_{t=\tau+1}^{T} a_{t,p_t} \leqslant \sqrt{\left(2C\overline{B}\right)^2 + \frac{L^2}{2}}\sqrt{dT\ln\frac{\lambda+T}{\lambda}} = \mathcal{O}\big(\sqrt{T\ln T\ln(T/\delta)}\big)\,,$$

where $\overline{B} \triangleq B_T(\delta/T^2) = \sqrt{d\lambda}\,C + \rho\sqrt{2\ln(T^2/\delta) + d\ln(1+T/\lambda)}$.

Remark 12. *This lemma follows from a straightforward adaptation/generalization of Lemma 19.1 of the monograph by Lattimore and Szepesvári [2020]; see also a similar result in Lemma 3 by Chu et al. [2011].*

Proof of Lemma 7. We consider the worst case when all summations would start at the round $\tau+1=2$. By definition, the quantity $\overline{B}$ upper bounds all the $B_{t-1}(\delta t^{-2})$. It therefore suffices to upper bound

$$\begin{aligned}\sum_{t=2}^{T}\min\Big\{L,\ 2C\overline{B}\,\|V_{t-1}^{-1/2}\varphi(x_t,p_t)\|\Big\} &\leqslant \sqrt{T}\sqrt{\sum_{t=2}^{T}\min\Big\{L^2,\ \left(2C\overline{B}\right)^2\|V_{t-1}^{-1/2}\varphi(x_t,p_t)\|^2\Big\}}\\ &= \sqrt{T}\sqrt{\sum_{t=2}^{T}\min\left\{L^2,\ \left(2C\overline{B}\right)^2\left(\frac{\det(V_t)}{\det(V_{t-1})}-1\right)\right\}}\end{aligned}$$

where we applied first the Cauchy-Schwarz inequality and used second the equality

$$1+\|V_{t-1}^{-1/2}\varphi(x_t,p_t)\|^2 = 1+\varphi(x_t,p_t)^{\mathrm{T}}V_{t-1}^{-1}\varphi(x_t,p_t) = \frac{\det(V_t)}{\det(V_{t-1})}\,,$$

that follows from a standard result in online matrix theory, namely, Lemma 8 below. Now, we get a telescoping sum with the logarithm function by using the inequality

$$\forall b>0,\quad \forall u>0,\qquad \min\{b,u\} \leqslant b\,\frac{\ln(1+u)}{\ln(1+b)}\,, \tag{4.30}$$

which is proved below. Namely, we further bound the sum above by

$$\begin{aligned}\sum_{t=2}^{T}\min\left\{L^2,\ \left(2C\overline{B}\right)^2\left(\frac{\det(V_t)}{\det(V_{t-1})}-1\right)\right\} &\leqslant \left(2C\overline{B}\right)^2\sum_{t=2}^{T}\min\left\{\frac{L^2}{\left(2C\overline{B}\right)^2},\ \frac{\det(V_t)}{\det(V_{t-1})}-1\right\}\\ &\leqslant \left(2C\overline{B}\right)^2\sum_{t=2}^{T}\frac{L^2/\left(2C\overline{B}\right)^2}{\ln\left(1+L^2/\left(2C\overline{B}\right)^2\right)}\ln\left(\frac{\det(V_t)}{\det(V_{t-1})}\right)\\ &= \frac{L^2}{\ln\left(1+L^2/\left(2C\overline{B}\right)^2\right)}\ln\left(\frac{\det(V_T)}{\det(V_2)}\right) \leqslant \frac{L^2}{\ln\left(1+L^2/\left(2C\overline{B}\right)^2\right)}\,d\ln\frac{\lambda+T}{\lambda}\end{aligned}$$

where we used Assumption 1 and one of its consequences to get the last inequality. Finally, we use $1/\ln(1+u)\leqslant 1/u+1/2$ for all $u\geqslant 0$ to get a more readable constant:

$$\frac{L^2}{\ln\left(1+L^2/\left(2C\overline{B}\right)^2\right)} \leqslant \left(2C\overline{B}\right)^2 + \frac{L^2}{2}\,.$$

The proof is concluded by collecting all pieces. Finally, we now provide the proofs of two either straightforward or standard results used above.

⋆ *Standard Result in Online Matrix Theory.* The following result is extremely standard in online matrix theory (see, among many others, Lemma 11.11 in Cesa-Bianchi and Lugosi, 2006 or the proof of Lemma 19.1 in the monograph by Lattimore and Szepesvári, 2020).

Lemma 8. *Let M a $d \times d$ full-rank matrix, let $u, v \in \mathbb{R}^d$ be two arbitrary vectors. Then*

$$1 + v^{\mathrm{T}} M^{-1} u = \frac{\det\big(M + uv^{\mathrm{T}}\big)}{\det(M)} \,.$$

The proof first considers the case $M = \mathrm{I}_d$. We are then left with showing that $\det\big(\mathrm{I}_d + uv^{\mathrm{T}}\big) = 1 + v^{\mathrm{T}} u$, which follows from taking the determinant of every term of the equality

$$\begin{bmatrix} \mathrm{I}_d & 0 \\ v^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{I}_d + uv^{\mathrm{T}} & u \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{I}_d & 0 \\ -v^{\mathrm{T}} & 1 \end{bmatrix} = \begin{bmatrix} \mathrm{I}_d & u \\ 0 & 1 + v^{\mathrm{T}} u \end{bmatrix} .$$

Now, we can reduce the case of a general M to this simpler case by noting that

$$\det\big(M + uv^{\mathrm{T}}\big) = \det(M)\ \det\Big(\mathrm{I}_d + \big(M^{-1}u\big)v^{\mathrm{T}}\Big) = \det(M)\ \big(1 + v^{\mathrm{T}} M^{-1} u\big) \,.$$

⋆ *Proof of Inequality* (4.30). This inequality is used in Lemma 19.1 of the monograph by Lattimore and Szepesvári [2020], in the special case $b = 1$. The extension to $b > 0$ is straightforward. We fix $b > 0$. We want to prove that

$$\forall u > 0, \qquad \min\{b, u\} \leqslant b\, \frac{\ln(1+u)}{\ln(1+b)} \,. \tag{4.31}$$

We first note that

$$\min\{b, u\} = b\, \frac{\ln(1+u)}{\ln(1+b)} \quad \text{for } u = b$$

and that $\min\{b, u\} = b$ for $u \geqslant b$, with the right-hand side of (4.31) being an increasing function of u. Therefore, it suffices to prove (4.31) for $u \in [0, b]$, where $\min\{b, u\} = u$. Now,

$$u \longmapsto b\, \frac{\ln(1+u)}{\ln(1+b)} - u$$

is a concave and (twice) differentiable function, vanishing at $u = 0$ and $u = b$, and is therefore non-negative on $[0, b]$. This concludes the proof. □

We are now ready to conclude the proof of Theorem 4. Using for the first $\tau \geqslant 2$ rounds that $L = C^2 + \Gamma$ upper bounds the (conditionally) expected losses ℓ_{t,p_t}, Proposition 1 and Lemmas 5 and 7 show that, with probability $1 - \delta$

$$\overline{R}_T \leqslant \tau L + T\upsilon + \sum_{t=\tau+1}^{T} a_{t,p_t} \leqslant \tau L + \mathcal{O}\left(T \ln^2\left(\frac{\tau}{\delta}\right) \sqrt{\frac{\ln(1/\delta)}{\tau}} + \sqrt{T\, \ln T \ln(T/\delta)} \right) .$$

Picking τ of order $T^{2/3}$ concludes the proof.

Dependence on the number of tariffs K and on the dimension of parameter vector d. The dependence of $\upsilon = \upsilon_\tau(\delta/2)$ on K and d is in K^2 and $\sqrt{d}$, respectively. Moreover, the dependence of $\bar{B}$ on θ-dimension is also in $\sqrt{d}$. Combining these dependencies with Lemma 7 and the inequality just above, we obtain that the regret bound is linear in d and cubic in K. However, we emphasize that d intrinsically depends on K.

Case of known covariance matrix Σ. We then have $\upsilon = 0$ in Proposition 1 and we may discard Lemma 5. Taking $\tau = 2$, the obtained regret bound is $2L + \sqrt{T \ln T \ln(T/\delta)}$. In this case, the regret bound is linear in d and does not depend directly on K.

Remark 13. *The algorithm of Theorem 4 depends on δ via the tuning (4.25) of α. But we can also define a regret with full expectations $\mathbb{E}\big[\ell_{t,p_t}\big]$ and $\min \mathbb{E}\big[\ell_{t,p}\big]$ – remember from Sections 3.1 and 3.2.1 that the losses $\ell_{t,p}$ are conditional expectations. In that case the algorithm can be made independent of δ. Only Step 3 of the proof of Proposition 1 is to be modified. The same rates in T are obtained.*

4 Fast rates, with global noise modeling

In this section, we consider Model 2 and show that under an attainability condition stated below, the order of magnitude of the regret bound in Theorem 4 can be reduced to a poly-logarithmic rate. This kind of fast rates already exist in the literature of linear contextual bandits (see, e.g., Abbasi-Yadkori et al., 2011, as well as Dani et al., 2008) but are not so frequent. We underline in the proof the key step where we gain orders of magnitude in the regret bound. Before doing so, we note that similarly to Section 3.1,

$$\mathbb{E}\big[(Y_{t,p} - c_t)^2\big] = \big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2 + \sigma^2\,, \tag{4.32}$$

which leads us to introduce a regret $\overline{R}_T$ defined by

$$\overline{R}_T = \sum_{t=1}^{T}\big(\varphi(x_t,p_t)^{\mathrm{T}}\theta - c_t\big)^2 - \sum_{t=1}^{T}\min_{p\in\mathcal{P}}\Big\{\big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2\Big\}\,.$$

Thus, as far as the minimization of the regret is concerned, Model 2 is a special case of Model 1, corresponding to a matrix Σ that can be taken as the null matrix $[0]$. Of course, as explained in Section 2.1, the covariance matrix Σ of Model 2 is $\sigma^2[1]$ in terms of real modeling, but in terms of regret-minimization it can be taken as $\Sigma = [0]$. Therefore, all results established above for Model 1 extend to Model 2, but under an additional assumption stated below, the $T^{2/3}$ rates (up to poly-logarithmic terms) obtained above can be reduced to poly-logarithmic rates only.

Assumption 2 – *Attainability.* For each time instance $t \geqslant 1$, the expected mean consumption is attainable, i.e.,

$$\exists\, p \in \mathcal{P} \,|\quad \varphi(x_t,p)^{\mathrm{T}}\theta = c_t\,.$$

We denote by $p_t^\star$ such an element of $\mathcal{P}$. In Model 2 and under this assumption, the expected losses $\ell_{t,p}$ defined in (4.32) are such that, for all $t \geqslant 1$ and all $x_t \in \mathcal{X}$,

$$\min_{p\in\mathcal{P}} \ell_{t,p} = \ell_{t,p_t^\star} = \sigma^2\,. \tag{4.33}$$

As in Model 2 the variance terms σ^2 cancel out when considering the regret, the variance σ^2 does not need to be estimated. Our optimistic algorithm thus takes a simpler form. For each $t \geqslant 2$ and $p \in \mathcal{P}$ we consider the same estimators (4.1) of θ as before and then define

$$\widetilde{\ell}_{t,p} = \big(\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1} - c_t\big)^2$$

(no clipping needs to be considered in this case). We set

$$\beta_{t,p} = B_{t-1}(\delta t^{-2})^2 \, \|V_{t-1}^{-1/2}\varphi(x_t,p)\|^2 \tag{4.34}$$

and then pick:

$$p_t \in \operatorname*{argmin}_{p\in\mathcal{P}} \Big\{ \widetilde{\ell}_{t,p} - \beta_{t,p} \Big\} \tag{4.35}$$

for $t \geqslant 2$ and p_1 arbitrarily. The tuning parameter $\lambda > 0$ is hidden in $B_{t-1}(\delta t^{-2})^2$. We get the following theorem, whose proof re-uses many parts of the proofs of Proposition 1 and Lemma 7. Without Assumption 2, a regret bound of order $\sqrt{T}$ up to logarithmic terms could still be proved.

Theorem 5. *In Model 2, assume that the boundedness and attainability assumptions (Assumptions 1 and 2) hold. Then, the optimistic algorithm* (4.35)*, tuned with $\lambda > 0$, ensures that for all $\delta \in (0,1)$,*

$$\overline{R}_T \leqslant d\left(4\overline{B}^2 + \frac{C^2}{2}\right) \ln \frac{\lambda + T}{\lambda} = \mathcal{O}\big(\ln^2 T\big)\,,$$

with probability at least $1-\delta$, where $\overline{B}$ is defined as in Lemma 7.

Proof of Theorem 5. The key observation lies in *Step 1* (and is tagged as such) and the rest is standard maths. Because of the expression for the expected losses (4.32) and the consequence (4.33) of attainability, the regret can be rewritten as

$$\overline{R}_T = \sum_{t=1}^{T} \ell_{t,p_t} = \sum_{t=1}^{T} \big(\varphi(x_t,p_t)^{\mathrm{T}}\theta - c_t\big)^2 .$$

We first successively prove (*Step 1*) that for $t \geqslant 2$, if the bound of Lemma 4 holds, namely,

$$\Big\| V_{t-1}^{1/2}\big(\theta - \widehat{\theta}_{t-1}\big) \Big\| \leqslant B_{t-1}(\delta t^{-2})\,, \tag{4.36}$$

then

$$\ell_{t,p_t} \leqslant 2\beta_{t,p_t} + 2\widetilde{\ell}_{t,p_t}\,, \tag{4.37}$$

$$\widetilde{\ell}_{t,p_t} \leqslant \beta_{t,p_t} + \widetilde{\ell}_{t,p_t^\star} - \beta_{t,p_t^\star}\,, \tag{4.38}$$

$$\widetilde{\ell}_{t,p_t^\star} \leqslant \beta_{t,p_t^\star}\,. \tag{4.39}$$

These inequalities collectively entail the bound $\ell_{t,p_t} \leqslant 4\beta_{t,p_t}$. Of course, because of Assumption 1, we also have $\ell_{t,p_t} \leqslant C^2$. It then suffices to bound the sum (*Step 2*) of the ℓ_{t,p_t} by the sum of the $\min\big\{C^2, 4\beta_{t,p_t}\big\}$ and control for the probability of (4.36).

$\star$ *Step 1: Proof of* (4.37)–(4.39). Inequality (4.38) holds by definition of the algorithm. For (4.39) and (4.37), we re-use the inequality (4.28) proved earlier: for all $p \in \mathcal{P}$,

$$\begin{aligned}\Big(\varphi(x_t,p)^{\mathrm{T}}\big(\theta - \widehat{\theta}_{t-1}\big)\Big)^2 &\leqslant \Big\| V_{t-1}^{1/2}\big(\theta - \widehat{\theta}_{t-1}\big)\Big\|^2 \, \|V_{t-1}^{-1/2}\varphi(x_t,p)\|^2 \\ &\leqslant B_{t-1}(\delta t^{-2})^2 \, \|V_{t-1}^{-1/2}\varphi(x_t,p)\|^2 \triangleq \beta_{t,p}\,, \end{aligned} \tag{4.40}$$

where we used the bound (4.36) for the last inequality. This inequality directly yields (4.39) by taking $p = p_t^\star$. Now comes the specific improvement and our key observation: using that $(u+v)^2 \leqslant 2u^2 + 2v^2$, we have

$$\begin{aligned}\ell_{t,p_t} &= \Big(\varphi(x_t,p_t)^{\mathrm{T}}\theta - \varphi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1} + \varphi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1} - c_t\Big)^2 \\ &\leqslant 2\big(\varphi(x_t,p_t)^{\mathrm{T}}\theta - \varphi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1}\big)^2 + 2\underbrace{\big(\varphi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1} - c_t\big)^2}_{=\widetilde{\ell}_{t,p_t}},\end{aligned}$$

which yields (4.37) via (4.40) used with $p = p_t$.

$\star$ *Step 2: Summing the bounds.* First, the bound (4.36) holds, by Lemma 4, with probability at least $1 - \delta t^{-2}$ for a given $t \geqslant 2$. By a union bound, it holds for all $t \geqslant 2$ with probability at least $1 - \delta$. By bounding ℓ_{t,p_t} by C^2 and the $B_{t-1}(\delta t^{-2})$ by $\overline{B}$, we therefore get, from Step 1, that with probability at least $1 - \delta$,

$$\overline{R}_T \leqslant C^2 + \sum_{t=2}^{T} \min\Big\{C^2,\ 4\overline{B}^2\|V_{t-1}^{-1/2}\varphi(x_t,p)\|^2\Big\}.$$

Now, as in the proof of Lemma 7 above,

$$\begin{aligned}&\sum_{t=2}^{T} \min\Big\{C^2,\ 4\overline{B}^2\|V_{t-1}^{-1/2}\varphi(x_t,p)\|^2\Big\} = \sum_{t=2}^{T} \min\left\{C^2,\ 4\overline{B}^2\left(\frac{\det(V_t)}{\det(V_{t-1})} - 1\right)\right\} \\ &\leqslant 4\overline{B}^2 \sum_{t=2}^{T} \frac{C^2/\big(4\overline{B}^2\big)}{\ln\Big(1 + C^2/\big(4\overline{B}^2\big)\Big)} \ln\left(\frac{\det(V_t)}{\det(V_{t-1})}\right) = \frac{C^2}{\ln\Big(1 + C^2/\big(4\overline{B}^2\big)\Big)} \ln\left(\frac{\det(V_T)}{\det(V_1)}\right) \\ &\leqslant \left(4\overline{B}^2 + \frac{C^2}{2}\right) d \ln\frac{\lambda + T}{\lambda}.\end{aligned}$$

This concludes the proof. □

Dependence on the dimension of parameter vector d. We recall that the dependence of $\bar{B}$ on θ-dimension is in $\sqrt{d}$. Thus, the regret bound is quadratic en d – which intrinsically depends on K.

5 Application to the Low Carbon London data set

Our simulations rely on a real data set of residential electricity consumption, in which different tariffs were sent to the customers according to some policy. But of course, we cannot test an alternative policy on historical data (we only observed the outcome of the tariffs sent) and therefore need to build first a data simulator.

5.1 The underlying real data set / the simulator

We consider open data published[3] by UK Power Networks and fully described in Section 2 of Chapter 3. It contains energy consumption (in kWh per half hour) at half hourly intervals of a thousand customers subjected to dynamic energy prices. We considered their

[3] *SmartMeter Energy Consumption Data in London Households* – see https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

mean[4] consumption. As far as contexts are concerned, we considered half-hourly temperatures τ_t in London, obtained from https://www.noaa.gov/ and the following calendar variables: the day of the week w_t (equal to 0 for Sunday, 1 for Monday, etc.), the half-hour of the day $h_t \in \{1, \dots, 48\}$, and the position in the year: $\pi_t \in [0, 1]$, linear values between $\pi_t = 0$ on January 1st at 00:00 and $\pi_t = 1$ on December the 31st at 23:59.

5.1.1 Realistic simulator

The simulator is based on the following additive model, which breaks down time by half hours:

$$\phi(x_t, j) = \sum_{h=1}^{48} \left[s_\tau^h(\tau_t) + s_\pi^h(\pi_t) + \eta_h \right] \mathbb{1}_{\{h_t = h\}} + \sum_{w=0}^{6} \zeta_w \mathbb{1}_{\{w_t = w\}} + \xi_j \,, \tag{4.41}$$

where the s_τ^h and s_π^h are functions catching the effect of the temperature and of the yearly seasonality. As explained in Example 4, the transfer parameter θ gathers coordinates of the s_τ^h and the s_π^h in bases of splines, as well as the coefficients η_h, ζ_w and ξ_j. Here, we work under the assumption that exogenous factors do not impact customers' reaction to tariff changes (which is admittedly a first step, and more complex models could be considered). Our algorithms will have to sequentially estimate the parameter θ, but we also need to set it to get our simulator in the first place. We do so by exploiting historical data together with the allocations of prices picked, of the form $(0, 1, 0)$, $(1, 0, 0)$ and $(0, 0, 1)$ only on these data (all customers were getting the same tariff), and apply the formula (7.8) through the `R`-package `mgcv`, see Wood [2020] (which replaces the λ identity matrix with a slightly more complex definite positive matrix S, see Wood, 2006). The deterministic part of the obtained model is realistic enough: its adjusted R-square on historical observations equals 92% while its mean absolute percentage of error equals 8.82%. Now, as far as noise is concerned, we take multivariate Gaussian noise vectors ε_t, where the covariance matrix Σ was built again based on realistic values. Namely, we considered the time series of residuals associated with our estimation of the consumption. The diagonal coefficients $\Sigma_{j,j}$ were given by the empirical variance of the residuals associated with tariff j, while non-diagonal coefficients $\Sigma_{j,j'}$ were given by the empirical covariance between residuals of tariffs j and j' at times t and $t \pm 48$. (A more realistic model might consider a noise which depends on the half-hour of the day).

Numerical expression obtained. More precisely, the variance terms $\Sigma_{1,1}$, $\Sigma_{2,2}$, and $\Sigma_{3,3}$ were computed with respectively 788, 15 072 and 1 660 observations, while the non-diagonal coefficients were based on fewer observations: 1 318 for $\Sigma_{2,3}$ and 620 for $\Sigma_{1,2}$, but only 96 for $\Sigma_{1,3}$. The resulting matrix Σ is

$$\Sigma = \sigma^2 \begin{pmatrix} 1.11 & 0.46 & 0.04 \\ 0.46 & 1.00 & 0.56 \\ 0.04 & 0.56 & 2.07 \end{pmatrix} \quad \text{with} \quad \sigma = 0.02.$$

To get an idea of the orders of magnitude at stake, we indicate that in the data set considered, the mean consumption remained between 0.08 and 0.21 kWh per half-hour and

[4] Only such a level of aggregation allows a proper estimation (individual consumptions are erratic); Sevlian and Rajagopal, 2018.

that its empirical average equals 0.46.

Off-diagonal coefficients are non-zero. We may test, for each $j \neq j'$, the null hypothesis $\Sigma_{j,j'} = 0$ using the Pearson correlation test; we obtain low p–values (smaller than something of the order of 10^{-13}), which shows that Σ is significantly different from a diagonal matrix. We may conduct a similar study to show that it is not proportional to the all-ones matrix, nor to any matrix with a special form.

5.2 Experiment design: learning added tariff effects

Target creation. We focus on attainable targets c_t, namely, $\phi(x_t, 1) \leqslant c_t \leqslant \phi(x_t, 3)$. To smooth consumption, we pick c_t near $\phi(x_t, 3)$ during the night and near $\phi(x_t, 1)$ in the evening. These hypotheses can be seen as an ideal configuration where targets and customers portfolio are in a way compatible.

Set of legible allocations of price levels. We assume that the electricity provider cannot send Low and High tariffs at the same round and that population can be split in $N = 100$ equal subsets. Thus, the set of price levels $\mathcal{P}$ is restricted to the grid

$$\left\{ \Big(\frac{i}{N}, 1 - \frac{i}{N}, 0\Big), \Big(0, \frac{i}{N}, 1 - \frac{i}{N}\Big), \quad i \in \Big\{0, \dots, N\Big\} \right\}.$$

Training period, testing period. We create one year of data using historical contexts and assume that only Normal tariffs are picked at first: $p_t = (0, 1, 0)$; this is a training period, which corresponds to what electricity providers are currently doing. As they can accurately estimate the covariance matrix Σ by ad-hoc methods, we assume that the algorithm knows the matrix Σ used by the simulator. Then the provider starts exploring the effects of tariffs for an additional month (a January month, based on the historical contexts) and freely picks the p_t according to our algorithm; this is the testing period. The estimation of θ in this testing period is still performed via the formula (7.8) and as indicated above (with the `mgcv` package), including the year when only $p_t = (0, 1, 0)$ allocations were picked. For learning to then focus on the parameters ξ_j, as other parameters were decently estimated in the training period, we modify the exploration term $\alpha_{t,p}$ of (4.3) into

$$\alpha_{t,p} = 2CB_{t-1}(\delta t^{-2}) \|\widetilde{V}_{t-1}^{-1/2} p_t\|, \quad \text{with} \quad \widetilde{V}_{t-1} = \lambda I_d + \sum_{s=1}^{t-1} p_s p_s^{\mathrm{T}}.$$

Finally, we pick a convenient λ.

5.3 Results

Algorithms were run 200 times each. The simplest set of results is provided in Figure 4.3: the regrets suffered on each run are compared to the theoretical orders of magnitude of the regret bounds. As expected, we observe a lower regrets for Model 2. The bottom parts of Figures 4.1–4.2 indicate, for a single run, which allocation vectors p_t were picked over time. During the first day of the testing period, the algorithms explore[5] the effect of

[5]Note that, over the first iterations, the exploration term for Model 2 is much larger than the exploitation term (but quickly vanishes), which leads to an initial quasi-deterministic exploration and an erratic consumption (unlike in Model 1).

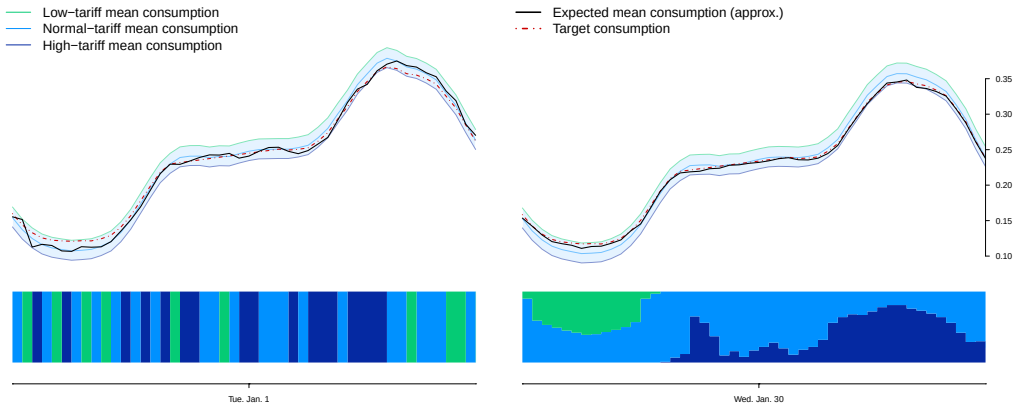


Figure 4.1 – *Left*: January 1st (first day of the testing set). *Right*: January 30th (last day of the testing set).

Top: 200 runs are considered. Plot: average of mean consumptions over 200 runs for the algorithm associated with Model 1 (full black line); target consumption (dashed red line); mean consumption associated with each tariff (Low–1 in green, Normal–2 in blue and High–3 in navy). The envelope of attainable targets is in pastel blue. *Bottom*: A single run is considered. Plot: proportions p_t used over time.

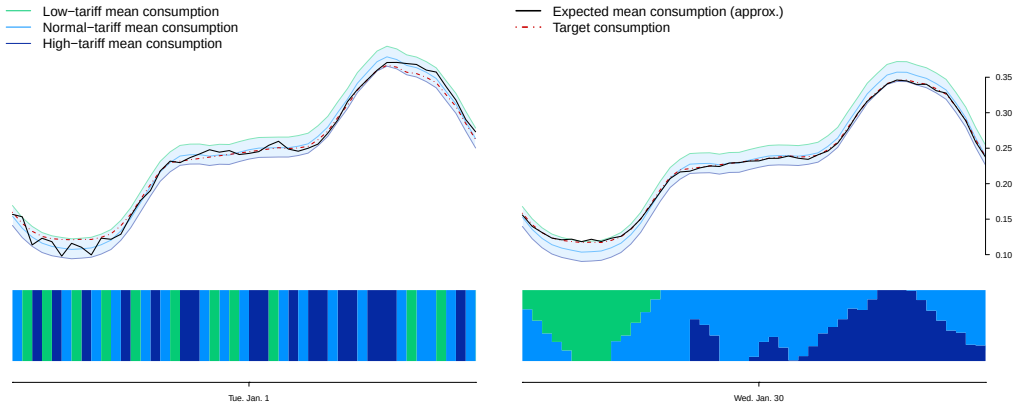


Figure 4.2 – Same legend, but with Model 2 (full black line).

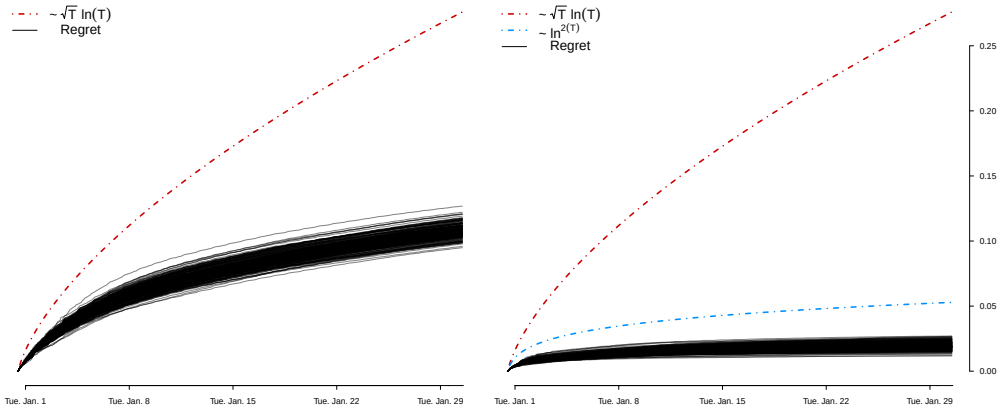


Figure 4.3 – Regret curves for each of the 200 runs for Model 1 (*left*) and Model 2 (*right*). We also provide plots of $c\sqrt{T}\ln T$ and $c'\ln^2(T)$ for some well-chosen constants $c, c' > 0$; these are the rates to be considered as the covariance matrix Σ is assumed to be known.

tariffs by sending the same tariff to all customers (the p_t vectors are Dirac masses) while at the end of the testing period, they cleverly exploit the possibility to split the population in two groups of tariffs. We obtain an approximation of the expected mean consumption $\phi(x_t, p_t)$ by averaging the 200 observed consumptions, and this is the main (black, solid) line to look at in the top parts of Figures 4.1–4.2. Four plots are depicted depending on the day of the testing period (first, last) and of the model considered. These (approximated) expected mean consumptions may be compared to the targets set (dashed red line). The algorithms seem to perform better on the last day of the testing period for Model 2 than for Model 1 as the expected mean consumption seems closer to the target. However, in Model 1, the algorithm has to pick tariffs leading to the best bias-variance trade-off (the expected loss features a variance term). This is why the average consumption does not overlap the target as in Model 2. This results in a slightly biased estimator of the mean consumption in Model 1.

6 Taking into account "rebound" and "side" effects

In our first modeling of the power consumption, we have assumed that it depends only on exogenous variables and on the current price levels. However, choosing a non-standard tariff may modify the power consumption over several hours even after the tariff is back to standard. For example, if some customers need to charge their electric vehicle, they likely do so if the tariff is low. Regardless of the price level after loading, they will probably not consume much. On the contrary, if a high rate is applied all day, they will connect the vehicle, whatever the price, because it needs to be charged. Similarly, a low tariff applied over a whole day will not lead to an increase in consumption at all half-hours of the day. Customer flexibility has some limits that were not taking into account so far. The effect of previous tariffs on consumption is known as "rebound effect". Moreover, for Low tariff, the effect lasts less than desired because consumers wait until they are well within the tariff window to be sure they will consume when prices are low. Conversely, for the High tariff, the effect lasts longer: consumers stop consuming before and resume after the tariff window to make sure they do not consume when prices are high. Therefore, the fall or rise of the energy consumption, induced by a special tariff, may occur a little bit before – for High tariff – or after – for Low tariff – the effective establishment of a special tariff; and it may last longer – for High tariff – or less long – for Low tariff – than the period in which the tariff is actually applied. This is called a "side effect". To take into account these effects, we propose the three modelings above. The first one is a straightforward extension of previous results which has however a high interest in practice. The second approach considers daily consumption profiles: the time step will refer to days and at each new day, the electricity provider will simultaneously observe 48 consumption records (one for each half-hour of the day before), some exogenous variables of the current day and a target profile – namely 48 consumption targets to reach. Therefore, it will choose all the 48 price levels – one convex vector per half-hour. This modeling takes intrinsically into account within-day rebound and side effects: the price levels for a given day are all picked and communicated to customers at the same time. The underlying consumption model will link the chosen tariffs to the 48 power consumption records. Thus, price levels picked for a given half-hour may influence the consumptions associated with all the half-hours of the considered day. Moreover, we detail below how some daily operational constraints can be taken into account with such a approach. The main drawback of this model is that the tariffs picked for a given day (even those chosen for the late evening), have no influence on

the consumption of the next day (even early in the morning). So we finally consider a third and final approach which combines the two previous ones for a daily profile management which takes into account rebound and side effects that can spread over several days.

6.1 A simple approach considering historical price levels

First, at a round $t > t_0$, we may simply consider the price level $p_{t-1}, p_{t-2}, \ldots, p_{t-t_0}$ as exogenous variables. In our previous modeling (Models 1 and 2), we have assumed the consumption to depend on some context $x_t \in \mathcal{X}$. There was no assumption on these variables, which could be deterministic or stochastic. The expected consumption associated with tariffs p_t was $\varphi(x_t, p_t)^{\mathrm{T}}\theta$, where $\varphi(x_t, p_t)$ and θ were vectors of dimension d. Here, we replace $x_t \in \mathcal{X}$ by $(x_t, p_{t-1}, \ldots, p_{t-t_0}) \in \mathcal{X} \times \mathcal{P}^{t_0}$. Since there is no assumption on contextual variables x_t all the previous results apply by adapting the mapping function $\varphi : \mathcal{X} \to \mathbb{R}^d$ into $\widetilde{\varphi} : \mathcal{X} \times \mathcal{P}^{t_0} \to \mathbb{R}^{\widetilde{d}}$. The only change is an increase of the dimension of the parameter vector to estimate – which was θ in our previous modeling and which we denote here by $\widetilde{\theta}$. For example, to take into account past price levels, we could add a linear effect to the previous consumption model, which would lead to $K \times t_0$ new coefficients to be estimated, so $\widetilde{d} = d + Kt_0$. Indeed, by considering the same underlying generalized additive model as before, and denoting by ξ_{-s} the K-vector of parameters associated with the effect of the tariffs chosen at $t - s$, the expected consumption at round $t > t_0$ is:

$$\widetilde{\varphi}(x_t, p_{t-1}, \ldots, p_{t-t_0}, p_t)^{\mathrm{T}}\widetilde{\theta} = \begin{bmatrix} \varphi(x_t, p_t) \\ p_{t-1} \\ \vdots \\ p_{t-t_0} \end{bmatrix} \left[\theta, \xi_{-1}^{\mathrm{T}}, \ldots, \xi_{-t_0}^{\mathrm{T}}\right] = \varphi(x_t, p_t)^{\mathrm{T}}\theta + \sum_{s=1}^{t_0} p_{t-s}^{\mathrm{T}}\xi_{-s}.$$

In the $\widetilde{d}$-vector $\widetilde{\theta}$ to estimate, there are some new coefficients: those of vectors $\xi_{-1}, \ldots, \xi_{-t_0}$, which model the impact of previous tariffs on current consumption. For the noise, we may consider the same terms as in Models 1 and 2 depending on whether the noise depends on chosen tariff or not.

Remark 14. *Note that the noise term does not depend on past tariffs. This kind of approach would belong to the family of autoregressive–moving-average (ARMA) models.*

A slight adaptation of algorithm initialization may be needed: to learn the coefficients linked to $p_{t-1}, \ldots, p_{t-t_0}$, we need to start bandit algorithms at $t > t_0$ rounds. Except in the case of Model 1, when Σ is unknown: the price levels are deterministic for the $\tau = o(T^{2/3})$ first rounds (pure exploration), no adaptation is required. Therefore, this extension of previous results ensures regret bounds of the orders as before – namely, $\mathcal{O}(T^{2/3}\ln^2 T)$ for Model 1 and $\mathcal{O}(\ln^2 T)$ for Model 2. We recall that the dependence in the dimension of the parameter vector – d or $\widetilde{d}$ – in the regret bound is linear for Model 1 and quadratic for Model 2.

Remark 15. *Since the consumption of the past day may have a significant impact on that of the current day, we can, in the same way, consider the power consumptions at* $t - 1$, $t - 2$, $\ldots$ *as exogenous variables.*

We highlight the notion of regret associated with this approach is

$$\overline{R}_T = \sum_{t=t_0+1}^{T} \left(\widetilde{\varphi}(x_t, p_{t-1}, \ldots, p_{t-t_0}, p_t)^{\mathrm{T}}\widetilde{\theta} - c_t\right)^2 + p_t^{\mathrm{T}}\Sigma p_t - \sum_{t=t_0+1}^{T} \min_{p\in\mathcal{P}}\left\{\left(\widetilde{\varphi}(x_t, p_{t-1}, \ldots, p_{t-t_0}, p)^{\mathrm{T}}\widetilde{\theta} - c_t\right)^2 + p^{\mathrm{T}}\Sigma p\right\},$$

Thus, at a round t, the expected loss of our strategy is compared to the expected loss of the best strategy, conditionally to the vectors $p_{t-1}, \ldots, p_{t-t_0}$ picked for our strategy. This latter has no reason to be equal to the expected loss of the "global" best strategy and thus

$$\overline{R}_T \neq \sum_{t=t_0+1}^{T} \left(\widetilde{\varphi}(x_t, p_{t-1}, \ldots, p_{t-t_0}, p_t)^{\mathrm{T}}\widetilde{\theta} - c_t\right)^2 + p_t^{\mathrm{T}}\Sigma p_t - \sum_{t=t_0+1}^{T} \left(\widetilde{\varphi}(x_t, p^\star_{t-1}, \ldots, p^\star_{t-t_0}, p)^{\mathrm{T}}\widetilde{\theta} - c_t\right)^2 + p_t^{\star\mathrm{T}}\Sigma p_t^\star,$$

where, for $t > t_0$, $p_t^\star \in \mathrm{argmin}_{p\in\mathcal{P}}\left\{\left(\widetilde{\varphi}(x_t, p^\star_{t-1}, \ldots, p^\star_{t-t_0}, p)^{\mathrm{T}}\widetilde{\theta} - c_t\right)^2 + p^{\mathrm{T}}\Sigma p\right\}$ and $p_t^\star = p_t$ otherwise.

6.2 A second approach considering daily profile management

6.2.1 Consumption modeling

Another modeling of rebound and side effects relies on the consideration of consumption profiles. Each day, customers consume a certain amount of electricity they distribute during the day according to their daily tariff profiles and to the flexibility of their equipment. Here, a new round t corresponds to a new day, and at each time step, we focus on the daily consumption profile, which gathers $H = 48$ half-hourly consumptions. In all that follows, days are broken down into H time ranges (we consider $H = 48$, this parameter can be changed according to the frequency of the consumption records and those of the possible price changes sent by the electricity provider). Thus, at a day t, prices levels are given for all the 48 half-hours of the day, which is much more realistic than our previous modeling. Indeed, in the Low Carbon London project – which provides the data set we used in the experiments of Section 5, the tariff prices were given a day ahead via the smart meter or text message to mobile phone.

At a day t, the electricity provider observes some context vector $\underline{x}_t \in \underline{\mathcal{X}}$, where $\underline{\mathcal{X}}$ is the new parametric space. This vector gathers contextual variables of the current day: for example, it may contain H records of temperature (one for each half-hour), of wind, the day of the week, the season etc. It also observes H consumption targets $c_c^1, \ldots, c_t^H$ to reach at each half-hour of the day. Then, it picks H convex vectors $p^1, \ldots, p^H$. These vectors are grouped in a $K \times H$ – matrix $\boldsymbol{p} = (p^1, \ldots, p^H)$ and we denote by $\boldsymbol{\mathcal{P}} \subset (\Delta_K)^H$ the new set of legible allocations of price levels.

For each half-hour $h = 1, \ldots, H$, the power consumption depends on the context $\underline{x}_t$ of the day and on all the vectors $p_t^1, \ldots p_t^H$ – but there is no dependence on the tariffs picked the day before. To model the H consumptions, we consider an approach similar to that

presented in Section 2.1: for a day t, at an half-hour h, with tariffs $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$ picked, the power consumption equals:

$$Y_{t,\boldsymbol{p}}^h = \underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \theta^h + \text{ noise}\,.$$

Thus, we consider H models that are linked by the same mapping function $\underline{\phi} : \underline{x} \times \mathbb{R}^{\underline{d}}$, but the H vectors of parameters θ^h differ from an half-hour model to another. The function $\underline{\phi}$ gathers the H power consumption models. For example, if we consider Model 1 for each half-hour, so that, for all $h = 1, \ldots, H$, we have $\mathbb{E}[Y_{t,p}^h] = \varphi(x_t^h, p^h)^{\mathrm{T}} \theta(h)$, where the contextual vector x_t^h contains the exogenous variables associated with half-hour h and $\theta(h)$ is the corresponding parameter vector of dimension d; we set

$$\underline{\phi}\big(\underline{x}_t, \boldsymbol{p}\big) = \Big(\varphi(x_t^1, p^1), \varphi(x_t^2, p^2), \ldots, \varphi(x_t^H, p^H)\Big) \quad \text{and} \quad \theta^h = \begin{pmatrix} \mathbf{0}_{d\times(h-1)} \\ \theta(h) \\ \mathbf{0}_{d\times(H-h)} \end{pmatrix},$$

where $\mathbf{0}_n$ is a n-dimensional vector of 0. The dimension $\underline{d}$ is then $\underline{d} = d \times H$. Note that when some of the components of the context vectors x_t^h are common from one half-hour to another, it is possible to reduce this dimension. As in Model 1, we assume that the noise term depends on the chosen tariffs. More precisely, for each half-hour h, the noise term will be of the form $(p^h)^{\mathrm{T}} \varepsilon_t^h$ – where p^h is the price levels of the considered half-hour. Exactly as in previous modeling (but now, t refers to days), the H-dimensional vectors of K noise vectors $(\varepsilon_1^1, \ldots, \varepsilon_1^H), (\varepsilon_2^1, \ldots, \varepsilon_2^H), \ldots$ will be independent and identically distributed. To model the possible correlations between the noise terms at different half-hours of the same day, we introduce the matrices of covariance $\Sigma^{hh'} = \mathrm{cov}(\varepsilon_1^h, \varepsilon_1^{h'})$, for $1 \leqslant h, h' \leqslant H$. Therefore, the noise term $(p^h)^{\mathrm{T}} \varepsilon_t^h$ depends only on the price levels of the considered half-hour, but noises at h and h' are linked through vectors ε_t^h and $\varepsilon_t^{h'}$. More precisely, for any tariffs k and k', we assume a correlation $\mathrm{cov}(\varepsilon_{t,k}^h, \varepsilon_{t,k'}^{h'}) = \Sigma_{k,k'}^{hh'}$ between residuals at the two half-hours of the same day h and h' when the tariffs k and k' are pickled. This modeling is summed up in Model 3 below.

Model 3: *Daily profile consumptions. At a day t, when the electricity provider observes the context variables $\underline{x}_t$ and picks the $K \times H$–matrix $\boldsymbol{p} = (p^1, \ldots, p^H)$, made of the H convex vectors p^h which correspond to the price levels chosen at the instants h, the daily consumption profile $\mathbf{Y}_{t,\boldsymbol{p}}$ is the H-vector:*

$$\mathbf{Y}_{t,\boldsymbol{p}} = \begin{bmatrix} Y_{t,\boldsymbol{p}}^1 \\ \vdots \\ Y_{t,\boldsymbol{p}}^h \\ \vdots \\ Y_{t,\boldsymbol{p}}^H \end{bmatrix} = \begin{bmatrix} \underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \theta^1 & + & (p^1)^{\mathrm{T}} \varepsilon_t^1 \\ \vdots & \vdots & \vdots \\ \underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \theta^h & + & (p^h)^{\mathrm{T}} \varepsilon_t^h \\ \vdots & \vdots & \vdots \\ \underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \theta^H & + & (p^H)^{\mathrm{T}} \varepsilon_t^H \end{bmatrix},$$

where $\underline{\phi} : \underline{x} \times \boldsymbol{\mathcal{P}} \mapsto \mathbb{R}^{\underline{d}}$ is a known mapping function and $\theta^1, \ldots \theta^H$ are unknown parameter $\underline{d}$–vectors. For any h, the noise K-vectors $\varepsilon_1^h, \varepsilon_2^h, \ldots$ are ρ–sub-Gaussian i.i.d. random variables with $\mathbb{E}[\varepsilon_1^h] = (0, \ldots, 0)^{\mathrm{T}}$. For two half-hours h and h', we denote by $\Sigma^{hh'} = \mathrm{Cov}\big(\varepsilon_1^h, \varepsilon_1^{h'}\big)$ the covariance matrix of vectors ε_1^h and $\varepsilon_1^{h'}$.

Notation. In what follows, we extend the results of Model 1 to Model 3. Since price levels and consumptions are now matrices and vectors (because the whole day is taken into account at each time step t), we write them in bold letters and we refer to a specific half hour by super-scripting the variable by h. All other variables to be updated with this new modeling (contextual vector, mapping function, dimension of parameter vector etc.) – but which remain of the same nature (vector, function, constant, etc.) – are written in underlined letters (mapping function, dimension, etc.). Finally, for losses, expected losses and deviation bounds we will simply replace the subscript p with $\boldsymbol{p}$.

Taking into account some operational constraints. By picking all the price levels of the day at the same time, some operational constraints may be taken into account. In the following examples, we consider that a single tariff is chosen for each half-hour for all the population – namely, the p_t^h vectors are Dirac masses. First, if we do not want that the customers receive the tariff k for more than n half-hours of the same day, we can restrict $\boldsymbol{\mathcal{P}}$ to the set of vectors $\boldsymbol{p}$ which satisfy condition (i) below. To be sure that the tariff k is not picked on too long a period (typically n half-hours), we can impose condition (ii). Finally, the electricity provider could stipulate in the contract of its customers that, if a high tariff (denoted by k) is chosen, a low tariff (denoted by k') will also be chosen on the same day and restricts $\boldsymbol{\mathcal{P}}$ to the set of vectors which satisfy condition (iii).

$$\text{(i)} : \sum_{h=1}^{H} \mathbb{1}_{\{p_k^h=1\}} \leqslant n, \quad \text{(ii)} : \sum_{h=0}^{H-n} \mathbb{1}_{\left\{\sum_{i=1}^{n} \mathbb{1}_{\{p_k^{h+i}=1\}} \geqslant n\right\}} = 0, \quad \text{(iii)} : \sum_{1 \leqslant h,h' \leqslant H} \mathbb{1}_{\{p_k^h=1\}} \mathbb{1}_{\{p_{k'}^{h'}=1\}} > 0.$$

We can also impose that tariffs are not changed too many times on the same day, etc. Therefore, lots of operational constraints may be taken into account in the set $\boldsymbol{\mathcal{P}}$. Note that it is less direct to write these constraints mathematically when the vectors p_t^h are not Dirac masses. By replacing terms $\mathbb{1}_{\{p_k^h=1\}}$ with $\mathbb{1}_{\{p_k^h \neq 0\}}$, we can be sure that the operational constraints are satisfied, but by choosing cleverly in the population the customers which receive tariffs k and k' at the different half-hours of the day, we may consider less restrictive constraints. For example, we may apply High tariff to 80% of the population (and Normal tariff to the 20% left) for as long as we want, without it being sent to one of the customers more than two consecutive hours. Indeed, by splitting the population into 5 groups and by changing the group which receives Normal tariff every half-hour, each customer will receive, High tariff for two hours or less, then Normal tariff for an half-hour break, and again High tariff for two extra hours etc. Finding optimal planning could be a difficult task that we do not deal with here.

Finally, for simplicity of notation (and exposition), we did not introduce any dependence on t in $\boldsymbol{\mathcal{P}}$ – or in $\mathcal{P}$ for Models 1 and 2. However, this would be possible as soon as the sets considered are independent of past price levels – namely, they are deterministic or depend solely on exogenous contextual variables (we denote by $\boldsymbol{\mathcal{P}}_1, \boldsymbol{\mathcal{P}}_2, \ldots$ these sets). Indeed, there would be no change in the regret analysis and this enables to specify different constraints depending on, for example, the day of the week – for example, no high tariff on Sundays or when the temperature is lower than 0°C. In this case, we assume that the knowledge of $\boldsymbol{\mathcal{P}}_t$ is acquired at the beginning of round t so that the regret is now

$$\sum_{t=1}^{T} \ell_{t,\boldsymbol{p}_t} - \sum_{t=1}^{T} \min_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}_t} \ell_{t,\boldsymbol{p}}\,,$$

where $\ell_{t,\boldsymbol{p}}$ is the expected (conditional) loss at t associated with the price levels $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}_t$.

6.2.2 Target profile, loss function and regret

For any day t, the electricity provider sets H targets $c_t^1, \dots, c_t^H$. Being close to the target may be more important at certain times of the day – such as peak hours – than at others. Thus, it can set some coefficients $\kappa_t^1, \dots, \kappa_t^H$ to weigh the losses. The higher κ_t^h is, the closer the consumption at the half-hour h has to be to c_t^h. With no loss of generality, we can assume that these weights are between 0 and 1 – indeed, it is enough to normalize them. Then, it observes the context vector $\underline{x}_t$ and chooses all the price levels for the day $\boldsymbol{p}_t \in \boldsymbol{\mathcal{P}}$ and suffers the loss $\sum_{h=1}^H \kappa_t^h (Y_{t,\boldsymbol{p}_t}^h - c_t^h)^2$. Protocol 5 sums up this online process.

Protocol 5 Target Daily Profile Tracking for Contextual Bandits

Input:
Breakdown of day into H time ranges
Set of legible vectors of convex vectors $\boldsymbol{\mathcal{P}}$
Bound on the consumption C
Transfer function $\underline{\phi} : \underline{\mathcal{X}} \times \boldsymbol{\mathcal{P}} \to \mathbb{R}^{\underline{d}}$
Unknown parameters:
Transfer parameters $\theta^1, \dots, \theta^H \in \mathbb{R}^{\underline{d} \times H}$
Covariance matrices $\Sigma^{11}, \Sigma^{12}, \dots, \Sigma^{HH} \in \left(\mathcal{M}_K(\mathbb{R})\right)^{H \times (H-1)/2}$
for days $t = 1, 2, \dots$ **do**
Observe a context $\underline{x}_t$, the targets $c_t^1, \dots, c_t^H$ and weights $\kappa_t^1, \dots, \kappa_t^H \in [0,1]^H$
Choose, for each time ranges h, an allocation of price levels p_t^h such that

$$\boldsymbol{p}_t = (p_t^1, \dots, p_t^H) \in \boldsymbol{\mathcal{P}}$$

Observe a resulting consumption profile $Y_{t,\boldsymbol{p}} = (Y_{t,\boldsymbol{p}}^1, \dots, Y_{t,\boldsymbol{p}}^H)^\intercal$, with

$$Y_{t,\boldsymbol{p}}^h = \underline{\phi}(\underline{x}_t, \boldsymbol{p})^\intercal \theta^h + (p^h)^\intercal \varepsilon_t^h \qquad \text{(Model 3)}$$

end for
Suffer a loss $\sum_{h=1}^H \kappa_t^h \left(Y_{t,\boldsymbol{p}_t}^h - c_t^h\right)^2$

For any day t, the contextual vector $\underline{x}_t$, the picked tariffs $\boldsymbol{p}_t$ and the targets $c_t^1, \dots, c_t^H$ and the weights κ_t^h are $\mathcal{F}_{t-1}$–measurable, where $\mathcal{F}_{t-1} = \sigma(\varepsilon_1^1, \dots, \varepsilon_1^H, \varepsilon_1^2, \dots, \varepsilon_{t-1}^H)$. Therefore, under Model 3, for an half-hour $h \in \{1, \dots, H\}$, calculations similar to those of Section 3.1 give

$$\mathbb{E}\left[(Y_t^h - c_t^h)^2 \,\middle|\, \mathcal{F}_{t-1}\right] = \left(\underline{\phi}(\underline{x}_t, \boldsymbol{p}_t)^\intercal \theta^h - c_t^h\right)^2 + (p^h)^\intercal \Sigma^{hh} p^h \,,$$

where we recall that $\Sigma^{hh} = \mathrm{Var}(\varepsilon_1^h)$. So, for any day t, by summing over the H time ranges, we obtain that the expected conditional loss is

$$\sum_{h=1}^H \kappa_t^h \, \mathbb{E}\left[(Y_t^h - c_t^h)^2 \,\middle|\, \mathcal{F}_{t-1}\right] = \sum_{h=1}^H \kappa_t^h \left(\left(\underline{\phi}(\underline{x}_t, \boldsymbol{p}_t)^\intercal \theta^h - c_t^h\right)^2 + (p^h)^\intercal \Sigma^{hh} p^h \right).$$

To ensure the minimization of the cumulative loss in expectation, we will compare, at each round t, our choices $\boldsymbol{p}_t \in \boldsymbol{\mathcal{P}}$ to the choices of the best possible strategy – namely the one

which minimizes the cumulative conditional expected loss. The new regret is then

$$\underline{\bar{R}}_T \triangleq \sum_{t=1}^{T} \sum_{h=1}^{H} \kappa_t^h \left(\left(\underline{\phi}(\underline{x}_t, \boldsymbol{p}_t)^{\mathrm{T}} \theta^h - c_t^h \right)^2 + (p_t^h)^{\mathrm{T}} \Sigma^{hh} p_t^h \right) - \sum_{t=1}^{T} \min_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \sum_{h=1}^{H} \kappa_t^h \left(\left(\underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \theta^h - c_t^h \right)^2 + (p^h)^{\mathrm{T}} \Sigma^{hh} p^h \right). \tag{4.42}$$

In Section 6.4, we provide an optimistic algorithm and a regret bound on $\underline{\bar{R}}_T$ of order $\mathcal{O}\big(T^{2/3}\big)$ which are similar to those of Section 3. Here we give a sketch on how we deduce this analyze from the previous one.

Proof sketch. We emphasize that, for two different half-hours of the same day $h \neq h'$, and two tariffs k, k', the correlation terms $\mathbb{E}\big[\varepsilon_k^h \varepsilon_{k'}^{h'}\big] = \Sigma_{kk'}^{hh'}$ do not appear in the expression of the expected loss. Thus, the algorithm of Section 6.4.1 below only need to estimate parameter vectors θ^h and the matrices Σ^{hh} – there is no need to estimate matrices $\Sigma^{hh'}$. We do so exactly – but it is H times slower – as in Sections 3.3.1 and 3.3.2, respectively and obtain some deviation inequalities on the estimations of parameter vectors θ^h and the matrices for Σ^{hh}. The deviation bounds associated with the estimations of θ^h do not depend on h because the H expected consumptions $\underline{\phi}(\underline{x}_t, \boldsymbol{p}_t)^{\mathrm{T}} \theta^h$ are linked by the same vector $\underline{\phi}(\underline{x}_t, \boldsymbol{p}_t)$. Thus, only a slight adaptation of the deviation bound $\alpha_{t,\boldsymbol{p}}$ (in the level risk δ and up to the multiplicative constant H) is required in the optimistic algorithm – details are provided in Section 6.4.1. Then, under boundedness assumption similar to Assumption 1, by union bound over the half-hours h, we will obtain a regret bound in high probability of the same order as Model 1 – namely, of order $\mathcal{O}\big(T^{2/3} \ln^2 T\big)$.

6.3 Final approach: daily profile model with historical price levels

This third approach combines the two previous ones. Indeed we consider a daily consumption model – namely, Model 3, but we include in the exogenous variables $\underline{x}_t$ price levels chosen during the last t_0 days $(\boldsymbol{p}_{t-1}, \dots \boldsymbol{p}_{t-t_0})$. This is exactly what we did in Section 6.1 where we included past tariffs in the contextual variables of Models 1 and 2. Thus, we now replace the vector $\underline{x}_t$ by $(\underline{x}_t, p_{t-1}^1, \dots, p_{t-1}^H, \dots, p_{t-t_0}^1, \dots, p_{t-t_0}^H)$, so the dimension of the contextual vectors is increased by KHt_0. We recall that there is no assumption on these exogenous variables, they can be deterministic as well as stochastic and that, therefore, all the previous results apply with an adaptation of the mapping function. For example, for a given half-hour h, the addition of linear effects for past price levels leads to this new modeling:

$$\widetilde{\underline{\phi}}(\underline{x}_t, p_{t-1}^1, \dots, p_{t-1}^H, \dots, p_{t-t_0}^1, \dots, p_{t-t_0}^H, \boldsymbol{p}_t)^{\mathrm{T}} \widetilde{\theta}^h = \varphi(\underline{x}_t, \boldsymbol{p}_t)^{\mathrm{T}} \theta^h + \sum_{s=1}^{t_0} \sum_{h'=1}^{H} (p_{t-s}^{h'})^{\mathrm{T}} \xi_{-s}^{hh'},$$

where the K-vector $\xi_{-s}^{hh'}$ models the effect of the price levels applied at half-hour h' on day $t-s$ on the consumption at half-hour h on day t. The only change is the increase of the dimension of the vector $\widetilde{\theta}^h$: there are KHt_0 new coefficients $\xi_{-s}^{hh'}$ to estimate. We can then generalize the definition of regret and get some bound to control it.

6.4 Regret bound for daily profile demand side management (*i.e.*, for the model discussed in Section 6.2)

The present section extends the results presented in Section 3 and follows the same structure. Indeed, we first generalize the optimistic algorithm (4.3) and then provide the analysis of the regret defined in Equation (4.42), which differs only in a few points from that of Section 3.3.3. We recall that we aim to minimize the regret, that can be rewritten:

$$\underline{\bar{R}}_T \triangleq \sum_{t=1}^{T} \ell_{t,\boldsymbol{p}_t} - \sum_{t=1}^{T} \min_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}}\,,$$

where the instantaneous expected loss associated with the choices $\boldsymbol{p}\in\boldsymbol{\mathcal{P}}$ is

$$\ell_{t,\boldsymbol{p}} \triangleq \sum_{h=1}^{H} \kappa_t^h \left(\left(\underline{\phi}(\underline{x}_t,\boldsymbol{p})^{\mathrm{T}}\theta_{t-1}^h - c_t^h\right)^2 + (p^h)^{\mathrm{T}}\Sigma^{hh}p^h \right),$$

where τ denotes the number of iterations used at the beginning of the algorithm to estimate the covariance matrices (pure deterministic exploration). We still assume that the power consumption at any day t and any half-hour h lies in $[0,C]$ and we update Assumption 1 of Model 1 to Model 3 by requiring that, no matter how the environment provides the $\underline{x}_t$,

$$\|\underline{\phi}\|_\infty \leqslant 1, \quad \|\theta^h\|_\infty \leqslant C, \quad \forall\, \boldsymbol{p}\in\boldsymbol{\mathcal{P}},\ \underline{\phi}(\underline{x}_t,\boldsymbol{p})^{\mathrm{T}}\theta^h \in [0,C] \quad \text{and} \quad (p^h)^{\mathrm{T}}\Sigma^h p^{hh} \leqslant \Gamma. \tag{4.43}$$

Since all the κ_t^h are bounded between 0 and 1, a consequence of all these boundedness assumptions is that $\underline{L} = H(C^2+\Gamma)$ upper-bounds the (conditionally) expected losses.

6.4.1 Optimistic Algorithm

In Section 3, at round t, we estimate θ and Σ to provide, for any $p\in\mathcal{P}$, an estimation $\widehat{\ell}_{t,p}$ for the expected loss $\ell_{t,p}$ with an associated deviation bound $\alpha_{t,p}$. Then, the optimistic algorithm (4.3) picks the price levels optimistically – namely, the ones which minimize $\{\widehat{\ell}_{t,p} - \alpha_{t,p}\}$. Here, in exactly the same way, the key is to estimate the parameter vectors θ^h and the correlation matrices Σ^{hh} – this will be done exactly as in Section 3.2. So, for any $\boldsymbol{p}\in\boldsymbol{\mathcal{P}}$, we will be able to estimate the expected loss $\ell_{t,\boldsymbol{p}}$ and an associated deviation bound $\alpha_{t,\boldsymbol{p}}$.

More precisely, for some parameter $\underline{\lambda} > 0$, at a day t, the H estimators $\widehat{\theta}_t^h$ of the $\underline{d}$-vectors θ^h are computed in parallel, with for $h\in\{1,\dots H\}$,

$$\widehat{\theta}_t^h \triangleq (\underline{V}_t)^{-1} \sum_{s=1}^{t} Y_{s,\boldsymbol{p}_s}^h \underline{\phi}(\underline{x}_s,\boldsymbol{p}_s)\,, \quad \text{where} \quad \underline{V}_t \triangleq \underline{\lambda}\mathrm{I}_{\underline{d}} + \sum_{s=1}^{t} \underline{\phi}(x_s,\boldsymbol{p}_s)\underline{\phi}(x_s,\boldsymbol{p}_s)^{\mathrm{T}}\,. \tag{4.44}$$

Moreover, for some fixed $\tau > 0$, we consider the H estimators $\widehat{\Sigma}_\tau^{hh}$ of the $K\times K$-matrices Σ^{hh}, with for $h\in\{1,\dots H\}$,

$$\widehat{\Sigma}_\tau^{hh} \in \underset{\widehat{\Sigma}^{hh}\in\mathcal{M}_K(\mathbb{R})}{\operatorname{argmin}} \sum_{s=1}^{\tau} \left(\left(\widehat{Z}_s^h\right)^2 - {p_s^h}^{\mathrm{T}}\widehat{\Sigma}^{hh}p_s^h \right)^2, \text{ where } \widehat{Z}_s^h \triangleq Y_{s,\boldsymbol{p}_s}^h - \left[\underline{\phi}(\underline{x}_s,\boldsymbol{p}_s)^{\mathrm{T}}\widehat{\theta}_\tau^h\right]_C. \tag{4.45}$$

Then, for $t\geqslant\tau+1$, given the mapping function $\underline{\phi}$, we compute the estimator $\widehat{\ell}_{t,\boldsymbol{p}}$ of the instantaneous expected loss $\ell_{t,\boldsymbol{p}}$ associated with the choices $\boldsymbol{p}\in\boldsymbol{\mathcal{P}}$:

$$\widehat{\ell}_{t,\boldsymbol{p}} \triangleq \sum_{h=1}^{H} \kappa_t^h \left(\left(\left[\underline{\phi}(\underline{x}_t,\boldsymbol{p})^{\mathrm{T}}\widehat{\theta}_{t-1}^h\right]_C - c_t^h \right)^2 + (p^h)^{\mathrm{T}}\widehat{\Sigma}_\tau^{hh}p^h \right).$$

We detail in the next section how the deviation bound below ensures some theoretical guarantees on the regret but here is an intuition of how we build it: we simply sum over h some deviation bounds of

$$\Big(\big[\underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \widehat{\theta}^h_{t-1} \big]_C - c^h_t \Big)^2 + (p^h)^{\mathrm{T}} \widehat{\Sigma}^{hh}_\tau p^h,$$

which are the estimators of the expected losses associated with the half-hours h. These H deviation bounds are adaptations of the one defined in Proposition 1 – the main change is the risk level δ that needs to be updated to ensure, by an application of the union bound, that each expected loss (one per half-hour) is correctly estimated. Therefore, we consider the following deviation bound:

$$\alpha_{t,\boldsymbol{p}} \triangleq \left(\sum_{h=1}^{H} \kappa^h_t \right) \left[\underline{\upsilon} + \min \left\{ \underline{L}, 2C \underline{B}_t(\delta/t^2 H) \left\| \underline{V}^{-1/2}_{t-1} \underline{\phi}(\underline{x}_t, \boldsymbol{p}) \right\| \right\} \right],$$

where $\underline{\upsilon}$ and $\underline{B}_t$ are adaptations of υ and B_t defined in the next section. We emphasize that the term in the brackets does not depend on h: all the deviation bounds are equal because of the same underlying mapping function $\underline{\phi}$ that models the consumptions. Finally, the optimistic algorithm for daily profile management picks, for $t \geqslant \tau + 1$,

$$\boldsymbol{p}_t \in \operatorname*{argmin}_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \big\{ \widehat{\ell}_{t,\boldsymbol{p}} - \alpha_{t,\boldsymbol{p}} \big\}. \tag{4.46}$$

6.4.2 Analysis of the regret

Theorem 6. *Fix a risk level $\delta \in (0,1)$ and a time horizon T. Assume that the boundedness assumptions* (4.43) *hold. The optimistic algorithm* (4.46) *with an initial exploration of length $\tau = \mathcal{O}(T^{2/3})$ rounds satisfies*

$$\bar{\underline{R}}_T \leqslant \mathcal{O}\left(H T^{2/3} \ln^2 \left(\frac{T}{\delta} \right) \sqrt{\ln \frac{1}{\delta}} \right),$$

with probability at least $1 - \delta$.

Remark 16. *If all the matrices Σ^{hh} are known, the regret bound is in $\mathcal{O}\big(\sqrt{T \ln T \ln(T/\delta)}\big)$.*

We emphasize that the regret bound depends linearly on the number of time ranges H. The analysis is really similar to the one of the regret associated with Model 1 and the proof below follows the same sketch as the proof of Proposition 1.

Proof of Theroem 6. First, by using Lemmas 4 and 5 and for an application of the union bound, we prove that with probability at least $1 - \delta$, the deviation inequalities on the $2H$ estimators $\widehat{\theta}^h_t$ and $\widehat{\Sigma}^{hh}_\tau$ below are true for all $t > \tau$, where τ the exploration budget to estimate covariance matrices. Then, all the instantaneous expected regrets $(\ell_{t,\boldsymbol{p}_t} - \min_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}})$ are bounded by $2\alpha_{t,\boldsymbol{p}}$. Finally, it only remains to sum these deviation bounds over t and to conclude the proof by applying Lemma 7.

For a given $t > \tau$ and for any $h \in \{1, \dots, H\}$, Lemmas 4 and 5 (see Sections 3.3.1 and 3.3.2, respectively) ensure that

$$\mathbb{P}\Big(\|(\underline{V}_{t-1})^{1/2}(\widehat{\theta}^h_{t-1} - \theta^h)\| \geqslant \underline{B}_t(\delta/t^2 H) \Big) \leqslant \frac{\delta}{t^2 H}$$

$$\text{where} \quad \underline{B}_t(\delta/t^2 H) = \sqrt{\underline{\lambda} d \underline{C}} + \rho \sqrt{2 \ln \frac{t^2 H}{\delta} + d \ln\big(1 + t/\underline{\lambda}\big)} \quad \text{and}$$

$$\mathbb{P}\bigg(\sup_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \Big| (p^h)^{\mathrm{T}} \big(\widehat{\Sigma}^{hh}_\tau - \Sigma^{hh} \big) p^h \Big| \geqslant \underline{\upsilon} \bigg) \leqslant \frac{\delta}{2H} \quad \text{where} \quad \underline{\upsilon} = \upsilon_\tau(\delta/2H)\,.$$

Therefore, by an application of the union bound, with probability at least $1 - \delta$, for all $h \in \{1, \dots, H\}$ and for all $t > \tau \geqslant 2$, we have

$$\big\| \underline{V}^{1/2}_{t-1} (\widehat{\theta}^h_{t-1} - \theta^h) \big\| \leqslant \underline{B}_t(\delta/t^2 H) \quad \text{and} \quad \sup_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \Big| (p^h)^{\mathrm{T}} \big(\widehat{\Sigma}^{hh}_\tau - \Sigma^{hh} \big) p^h \Big| \leqslant \underline{\upsilon}\,. \tag{4.47}$$

In the same way as in Step 1 of the proof of Proposition 1, as soon as these $2H$ inequalities are true, we get

$$\max_{h \in \{1, \dots, H\}} \bigg(\Big| \Big(\underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \theta^h - c^h_t \Big)^2 - \Big(\big[\underline{\phi}(\underline{x}_t, \boldsymbol{p})^{\mathrm{T}} \widehat{\theta}^h_{t-1} \big]_C - c^h_t \Big)^2 \Big| + \Big| (p^h)^{\mathrm{T}} \Sigma^{hh} p^h - (p^h)^{\mathrm{T}} \widehat{\Sigma}^h_\tau p^h \Big| \bigg)$$

$$\leqslant 2C \underline{B}_t\big(\delta/t^2 H\big) \big\| \underline{V}^{-1/2}_{t-1} \underline{\phi}(\underline{x}_t, \boldsymbol{p}) \big\| + \underline{\upsilon}\,.$$

By summing over all the half-hours, for any $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$, we obtain that

$$\Big| \ell_{t,\boldsymbol{p}} - \widehat{\ell}_{t,\boldsymbol{p}} \Big| \leqslant \Big(\sum_{h=1}^H \kappa^h \Big) \Big[2C \underline{B}_t(\delta/t^2 H) \big\| \underline{V}^{-1/2}_{t-1} \underline{\phi}(\underline{x}_t, \boldsymbol{p}) \big\| + \underline{\upsilon} \Big]\,.$$

Exactly as in Step 2 of the proof of Proposition 1, by using the inequality above for $\boldsymbol{p} = \boldsymbol{p}_t$ and $\boldsymbol{p} = \boldsymbol{p}^\star_t$ (where $\boldsymbol{p}^\star_t \triangleq \min_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}}$), combined with the definition of Algorithm (4.46), we obtain that, with probability at least $1 - \delta$,

$$\sum_{t=\tau+1}^T \ell_{t,\boldsymbol{p}_t} - \sum_{t=\tau+1}^T \min_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}} \leqslant 2 \sum_{t=\tau+1}^T \alpha_{t,\boldsymbol{p}_t}.$$

Lemma 7 ensures that no matter how the environment and provider pick the $\underline{x}_t$ and $\boldsymbol{p}_t$,

$$\sum_{t=\tau+1}^T \min \Big\{ \underline{L}, 2C \underline{B}_t(\delta/t^2 H) \big\| \underline{V}^{-1/2}_{t-1} \underline{\phi}(\underline{x}_t, \boldsymbol{p}) \big\| \Big\} \leqslant \bigg(\sqrt{\big(2C \underline{B}_T(\delta/T^2 H)\big)^2 + \frac{\underline{L}^2}{2}} \sqrt{\underline{d} T \ln \frac{\underline{\lambda} + T}{\underline{\lambda}}} \bigg)\,.$$

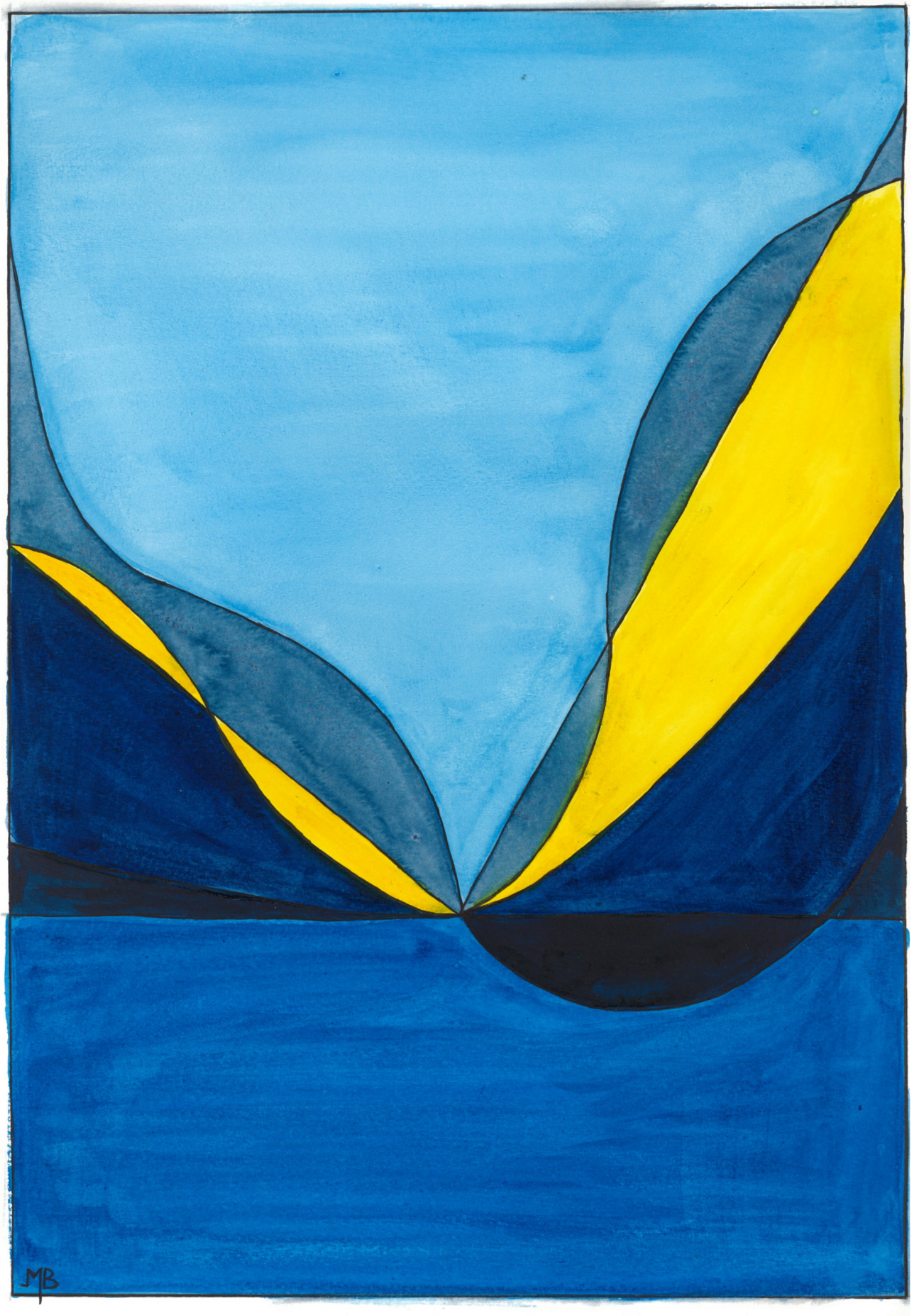
So, by using the definition of $\alpha_{t,\boldsymbol{p}_t}$ and summing over h, as all the κ^h_t lie in $[0,1]$, we get

$$\sum_{t=\tau+1}^T \alpha_{t,\boldsymbol{p}_t} \leqslant H \bigg(T \underline{\upsilon} + \sqrt{\big(2C \underline{B}_T(\delta/T^2 H)\big)^2 + \frac{\underline{L}^2}{2}} \sqrt{\underline{d} T \ln \frac{\underline{\lambda} + T}{\underline{\lambda}}} \bigg)\,.$$

Finally, we recall that $\underline{\upsilon} = \upsilon_\tau(\delta/2H) = \mathcal{O}\big(\frac{1}{\tau} \ln^2(H\tau/\delta) \sqrt{H/\delta}\big)$ – see Lemma 5 – and we bound the instantaneous expected regret $\big(\ell_{t,\boldsymbol{p}_t} - \min_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}}\big)$ by $\underline{L}$ for the τ first rounds and we get

$$\bar{R}_T \leqslant \tau \underline{L} + \sum_{t=\tau+1}^T \alpha_{t,\boldsymbol{p}_t} = \mathcal{O}\bigg(\tau \underline{L} + H \Big(\sqrt{T \ln T \ln(TH/\delta)} + \frac{T}{\sqrt{\tau}} \ln^2(\tau H/\delta) \sqrt{\ln(H/\delta)} \Big) \bigg).$$

As $\underline{L} = H(C^2 + \Gamma)$, we obtain the linear dependence on H and picking τ of order $T^{2/3}$ concludes the proof. □


MB

5

Target tracking for contextual bandits: a generalization to general loss functions

This chapter generalizes the contextual bandit approach to demand management presented in Chapter 4. The target mean consumption is still fixed at each round and the average consumption still modeled as a function of the distribution of prices sent and of some contextual variables; but the performance of our strategies is now measured with any loss function - instead of a quadratic loss function. More precisely, we first show how to extend the previous algorithm and regret bound in the case of a Gaussian model and a polynomial loss function and we then briefly present other possible extensions. Finally, we illustrate the interest of this work with a practical example of a non-quadratic loss function.

1 Introduction

Chapter 4 proposed a contextual-bandit approach for demand side management by offering price incentives. More precisely, a target mean consumption was set at each round and the mean consumption was modeled as a complex function of the distribution of prices sent and of some contextual variables such as the temperature, weather, and so on. The performance of our strategies was measured in quadratic losses through a regret criterion. We offer $T^{2/3}$ upper bounds on this regret (up to poly-logarithmic terms). This chapter extends the previous work (see Sections 2 and 3 of Chapter 4) to loss functions more general than the quadratic loss. This is good news for the potential applications of bandits for demand side managements in real life, as losses are not generally quadratic. For example, over-consumption can have a much more devastating effect on the power grid than under-consumption. Therefore, taking into account non-symmetrical losses is a major issue. Section 5 discusses how this result could be applied in practice and how to define the losses according to the objectives of the electricity supplier. In this chapter, we consider any loss function regular enough and known in advance (it may not be always the case in practice). As any function may be approximated by an interpolation of Lagrange polynomials, we first focus on polynomial loss functions, which have some regularity properties useful for establishing regret bounds. Therefore we will obtain theoretical guarantees on the regret of a more practical modeling of demand response policy.

We retain the two models of the power consumption introduced in Chapter 4, so conditionally to the tariff picked by the electricity provider, the expected power consumption is still modeled by a semi-parametric function (to be estimated) of some contextual variables and of the tariffs chosen. The two models differ in their noise terms: for the first, noise depends on chosen tariffs, whereas for the second, it does not. We assume that noises are Gaussian – instead of sub-Gaussian – and that we know their variance or covariance matrix (depending on the power consumption model considered). In addition, we consider a loss function which is a polynomial function of the difference between the power consumption and the target consumption. With no loss of generality, this loss function is firstly fix over time in Sections 2 and 2.3 and we extend this framework to loss functions which may change at each round in Section 4. Then, we show how to control the cumulative loss though a $\sqrt{T}\ln T$ regret bound.

In Section 2, we present the two power consumption modelings and the new target tracking protocol; we also show how it is possible to compute the expected losses (under the normality assumptions on the noises). Section 3 states the regret bound. Its analysis is similar to that of Theorem 4 in Chapter 4: the key is to prove that a good estimation of power consumption provides a good estimation of expected losses. Section 4 discusses other possible extensions. Finally, Section 5 presents a practical application in which a non-quadratic loss may be useful.

2 Modeling of the power consumption and expectation of the polynomial losses

First of all, we update Models 1 and 2 presented in Chapter 4: from now on, we consider Gaussian noises for which we know the variance or the covariance matrix (depending on whether or not the noise depends on the chosen price levels). These models are defined

below and we refer to them as Models 1G and 2G. Next, we introduce the loss function, which is a polynomial function of the difference between the observed and the target power consumption. We also define the new protocol (see Protocol 6). In Section 2.3, we present the calculations, for both models, of the expected losses.

2.1 Gaussian modelings of the power consumption

We have retained the notations of Chapter 4. Therefore, the power consumption associated with a contextual vector $x_t \in \mathcal{X}$ and some price levels $p_t \in \mathcal{P}$ is denoted by Y_{t,p_t}; furthermore the electricity provider, which chooses vectors p_t, aims to approach a target consumption c_t. For both models, this consumption will still be of the form $Y_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta +$ noise, where the mapping function $\varphi : \mathcal{X} \to \mathbb{R}^d$ is known and the parameter $\theta \in \mathbb{R}^d$ has to be estimated. As in Chapter 4, the first model considers a noise term that depends on the chosen vector p_t, while the second model considers a global noise term.

Model 1G: *Tariff-dependent Gaussian noise. When the electricity provider picks the convex vector p, the mean consumption obtained at time instance t equals*

$$Y_{t,p} = \varphi(x_t, p)^{\mathrm{T}}\theta + p^{\mathrm{T}}\varepsilon_t \,.$$

The noise vectors $\varepsilon_1, \varepsilon_2, \ldots$ are Gaussian i.i.d. random variables with $\mathbb{E}[\varepsilon_1] = (0, \ldots, 0)^{\mathrm{T}}$. We denote by $\Sigma = \mathrm{Var}(\varepsilon_1)$ their covariance matrix.

We point out that the only difference with Model 1 in Chapter 4 is the Gaussian assumption: in the previous modeling, noises were assumed to be only sub-Gaussian. Similarly, we introduce the following global noise model, which is the Gaussian version of Model 2 of Chapter 4.

Model 2G: *Global Gaussian noise. When the electricity provider picks the convex vector p, the mean consumption obtained at time instance t equals*

$$Y_{t,p} = \varphi(x_t, p)^{\mathrm{T}}\theta + \mathrm{e}_t \,.$$

The scalar noises $\mathrm{e}_1, \mathrm{e}_2, \ldots$ are Gaussian i.i.d. random variables, with $\mathbb{E}[\mathrm{e}_1] = 0$. We denote by $\sigma^2 = \mathrm{Var}(\mathrm{e}_1)$ the variance of the random noises e_t.

In what follows, we assume that Σ or σ^2 are known. Therefore, unlike the algorithm in Chapter 4, here, it will not be necessary to estimate Σ and initial exploration will be required to do so.

Assumption 3 – *Knowledge of Σ or σ^2.* The covariance matrix Σ (for Model 1G) and the variance σ^2 (for Model 2G) are known.

In the next section, we introduce the loss function, which is no longer the quadratic difference between energy consumption and its target. Protocol 6,which is almost identical to Protocol 4 in Chapter 4, recalls the steps involved in the on-line tracking of a power consumption target c_t.

2.2 Target tracking with polynomial losses

In the following, we introduce the concept of polynomial loss function. We emphasize that focusing on the family of polynomial loss functions comes with some generality: we could approximate any nonpolynomial loss function with a polynomial interpolation in the Lagrange form.

Assumption 4 – *Polynomial loss function (of degree q).* For a power consumption y and a target consumption c, the loss function $\ell : \mathbb{R}^2 \to \mathbb{R}$ satisfies $\ell(y, c) = P(y - c)$, with P a polynomial function of degree q, which can be written in the form

$$P : x \mapsto \sum_{n=0}^{q} a_n x^n \,.$$

We could also have defined the loss function $\ell(y, c) = P(c - y)$ instead of $P(y - c)$. Both choices are possible and symmetrical. Moreover, we make no assumption about the non-negativity of the coefficients a_n, for $n = 0, \ldots, q$, as even if it seems counter-intuitive, the losses can be non-positive. Indeed, to approach some real loss function, the Lagrange polynomial interpolation could provide a polynomial function which would be negative in some places.

Since the loss function is now defined, the online protocol that models the online tracking of a power consumption target is stated in Protocol 6. We also recall that the choices x_t, c_t and p_t need to be $\mathcal{F}_{t-1}$–measurable, where $\mathcal{F}_{t-1} \triangleq \sigma(\varepsilon_1, \ldots, \varepsilon_{t-1})$ for Model 1G and $\mathcal{F}_{t-1} \triangleq \sigma(\mathrm{e}_1, \ldots, \mathrm{e}_{t-1})$ for Model 2G.

2.3 Conditional expectation of the losses

In this section we show that, for any round t, the conditional expected loss is a polynomial function of the distance between the expected power consumption $\varphi(x_t, p_t)^{\mathrm{T}}\theta$ and its target c_t. Indeed, it is possible to obtain an equality of the form:

$$\ell_{t,p_t} = \mathbb{E}\Big[\ell(Y_{t,p_t}, c_t) \,\Big|\, \mathcal{F}_{t-1}\Big] = \sum_{m=0}^{q} \kappa_m(p_t)\big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^m \,,$$

where the coefficients $\kappa_m(p_t)$ can be computed and depend on the variance of the noise, namely on the covariance matrix Σ or on σ^2. We highlight that, exactly as in Chapter 4, we aim to minimize the cumulative conditional expected loss and that the regret $\bar{R}_T$ further defined will be written with the conditional expected $\ell_{t,p}$, for any $p \in \mathcal{P}$. This equation is actually crucial to obtain regret bounds. Indeed by replacing θ by its estimator, we may consider an estimator of ℓ_{t,p_t}, then the confidence interval on the estimation of θ previously obtained (see Section 3 of Chapter 4) will lead to confidence intervals on the conditional expected loss estimations and therefore to the inequalities required to bound the regret.

To obtain this equality, we first introduce the distance between the expected power consumption and its target

$$d_{t,p_t} \triangleq \varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t \,.$$

Protocol 6 Target Tracking for Contextual Bandits with Polynomial Loss Function

Input

Parametric context set $\mathcal{X}$

Set of legible convex weights $\mathcal{P}$

Bound on mean consumptions C

Transfer function $\varphi : \mathcal{X} \times \mathcal{P} \to \mathbb{R}^d$

Polynomial $P : x \mapsto \sum_{n=0}^{q} a_n x^n$ and loss function $\ell : (y, c) \mapsto P(y - c)$

Unknown parameters

Transfer parameter $\theta \in \mathbb{R}^d$

Covariance matrix Σ of size $K \times K$ (Model 1G)

Variance σ^2 (Model 2G)

for $t = 1, 2, \ldots$ **do**

Observe a context $x_t \in \mathcal{X}$ and a target $c_t \in (0, C)$

Choose an allocation of price levels $p_t \in \mathcal{P}$

Observe a resulting mean consumption

$$Y_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t \quad \text{(Model 1G)}$$

$$Y_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta + \mathrm{e}_t \quad \text{(Model 2G)}$$

Suffer a loss $\ell(Y_{t,p_t}, c_t) = P(Y_{t,p_t} - c_t) = \sum_{n=0}^{q} a_n (Y_{t,p_t} - c_t)^n$

end for

Aim

Minimize the cumulative loss $L_T = \displaystyle\sum_{t=1}^{T} \sum_{n=0}^{q} a_n (Y_{t,p_t} - c_t)^n$

Moreover, we denote by e_{t,p_t} the noise term associated with the price levels $p_t \in \mathcal{P}$, which is equal to $p_t^{\mathrm{T}}\varepsilon_t$ (Model 1G) or to e_t (Model 2G), so we get that $Y_{t,p_t} - c_t = d_{t,p_t} + e_{t,p_t}$. As in the case of quadratic loss, we will develop $\ell(Y_{t,p_t}, c_t) = P(Y_{t,p_t} - c_t) = P(d_{t,p_t} + e_t)$ to obtain the polynomial $d \mapsto \sum_{m=0}^{q} \kappa_m(p_t) d^m$.

Because of the expression of the polynomial P (see Assumption 4), for any $(d, e) \in \mathbb{R}^2$, by expanding powers of $d + e$ with the binomial formula, $P(d + e)$ is equal to

$$P(d+e) = \sum_{n=0}^{q} a_n (d+e)^n = \sum_{n=0}^{q} a_n \sum_{k=0}^{n} \binom{n}{k} d^k e^{n-k} = \sum_{k=0}^{q} \left(\sum_{n=k}^{q} a_n \binom{n}{k} e^{n-k} \right) d^k \,.$$

For any round t, with the expression above, as p_t, x_t and c_t are $\mathcal{F}_{t-1}$-measurable, the expected loss associated with price vector p_t can new be written, by linearity of conditional expectation, as

$$\begin{aligned} \ell_{t,p_t} &= \mathbb{E}\big[\ell(Y_{t,p_t}, c_t) \,\big|\, \mathcal{F}_{t-1}\big] = \mathbb{E}\big[P(d_{t,p_t} + e_{t,p_t}) \,\big|\, \mathcal{F}_{t-1}\big] \\ &= \sum_{k=0}^{q} \left(\sum_{n=k}^{q} a_n \binom{n}{k} \mathbb{E}\big[e_{t,p_t}^{n-k} \,\big|\, \mathcal{F}_{t-1}\big] \right) d_{t,p_t}^k \,. \end{aligned}$$

By a mere change in the summation index, we obtain that

$$\sum_{n=k}^{q} a_n \binom{n}{k} \mathbb{E}\big[e_{t,p_t}^{n-k} \,\big|\, \mathcal{F}_{t-1}\big] = \sum_{i=0}^{q-k} a_{k+i} \binom{k+i}{k} \mathbb{E}\big[e_{t,p_t}^{i} \,\big|\, \mathcal{F}_{t-1}\big] \,,$$

so the expected loss is

$$\ell_{t,p_t} = \sum_{k=0}^{q} \left(\sum_{i=0}^{q-k} a_{k+i} \binom{k+i}{k} \mathbb{E}\big[e_{t,p_t}^i \,\big|\, \mathcal{F}_{t-1}\big] \right) d_{t,p_t}^k$$
$$= \sum_{n=0}^{q} \left(\sum_{k=0}^{q-n} a_{n+k} \binom{n+k}{n} \mathbb{E}\big[e_{t,p_t}^k \,\big|\, \mathcal{F}_{t-1}\big] \right) d_{t,p_t}^n \,.$$

Therefore, we can rewrite the expected loss associated with price levels p in the following way:

$$\ell_{t,p_t} = \mathbb{E}\big[P(d_{t,p_t} + e_{t,p_t}) \,\big|\, \mathcal{F}_{t-1}\big] = \sum_{m=0}^{q} \kappa_m(p_t) d_{t,p_t}^m = \sum_{m=0}^{q} \kappa_m(p_t) \Big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\Big)^m$$
$$\text{where} \quad \kappa_m(p_t) \triangleq \sum_{k=0}^{q-m} a_{m+k} \binom{m+k}{m} \mathbb{E}\big[e_{t,p_t}^k \,\big|\, \mathcal{F}_{t-1}\big] \,.$$

To compute the coefficients $\kappa_m(p_t)$, it is sufficient to explicitly calculate the expectations $\mathbb{E}[e_{t,p_t}^k \,|\, \mathcal{F}_{t-1}]$, for $k = 0, \dots, q$, depending on the chosen model. We start with Model 2G, for which these (conditional) expectations take an elementary form.

⋆ **Expression of $\mathbb{E}[e_{t,p_t}^k \,|\, \mathcal{F}_{t-1}]$ for Model 2G.** The noise terms $e_{t,p_t} = \mathrm{e}_t$ are scalar centered Gaussian i.i.d random variables of variance σ^2 independent on p_t; therefore for any integer $k \geqslant 0$, by using the expression of the moments of a centered normal distribution of variance σ^2, we have

$$\mathbb{E}\big[e_{t,p_t}^{2k+1} \,\big|\, \mathcal{F}_{t-1}\big] = 0 \quad \text{and} \quad \mathbb{E}\big[e_{t,p_t}^{2k} \,\big|\, \mathcal{F}_{t-1}\big] = \frac{(2k)!}{2^k k!} \sigma^{2k} \,.$$

In Model 1G, the noise term depends on both vectors p_t and ε_t, so the calculations of $\mathbb{E}\big[e_{t,p_t}^k \,\big|\, \mathcal{F}_{t-1}\big]$ are less easy and the calculations are provided in Appendix. The key is the use of Isserlis' theorem (see Isserlis [1918] for further details), a formula that allows the computation of higher-order moments of the multivariate normal distribution in terms of its covariance matrix.

Therefore, for both models, we can compute $\mathbb{E}[e_{t,p_t}^k \,|\, \mathcal{F}_{t-1}]$ and the calculations are true for any $\mathcal{F}_{t-1}$-measurable vector $p \in \mathcal{P}$ and all the coefficients $\kappa_m(p)$ can be computed. We emphasize that in the case of Model 2G, $\kappa_m(p)$ does not depend on p – we will write κ_m – and we have the closed form

$$\kappa_m = \kappa_m(p) = \sum_{k=0}^{\lfloor (q-m)/2 \rfloor} a_{m+2k} \binom{m+2k}{m} \frac{(2k)!}{2^k k!} \sigma^{2k}.$$

For Model 1G, $\kappa_m(p)$ does depend on p. We do not provide any closed form here, but we highlight that $\kappa_m(p)$ is a sum of some products of the coefficients $\sigma_{ii'}$ of matrix Σ and of the price levels p_i and $p_{i'}$, for $(i, i') \in \{1, \dots, K\}$, and it can be computed in practice. Moreover, since all price levels $p \in \mathcal{P}$ lie in the K-dimensional simplex, for any $j \in \{1, \dots, K\}$, the component p_j is bounded by 1 and using the expression of Equation (5.14), we get that for $p \in \mathcal{P}$, the coefficient $\kappa_m(p)$ is bounded, by a constant which depends on $\max_{i,i'} \sigma_{i,i'}$. Therefore, we introduce, for Model 1G,

$$\bar{\kappa}_m \triangleq \max_{p \in \mathcal{P}} |\kappa_m(p)| \,.$$

For Model 2G, we set $\bar{\kappa}_m = |\kappa_m|$. As we made no assumption on the non-negativity of the coefficients a_n, for $n = 0, \ldots, q$, the coefficients $\kappa_m(p)$ may be negative.

Remark 17. *Considering the quadratic loss leads to $q = 2$ with $a_2 = 1$ and $a_0 = a_1 = 0$. In that case, we get that*

$$\kappa_0(p) = \mathbb{E}\big[e_{t,p_t}^2 \,\big|\, \mathcal{F}_{t-1}\big] = \begin{cases} p_t^{\mathrm{T}} \Sigma p_t & \text{for Model 1G} \\ \sigma^2 & \text{for Model 2G} \end{cases}$$

$$\kappa_1(p_t) = 2\mathbb{E}\big[e_{t,p_t} \,\big|\, \mathcal{F}_{t-1}\big] = 0 \quad \text{and} \quad \kappa_2(p_t) = 1\,.$$

The corresponding conditionally expected losses are given by

$$\ell_{t,p_t} = \begin{cases} \big(\varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}} \Sigma p_t & \text{for Model 1G} \\ \big(\varphi(x_t, p)^{\mathrm{T}}\theta - c_t\big)^2 + \sigma^2 & \text{for Model 2G} \end{cases}$$

as defined in Chapter 4 for Models 1 and 2, respectively.

⋆ **Summary of the section.** The above shows that for a polynomial loss, the conditional expectation of the loss associated with price levels tariff $p_t \in \mathcal{P}$ at a round t is of the form

$$\ell_{t,p} \triangleq \sum_{m=0}^{q} \kappa_m(p) \Big(\varphi(x_t, p)^{\mathrm{T}}\theta - c_t\Big)^m, \tag{5.1}$$

where the coefficients $\kappa_m(p)$ can be computed explicitly, when the power consumption follows Models 1G or 2G. This formula is also valid for the chosen vector p_t which are $\mathcal{F}_{t-1}$-measurable. For a time budget $T \geqslant 1$, we can now introduce the regret

$$\overline{R}_T \triangleq \sum_{t=1}^{T} \ell_{t,p_t} - \sum_{t=1}^{T} \min_{p \in \mathcal{P}} \ell_{t,p}\,. \tag{5.2}$$

3 A regret bound with Gaussian noises

The closed form of the conditionally expected loss $\ell_{t,p}$ associated with the choice $p \in \mathcal{P}$ $\ell_{t,p}$ suggests to estimate it by simply replacing θ by its estimator in Equation (5.1). We will show how a good estimation of θ induces, for any $p \in \mathcal{P}$, a good estimation of the expected loss $\ell_{t,p}$, which we denote by $\widehat{\ell}_{t,p}$. Given a confidence level on $\widehat{\ell}_{t,p}$, we will then define an optimistic bandit algorithm very similar to the one defined in Chapter 4. We will also prove some regret bounds using the same arguments as the ones of the proof of Theorem 4 of Chapter 4.

3.1 Good estimation of the losses

For $\lambda > 0$, at any round t, we consider the estimation of the parameter θ defined in Chapter 4:

$$\widehat{\theta}_t = V_t^{-1} \sum_{s=1}^{t} Y_{s,p_s} \varphi(x_s, p_s) \quad \text{where} \quad V_t = \lambda \mathrm{I}_d + \sum_{s=1}^{t} \varphi(x_s, p_s)\varphi(x_s, p_s)^{\mathrm{T}}\,.$$

Lemma 4 (of the same chapter) ensures that, for a risk level $\delta \in (0,1)$, no matter how the provider picks the p_t, we have, for all $t \geqslant 1$, with probability at least $1-\delta$,

$$\left\| V_t^{1/2}\big(\widehat{\theta}_t - \theta\big)\right\| \leqslant B_t(\delta) = \sqrt{\lambda d}\, C + \rho\sqrt{2\ln\frac{1}{\delta} + d\ln\left(1+\frac{t}{\lambda}\right)}\,.$$

Here, we also considered the same boundedness assumptions (see Chapter 4) which are all linked to the knowledge that the average consumptions lie in $[0,C]$ and indicate some normalization of the modeling.

Assumption 5 – *Boundedness assumptions.* For any round $t \geqslant 1$ and any $p \in \mathcal{P}$, we assume that

$$\varphi(x_t,p)^{\mathrm{T}}\theta \in [0,C]\,, \quad \text{with} \quad \|\varphi\|_\infty \leqslant 1 \quad \text{and} \quad \|\theta\|_\infty \leqslant C\,.$$

A consequence of these boundedness assumptions is that, for any contextual vector x_t, target c_t and price levels $p \in \mathcal{P}$, the expected difference between the consumption and the target $|\,\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\,|$ is bounded by C. Therefore, for any $m = 0,\dots,n$, with $\bar{\kappa}_m = \max_{p\in\mathcal{P}}|\kappa_m(p)|$ the expected losses $\ell_{t,p}$ can be bounded by

$$\begin{aligned}|\ell_{t,p}| &\leqslant \sum_{m=0}^{q} |\kappa_m(p)\big(\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^m| \leqslant \sum_{m=0}^{q} \bar{\kappa}_m\big|\varphi(x_t,p)^{\mathrm{T}}\theta - c_t\big|^m \\ &\leqslant L \qquad \text{where} \quad L \triangleq \sum_{m=0}^{q} \bar{\kappa}_m C^m\,. \end{aligned} \tag{5.3}$$

For any $p \in \mathcal{P}$, with $[x]_C$ is the clipped part of a real number x (clipping between 0 and C), we consider the estimation of the loss:

$$\widehat{\ell}_{t,p} \triangleq \sum_{m=0}^{q} \kappa_m(p)\Big(\big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t\Big)^m\,. \tag{5.4}$$

The following lemma shows that the quality of this estimation depends on the equality of the estimation of θ.

Lemma 9. *For any round t and any allocation of price levels $p \in \mathcal{P}$, the difference between the expected loss and its estimation satisfies*

$$\begin{aligned}\big|\,\ell_{t,p} - \widehat{\ell}_{t,p}\,\big| &\leqslant M\,\big|\varphi(x_t,p)^{\mathrm{T}}\theta - \varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big| \\ \text{where} \quad M &\triangleq \bar{\kappa}_1 + \sum_{m=2}^{q} \bar{\kappa}_m m C^{m-1} = \sum_{m=1}^{q} \bar{\kappa}_m m C^{m-1}\,.\end{aligned}$$

Proof of Lemma 9. For any $p \in \mathcal{P}$, the distance between the expected power consumption and its target is $d_{t,p} = \varphi(x_t,p)^{\mathrm{T}}\theta - c_t$ and $\widehat{d}_{t,p} = [\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}]_C - c_t$ denotes its estimation. Using the polynomial expressions of $\ell_{t,p}$ and $\widehat{\ell}_{t,p}$, we get that

$$\begin{aligned}\big|\,\ell_{t,p} - \widehat{\ell}_{t,p}\,\big| &= \Big|\sum_{m=0}^{q}\kappa_m(p)d_{t,p}^m - \sum_{m=0}^{q}\kappa_m(p)\widehat{d}_{t,p}^m\Big| = \Big|\sum_{m=1}^{q}\kappa_m(p)\left(d_{t,p}^m - \widehat{d}_{t,p}^m\right)\Big| \\ &\leqslant \sum_{m=1}^{q}|\,\kappa_m(p)\,| \times |d_{t,p}^m - \widehat{d}_{t,p}^m| \leqslant \sum_{m=1}^{q}\bar{\kappa}_m\,|d_{t,p}^m - \widehat{d}_{t,p}^m|\,.\end{aligned}$$

For any $m \geqslant 2$, as $a^m - b^m = (a-b)\sum_{k=0}^{m-1} a^{m-1-k}b^k$, we can factorize $d_{t,p}^m - \widehat{d}_{t,p}^m$ this way:

$$\begin{aligned}\left|\, d_{t,p}^m - \widehat{d}_{t,p}^m \,\right| &= \left|\, d_{t,p} - \widehat{d}_{t,p} \,\right| \times \left| \sum_{k=0}^{m-1} d_{t,p}^{m-1-k} \widehat{d}_{t,p}^k \right| \\ &\leqslant |d_{t,p} - \widehat{d}_{t,p}| \sum_{k=0}^{m-1} C^{m-1} = mC^{m-1} \left|\varphi(x_t,p)^{\mathrm{T}}\theta - \left[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\right]_C\right|\end{aligned}$$

The second inequality comes from the boundedness assumptions 5. Indeed, as c_t, $\varphi(x_t,p)^{\mathrm{T}}\theta$ and $[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}]_C$ lie in $[0,C]$, the differences $d_{t,p}$ and $\widehat{d}_{t,p}$ are bounded by C; and for any $k = 0,\dots,m-1$, we obtain that $\left|d_{t,p}^{m-1-k}\widehat{d}_{t,p}^k\right| \leqslant C^{m-1-k}C^k = C^{m-1}$. Then, it only remains to sum over m to conclude the proof. □

Remark 18. *Considering the quadratic loss, namely $q = 2$, $\bar{\kappa}_1 = 0$ and $\bar{\kappa}_2 = 1$, we get $M = 2C$. This is exactly the inequality obtained in Chapter 4 (see Step 1 of the proof of Proposition 1).*

3.2 Statement of the regret bound

As in Chapter 4, for $t \geqslant 2$ the optimistic algorithm picks

$$p_t \in \operatorname*{argmin}_{p\in\mathcal{P}} \left\{\widehat{\ell}_{t,p} - a_{t,p}\right\}. \tag{5.5}$$

and p_1 arbitrarily, with $a_{t,p}$ the new deviation bounds to be set in the analysis. The following theorem states a regret bound of order $\mathcal{O}\big(\sqrt{T}\ln T\big)$ (in Theorem 4 of Chapter 4, when the covariance matrix Σ is known we obtained a regret bound of the same order). Only the constants need to be updated, they will depend on bounds $\bar{\kappa}_m$ and on some powers of C, up to C^q.

Theorem 7. *Fix a risk level $\delta \in (0,1)$ and a time horizon $T \geqslant 1$. Assume that Assumption 5 holds. The optimistic algorithm* (5.5) *satisfies*

$$\overline{R}_T = \mathcal{O}\left(\sqrt{T}\ln T\sqrt{\ln\frac{1}{\delta}}\right)$$

with probability at least $1 - \delta/2$.

The regret analysis follows the same steps as the one provided in Section 3 of Chapter 4. First, we show that deviation bounds $a_{t,p}$ of the form

$$a_{t,p} = \min\left\{\square, \triangle B_{t-1}(\delta t^{-2})\left\|V_{t-1}^{-1/2}\varphi(x_t,p)\right\|\right\}$$

(by replacing the constants to be set below by the symbols $\square$ and $\triangle$), ensures that, with probability $1-\delta/2$, for all $t = 2,\dots,T$, the instantaneous regrets $r_t = \ell_{t,p_t} - \min_{p\in\mathcal{P}}\ell_{t,p}$ are bounded by $2a_{t,p}$ (see Proposition 2 below). It exploits how well each $\widehat{\theta}_{t-1}$ estimates θ. Then, Lemma 7 of Chapter 4 is used to provide a bound on $\sum_{t=2}^T 2a_{t,p}$ of order $\mathcal{O}(\sqrt{T}\ln T)$.

Proposition 2. *Fix a risk level $\delta \in (0,1)$. Under Assumption 5, by choosing*

$$a_{t,p} = \min\left\{2L,\, MB_{t-1}(\delta t^{-2})\left\|V_{t-1}^{-1/2}\varphi(x_t,p)\right\|\right\},$$

the optimistic algorithm (5.5) *ensures that with probability* $1-\delta/2$,

$$\sum_{t=1}^{T} \ell_{t,p_t} - \sum_{t=1}^{T} \min_{p\in\mathcal{P}} \ell_{t,p} \leqslant 2L + 2\sum_{t=2}^{T} a_{t,p_t}\,.$$

We recall that $L = \sum_{m=0}^{q} \bar{\kappa}_m C^m$ *and* $M = \sum_{m=1}^{q} \bar{\kappa}_m m C^{m-1}$.

Proof of Proposition 2. We show below that for all $t \geqslant 2$, if the estimation $\widehat{\theta}_{t-1}$ of θ is good enough to ensure

$$\big\|V_{t-1}^{1/2}\big(\widehat{\theta}_{t-1}-\theta\big)\big\| \leqslant B_{t-1}(\delta t^{-2})\,, \tag{5.6}$$

then the instantaneous regret is bounded:

$$r_t = \ell_{t,p_t} - \min_{p\in\mathcal{P}} \ell_{t,p} \leqslant 2a_{t,p_t}\,.$$

The conditionally expected losses $\ell_{t,p}$ are bounded by L (see Equation (5.3)). The clipping in the definition of $\widehat{\ell}_{t,p}$ in Equation (5.4) ensures that L also bounds these estimators. Therefore, for any round t and any vector $p \in \mathcal{P}$, $\big|\,\ell_{t,p} - \widehat{\ell}_{t,p}\,\big| \leqslant 2L$. We emphasize that we made no assumption on the non-negativity of the loss function and this is why the factor 2 arises (compared to the deviation bounds introduced for Theorem 4 of Chapter 4). Moreover, by using Lemma 9 and Equation (5.6), we obtain

$$\begin{aligned}\big|\,\ell_{t,p} - \widehat{\ell}_{t,p}\,\big| &\leqslant M\,\big|\varphi(x_t,p)^{\mathrm{T}}\theta - \varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big| \leqslant M\big\|V_{t-1}^{1/2}\varphi(x_t,p)\big\|\big\|V_t^{1/2}\big(\widehat{\theta}_{t-1}-\theta\big)\big\| \\ &\leqslant MB_{t-1}(\delta t^{-2})\big\|V_{t-1}^{1/2}\varphi(x_t,p)\big\|\,.\end{aligned}$$

Therefore, if the event (5.6) holds,

$$\forall p\in\mathcal{P}, \qquad \big|\ell_{t,p} - \widehat{\ell}_{t,p}\,\big| \leqslant a_{t,p}\,. \tag{5.7}$$

With $p_t^\star \in \operatorname{argmin}_{p\in\mathcal{P}}\big\{\ell_{t,p}\big\}$ an optimal convex vector to be used at round t, by definition (5.5) of the optimistic algorithm, it played an allocation p_t that satisfies

$$\widehat{\ell}_{t,p_t} - a_{t,p_t} \leqslant \widehat{\ell}_{t,p_t^\star} - a_{t,p_t^\star}\,, \qquad \text{that is,} \qquad \widehat{\ell}_{t,p_t} - \widehat{\ell}_{t,p_t^\star} \leqslant a_{t,p_t} - a_{t,p_t^\star}\,.$$

Now, for those t for which event (5.6) holds, the property (5.7) also holds and yields, respectively for $p = p_t$ and $p = p_t^\star$:

$$\ell_{t,p_t} - \widehat{\ell}_{t,p_t} \leqslant a_{t,p_t} \qquad \text{and} \qquad \widehat{\ell}_{t,p_t^\star} - \ell_{t,p_t^\star} \leqslant a_{t,p_t^\star}\,.$$

Combining all these three inequalities together, we get

$$\begin{array}{rcccccccl} r_t = \ell_{t,p_t} - \ell_{t,p_t^\star} & = & \big(\ell_{t,p_t} - \widehat{\ell}_{t,p_t}\big) & + & \big(\widehat{\ell}_{t,p_t} - \widehat{\ell}_{t,p_t^\star}\big) & + & \big(\widehat{\ell}_{t,p_t^\star} - \ell_{t,p_t^\star}\big) & \\ & \leqslant & a_{t,p_t} & + & (a_{t,p_t} - a_{t,p_t^\star}) & + & a_{t,p_t^\star} & = 2a_{t,p_t}\,. \end{array}$$

With $r_1 \leqslant 2L$, by summing the instantaneous regrets for $t \geqslant 2$, if the event 5.6 holds for all $t \geqslant 2$, we get the claimed bound.

We do as in Step 3 of the proof of Proposition 1 of Chapter 4 to deal with the time steps $t \geqslant 2$ when the event (5.6) does not hold. Using Lemma 4 of Chapter 4 recalled

in Equation (3.1), each of these events happens with probability at least $1 - \delta t^{-2}$. By a union bound, this happens for some $t \geqslant 1$ with probability at most

$$\delta \sum_{t \geqslant 2} t^{-2} \leqslant \delta \int_2^\infty \frac{1}{t^2} \mathrm{d}t = \delta/2 \,,$$

These special cases thus account for the claimed $1 - \delta/2$ confidence level. □

By replacing L by $2L$ and $2C$ by M in Lemma 7 of Chapter 4, we get that no matter how the environment and provider pick the x_t and p_t,

$$\begin{aligned} \sum_{t=2}^{T} a_{t,p} &= \sum_{t=2}^{T} \min\Big\{2L,\ M\overline{B}\,\|V_{t-1}^{-1/2}\varphi(x_t,p_t)\|\Big\} \\ &\leqslant \sqrt{\big(M\overline{B}\big)^2 + 2L^2}\sqrt{dT\ln\frac{\lambda+T}{\lambda}} = \mathcal{O}\big(\sqrt{T\ln T\ln(T/\delta)}\big)\,, \end{aligned}$$

where $\overline{B} \triangleq B_T(\delta/T^2) = \sqrt{d\lambda}\,C + \rho\sqrt{2\ln(T^2/\delta) + d\ln(1+T/\lambda)}$. This concludes the proof of Theorem 7.

4 Some other possible extensions

This section briefly outlines other possible extensions. First, we explain that a regret bound of order $\sqrt{T}\ln T$ may be obtained for complex loss function beyond polynomial functions for Models 1G and 2G. We also consider the case of unknown variance or covariance matrix and show that it is possible to obtain sub-linear regret bounds through a initial exploration step that estimates these variance terms. The two lasts subsections presents very quickly the cases of non-Gaussian noise and loss functions that vary over iterations. Finally, we emphasize that we could also consider a daily profile Gaussian model to take into account rebound and side effects in the power consumption profiles, exactly as the daily profile Model 3 associated with Model 1 (see Section 6 of Chapter 4).

4.1 Non-polynomial loss functions

This section explains how to generalize the previous regret analysis to any loss function $\ell : (y,c) \mapsto \ell(y,c)$ which satisfies Assumption 6 below. This assumption held for polynomial loss functions $\ell(y,c) = P(y-c)$ and it is actually sufficient that it is true for the considered loss function (polynomial or not) to control the regret (under boundedness assumptions, see Assumption 5). A basic approach could consider an approximation of the loss function with polynomials (we will show how to do it in the last paragraph of the section) but this would lead to worst constants in the regret bound than in the one obtained below.

Assumption 6 – *Properties of the loss function.* For the loss function ℓ, there exists a function $f : [-C,C] \times \mathcal{P} \to [-L,L]$ such that, for any round t, no matter how the environment picks vectors x_t and p_t and the target c_t, the conditional expectation of the loss $\ell\big(Y_{t,p_t}, c_t\big)$ can be rewritten

$$\mathbb{E}\Big[\ell\big(Y_{t,p_t}, c_t\big)|\mathcal{F}_{t-1}\Big] = f(d_{t,p_t}, p_t), \quad \text{with} \quad d_{t,p_t} = \varphi(x_t,p_t)^{\mathrm{T}}\theta - c_t\,.$$

Moreover, for all $p \in \mathcal{P}$, the functions $f(\,\cdot\,, p)$ are M-Lipschitz.

We point out that the function f may depend on the covariance matrix Σ for Model 1G or on the variance σ^2 for Model 2G.

Remark 19. *For polynomial loss functions, for any $d \in [-C, C]$ and any $p \in \mathcal{P}$, we had*

$$f(d,p) = \sum_{k=0}^{q} \kappa_m(p) d^m \,.$$

Under Assumption 6, exactly as for polynomial losses, we introduce the estimators of the conditional expected losses

$$\widehat{\ell}_{t,p} = f(\widehat{d}_{t,p}, p) \quad \text{with} \quad \widehat{d}_{t,p} = \big[\varphi(x_t,p)^{\mathrm{T}} \widehat{\theta}_{t-1}\big]_C - c_t \,.$$

The M-Lipschitz property of $f(\,\cdot\,, p)$ ensures that a good estimation of θ induces a good estimation of the expected losses. Indeed, under Assumption 6, for any round t and any $p \in \mathcal{P}$, we have

$$|\ell_{t,p} - \widehat{\ell}_{t,p}| = |f\big(d_{t,p},\, p\big) - f\big(\widehat{d}_{t,p},\, p\big)| \leqslant M |d_{t,p} - \widehat{d}_{t,p}| = M \big|\varphi(x_t,p)^{\mathrm{T}}\theta - \big[\varphi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C\big| \,.$$

This inequality corresponds to Lemma 9 in the case of polynomial losses. By considering the optimistic algorithm

$$p_t \in \operatorname*{argmin}_{p\in\mathcal{P}} \big\{\widehat{\ell}_{t,p} - a_{t,p}\big\} \quad \text{with} \quad a_{t,p} = \min\Big\{2L,\, MB_{t-1}(\delta t^{-2}) \,\big\|V_{t-1}^{-1/2}\varphi(x_t,p)\big\|\Big\} \,, \tag{5.8}$$

we can then redo exactly the same analysis of regret as in the previous section. Proposition 2 and Theorem 7 are still true and we obtain a regret bound of the order of $\mathcal{O}\big(\sqrt{T}\ln T\big)$.

In the next paragraphs, we focus on Model 2G and give some examples of loss functions that may or may not satisfy Assumption 6. Finally, in the last paragraph of this section we explain how we can still control the regret when we cannot find any closed form of the expected losses, namely when the function f is unknown.

⋆ **General loss function for Model 2G.** In what follows, we consider Model 2G and we assume that there exists a function $g : \mathbb{R} \to \mathbb{R}$, such that

$$\forall\, (y,c) \in \mathbb{R} \times [0,C] \,, \quad \ell(y,c) = g(y-c) \,.$$

Then, for any round t, as p_t and c_t are $\mathcal{F}_{t-1}$-measurable, with $d_{t,p_t} = Y_{t,p_t} - c_t$, the conditionally expected loss satisfies

$$\begin{aligned} \ell_{t,p_t} \quad &= \quad \mathbb{E}\Big[g(Y_{t,p_t} - c_t)\big|\mathcal{F}_{t-1}\Big] \\ &\overset{\text{(Model 2G)}}{=} \mathbb{E}\big[g(X)\big] \quad \text{with} \quad X \sim \mathcal{N}(d_{t,p_t}, \sigma^2) \,. \end{aligned}$$

Thus, as soon as the function $x \mapsto g(x)\exp\big(-(x-d_{t,p_t})^2/2\sigma^2\big)$ can be integrated over $\mathbb{R}$, the conditionally expected loss is

$$\ell_{t,p_t} = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} g(x) \exp\Big(-\frac{(x-d_{t,p_t})^2}{2\sigma^2}\Big) \mathrm{d}x \,.$$

For Model 2G, the noise term does not depend of the allocation of price levels picked so the functions $f(\cdot\,, p)$ are all equal, for the ease of notation, we thus write $f(d_{t,p_t})$ instead of $f(d_{t,p_t}, p_t)$ (this was also the case for polynomial losses, the coefficients $\kappa_m(p)$ did not depend on tariffs chosen). Therefore, if the loss is a function of the distance between the power consumption and its target, namely if the function g exists, and if for any $d \in [-C, C]$, the function $x \mapsto g(x) \exp\left(-(x-d)^2/2\sigma^2\right)$ can be integrated over $\mathbb{R}$ (so if g does not grow too fast), the function f defined in Assumption 6 is

$$\begin{array}{rcl} f : [-C, C] & \to & \mathbb{R} \\ d & \mapsto & \frac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{\mathbb{R}} g(x) \exp\Big(-\frac{(x-d)^2}{2\sigma^2}\Big) \mathrm{d}x\,. \end{array} \tag{5.9}$$

The counter-example below presents a loss function for which the function f is not defined.

⋆ **A loss function which does not satisfy Assumption 6 for Model 2G.** If we consider the exponential loss

$$\ell : (y, c) \mapsto \exp\Big(\frac{(y-c)^2}{2\sigma^2}\Big)\,,$$

we get $g(x) = \exp(x^2/2\sigma^2)$. Then, for any round t, to compute the conditionally expected loss ℓ_{t,p_t}, the function $x \mapsto \exp\big((2x - d_{t,p_t})d_{t,p_t}\big)$ has to be integrated over $\mathbb{R}$, which is not possible. Therefore Assumption 6 does not hold.

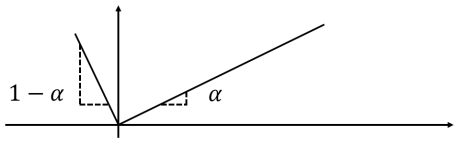
We now detail the pinball loss case, a function which will be used in the practical application of Section 5 to model the real costs suffered by the electricity providers when the balance between consumption and production is not guaranteed.

⋆ **The case of the pinball loss for Model 2G.** Each of the electricity providers is responsible of the balance between its electricity production and the power consumption of its customers. To maintain the global balance, electricity network managers impose some financial penalties which are proportional to the absolute difference between the announced production c_t and the power consumption Y_{t,p_t} of the electricity provider customers. The multiplying coefficient is different depending on whether the electricity provider is in a situation of over- or under-consumption, compared to its production (this practical application is fully detailed in Section 5). We highlight that pinball losses are generally used to obtain quantile forecasts (see Steinwart et al., 2011) and we can also see it from this point of view: if the electricity provider does not want to exceed a target too much (no more than 95% of the time, for example), we may consider the pinloss associated with the corresponding quantile (the quantile at level $\alpha = 0.95$). Here, we introduce $\alpha \in [0, 1]$ and consider a loss equal to $\alpha|Y_{t,p_t} - c_t|$ in the case of an over-consumption and to $(1-\alpha)|Y_{t,p_t} - c_t|$ for an over-production. Therefore, at each round t, the loss suffered is

$$\ell(Y_{t,p_t}, c_t) = (1-\alpha)|Y_{t,p_t} - c_t|\mathbb{1}_{\{Y_{t,p_t} \leqslant c_t\}} + \alpha|Y_{t,p_t} - c_t|\mathbb{1}_{\{Y_{t,p_t} \geqslant c_t\}}\,.$$

With g the function such that $\ell(y, c) = g(y - c)$ defined by

$$g(x) = \begin{cases} (1-\alpha)|x| & \text{if} \quad x \leqslant 0 \\ \alpha\,|x| & \text{if} \quad x \geqslant 0\,, \end{cases}$$



it is possible to compute, for any $d \in [-C, C]$, the function f defined in Equation (5.9), which is equal to

$$\begin{aligned} f(d) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \big((1-\alpha)|x|\mathbb{1}_{\{x\leqslant 0\}} + \alpha|x|\mathbb{1}_{\{x\geqslant 0\}}\big) \exp\Big(-(x-d)^2/2\sigma^2\Big)\mathrm{d}x \\ &= (1-\alpha)\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} -x \exp\Big(-(x-d)^2/2\sigma^2\Big)\mathrm{d}x + \alpha\frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{+\infty} x \exp\Big(-(x-d)^2/2\sigma^2\Big)\mathrm{d}x\,. \end{aligned}$$

With the variable change $t = (x-d)/\sigma$ in the integrals and by denoting the normal cumulative distribution function by $\Phi(x) = \int_{-\infty}^{x} \exp\big(-t^2/2\big)\mathrm{d}t$, we obtain

$$\begin{aligned} f(d) &= (1-\alpha)\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-d/\sigma} -(d+t\sigma)\exp\big(-t^2/2\big)\mathrm{d}t + \alpha\frac{1}{\sqrt{2\pi}} \int_{-d/\sigma}^{+\infty} (d+t\sigma)\exp\big(-t^2/2\big)\mathrm{d}t \\ &= d\Big(-(1-\alpha)\Phi(-d/\sigma) + \alpha\big(1-\Phi(-d/\sigma)\big)\Big) \\ &\qquad + (1-\alpha)\frac{\sigma}{\sqrt{2\pi}}\Big[\exp(-t^2/2)\Big]_{-\infty}^{-d/\sigma} - \alpha\frac{\sigma}{\sqrt{2\pi}}\Big[\exp(-t^2/2)\Big]_{-d/\sigma}^{+\infty} \\ &= d\big(\alpha - \Phi(-d/\sigma)\big) + \frac{\sigma}{\sqrt{\pi}}\exp\big(-d^2/2\sigma\big)\,. \end{aligned}$$

For any real number x, $\Phi(x) \in [0,1]$ and $\exp(-x^2) \leqslant 1$, so we get that for any $d \in [-C, C]$, $|f(d)|$ is bounded by $C + \sigma\sqrt{2/\pi}$. Moreover, f is derivable and as $\Phi'(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$, we get

$$\begin{aligned} f'(d) &= \alpha - \Phi\big(-d/\sigma\big) + \frac{d}{\sigma}\Phi'\big(-d/\sigma\big) - \frac{d}{\sigma\sqrt{\pi}}\exp\Big(-d^2/2\sigma\Big) \\ &= \alpha - \Phi\big(-d/\sigma\big) + \frac{d}{\sigma\sqrt{\pi}}\exp\big(-d^2/2\sigma\big) - \frac{d}{\sigma\sqrt{\pi}}\exp\big(-d^2/2\sigma\big) \\ &= \alpha - \Phi\big(-d/\sigma\big) \in [\alpha - 1, \alpha]\,. \end{aligned}$$

With $0 \leqslant \alpha \leqslant 1$, the derivative of f is bounded by 1, so f is 1-Lipschitz. Therefore, for any round t and any vector $p_t \in \mathcal{P}$, we obtain

$$\mathbb{E}\Big[|Y_{t,p_t} - c_t| \,\Big|\, \mathcal{F}_{t-1}\Big] = f(d_{t,p_t}), \quad \text{with} \quad d_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\,,$$

where the function f is bounded by $C + \sigma\sqrt{2/\pi}$ on $[-C, C]$ and 1-Lipschitz. So, if we consider Model 2G and the pinball loss, Assumption 6 holds and we can control the regret associated with the optimistic algorithm (5.8), by taking $L = C + \sigma\sqrt{2/\pi}$ and $M = 1$ in the definition of the deviation bounds $a_{t,p}$.

Finally, if Assumption 6 does not hold, we show how it is still possible to obtain some regret bound, by using a polynomial interpolation of the loss function.

⋆ **Polynomial interpolation of the loss function.** We now consider a non-polynomial loss function ℓ^{NP}. Let us denote by ℓ a polynomial interpolation of this function (e.g. computed with Lagrange polynomials). If the approximation ℓ satisfies, for any $d \in [-C, C]$,

$$\ell^{\mathrm{NP}}(d) = \ell(d) + \gamma_d, \quad \text{with} \quad |\gamma_d| \leqslant \gamma\,,$$

where γ_d denotes the approximation error; we can still control the regret by a bound of order $\mathcal{O}\big(2T\gamma + \sqrt{T}\ln T\big)$. Indeed, for any round t, we have

$$\mathbb{E}\Big[\,\ell^{\mathrm{NP}}\big(Y_{t,p_t} - c_t\big)\,|\,\mathcal{F}_{t-1}\Big] = \mathbb{E}\Big[\ell\big(Y_{t,p_t} - c_t\big)\,|\,\mathcal{F}_{t-1}\Big] + \mathbb{E}\big[\gamma_{d_{t,p_t}}\,|\,\mathcal{F}_{t-1}\big]\,,$$

with $\gamma_{d_{t,p_t}}$ a random approximation error. Under the assumption on the boudedness of this error, $\gamma_{d_{t,p_t}}$ is bounded by γ, and so is its expectation. Then, by denoting by $\overline{R}_T(\ell^{\mathrm{NP}})$ the regret associated with the loss function ℓ^{NP} and by $\overline{R}_T(\ell)$ the one associated with the polynomial approximation of ℓ^{NP}, we obtain

$$\begin{aligned}
\overline{R}_T(\ell^{\mathrm{NP}}) &= \sum_{t=1}^{T}\mathbb{E}\Big[\ell^{\mathrm{NP}}\big(Y_{t,p_t} - c_t\big)\,|\,\mathcal{F}_{t-1}\Big] - \sum_{t=1}^{T}\min_{p\in\mathcal{P}}\mathbb{E}\Big[\ell^{\mathrm{NP}}\big(Y_{t,p} - c_t\big)|\mathcal{F}_{t-1}\Big] \\
&\leqslant \sum_{t=1}^{T}\mathbb{E}\Big[\ell\big(Y_{t,p_t} - c_t\big)|\mathcal{F}_{t-1}\Big] + \gamma - \sum_{t=1}^{T}\min_{p\in\mathcal{P}}\mathbb{E}\Big[\ell\big(Y_{t,p} - c_t\big)|\mathcal{F}_{t-1}\Big] + \gamma \\
&\leqslant 2\gamma T + \overline{R}_T(\ell)\,.
\end{aligned}$$

Since we can control $\overline{R}_T(\ell)$, we can also control $\overline{R}_T(\ell^{\mathrm{NP}})$. Moreover, if the approximation is good enough to ensure an approximation error of order $\mathcal{O}(T^{-\alpha})$, we can still obtain a sub-linear regret bound. We point out that to obtain such an approximation, especially if the loss function has irregularities, the degree q of the interpolation polynomial may be high. This leads to very large constants, proportional to C^q, within the regret bound.

4.2 Unknown variance or covariance matrix

If the covariance matrix Σ or the variance σ^2 is unknown, it is still possible to obtain some regret bound. As in Section 3 of Chapter 4, we can use the first τ rounds of the algorithm to estimate these quantities. By keeping the notations of the above section, we consider the function f defined in Assumption 6. This function may depend on the unknown variance σ^2 (for Model G2) or covariance matrix Σ (for Model G1), which are, from now on, unknown. By using the estimation of Σ or σ^2 computed at round $\tau+1$, we consider an estimator of the function f obtained by replacing Σ or σ^2 by its estimator in the definition of f. Let us denote it by $\widehat{f}_\tau$. Assumption 7 below states some guarantee on the estimator f_τ.

Assumption 7 – *Estimation of the variance or covariance matrix.* For any round t, no matter how the environment picks the vectors x_t and p_t and the target c_t, the conditional expectation of the losses may be written with a function $f : [C, C] \times \mathcal{P} \to [-L, L]$ such that

$$\mathbb{E}\big[\ell(Y_{t,p_t}, c_t)\,|\,\mathcal{F}_{t-1}\big] = f(d_{t,p_t}, p_t) \quad \text{with} \quad d_{t,p_t} = \varphi(x_t, p_t)^{\mathrm{T}}\theta - c_t\,.$$

For $\tau \geqslant 1$ initial explorations rounds, with $\widehat{f}_\tau$ the estimator of f, computed at round $\tau+1$, there exists a real number $\alpha > 0$, such that with probability at least $1 - \delta/2$, the estimator satisfies for any $d \in [-C, C]$,

$$\big|\widehat{f}_\tau(d,p) - f(d,p)\big| \leqslant \nu, \quad \text{with} \quad \nu = \mathcal{O}(\tau^{-\alpha})\,. \tag{5.10}$$

This assumption actually held in Chapter 4, see Example 5 below for further details.

From now on, for any round t and any allocation of price levels $p \in \mathcal{P}$, we will consider this new estimation of the expected loss

$$\widetilde{\ell}_{t,p} = \widehat{f}\big(\widehat{d}_{t,p}, p\big)\,.$$

In this new definition, estimators of the distance $d_{t,p}$ but also of the function f, in which the estimation of the variance or covariance matrix is involved, come into play. If the event (5.10) holds, for any round $t \geqslant \tau + 1$ and any vector $p \in \mathcal{P}$, the following deviation bound on the expected loss estimation holds:

$$\begin{aligned}
|\ell_{t,p} - \widetilde{\ell}_{t,p}| &= \big|f(d_{t,p}) - \widehat{f}(\widehat{d}_{t,p}, p)\big| = \big|f(d_{t,p}) - f(\widehat{d}_{t,p}, p) + f(\widehat{d}_{t,p}, p) - \widehat{f}(\widehat{d}_{t,p}, p)\big| \\
&\leqslant \big|f(d_{t,p}) - f(\widehat{d}_{t,p}, p)\big| + \big|f(\widehat{d}_{t,p}, p) - \widehat{f}(\widehat{d}_{t,p}, p)\big| \\
&\leqslant M\big|\varphi(x_t, p)^{\mathrm{T}}\theta - \big[\varphi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C\big| + \nu\,.
\end{aligned}$$

Then, we consider the new optimistic algorithm which picks, for $t \geqslant \tau + 1$, the price levels

$$p_t \in \operatorname*{argmin}_{p \in \mathcal{P}} \Big\{\widetilde{\ell}_{t,p} - \alpha_{t,p}\Big\}, \quad \text{with} \quad \alpha_{t,p} = a_{t,p} + \nu\,. \tag{5.11}$$

The first τ vectors $p_1, \ldots, p_\tau$ are chosen to provide the estimator of Σ or σ^2.

Theorem 8. *Fix a risk level $\delta \in (0,1)$ and a time horizon $T \geqslant 1$. Assume that Assumption 7 holds. With probability at least $1 - \delta$, the regret associated with the optimistic algorithm (5.11) with an initial exploration of length $\tau = \mathcal{O}\big(T^{1/(1+\alpha)}\big)$ is controlled by a sub-linear regret bound of order*

$$\mathcal{O}\Big(\max\big(T^{1/(1+\alpha)}, \sqrt{T}\big)\Big)\,,$$

up to poly-logarithmic terms.

Proof of Theorem 8. We can redo the same analysis on the regret by replacing $\widehat{\ell}_{t,p}$ and $a_{t,p}$ by $\widetilde{\ell}_{t,p}$ and $\alpha_{t,p}$, respectively. Indeed, for any $t \geqslant \tau + 1$, if both events (5.6) and (5.10) hold (that is if $\widehat{\theta}_{t-1}$ and $\widehat{f}$ estimate correctly θ and f, respectively), we get, for any $p \in \mathcal{P}$,

$$\forall p \in \mathcal{P}, \quad |\,\ell_{t,p} - \widetilde{\ell}_{t,p}\,| \leqslant \alpha_{t,p}\,.$$

Next, by choosing p_t according to the optimistic algorithm (5.11), we obtain

$$\ell_{t,p_t} - \min_{p \in \mathcal{P}} \ell_{t,p} \leqslant 2\alpha_{t,p_t}\,.$$

Therefore, under event (5.10) and if for all $t \geqslant \tau + 1$, events (5.6) also hold, by bounding the first τ instantaneous regrets by $2L$ (expected losses are all bounded by L), we obtain

$$\begin{aligned}
\overline{R}_T = \sum_{t=1}^{T} \ell_{t,p_t} - \min_{p \in \mathcal{P}} \ell_{t,p} &\leqslant 2L\tau + \sum_{t=\tau+1}^{T} \alpha_{t,p_t} \\
&\leqslant 2L\tau + T\nu + \sum_{t=\tau+1}^{T} a_{t,p_t}\,.
\end{aligned}$$

We already showed that the sum $\sum_{t=\tau+1}^{T} a_{t,p_t}$ is of order $\mathcal{O}\big(\sqrt{T}\ln T\big)$. By taking τ of order $T^{1/(1+\alpha)}$ we obtain a bound of order $\mathcal{O}\big(\max(T^{1/(1+\alpha)}, \sqrt{T}\big)$, up to poly-logarithmic terms.

It only remains to deal with the probability that at least one of the events (5.6) does not hold (which is bounded by $\delta/2$, see the end of the proof of Proposition 2) or that the event

$$\exists\, d \in [-C, C] \text{ and } p \in \mathcal{P} \quad \Big| \quad \big|\widehat{f}_\tau(d,p) - f(d,p)\big| \leqslant \nu\,,$$

does not hold (which is bounded by $\delta/2$ by Assumption 7). Finally, at least on the previous events does not hold with probability at most δ. □

Example 5: *The case of the quadratic loss function.* We highlight that we can recover the regret bound we had in Section 3 of Chapter 4. Indeed, with the quadratic loss, we get that for all $p \in \mathcal{P}$,

$$f(d,p) = d^2 + p^{\mathrm{T}}\Sigma p\,.$$

By using the estimation $\widehat{\Sigma}_\tau$ of the covariance matrix provided in Section 3.3.2 of Chapter 4, after τ first rounds, we define the following estimator of f for any $d \in [-C, C]$ and any $p \in \mathcal{P}$,

$$\widehat{f}(d,p) = d^2 + p^{\mathrm{T}}\widehat{\Sigma}_\tau p\,.$$

In Lemma 5 page 116, we proved that with probability at least $1 - \delta/2$, this estimator ensures

$$\big|\widehat{f}(d,p) - f(d,p)\big| = |p^{\mathrm{T}}\big(\widehat{\Sigma}_\tau - \Sigma\big)p| \leqslant \nu_\tau(\delta/2), \quad \text{with} \quad \nu_\tau = \mathcal{O}(\tau^{-1/2})\,.$$

Therefore, Assumption 7 holds with $\alpha = 1/2$ and we obtain a regret bound of order

$$\mathcal{O}\big(\max\big(\sqrt{T}, T^{\frac{1}{1+1/2}}\big)\big) = \mathcal{O}\big(T^{2/3}\big)\,,$$

as claimed in Chapter 4.

4.3 Non-Gaussian noises

We point out that, for the polynomial loss functions, we used the Gaussian assumption on the noises only to compute the moments $\mathbb{E}\big[e_t^k|\mathcal{F}_{t-1}\big]$ (see Section 2.3). As soon as the law of the noises is such that we can compute the functions $f(d,p)$ for $d \in [-C, C]$ and $p \in \mathcal{P}$, we can control the regret.

4.4 Time-dependent loss functions

Finally, in all this chapter, we consider, at each round t, the loss function ℓ. But we could very well imagine a loss function that changes from moment to moment. We will see in the applications of Section 5 that these functions can depend in particular on the electricity market, which varies over time. Therefore we could introduce a time dependence and consider the loss functions ℓ_t. Such changes have no impact on the regret bound as long as we can find some constants M (which is the Lipschitz-constant of functions $f(\cdot\,, p)$, with $p \in \mathcal{P}$) and L (which bounds the expected losses) that work for all the losses considered between $t = 1$ and $t = T$.

5 A practical application: cost of an over or under production of electricity

This section briefly presents how the grid manager RTE (*Réseau de Transport d'Électricité*) maintains the balance between the load of France and the amount of electricity generated by the French providers. We will not go into detail on the mechanisms of the electricity market and we will focus on positive spot electricity prices (as electricity cannot be stored, when a too large amount has been generated, spot price of the electricity may become negative).

Electricity providers like EDF inject the electricity from power plants into one end of the grid, while their customers consume it at the other end, at home or at work. Each provider has to manage its own balance. To maintain the system security and the electricity quality, the grid manager is in charge of the global balance. In France, RTE manages both the transport of the electricity (through high-voltage lines) and the balance of the French power system. Every half-hour, it penalizes the electricity providers which imbalance the grid. The amounts of these penalties are calculated to incite the electricity providers to correct their imbalances and reflect the price of the actions operated by the grid manager to re-balance the power system (see the RTE website[1]). If the global amount of electricity generated by the providers is greater than the French consumption, the trend in the power system is said to be "upward". Then, RTE may, for example, ask the owners of hydraulic dams to pump the water and refill dams in order to consume the surplus of electricity or sell it to grid managers of border countries. If there is not enough electricity, the trend is "downward". Consequently, RTE has to buy electricity from border countries or on the capacity market (namely, from producers of powers plants that can be turned on very quickly) and inject it on the grid; it may also pay big companies to stop consuming.

From now on, for a half-hour t, we denote by P_t^{UP} and P_t^{DOWN} the cost, per kWh, paid by RTE to re-balance the power system, depending on whether the trend is upward or downward, respectively. We consider an electricity provider and we denote by c_t the amount of electricity it generates at the half-hour t. We assume that c_t is known at the half-hour $t-1$, which makes sense since the production is generally scheduled. In the case of intermittent energies, like solar or wind power, we could replace c_t by power generation forecasts. This amount of electricity defines the consumption target. At $t-1$, the electricity provider has also chosen the price allocations p_t for its customers, then its observes their consumption Y_{t,p_t}. If Y_{t,p_t} is higher than c_t, RTE bills the consumed electricity overage, while if c_t is higher than Y_{t,p_t}, RTE buys the produced electricity overage. The costs associated depends on the trend of the power system and are detailed in Table 5.1. The coefficient k is a penalty (around 5%) imposed by RTE. In each case, the electricity provider has no interest in being imbalanced. Indeed, in case of a positive imbalance, that is $Y_{t,p_t} < c_t$, RTE buys its excess of electricity at very low tariffs; the provider has thus missed an opportunity to sell it to its own customers (or to other providers), at a much higher price. On the contrary, for a negative imbalance, that is $Y_{t,p_t} > c_t$, RTE bills the excess of electricity consumed by the provider's customers at much higher prices than the market. If

[1] `https://www.services-rte.com/en/learn-more-about-our-services/becoming-a-balance-responsible-party.html`

the electricity provider is in the trend of the power system, namely if it contributes to the global imbalance, the losses suffered are even higher.

Electricity provider	Power system imbalance	trend	
	$(c_t > Y_{t,p_t})$	$+ P_t^{\text{DOWN}}(1-k) \times \lvert Y_{t,p_t} - c_t \rvert$	$+ P_t^{\text{UP}}(1-k) \times \lvert Y_{t,p_t} - c_t \rvert$
	$(c_t < Y_{t,p_t})$	$- P_t^{\text{DOWN}}(1+k) \times \lvert Y_{t,p_t} - c_t \rvert$	$- P_t^{\text{UP}}(1+k) \times \lvert Y_{t,p_t} - c_t \rvert$

Table 5.1 – Financial regularizations imposed, at an half-hour t, by electricity network manager to the electricity provider, depending on the trend of the power system (down on the left, up on the right) and on the imbalance of the electricity provider (positive at the top, negative at the bottom), which produces an amount c_t of electricity and whom customers consume Y_{t,p_t}.

Therefore, in all the imbalanced situations, the electricity provider suffers a financial loss. This loss is proportional to $|Y_{t,p_t} - c_t|$; the difference between the amount of electricity it generates on the grid c_t and the amount of electricity its customers consume Y_{t,p_t}. The proportional coefficients P_t^{UP} and P_t^{DOWN} depend on the cost suffered by the grid manager to re-balance the power system; and it is higher when the electricity provider accentuates the imbalance (when both balances are on the same side), than when the provider "helps" the grid manager (when both balances are on opposite sides). We denote by τ these coefficients and super-script them with $+$ or $-$ depending on whether the provider balance is positive, namely for an over-production $c_t > Y_{t,p_t}$, or negative, namely for an over-consumption $c_t < Y_{t,p_t}$; and by UP or DOWN depending on whether the trend is upward or downward, respectively. With these notations, at each half-hour t, the electricity provider suffers a financial loss:

$$\ell_t(Y_{t,p_t}, c_t) = \left(\tau_t^{\text{UP}^-} \lvert Y_{t,p_t} - c_t \rvert \mathbb{1}_{\{c_t < Y_{t,p_t}\}} + \tau_t^{\text{UP}^+} \lvert Y_{t,p_t} - c_t \rvert \mathbb{1}_{\{c_t > Y_{t,p_t}\}}\right) \mathbb{1}_{\{\text{TREND} = \text{UP}\}} + \left(\tau_t^{\text{DOWN}^-} \lvert Y_{t,p_t} - c_t \rvert \mathbb{1}_{\{c_t < Y_{t,p_t}\}} + \tau_t^{\text{DOWN}^+} \lvert Y_{t,p_t} - c_t \rvert \mathbb{1}_{\{c_t > Y_{t,p_t}\}}\right) \mathbb{1}_{\{\text{TREND} = \text{DOWN}\}} . \quad (5.12)$$

We note that if the provider could have sold the electricity at the prices P_t^{UP} or P_t^{DOWN} (depending on the trend), the coefficients $\tau_t^{\text{UP}^-}$ and $\tau_t^{\text{UP}^+}$ are equal to kP_t^{UP} and $\tau_t^{\text{DOWN}^-}$ and $\tau_t^{\text{DOWN}^+}$ are equal to kP_t^{DOWN}. When the trend and the prices of re-balancing are known in advance (namely at the half-hour $t-1$), the loss suffered at t is made of two pieces of positive linear functions and is null when the consumption equals the target. Thus, it may be written as:

$$\ell_t(Y_{t,p_t}, c_t) = \tau_t^+ \lvert Y_{t,p_t} - c_t \rvert \mathbb{1}_{\{c_t > Y_{t,p_t}\}} + \tau_t^- \lvert Y_{t,p_t} - c_t \rvert \mathbb{1}_{\{c_t < Y_{t,p_t}\}} .$$

By considering Model 2G, for any $p \in \mathcal{P}$, with the calculations presented in Section 4 for the case of the pin-ball loss, it is possible to compute the expected loss:

$$\ell_{t,p} = \mathbb{E}\big[\ell_t(Y_{t,p}, c_t) | \mathcal{F}_{t-1}\big] ,$$

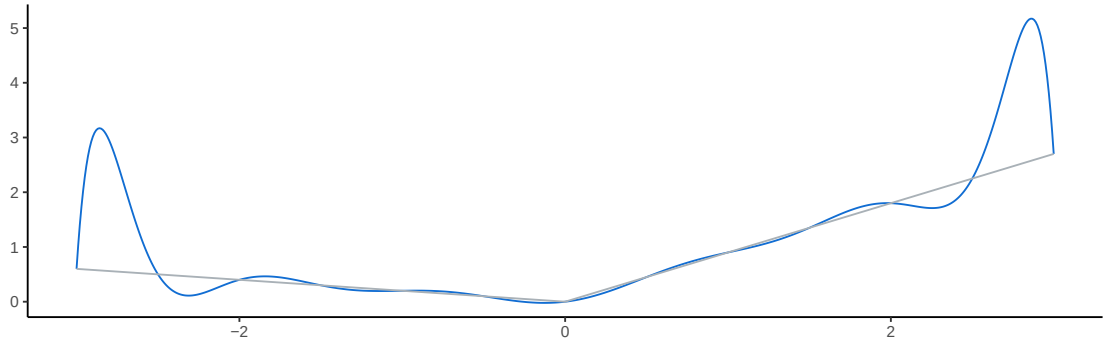


Figure 5.1 – Illustration of an approximation of a piece-wise linear loss function with a 12-degree Lagrange polynomial on $[-2, 2]$. The approximation is close to the loss function between -2 and 2. Outside this interval, the points chosen for the Lagrange interpolation are too far apart to provide a fair approximation. The increase in the number of points leads to an increase in the polynomial degree.

and consequently to apply Theorem 7. For Model 2, we do not know any closed form of the expected loss but we can still consider a polynomial approximation and apply the results of Section 3. Figure 5.1 gives an illustration of such an approximation. Therefore, no matter the model we consider, we can apply an optimistic algorithm and obtain regret bounds.

The main drawback of this modeling is that, in reality, the trend of the electrical system is not deterministic. By predicting it, we could obtain probabilistic regret bounds (namely, that are true under the events "the prediction of the trend at t is correct"). We could also weight the loss function with the forecasts of the trend and consider:

$$\begin{aligned}\ell_t(Y_{t,p_t}, c_t) = \Big(&\tau_t^{\text{UP}^-}\big|Y_{t,p_t} - c_t\big|\mathbb{1}_{\{c_t<Y_{t,p_t}\}} + \\ &\tau_t^{\text{UP}^+}\big|Y_{t,p_t} - c_t\big|\mathbb{1}_{\{c_t>Y_{t,p_t}\}}\Big)\mathbb{P}\big(\textsc{Trend} = \textsc{down}\big) + \\ \Big(&\tau_t^{\text{DOWN}^-}\big|Y_{t,p_t} - c_t\big|\mathbb{1}_{\{c_t<Y_{t,p_t}\}} + \\ &\tau_t^{\text{DOWN}^+}\big|Y_{t,p_t} - c_t\big|\mathbb{1}_{\{c_t>Y_{t,p_t}\}}\Big)\Big(1 - \mathbb{P}\big(\textsc{Trend} = \textsc{down}\big)\Big).\end{aligned}$$

We highlight that this loss is not the conditional expectation on the loss defined in Equation (5.12) because the trend and Y_{t,p_t} are not independant. Finally, we also not that the coefficients $\tau_t^{\text{UP}^-}$, $\tau_t^{\text{UP}^+}$, etc. are not known in advance and must also be estimated; which further complicates the problem.

These examples show that it makes sense to look at non-quadratic losses, and that it is possible to link the loss functions considered in the bandit algorithm with the true losses suffered by the electricity provider. Indeed, with the piece-wise linear loss functions introduced above, we modeled the financial penalties imposed by the grid manager. We highlight that our example is just a simplification of the real production-consumption management processes. By considering a modeling of the production costs and by integrating some mechanisms of the electricity market, the loss functions would certainly become more complex.

Appendix

Expression of $\mathbb{E}[e_{t,p_t}^k \,|\, \mathcal{F}_{t-1}]$ for Model 1G

The noise term e_{t,p_t} is $p_t^{\mathrm{T}}\varepsilon_t = \sum_{j=1}^K p_{t,j}\varepsilon_{t,j}$. Applying the multinomial theorem, we get

$$\begin{aligned} e_{t,p_t}^k &= \Big(p_{t,1}\varepsilon_{t,1} + p_{t,2}\varepsilon_{t,2} + \cdots + p_{t,K}\varepsilon_{t,K}\Big)^k \\ &= \sum_{m_1+m_2+\cdots+m_K=k} \binom{k}{m_1, m_2, \ldots, m_K} \prod_{j=1}^K (p_{t,j}\varepsilon_{t,j})^{m_j} \end{aligned}$$

with the multinomial coefficient $\displaystyle\binom{k}{m_1, m_2, \ldots, m_K} = \frac{k!}{m_1! m_2! \ldots m_K!}$.

As the vector p_t is $\mathcal{F}_{t-1}$-measurable and the vector ε_t is independent on $\mathcal{F}_{t-1}$, we obtain

$$\mathbb{E}\big[e_{t,p_t}^k \,\big|\, \mathcal{F}_{t-1}\big] = \sum_{m_1+m_2+\cdots+m_K=k} \binom{k}{m_1, m_2, \ldots, m_K} \Big(\prod_{j=1}^K p_{t,j}^{m_j}\Big)\, \mathbb{E}\Big[\prod_{j=1}^K \varepsilon_{t,j}^{m_j}\Big].$$

Isserlis' theorem states that if $(X_1, X_2, \ldots, X_k)$ is a zero-mean multivariate normal random vector, then

$$\mathbb{E}\big[X_1 X_2 \ldots X_k\big] = \sum_{\gamma \in \Gamma_k} \prod_{\{i,j\}\in\gamma} \mathrm{Cov}\big(X_i, X_{i'}\big), \tag{5.13}$$

where Γ_k denotes all the pairings of $\{1, \ldots, k\}$, namely, all distinct ways of partitioning $\{1, \ldots, k\}$ into pairs $\{i, i'\}$. Note that if k is odd, Γ_k is empty and thus $\mathbb{E}\big[X_1 X_2 \ldots X_k\big] = 0$. We consider the k-dimensional zero-mean multivariate normal random vector

$$\Big(\underbrace{\varepsilon_{t,1}, \ldots, \varepsilon_{t,1}}_{m_1 \text{ times}}, \underbrace{\varepsilon_{t,2}, \ldots., }_{m_2 \text{ times } \ldots} \ldots\ldots, \underbrace{\varepsilon_{t,K}}_{m_K \text{ times}}\Big)$$

and the vector I_m associated with this re-indexation

$$\mathrm{I}_m = \Big(\underbrace{1, \ldots, 1}_{m_1 \text{ times}}, \underbrace{2, \ldots\ldots., }_{m_2 \text{ times } \ldots} \ldots\ldots\ldots, \underbrace{K}_{m_K \text{ times}}\Big).$$

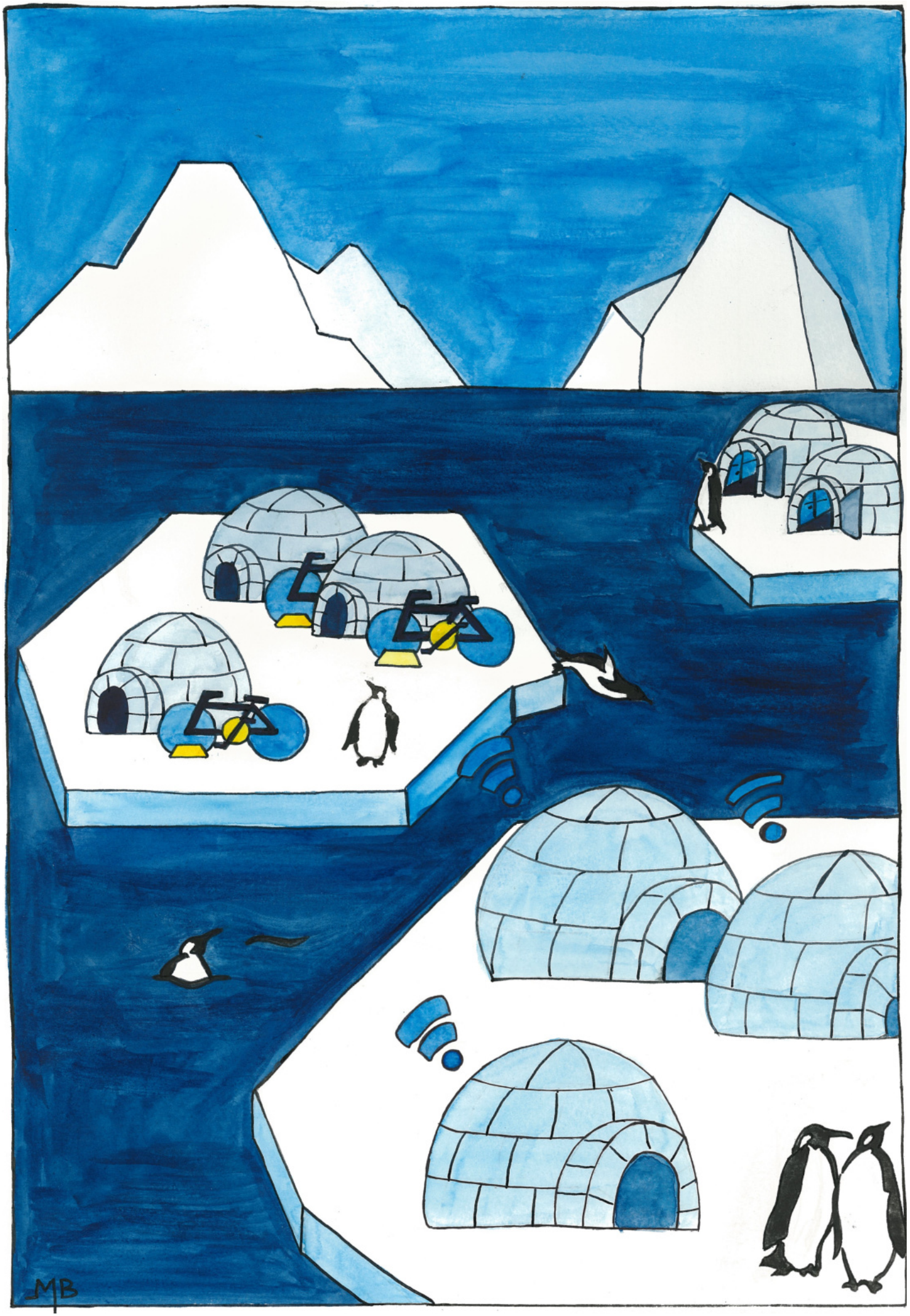
We denote by $\mathrm{I}_m(i)$ the i^{th} integer of vector I_m. By applying Equation 5.13, we obtain that, for all integers $m_1 + m_2 + \cdots + m_K = k$,

$$\mathbb{E}\Big[\prod_{j=1}^K \varepsilon_{t,j}^{m_j}\Big] = \mathbb{E}\Big[\prod_{j=1}^k \varepsilon_{t,\mathrm{I}_m(j)}\Big] = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \sum_{\gamma\in\Gamma_k} \prod_{\{i,i'\}\in\gamma} \sigma_{\mathrm{I}_m(i)\mathrm{I}_m(i')} & \text{if } k \text{ is even}, \end{cases}$$

where $\sigma_{\mathrm{I}_m(i)\mathrm{I}_m(i')}$ is the coefficient $(\mathrm{I}_m(i), \mathrm{I}_m(i'))$ of the covariance matrix Σ. Therefore, if k is odd, $\mathbb{E}\big[(p_t^{\mathrm{T}}\varepsilon_t)^k \,\big|\, \mathcal{F}_{t-1}\big] = 0$; and otherwise

$$\begin{aligned} \mathbb{E}\big[e_{t,p_t}^k \,|\, \mathcal{F}_{t-1}\big] &= \mathbb{E}\Big[\big(p_t^{\mathrm{T}}\varepsilon_t\big)^k \,\Big|\, \mathcal{F}_{t-1}\Big] \\ &= \sum_{m_1+m_2+\cdots+m_K=k} \binom{k}{m_1, m_2, \ldots, m_K} \Big(\prod_{j=1}^K p_{t,j}^{m_j}\Big) \Big(\sum_{\gamma\in\Gamma_k} \prod_{\{i,i'\}\in\gamma} \sigma_{\mathrm{I}_m(i)\mathrm{I}_m(i')}\Big), \end{aligned} \tag{5.14}$$

where the vector of index I_m depends on the integers $m_1, \ldots, m_K$.


MB

6

Online hierarchical forecasting

We study the forecasting of the power consumptions of a population of households and of subpopulations thereof. These subpopulations are built according to location, to exogenous information and/or to profiles we determined from historical households consumption time series. Thus, we aim to forecast the electricity consumption time series at several levels of households aggregation. These time series are linked through some summation constraints which induce a hierarchy. Our approach consists in three steps: feature generation, aggregation and projection. Firstly (feature generation step), we build, for each considering group for households, a benchmark forecast (called features), using random forests or generalized additive models. Secondly (aggregation step), aggregation algorithms, run in parallel, aggregate these forecasts and provide new predictions. Finally (projection step), we use the summation constraints induced by the time series underlying hierarchy to re-conciliate the forecasts by projecting them in a well-chosen linear subspace. We provide some theoretical guaranties on the average prediction error of this methodology, through the minimization of a quantity called regret. We also test our approach on households power consumption data collected in Great Britain by multiple energy providers in the "*Energy Demand Research Project*" context. We build and compare various population segmentations for the evaluation of our approach performance.

This chapter was written in collaboration with Malo Huard and is currently submitted for journal publication; it is available as arXiv preprint number 2003.00585.

1 Introduction

New opportunities come with the recent deployment of smart grids and the installation of meters: they record consumption quasi instantaneously in households. From these records, time series of demand are obtained at various levels of aggregation, such as consumption profiles and regions. For privacy reasons, household records may not be used directly. Moreover, consumption at individual level is erratic and unpredictable. This is why we focus on household aggregations. For demand management, it is useful to predict the global consumption. Furthermore, to dispatch correctly the electricity into the grid, forecasting demand at a regional level is also an important goal. Finally, a good estimation of the consumption of some groups of consumers (with the same profile) may be helpful for the electricity provider which may adapt its offer to perform effective demand side management. Thus, forecasts at various aggregated levels (entire population, geographical areas, groups of same consumption profiles) are useful for an efficient management of consumption.

In this work, we consider a large population that we first split into sub-groups thanks to some clustering methods. We consider the aggregated power consumption of each cluster and also the aggregated power consumption of higher aggregation levels (larger regions, entire population, etc.). Then, we build at each aggregation level, and independently,

benchmark forecasts (called features) using random forests or generalized additive models. Noticing that these time series may be correlated (the consumption of a given region may be close to the one of a neighboring region) and connected to each other through summation constraints (the global consumption is the sum of the region consumptions, e.g.), the problem considered falls under the umbrella of hierarchical time series forecasting. Using these hierarchical relationships may improve the benchmark forecasts that were generated. Our approach consists in combining two methods: feature aggregation and projection in a constrained space. Our aim is to improve forecasts both at the global and at the local levels.

⋆ *Literature discussion for clustering methods.* Different clustering approaches were already proposed in the literature to segment consumers according to their energy consumption behavior. Generally, they relied on the construction of individual features from the average/total consumption and demographic factors. With the recent smart meter deployment, individual consumption records at higher temporal resolutions are now available and allow to consider energy consumption time series in consumers segmentation. Therefore, more complex features may be extracted and used to cluster consumers with classical algorithms. Among others, Chicco *et al.* compared the results obtained by using various unsupervised clustering algorithms (*i.e.*, modified follow-the-leader, hierarchical clustering, k-means, fuzzy k-means) to group together customers with similar consumption behavior (see Chicco et al., 2006); Le Ray and Pinson proposed an adaptive and recursive clustering method that creates typical load profiles updated with newly collected data see(Le Ray and Pinson, 2019); Rodrigues *et al.* described an online hierarchical clustering algorithm, which was applied to cluster energy consumption time series in a load forecasting task (see Rodrigues et al., 2008); Fidalgo *et al.* described a clustering approach based on simulated annealing that tries to reconcile billing processes that use 15 min meter data and monthly total consumption and derive typical profiles for consumers classes (see Fidalgo et al., 2012); Sun *et al.* proposed a copula-based mixture model clustering algorithm that captures complex dependency structures present in energy consumption profiles and detects outliers (see Sun et al., 2017).

⋆ *Literature discussion for hierarchical forecasting.* Traditionally two types of methods have been used for hierarchical forecasting: bottom-up and top-down approaches. In the bottom-up approaches (see Dunn et al., 1976) forecasts are constructed for lower-level quantities and are then summed up to obtain forecasts at the upper levels. In contrast, top-down approaches (see Gross and Sohl, 1990) work by forecasting aggregated quantities and then by determining dis-aggregate proportions to compute lower level predictions. Shlifer and Wolff [1979] compare these two families of methods and conclude that bottom-up approaches work better. The problem of hierarchical forecasting for energy demand was first explored in the GEFCom 2017 forecasting competition, even if none of the top six teams took advantage of the hierarchy information, the different contributions are presented in Hong et al. [2019]. Since then, there has been growing interest in research into hierarchical probabilistic load forecasting. Recently, it has indeed proven successful for load forecasting to improve the global consumption prediction error (see among others Auder et al., 2018). Other approaches (neither bottom-up nor top-down) were recently introduced. For example, Wickramasuriya et al. [2019] forecast all nodes in the hierarchy and reconcile (i.e. impose the respect of hierarchical constraints) them by projection. Their general MinT (for Minimum Trace) approach attempts to capture some

cross-sectional information between times series via the covariance matrix of the errors of the base forecasts. It includes both oblique and orthogonal projections (this is discussed from a geometric perspective in Panagiotelis et al., 2020). Moreover, Van Erven and Cugliari [2015] introduce a game-theoretically optimal reconciliation method to improve a given set of forecasts. Firstly, one comes up with some forecasts for the time series without worrying about hierarchical constraints and then a reconciliation procedure is used to make the forecasts aggregate consistent. This generalizes the previous orthogonal projection to other possible projections in the constrained space (which ensures that the forecasts satisfy the hierarchy). Most work on hierarchical forecasting concentrates on the mean, but some recent work has addressed probabilistic forecasting including that of Ben Taieb et al. [2017], Ben Taieb et al. [2020] and Panagiotelis et al. [2020].

⋆ *Literature discussion for aggregation Methods.* Aggregation methods (also called ensemble methods) for individual sequences forecasting originate from theoretical works by Vovk [1990], Cover [1991] and Littlestone and Warmuth [1994]; their distinguishing feature with respect to classical ensemble methods is that they do not rely on any stochastic modeling of the observations and thus, are able to combine forecasts independently of their generating process. They have been proved to be very effective to predict time series (see for instance Mallet et al., 2009 and Devaine et al., 2013) and those methods were used to win forecasting competitions (see Gaillard et al., 2016). This aggregation approach has recently been extended to the hierarchical setting by Goehry et al. [2019]; they used a bottom-up forecasting approach which consists in aggregating the consumption forecasts of small customers clusters.

In this chapter, we combine the reconciliation approach based on orthogonal projection with some aggregation algorithm to propose a three stage meta-algorithm which is as follows:

1. Generate base forecasts for all times series in the hierarchy,
2. Apply, for each series, an aggregation algorithm that finds an optimal linear combination of the base forecasts
3. Project the combination forecasts onto a coherent subspace to ensure the final forecasts satisfy the hierarchical constraints.

The second step here provides the innovation (Steps 1 and 3 on their own are equivalent to the Ordinary Least Squares version of the MinT algorithm – see Wickramasuriya et al., 2019). By including an aggregation algorithm between these steps, much more of the cross-sectional information is able to be captured, thus improving the forecasts. A theoretical result is provided for the regret bound of the meta-algorithm. We then illustrate the proposed methods using smart meter data collected in Great Britain by multiple energy providers (see Schellong, 2011 and **?**). '*Energy Demand Research Project*' data gathers multiple households power consumption data. We use the polynomially weighted average forecaster with multiple learning rates (ML-Pol, see Gaillard, 2015) aggregation algorithm and consider two population segmentations: a spatial segmentation based on the location of the households and a behavioral one based on household consumption profiles. We evaluate the performance of four strategies for the forecasting of the electricity consumption time series at the several aggregation levels: features, aggregated features, projected features and finally aggregated and projected features.

Notation. In this chapter, vectors will be in bold type and unless stated otherwise, they

are column vectors, while matrices will be in bold underlined. Moreover, we denote the inner product of two vectors $\mathbf{x}$ and $\mathbf{y}$ of the same size by $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{y}$.

2 Methodology

We consider a set of time series $\left\{(y_t^{\mathrm{g}})_{t>0}, \mathrm{g} \in \mathcal{G}\right\}$ connected to each other by some summation constraints: a few of them are equal to the sum of several others – see further for a definition of $\mathcal{G}$. To forecast these time series, a set of features is generated. At any time step t, we want to forecast the vector of the values of the $|\mathcal{G}|$ times series at t, denoted by $\mathbf{y}_t \stackrel{\Delta}{=} (y_t^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}}$. We propose a three-step method to obtain relevant forecasts from these features.

2.1 Modeling of the hierarchical relationships

The relationships between the time series induce a hierarchy which should be exploited to improve forecasts. These summation constraints may be represented by one or more trees, the value at each node being equal to the sum of the ones at its leaves. Let us denote by $\mathcal{G}$ the set of the tree's nodes and $|\mathcal{G}|$ its cardinality. There are as many summation constraints as there are nodes with leaves. Subsequently, we will introduce a matrix $\underline{\mathbf{K}}$ to encode these relationships. Each line of $\underline{\mathbf{K}}$ is related to one of the summation constraints with -1 at the associated node and 1 at its leaves. Thus, for any time step t, the vector of the values of the $|\mathcal{G}|$ times series at t, denoted by $\mathbf{y}_t$, is in the kernel of $\underline{\mathbf{K}}$. Details on and examples of $\underline{\mathbf{K}}$ are provided below. Example 6 treats a single summation constraint. Examples 7 and 8 present more complex relationships between the time series, considering a hierarchy with two levels and two different partitions of the same time series, respectively. Finally, Example 9 combines the two previous cases. In our experiments of Section 7, the underlying hierarchies will be of the form of the ones of Examples 6 and 9.

Remark 20. *The existing hierarchical forecasting literature (e.g., Wickramasuriya et al., 2019) uses the first rows of a "summing matrix" to encode the aggregation constraints with each row corresponding to one of the series, and each column corresponding to the leaves of the hierarchy. We propose here an alternative way of encoding the aggregation constraints using the matrix* $\underline{\mathbf{K}}$ *which is possibly a more parsimonious approach.*

Example 6: *Two-level Hierarchy.* The simplest approach consists in considering a single equation connecting the time series. Here, y^{Tot} stands for the one which is the sum of the N others which are denoted by $y^1, \ldots, y^N$. The underlying hierarchy is represented in Figure 6.1 by a tree with a single root directly connected to N leaves. For any time step t, the time series satisfy $y_t^{\mathrm{Tot}} = y_t^1 + y_t^2 + \cdots + y_t^N$ and the vector $\mathbf{y}_t = \left(y_t^{\mathrm{Tot}}, y_t^1, \ldots, y_t^N\right)^{\mathrm{T}}$ respects the hierarchy if and only if $\underline{\mathbf{K}}\mathbf{y}_t = \mathbf{0}$ with $\underline{\mathbf{K}} = \big(-1, 1, 1, \ldots, 1\big)$.

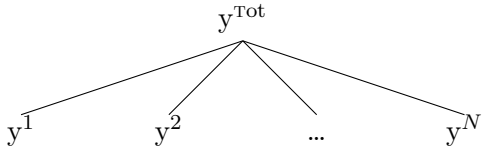


Figure 6.1 – Representation of a two-level hierarchy.

In Section 7, we consider the power consumption of a population of households which are distributed in N regions. This setting will correspond to the present example.

Example 7: *Three-level Hierarchy.*

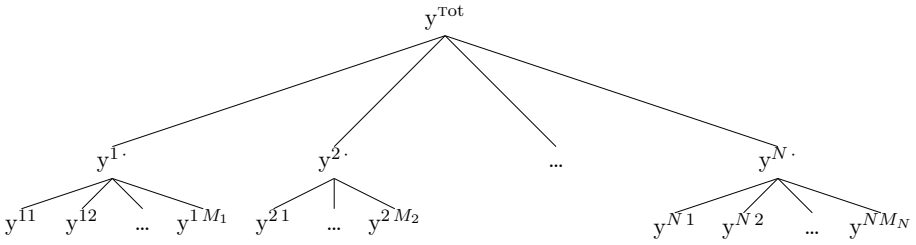


Figure 6.2 – Representation of a three-level hierarchy.

A few leaves of the tree of Example 6 may be broken down into new time series and so on. Figure 6.2 represents a complete three-level hierarchy (although we could consider any multilevel hierarchy) leading to the following summation equations, for each time step t,

$$y_t^{\text{Tot}} = y_t^{1\cdot} + y_t^{2\cdot} + \cdots + y_t^{N\cdot} \tag{1}$$
$$y_t^{i\cdot} = y_t^{i1} + y_t^{i2} + \cdots + y_t^{iM_i}, \quad \forall\, i = 1, \ldots, N. \tag{2i}$$

We order the time series in lexicographical order:

$$\mathbf{y}_t = (y_t^{\text{Tot}}, y_t^{1\cdot}, y_t^{2\cdot}, \ldots, y_t^{N\cdot}, y_t^{11}, y_t^{12}, \ldots, y_t^{1M_1}, y_t^{21}, \ldots, y_t^{NM_N}),$$

and define the constraint matrix $\underline{\mathbf{K}}$ below; each line of $\underline{\mathbf{K}}$ corresponds to one of the constraints mentioned above, either (1) or one of the N constraints $(2i)$ in a way that $\underline{\mathbf{K}}\mathbf{y}_t = \mathbf{0}$ if and only if $\mathbf{y}_t$ respects the hierarchy.

$$\underline{\mathbf{K}} = \begin{pmatrix} -1 & \overbrace{1 \cdots 1}^{N} & & & \\ & -1 & \ddots & \overbrace{1 \cdots 1}^{M_1} & \ddots \\ & & & & \overbrace{1 \cdots 1}^{M_N} \\ & & & -1 & \end{pmatrix} \begin{matrix} \\ \leftarrow \quad (1) \\ \\ \leftarrow \quad (2\text{i}) \end{matrix}$$

This hierarchy corresponds to the concrete example above where, for $1 \leqslant n \leqslant N$, the n region would be further divided into $M_1, \ldots, M_n$ municipalities.

Example 8: *Two Hierarchies of the Same Time Series.* It is also possible to consider two partitions of the same time series y_t^{Tot}. For example, in our experiments of Section 7, in addition to the geographical clustering, we introduce a segmentation of the households based on their profiles. Indeed, they are distributed in N_1 regions but also in N_2 groups depending on their consumption habits. These two different partitions induce the following two equations

$$y_t^{\text{Tot}} = y_t^{1\cdot} + y_t^{2\cdot} + \cdots + y_t^{N_1\cdot}$$
$$y_t^{\text{Tot}} = y_t^{\cdot 1} + y_t^{\cdot 2} + \cdots + y_t^{\cdot N_2},$$

and the two trees associated with these constraints which share the same root and are represented on Figure 6.3.

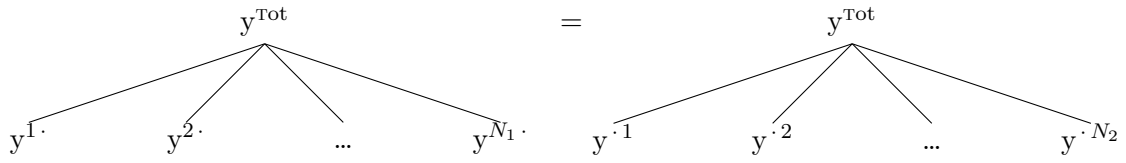


Figure 6.3 – Representation of two two-level hierarchies.

For any time step t, the vector of times series $\mathbf{y}_t = (y_t^{\text{Tot}}, y_t^{1\cdot}, y_t^{2\cdot}, \ldots, y_t^{N_1\cdot}, y_t^{\cdot 1}, y_t^{\cdot 2}, \ldots, y_t^{\cdot N_2})$ satisfies the above equations if and only if $\underline{\mathbf{K}}\mathbf{y}_t = \mathbf{0}$ with

$$\underline{\mathbf{K}} = \begin{pmatrix} \text{-1} & \overbrace{1 \quad \cdots \quad 1}^{N_1} & & \\ \text{-1} & & \overbrace{1 \quad \cdots \quad 1}^{N_2} \end{pmatrix}.$$

The equality of the roots of the two trees is always satisfied in this model. Indeed there is a single time series y_t^{Tot} to forecast and there are therefore only two summation constraints to take into account.

Example 9: *Two Crossed Hierarchies.* Considering two partitions, the time series

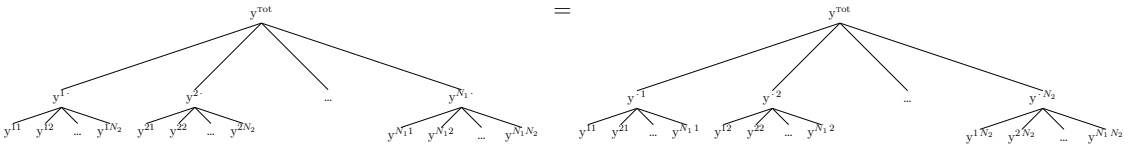


Figure 6.4 – Representation of two crossed hierarchies.

can be represented with two three-level trees sharing the same root and leaves. Only the intermediate levels differs according to which partition is firstly taking into account. The leaves of the trees form a $N_1 \times N_2$-matrix $\left(y^{i\,j}\right)_{1 \leqslant i \leqslant N_1,\, 1 \leqslant j \leqslant N_2}$. An intermediate node of the first tree $y^{i\cdot}$ is the sum of the line i while a node $y^{\cdot j}$ of the second tree is the sum of the column j. Whether we sum rows or columns first, the sum of all coefficients is y_t^{Tot}. In the experiments of Section 7, one partition refers to a geographic distribution of the households while the other classifies them according to their consumption behaviors. The first tree considers breaks down consumption firstly by the N_1 regions and then by the N_2 household profiles. The second one divides the households according to their habits before splitting them geographically. Both trees are represented in Figure 6.4. For any

time step t, the time series satisfy the $2 + N_1 + N_2$ equations

$$y_t^{\mathrm{Tot}} = y_t^{1\,\cdot} + y_t^{2\,\cdot} + \cdots + y_t^{N_1\,\cdot} \tag{1}$$
$$y_t^{i\,\cdot} = y_t^{i\,1} + y_t^{i\,2} + \cdots + y_t^{i\,N_2}, \quad \forall\, i = 1, \ldots, N_1 \tag{2i}$$
$$y_t^{\mathrm{Tot}} = y_t^{\cdot\,1} + y_t^{\cdot\,2} + \cdots + y_t^{\cdot\,N_2} \tag{3}$$
$$y_t^{\cdot\,j} = y_t^{1\,j} + y_t^{2\,j} + \cdots + y_t^{N_1\,j}, \quad \forall\, j = 1, \ldots, N_2. \tag{4j}$$

Equations (1) and (3) refer to the first level of the trees while the $N_1 + N_2$ Equations (2i) and (4j) refer to second levels. At a time step t, by ordering the time series in the vector as

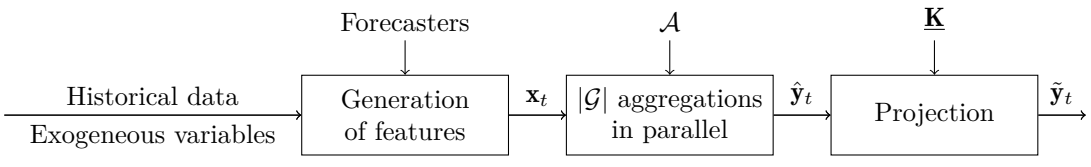
$$\mathbf{y}_t = (y_t^{\mathrm{Tot}}, y_t^{1\,\cdot}, \ldots, y_t^{N_1\,\cdot}, y_t^{\cdot\,1}, \ldots, y_t^{\cdot\,N_1}, y_t^{1\,1}, y_t^{1\,2}, \ldots, y_t^{N_1 N_2}),$$

it may be seen that they respect the hierarchy if and only if $\underline{\mathbf{K}}\mathbf{y}_t = \mathbf{0}$ with

$$\underline{\mathbf{K}} = \begin{pmatrix}
-1 & \overbrace{1 \;\cdots\; 1}^{N_1} & & & & & & \leftarrow\ (1) \\
 & -1 & & & \overbrace{1 \;\cdots\; 1}^{N_2} & & & \\
 & & \ddots & & & \ddots & \overbrace{1 \;\cdots\; 1}^{N_2} & \leftarrow\ (2\mathrm{i}) \\
 & & & -1 & & & & \\
-1 & & & & \overbrace{1 \;\cdots\; 1}^{N_2} & & & \leftarrow\ (3) \\
 & & & & -1 & 1 & 1 & \\
 & & & & \ddots & \ddots \;\cdots\; & \ddots & \leftarrow\ (4\mathrm{j}) \\
 & & & & & -1 & 1 &
\end{pmatrix}.$$

2.2 A three-step forecast

Step 1: For each node g $\in \mathcal{G}$, at each time step t, thanks to an historical data set of the g time series and to some exogenous variables proper to the node g, a forecaster makes the prediction x_t^{g}. These $|\mathcal{G}|$ benchmark forecasts are then collected into the feature vector $\mathbf{x}_t \triangleq \left(x_t^{\mathrm{g}}\right)_{\mathrm{g}\in\mathcal{G}}$. We propose to use the knowledge of all of the features, namely the $|\mathcal{G}|$ benchmark forecasts, and of the summation constraints to improve these $|\mathcal{G}|$ predictions. Step 2: For each node g and each time step t, we form our prediction $\widehat{y}_t^{\mathrm{g}}$ by linearly combining the components of the feature vector $\mathbf{x}_t$ thanks to a so-called aggregation algorithm (a copy $\mathcal{A}^{\mathrm{g}}$ of an aggregation algorithm $\mathcal{A}$ is run separately for each node g). That is, we use all $|\mathcal{G}|$ benchmark forecasts to predict y_t^{g}, not only x_t^{g}. We explain below why this is a good idea – the main reason is given by correlations between time series. The forecasts thus obtained are then gathered into a vector $\hat{\mathbf{y}}_t \triangleq \left(\widehat{y}_t^{\mathrm{g}}\right)_{\mathrm{g}\in\mathcal{G}}$. Step 3: Finally, a re-conciliation step will update the forecast vector so that it is in the kernel $\underline{\mathbf{K}}$. Let us denote by $\tilde{\mathbf{y}}_t$ the final vector of forecasts. We detail below each step of our procedure.



⋆ *First Step: Generation of features.* At a fixed node $\mathrm{g} \in \mathcal{G}$, for any time step t, a forecasting method, which may depend on g, predicts x_t^{g} with the historical data and the exogenous variables of the node g. The forecasting methods we use in the experiments of Section 7 are described in Section 5 and include non linear sequential ridge regression, fully adaptive Bernstein online aggregation and polynomially weighted average forecaster with multiple learning rates. These benchmark forecasts are henceforth called features and are gathered in $\mathbf{x}_t = \left(x_t^{\mathrm{g}}\right)_{\mathrm{g}\in\mathcal{G}}$. This feature vector is used in the aggregation step that comes next to predict again each time series; we discuss below and in Subsection 2.4 why we do so (the main reasons being that it is a good idea because of the correlations between the times series and also because it eases the description of our method). We focus here on $|\mathcal{G}|$ benchmark forecasts – one for each of the nodes; however, we could also have considered several predictions per nodes.

⋆ *Second Step: Aggregation.* The above features are generated independently with different exogenous variables and possibly different methods. Yet, the observations $\left(y_t^{\mathrm{g}}\right)_{\mathrm{g}\in\mathcal{G}}$ may be correlated. For example, considering load forecasting, the consumptions associated with two nearby regions can be strongly similar. Furthermore, the observations are related though the summations constraints (although we disregard these equations here). This is why linearly combining the features may refine some forecasts – this is exactly what this step does. Formally, an aggregation algorithm outputs at each round a vector of weights $\hat{\mathbf{u}}_t^{\mathrm{g}}$ and returns the forecast $\widehat{y}_t^{\mathrm{g}} \triangleq \hat{\mathbf{u}}_t^{\mathrm{g}} \cdot \mathbf{x}_t$. It does so based on the information available, that is, the feature vector $\mathbf{x}_t$ and past data. We consider an aggregation algorithm $\mathcal{A}$ (see Section 6) and form a copy $\mathcal{A}^{\mathrm{g}}$ for each node g, which we feed with an input parameter vector $\mathbf{s}_0^{\mathrm{g}}$. These predictions are then gathered into the vector $\hat{\mathbf{y}}_t = (y_t^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}}$. This algorithm aims for the best linear combination of features and there are theoretical performance guaranties associated with these aggregation algorithms, see Section 6 for details.

Instead of this approach based on benchmark forecasting and aggregation node by node, we could have considered a meta-model to directly predict the time series vector $\left(y_t^{\mathrm{g}}\right)_{\mathrm{g}\in\mathcal{G}}$ at each round t (with a common forecaster and therefore without any aggregation step). Once this global forecast would have been obtained, we would have gone straight to the projection stage. In such a model, the number of variables to be taken into account (the historical data of the time series but also the exogenous variables specific to each node) would have been considerable and getting relevant forecasts would have not been an easy task. But actually, a practical choice motivated our method for the most. Indeed, the forecasters may be black boxes proper to each node and the exogenous variables of a node g may be unknown at a node g′. In our experiments, we followed this three-step approach. However, our method totally operates if, for each node g and at each time step t, an external expert provides the forecast x_t^{g}. How these features have been obtained is no longer an issue and the aim is to improve these benchmark forecasts with aggregation and reconciliation steps. Thus, at each time step t, only the features are reveal at time t and by skipping the generation of features step, we go straight to the aggregation step.

⋆ *Third Step: Projection.* As the $|\mathcal{G}|$ executions of Algorithm $\mathcal{A}$ are run in parallel and independently, the obtained forecast vector $\hat{\mathbf{y}}_t$ does not necessary respect hierarchical constraints. To correct that, we consider the orthogonal projection of $\hat{\mathbf{y}}_t$ onto the kernel of $\underline{\mathbf{K}}$, which we denote by $\Pi_{\underline{\mathbf{K}}}(\hat{\mathbf{y}}_t)$. This updated forecast $\tilde{\mathbf{y}}_t \triangleq \Pi_{\underline{\mathbf{K}}}(\hat{\mathbf{y}}_t)$ fulfills the hierarchical constraints.

To sum up, at each round t, we first generate benchmark forecasts – also called features – $\mathbf{x}_t$. These predictions are then aggregated to form a new vector of forecast $\hat{\mathbf{y}}_t$, which is itself updated in the projection step in $\tilde{\mathbf{y}}_t$. This procedure is stated in Meta-algorithm 7. Moreover, we can also directly project the features, skipping the aggregation step; this leads to the forecasts $\Pi_{\underline{\mathbf{K}}}(\mathbf{x}_t)$ – they are identical to the OLS version of the MinT algorithm proposed in Wickramasuriya et al. [2019]. Thus, we get four forecasts ($\mathbf{x}_t$, $\Pi_{\underline{\mathbf{K}}}(\mathbf{x}_t)$, $\hat{\mathbf{y}}_t$ and $\tilde{\mathbf{y}}_t$) for each node and each round. The performance of our strategies is measured in mean squared error. In Section 7, we compare these four methods in the scope of power consumption forecasting.

Protocol 7 Aggregation and projection of features with summation constraints

Input
Set of nodes $\mathcal{G}$ and constraint matrix $\underline{\mathbf{K}}$
Feature generation technique, see Section 5
Aggregation algorithm $\mathcal{A}$ taking parameter vector $\mathbf{s}_0$, see Section 6
Compute the orthogonal projection matrix $\Pi_{\underline{\mathbf{K}}} = \left(I_{|\mathcal{G}|} - \underline{\mathbf{K}}^{\mathrm{T}}(\underline{\mathbf{K}}\underline{\mathbf{K}}^{\mathrm{T}})^{-1}\underline{\mathbf{K}}\right)$
for $\mathrm{g} \in \mathcal{G}$ **do**
Create a copy of $\mathcal{A}$ denoted by $\mathcal{A}^{\mathrm{g}}$ and run with $\mathbf{s}_0^{\mathrm{g}} = \mathbf{s}_0$
end for
for $t = 1, \ldots$ **do**
Generate features $\mathbf{x}_t$
for $\mathrm{g} \in \mathcal{G}$ **do**
$\mathcal{A}^{\mathrm{g}}$ outputs $\mathbf{u}_t^{\mathrm{g}}$
end for
Collect forecasts: $\hat{\mathbf{y}}_t = (\widehat{y}_t^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}}^{\mathrm{T}}$, where $\widehat{y}_t^{\mathrm{g}} = \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$
Project forecasts: $\tilde{\mathbf{y}}_t = \Pi_{\underline{\mathbf{K}}}(\hat{\mathbf{y}}_t)$
for $\mathrm{g} \in \mathcal{G}$ **do**
$\mathcal{A}^{\mathrm{g}}$ observes y_t^{g}
end for
Suffer a prediction error $\frac{1}{|\mathcal{G}|}\sum_{\mathrm{g}\in\mathcal{G}}\left(y_t^{\mathrm{g}} - \widetilde{y}_t^{\mathrm{g}}\right)^2$
end for
aim
Minimize the average prediction error

$$\widetilde{L}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{G}|}\left\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\right\|^2 = \frac{1}{T\,|\mathcal{G}|}\sum_{t=1}^{T}\sum_{\mathrm{g}\in\mathcal{G}}\left(y_t^{\mathrm{g}} - \widetilde{y}_t^{\mathrm{g}}\right)^2.$$

2.3 Forecast assessment – form of the theoretical guaranties achieved

Our forecasts are linear combinations of the features and are evaluated by the average prediction error

$$\widetilde{L}_T \triangleq \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{G}|}\sum_{\mathrm{g}\in\mathcal{G}}\left(y_t^{\mathrm{g}} - \widetilde{y}_t^{\mathrm{g}}\right)^2. \tag{6.1}$$

We want to compare our method to constant linear combinations of features. For example, recalling that, for $\mathrm{g} \in \mathcal{G}$, x_t^{g} is the benchmark prediction of y_t^{g}, using $\boldsymbol{\delta}^{\mathrm{g}} \triangleq \mathbf{1}_{\{i=\mathrm{g}\}}$ (the standard basis vector that points in the g direction) as weights should be a good first choice to define a constant linear combination (for any $\mathrm{g} \in \mathcal{G}$, this strategy provides $\boldsymbol{\delta}^{\mathrm{g}} \cdot \mathbf{x}_t = x_t^{\mathrm{g}}$ as forecast for y_t^{g}). Thus, the matrix $\big(\boldsymbol{\delta}^{\mathrm{g}}\big)_{\mathrm{g}\in\mathcal{G}}$ defines a constant benchmark strategy and its cumulative prediction error is

$$L_T\Big(\big(\boldsymbol{\delta}^{\mathrm{g}}\big)_{\mathrm{g}\in\mathcal{G}}\Big) \triangleq \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{G}|}\sum_{\mathrm{g}\in\mathcal{G}}\big(y_t^{\mathrm{g}} - x_t^{\mathrm{g}}\big)^2.$$

As soon as the features $(x_t^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}}$ are well-chosen, this quantity is small. But, these benchmark predictions do not satisfy the summation constraints *a priori* and it won't be fair to compare our forecasts (which do respect to hierarchy – projection step ensures it) to these benchmark forecasts – or any other constant linear combinations of features. Thus, we introduce, in paragraph 2.3.1, the set $\mathcal{C}$ which contains all the constant strategies which satisfy the hierarchical constraints and we also detail how a such strategy can be represented by a $|\mathcal{G}| \times |\mathcal{G}|$-matrix $\underline{\mathbf{U}} \in \mathcal{C}$. In paragraph 2.3.2, we decompose, for any $\underline{\mathbf{U}} \in \mathcal{C}$, the average prediction error into an approximation error $L_T(\underline{\mathbf{U}})$ – the average prediction error of $\underline{\mathbf{U}}$ – and a sequential estimation error $\mathcal{E}_T(\underline{\mathbf{U}})$. To achieve almost as well as the best constant combination of features, we want to obtain some guarantee of the form:

$$\widetilde{L}_T \leqslant \inf_{\underline{\mathbf{U}}\in\mathcal{C}}\Big\{L_T(\underline{\mathbf{U}}) + \mathcal{E}_T(\underline{\mathbf{U}})\Big\}, \quad \text{where} \quad \mathcal{E}_T(\underline{\mathbf{U}}) = \mathcal{O}\Big(\frac{1}{\sqrt{T}}\Big). \tag{6.2}$$

Indeed, if $\mathcal{E}_T(\underline{\mathbf{U}}) \underset{T\to+\infty}{\longrightarrow} 0$, the average prediction error of our strategy tends to $L_T(\underline{\mathbf{U}})$ – and classical convergence rate are in $\frac{1}{\sqrt{T}}$ (see sections 2.3.2 and 6). We will explain how this aim is equivalent to minimizing the quantity called regret that we define below.

2.3.1 Class of comparison

We consider here a constant strategy, namely $|\mathcal{G}|$ linear combinations of the features. More formally, let us denote by $\mathbf{u}^{\mathrm{g}}$ a constant weight vector which provides, for any time step t, the forecast $\mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t$ for the time series y_t^{g}. By batching these $|\mathcal{G}|$ vectors into a matrix $\underline{\mathbf{U}} \triangleq (\mathbf{u}^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}} \in \mathcal{M}_{|\mathcal{G}|}$, predictions satisfy the constraints for a time step t if $\underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t \in \mathrm{Ker}(\underline{\mathbf{K}})$. For it to be true for any t (except for a few particular case – for instance if all features vector are null), this requires that the image of $\underline{\mathbf{U}}^{\mathrm{T}}$ is in the kernel of $\underline{\mathbf{K}}$. We introduce the following set of matrices, for which associated forecasts necessarily satisfy the hierarchical constraints

$$\mathcal{C} \triangleq \Big\{\underline{\mathbf{U}} = \big(\mathbf{u}^1 \,\big|\, \ldots \,\big|\, \mathbf{u}^{|\mathcal{G}|}\big) \,\Big|\, \mathrm{Im}\big(\underline{\mathbf{U}}^{\mathrm{T}}\big) \subset \mathrm{Ker}\big(\underline{\mathbf{K}}\big)\Big\}.$$

Note that, for any matrix $\underline{\mathbf{U}} \in \mathcal{M}_{|\mathcal{G}|}$, by definition of the orthogonal projection $\Pi_{\underline{\mathbf{K}}}$, the forecast vector $\Pi_{\underline{\mathbf{K}}}\underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t$ satisfies the hierarchical relationships so the set $\mathcal{C}$ contains the matrix $\underline{\mathbf{U}}\Pi_{\underline{\mathbf{K}}}^{\mathrm{T}}$. This implies that the set $\mathcal{C}$ is not empty. To compare our methods to any constant strategy $\underline{\mathbf{U}} \in \mathcal{C}$, we now introduce the common notion of regret.

2.3.2 Aim: regret minimization

We want to compare the average prediction error $\widetilde{L}_T$ to $L_T(\underline{\mathbf{U}})$, where $\underline{\mathbf{U}} \in \mathcal{C}$ so the forecasts associated with $\underline{\mathbf{U}}$ satisfy the hierarchical constraints – otherwise, the two strategies would

not be comparable because our predictions do respect the hierarchy. Good algorithms should ensure that $\widetilde{L}_T$ is not too far from the best $L_T(\underline{\mathbf{U}})$. We thus define, for any $\underline{\mathbf{U}} = (\mathbf{u}^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}} \in \mathcal{C}$, the cumulative prediction error of the associated constant linear combinations of features by

$$L_T(\underline{\mathbf{U}}) \triangleq \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{G}|}\sum_{\mathrm{g}\in\mathcal{G}}\left(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\right)^2 = \frac{1}{T|\mathcal{G}|}\sum_{t=1}^{T}\left\|\mathbf{y}_t - \underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t\right\|^2.$$

In order to obtain a theoretical guarantee of the form of Equation (6.2), we decompose the average prediction error as

$$\widetilde{L}_T = L_T(\underline{\mathbf{U}}) + \frac{R_T(\underline{\mathbf{U}})}{T|\mathcal{G}|}\,, \tag{6.3}$$

where, the quantity $R_T(\underline{\mathbf{U}})$, commonly called regret is defined as the difference between the cumulative prediction error of our method and the one for weights $\underline{\mathbf{U}}$:

$$R_T(\underline{\mathbf{U}}) \triangleq T|\mathcal{G}| \times \left(\widetilde{L}_T - L_T(\underline{\mathbf{U}})\right) = \sum_{t=1}^{T}\left\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\right\|^2 - \sum_{t=1}^{T}\left\|\mathbf{y}_t - \underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t\right\|^2.$$

In the light of Equation (6.3), the average prediction error $\widetilde{L}_T$ we attempt to minimize breaks down into an approximation error $L_T(\underline{\mathbf{U}})$ (the best prediction error we can hope for) and a sequential estimation error (dependent of how quickly the model estimate $\underline{\mathbf{U}}$), proportional to the regret $R_T(\underline{\mathbf{U}})$. As stated before, the aim for algorithms is that $\widetilde{L}_T$ is as close as possible to $\min_{\underline{\mathbf{U}}\in\mathcal{C}} L_T(\underline{\mathbf{U}})$ (with $\mathcal{C}$ the class of comparison defined above), which is equivalent to $\max_{\underline{\mathbf{U}}\in\mathcal{C}} R_T(\underline{\mathbf{U}})$ being small. This point of view is very common for online forecasting methods (see, among others, Devaine et al., 2013 and Mallet et al., 2009), and for an algorithm to be useful, $\max_{\underline{\mathbf{U}}\in\mathcal{C}} R_T(\underline{\mathbf{U}})$ need to be sub-linear in T (otherwise the error remains constant – or even worst: it increases with time). Typical theoretical guaranties provide bounds of order $\sqrt{T}$ (see for example, Deswarte et al., 2019 and Amat et al., 2018).

2.4 Technical discussion: why we require the same features at each node.

In this section, we explain why we consider the same features vector for each nodes. *A priori*, we could have a different set of features at each node $\mathbf{x}_t^{\mathrm{g}}$, created with methods specific to this node. Also the size of feature vector d^{g} associated with the node g could vary. Prediction of a time series y_t^{g} associated to a d^{g}-vector $\mathbf{u}^{\mathrm{g}}$ is $\widehat{y}_t^{\mathrm{g}} = \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t^{\mathrm{g}}$. Therefore, a global constant strategy is a set $\left\{\mathbf{u}^1,\ldots,\mathbf{u}^{|\mathcal{G}|}\right\} \subset \mathbb{R}^{d^1\times\cdots\times d^{|\mathcal{G}|}}$. First it is a little less practical because unlike the previous setting, the vectors $\mathbf{u}^1,\ldots,\mathbf{u}^{|\mathcal{G}|}$ and $\mathbf{x}^1,\ldots,\mathbf{x}^{|\mathcal{G}|}$ are of different sizes, so it is less easy to use matrix notations. Moreover, it becomes tricky to specify the class of constant strategies to compare to. As said before, the forecast vector $\hat{\mathbf{y}}_t = (\widehat{y}_t^{\mathrm{g}})_{\mathrm{g}\in\mathcal{G}}$ satisfies the summation constraints if and only if it is in the kernel of $\underline{\mathbf{K}}$. Thus, the following set, which contains the constant strategies fulfilling the hierarchical constraints for all $t > 0$,

$$\left\{\left(\mathbf{u}^1,\ldots,\mathbf{u}^{|\mathcal{G}|}\right) \in \mathbb{R}^{d^1\times\cdots\times d^{|\mathcal{G}|}} \,\middle|\, \forall t>0,\ \left(\mathbf{u}^1\cdot\mathbf{x}_t^1,\ldots,\mathbf{u}^{|\mathcal{G}|}\cdot\mathbf{x}_t^{|\mathcal{G}|}\right)^{\mathrm{T}} \in \mathrm{Ker}\left(\underline{\mathbf{K}}\right)\right\},$$

is not explicitly defined and may be empty because of the number of constraints on $(\mathbf{u}^1,\ldots,\mathbf{u}^{|\mathcal{G}|}) \in \mathbb{R}^{d^1\times\cdots\times d^{|\mathcal{G}|}}$ which increases at each time step. If there is no restrictions on

the feature vectors, these constraints could be linearly independent, leading to an empty set. Indeed, if we consider that the times series are connected by K summations relationships, at each time step t, the $d^1 + \cdots + d^{|\mathcal{G}|}$ coefficients of vectors $\mathbf{u}^1, \ldots, \mathbf{u}^{|\mathcal{G}|}$ are linked by K equations. As features are proper to each node, theses constraints have no reason to be dependent, so as soon as $T \times K > d^1 + \cdots + d^{|\mathcal{G}|}$, the above set may likely be empty. Because of that, it is not clear how to define the regret in this setting. For this reason, we decided to use the same features vectors $\mathbf{x}_t$ for all nodes of $\mathcal{G}$; which has also the benefit of allowing a simpler presentation.

3 Main theoretical result

From now on, let us introduce the following notation concerning the regret bound of Algorithm $\mathcal{A}$.

Notation 1. *We assume that, for any* $\mathrm{g} \in \mathcal{G}$ *with the initialization parameter vector* $\mathbf{s}_0^{\mathrm{g}}$*, Algorithm* $\mathcal{A}^{\mathrm{g}}$ *ensures, for* $T > 0$ *and for any* $\mathbf{u}^{\mathrm{g}} \in \mathbb{R}^{|\mathcal{G}|}$*, any* $\mathbf{x}_{1:T} = \mathbf{x}_1, \ldots \mathbf{x}_T$ *and any* $y_{1:T}^{\mathrm{g}} = y_1^{\mathrm{g}}, \ldots, y_T^{\mathrm{g}}$*,*

$$R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}}) \triangleq \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \hat{y}_t^{\mathrm{g}}\right)^2 - \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\right)^2 \leqslant \mathrm{B}(\mathbf{x}_{1:T}, y_{1:T}^{\mathrm{g}}, \mathbf{s}_0^{\mathrm{g}}, \mathbf{u}^{\mathrm{g}}), \tag{6.4}$$

where $\mathrm{B}(\ldots)$ *is some regret bound obtained on the aggregation algorithm regret and which may depend on features* $\mathbf{x}_{1:T}$*, observations* $y_{1:T}^{\mathrm{g}}$*, the vector of weights* $\mathbf{u}^{\mathrm{g}}$ *and on the algorithm parameters* $\mathbf{s}_0^{\mathrm{g}}$*.*

Details and examples of these regret bounds are provided in Section 6 that describes the aggregation algorithms considered in the experiments of Section 7. As getting a linear bound is trivial (by using the common assumption that prediction errors are bounded), these bounds have to be sub-linear to be of interest. Referring to the average prediction error decomposition of Equation (6.3), the sub-linearity ensures that the sequential estimation error $R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}})/T$ tends to 0. This notation makes it possible to establish a bound of the cumulative regret.

Theorem 9. *Under Notation 1, for any matrix* $\underline{\mathbf{U}} \in \mathcal{C}$ *and any* $T \geqslant 1$*,*

$$R_T(\underline{\mathbf{U}}) = \sum_{t=1}^{T} \left\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\right\|^2 - \sum_{t=1}^{T} \left\|\mathbf{y}_t - \underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t\right\|^2 \leqslant \sum_{\mathrm{g}\in\mathcal{G}} \mathrm{B}\left(\mathbf{x}_{1:T}, y_{1:T}^{\mathrm{g}}, \mathbf{s}_0^{\mathrm{g}}, \mathbf{u}^{\mathrm{g}}\right).$$

The regret $R_T(\underline{\mathbf{U}})$ is not just the sum over all the nodes of the regrets $R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}})$ of Equation (6.4). Indeed, we do not evaluate here the forecasts $\hat{\mathbf{y}}_t$ but those obtained after the projection step: $\tilde{\mathbf{y}}_t$. The projection step provides a diminishing of the square prediction error and we just have to sum Equation (6.4) on all nodes to get the bound.

Proof. This regret bound results from two main arguments: Pythagorean theorem, on the one hand, and Notation 1, on the other hand. For any $t \geqslant 1$, as $\mathbf{y}_t \in \mathrm{Ker}(\underline{\mathbf{K}})$, the Pythagorean theorem ensures

$$\left\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\right\|^2 = \left\|\mathbf{y}_t - \Pi_{\underline{\mathbf{K}}}(\hat{\mathbf{y}}_t)\right\|^2 \leqslant \left\|\mathbf{y}_t - \hat{\mathbf{y}}_t\right\|^2. \tag{6.5}$$

Let us fix a matrix $\underline{\mathbf{U}} = \left(\mathbf{u}^1 | \ldots | \mathbf{u}^{|\mathcal{G}|}\right) \in \mathcal{C}$. Firstly, the application of Pythagorean theorem ensures that the projection step reduces regret. Rewriting the regret as a sum over the nodes, we then use Notation 1 independently for each node of $\mathcal{G}$ to conclude the proof.

$$\begin{aligned}
R_T(\underline{\mathbf{U}}) &= \sum_{t=1}^{T} \left\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\right\|^2 - \sum_{t=1}^{T} \left\|\mathbf{y}_t - \underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t\right\|^2 \\
&\overset{(6.5)}{\leqslant} \sum_{t=1}^{T} \left\|\mathbf{y}_t - \hat{\mathbf{y}}_t\right\|^2 - \sum_{t=1}^{T} \left\|\mathbf{y}_t - \underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}_t\right\|^2 \\
&= \sum_{\mathrm{g}\in\mathcal{G}} \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}}\right)^2 - \sum_{\mathrm{g}\in\mathcal{G}} \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\right)^2 = \sum_{\mathrm{g}\in\mathcal{G}} R_T^{\mathrm{g}}\left(\mathbf{u}^{\mathrm{g}}\right) \\
&\overset{(6.4)}{\leqslant} \sum_{\mathrm{g}\in\mathcal{G}} \mathrm{B}\left(\mathbf{x}_{1:T}, y_{1:T}^{\mathrm{g}}, \mathbf{s}_0^{\mathrm{g}}, \mathbf{u}^{\mathrm{g}}\right).
\end{aligned}$$

□

Note that similar results, also based on Pythagorean theorem, have been obtained in Panagiotelis et al. [2020] (see Theorem 3.1).

Remark 21. *For an initialization parameter vector $\mathbf{s}_0^{\mathrm{g}}$, and a subset $\mathcal{D} \subset \mathbb{R}^{|\mathcal{G}|}$, some aggregation algorithms provide a uniform regret bound of the following form:*

$$R_T^{\mathrm{g}}(\mathcal{D}) \triangleq \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}}\right)^2 - \min_{\mathbf{u}^{\mathrm{g}}\in\mathcal{D}} \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\right)^2 \leqslant \mathrm{B}\left(\mathbf{x}_{1:T}^{\mathrm{g}}, y_{1:T}^{\mathrm{g}}, \mathbf{s}_0^{\mathrm{g}}\right).$$

In this case, let us introduce, for any subset $\mathcal{B} \subset \mathcal{M}_{|\mathcal{G}|}$, the subset $\mathcal{B}_{|\mathcal{D}} \triangleq \left\{\underline{\mathbf{U}} \in \mathcal{B} \,|\, \forall \mathrm{g} \in \mathcal{G}, \mathbf{u}^{\mathrm{g}} \in \mathcal{D}\right\}$. Then, we bound the cumulative regret $R_T(\mathcal{D})$ defined just below with

$$R_T(\mathcal{D}) \triangleq \max_{\underline{\mathbf{U}}\in\mathcal{C}_{|\mathcal{D}}} R_T(\underline{\mathbf{U}}).$$

With the same previous arguments we get the uniform regret bound

$$R_T(\mathcal{D}) = \sum_{t=1}^{T} \left\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\right\|^2 - \min_{\underline{\mathbf{U}}\in\mathcal{C}_{|\mathcal{D}}} \sum_{t=1}^{T} \left\|\mathbf{y}_t - \underline{\mathbf{U}}^{\mathrm{T}}\mathbf{x}\right\|^2 \leqslant \sum_{\mathrm{g}\in\mathcal{G}} \mathrm{B}\left(\mathbf{x}_{1:T}^{\mathrm{g}}, y_{1:T}^{\mathrm{g}}, \mathbf{s}_0^{\mathrm{g}}\right).$$

4 On one operational constraint: half-hourly predictions with one-day-delayed observations

In this section, we highlight the differences between the previous theoretical setting and the practical setting of our experiments and how these changes affect the regret bound. In Section 7, we aim to forecast power consumptions at half-hourly intervals. Meta-algorithm 7 makes the implicit assumption that historical time series values are available and to forecast at a time step t, we can use $\mathbf{y}_{1:t-1}$. We thus assume that very recent past observations, up to half an hour ago, would be available – and it is not realistic at all. Indeed, there is some operational constraints on the power network and on meters that make it difficult to

instantly access the data: it is common to obtain load records with a delay of a few hours or even a few days. Although this delay is becoming shorter with the deployment of smart meters and the evolution of grids, we cannot consider we have access to the consumption of the previous half-hour. To take into account these operational constraints and to carry out experiments under practical conditions, we make the classic assumption that we have access to consumptions with a delay of 24 hours (see among others Fan and Hyndman, 2012 and Gaillard et al., 2016). As now, only past observations $\mathbf{y}_{1:t-H}$ are available at a time step t, with $H = 48$, we adapt the previous method a bit.

As we will see in Section 5, the half-hour of the day is a crucial variable for power consumption forecasting and to obtain relevant forecasts, we will consider the consumption of the previous day at the same half-hour (but never the one of the previous half-hour). Thus the delay in the access to consumption observation is not an issue for feature generation. But it becomes especially problematic for online learning (in our experiments, features are generated offline with models trained on historical data). Indeed, in the aggregation step of our method, we assume to observe, for each node g and at each time step t, the consumption y^{g}_{t-1} – that is not possible anymore. To deal with this issue we initially considered two solutions. In our first approach, for any $\mathrm{g} \in \mathcal{G}$, the time series (y^{g}_t) is divided into H time series with daily time steps. Then, H aggregations are done in parallel and, as $t-1$ now refers to the previous day, there is no more delay issue. The H series are then collected to reconstruct a time series at half-hour time step. For a constant strategy $\mathbf{u}^{\mathrm{g}}$, the regret of the global aggregation $R^{\mathrm{g}}_T(\mathbf{u}^{\mathrm{g}})$ is simply the sum of the H regrets – that refer to the H aggregation run in parallel on the H daily time series – denoted by $\left(R^{\mathrm{g}\,h}_{T/H}(\mathbf{u}^{\mathrm{g}})\right)_{1\leqslant h\leqslant H}$, so we have

$$R^{\mathrm{g}}_T(\mathbf{u}^{\mathrm{g}}) = \sum_{h=1}^{H} R^{\mathrm{g}\,h}_{T/H}(\mathbf{u}^{\mathrm{g}}).$$

If we consider an aggregation algorithm that ensures a bound of the form of Notation 1 where the bound B depends only on the horizon time – namely, $R \leqslant B$ for all h – the regret associated with the half-hourly time series (y^{g}_t) satisfies:

$$R^{\mathrm{g}}_T(\mathbf{u}^{\mathrm{g}}) \leqslant H \times B(T/H).$$

Joulani et al. [2013] provide an overview of work on online learning under delayed feedback and for our framework, which refers to full information setting with general feedback. The bound above matches their results. In a second approach, we "ignore" the delay in a sense that we apply the aggregation algorithms as if the delayed observations $\mathbf{y}_{t-H}$ were $\mathbf{y}_t$. Thus, in Meta-algorithm 7, at each node g and any time step t, instead of outputting the forecast $y^{\mathrm{g}}_t = \mathbf{u}^{\mathrm{g}}_t \cdot \mathbf{x}_t$, algorithm $\mathcal{A}^{\mathrm{g}}$ outputs $y^{\mathrm{g}}_t = \mathbf{u}^{\mathrm{g}}_{t-H} \cdot \mathbf{x}_t$. For simplicity of notation, the aggregation algorithms of Section 6 are presented in their original version, namely assuming that observations at $t-1$ are available at a time step t. Such adaptations have already been tested: Algorithm 15 of Gaillard [2015] gives a delayed version of Algorithm 7 that we also use in Section 7. After testing both approaches, we kept the second one, which achieves a much better performance. Our choice was also supported by Chapter 9 of Gaillard [2015] experiments, which drew similar conclusions.

5 Generation of the features

Here we describe the forecasting methods we use in the experiments of Section 7 to generate the benchmark predictions that will be used as features in the sequel. We recall (see Section 2) that throughout this work, we consider that, at each node $g \in \mathcal{G}$ and for any time step t, a forecaster provides a benchmark prediction x_t^g based on historical data of the time series $(y_t^g)_{g\in\mathcal{G}}$ and on exogenous variables relative to the node g. These $|\mathcal{G}|$ forecasters independently generate the $|\mathcal{G}|$ forecasts (x_t^g) in parallel and the set of features $\mathbf{x}_t$ is made up of the above $|\mathcal{G}|$ benchmark predictions. Forecasts can be the output of any predictive model. In the experiments of Section 7, we consider three forecasting methods, that are described in the following Subsections 5.1, 5.2 and 5.3.

Notation. Subsections 5.1 and 5.2 present parametric methods. For any parameter a of the model, we will denote by $\widehat{a}$ its estimation (no matter the method we use).

5.1 Auto-regressive model

A simple approach consists in considering an auto-regressive model. Let us fix $g \in \mathcal{G}$ and assume that, to predict the time series $(y_t^g)_{t>0}$, we have access to historical observations. For a time step t, the model specifies that the output variable y_t^g depends linearly on its own previous values. In Section 7, we consider the power consumption at half-hourly intervals. For a time step t, to forecast the time series y_t^g we assume to have access to the power consumption at D-1 and D-7, which correspond to y_{t-H}^g and $y_{t-7\times H}^g$, respectively. We predict the consumption half-hour by half-hour thanks to linear models taking as explanatory variables its values at D-1 and D-7. We assume that these H auto-regressive models have the same coefficients. Thus, for this modeling, the power consumption associated with the node g equals

$$y_t^g = a_1^g y_{t-H}^g + a_7^g y_{t-7\times H}^g + \text{noise}\,.$$

For each $g \in \mathcal{G}$, we estimate the coefficients a_1^g and a_7^g using ordinary least squares regression on a training data set. Therefore, at a new round t, we predict

$$x_t^g = \widehat{a}_1^g y_{t-H}^g + \widehat{a}_7^g y_{t-7\times H}^g\,.$$

5.2 General additive model

Generalized additive models are effective semi-parametric approaches to forecast electricity consumption (see Chapter 3). They model the power demand as a sum of independent exogenous (possibly non-linear) variable effects. We describe this model using the specification we chose in our experiments. In Section 7, for a node $g \in \mathcal{G}$, we take into account some local meteorological variables at the half-hour time step: the temperature τ^g and the smoothed temperature $\bar{\tau}^g$, the visibility ν^g, and the humidity υ^g. For round t, we also introduce calendar variables: the day of the week d_t (equal to 1 for Monday, 2 for Tuesday, etc.), the half-hour of the day $h_t \in \{1, ..., H\}$ and the position in the year $\pi_t \in [0, 1]$, which takes linear values between $\pi_t = 0$ on January 1st at 00:00 and $\pi_t = 1$ on December the 31st at 23:59. As the effect of the half-hour h_t is crucial to forecast load, it is often more efficient to consider a model per half-hour (see Fan and Hyndman, 2012 and Goude et al., 2014). The global model is then the sum of H daily models, one for each half-hour of the

day. More precisely, we consider the following additive model for the load, which breaks down time by half hours:

$$y_t^{\mathrm{g}} = \sum_{h=1}^{H} \mathbf{1}_{h_t=h} \Big[a_h^{\mathrm{g}} y_{t-7\times H}^{\mathrm{g}} + s_{1,h}^{\mathrm{g}}\big(y_{t-H}^{\mathrm{g}}\big) + s_{\tau,h}^{\mathrm{g}}(\tau_t^{\mathrm{g}}) + s_{\bar{\tau},h}^{\mathrm{g}}(\bar{\tau}_t^{\mathrm{g}}) + s_{\nu,h}^{\mathrm{g}}(\nu_t^{\mathrm{g}}) + s_{v,h}^{\mathrm{g}}(v_t^{\mathrm{g}}) + \sum_{d=1}^{7} w_{d,h}^{\mathrm{g}} \mathbf{1}_{d_t=d} + s_{\pi,h}^{\mathrm{g}}(\pi_t) \Big] + \text{noise}.$$

The $s_{1,h}^{\mathrm{g}}$, $s_{\tau,h}^{\mathrm{g}}$, $s_{\bar{\tau},h}^{\mathrm{g}}$, $s_{\nu,h}^{\mathrm{g}}$, $s_{v,h}^{\mathrm{g}}$ and $s_{\pi,h}^{\mathrm{g}}$ functions catch the effect of the consumption lag, the meteorological variables and of the yearly seasonality. They are cubic splines: $\mathcal{C}^2$-smooth functions made up of sections of cubic polynomials joined together at points of a grid. The coefficients a_h^{g} and $w_{d,h}^{\mathrm{g}}$ model the influence of the consumption at D-7 and of the day of the week. Indeed, we consider a linear effect for the consumption at D-7 (it achieved a better performance than a spline effect in our experiments) and as the day of the week takes only 7 values, we write its effect as a sum of indicator functions, and thus 7 coefficients $w_{d,h}^{\mathrm{g}}$ are considered. As we consider a model per half-hour, all the coefficients and splines are indexed by h. To estimate each model, we use the Penalized Iterative Re-Weighted Least Square (P-IRLS) method Wood, 2006, implemented in the `mgcv` R-package (see Wood, 2020), on a training data set. At any node $\mathrm{g} \in \mathcal{G}$, for a new round t, we then output the forecast

$$x_t^{\mathrm{g}} = \sum_{h=1}^{H} \mathbf{1}_{h_t=h} \Big[\widehat{a}_h^{\mathrm{g}} y_{t-7\times H}^{\mathrm{g}} + \widehat{s}_{1,h}^{\mathrm{g}}\big(y_{t-H}^{\mathrm{g}}\big) + \widehat{s}_{\tau,h}^{\mathrm{g}}(\tau_t^{\mathrm{g}}) + \widehat{s}_{\bar{\tau},h}^{\mathrm{g}}(\bar{\tau}_t^{\mathrm{g}}) + \widehat{s}_{\nu,h}^{\mathrm{g}}(\nu_t^{\mathrm{g}}) + \widehat{s}_{v,h}^{\mathrm{g}}(v_t^{\mathrm{g}}) + \sum_{d=1}^{7} \hat{w}_{d,h} \mathbf{1}_{d_t=d} + \widehat{s}_{\pi,h}^{\mathrm{g}}(\pi_t) \Big].$$

5.3 Random forests

Random forests form a powerful learning method for classification and regression that constructs a collection of decision trees from training data and output, for each new data point, the mean prediction of the individual trees. Introduced by Breiman [2001], theses approaches operate well on many applications. Recent work demonstrates their efficiency in forecasting power consumption (see, among others Goehry et al., 2019 and Fan and Hyndman, 2012). A random forest is made up of a set $\big(T_k^{\mathrm{g}}\big)_{1\leqslant k\leqslant K}$ of decision trees grown in the following way (see Breiman et al., 1984 for further details). For each $k = 1, \ldots, K$, we first randomly draw, with replacement, n points from the training data set and start at the root, that contains all the points of the sub-sample. At each node $\mathcal{N}$ with more than m data points, V variables are randomly selected among the exogenous variables. Given a variable $v \in V$ and a threshold s, each point of the node $\mathcal{N}$ is assigned to the left daughter node $\mathcal{N}_L$ if its value in v is lower than s or to the right daughter node $\mathcal{N}_R$ otherwise. Considering only these V variables, the best split – given by a pair (v, s) of variable and an associated threshold – to separate the points into two set $\mathcal{N}_L$ and $\mathcal{N}_R$ is determined by minimizing the variance criterion indicated below. For any node $\mathcal{N}$ let us define the variance $\mathrm{Var}(\mathcal{N})$ by

$$\mathrm{Var}(\mathcal{N}) \triangleq \frac{1}{|\mathcal{N}|} \sum_{i\in\mathcal{N}} \big(y_i^{\mathrm{g}} - \bar{y}_{\mathcal{N}}^{\mathrm{g}}\big)^2, \quad \text{with} \quad \bar{y}_{\mathcal{N}}^{\mathrm{g}} \triangleq \frac{1}{|\mathcal{N}|} \sum_{i\in\mathcal{N}} y_i^{\mathrm{g}}.$$

Each node $\mathcal{N}$ is split in the two daughter nodes $\mathcal{N}_R^\star$ and $\mathcal{N}_L^\star$ (determined by the choice of v and s) minimizing the following criterion

$$\big(\mathcal{N}_R^\star, \mathcal{N}_L^\star\big) \in \underset{\mathcal{N}_R, \mathcal{N}_L}{\operatorname{argmin}} \frac{|\mathcal{N}_R|}{n} \operatorname{Var}\big(\mathcal{N}_R\big) + \frac{|\mathcal{N}_L|}{n} \operatorname{Var}\big(\mathcal{N}_L\big). \tag{6.6}$$

Thus, we create a binary test to split the points of the node. When all the leaves contain fewer than m points, we associate with each leaf the mean of its data points. For a new point, we look at the values of its variables. For each $k = 1, \ldots, K$, we browse the tree T_k^{g} and predict the value of the corresponding leaf. The K resulting forecasts are then averaged out. Algorithm 4 describes the above procedure and is implemented in the `ranger` R-package. In the experiments of Section 7, we take n equal to the number of

Algorithm 4 Random Forest for Regression

Parameters
Number of trees K
Sample size n
Minimal node size m
Number of variables to possibly split at in each node V
for $k = 1, \ldots, K$ **do**
Draw a sample (with replacement) of size n from training data
Construct the tree T_k starting at the root with all the n data points
while a leaf contains more than m data points **do**
for each leaf of more than m data points **do**
Select V variables
Split the node into two nodes using the variance criterion (6.6) among the chosen variables
end for
end while
end for
Output $\big(T_k^{\mathrm{g}}\big)_{1 \leqslant k \leqslant K}$
Prediction at a new data point
Mean of the K forecasts output by the trees $\big(T_k^{\mathrm{g}}\big)_{1 \leqslant k \leqslant K}$

data points in the training set, $m = 5$ and $K = 500$ (default parameters of `ranger`). The number V has been optimized by grid search; what we obtained is that, for each node, we keep two-thirds of the variables to split it (these variables are the same as the ones described in the previous section). With $\big(T_k\big)_{1 \leqslant k \leqslant K}$, the trees constructed by Algorithm 4 run on a training data set, the forecast of any node $\mathrm{g} \in \mathcal{G}$, at a new round t, is then

$$x_t^{\mathrm{g}} = \frac{1}{K} \sum_{k=1}^{K} T_k^{\mathrm{g}}\big(y_{t-7\times H}^{\mathrm{g}}, y_{t-H}^{\mathrm{g}}, \tau_t^{\mathrm{g}}, \bar{\tau}_t^{\mathrm{g}}, \nu_t^{\mathrm{g}}, \upsilon_t^{\mathrm{g}}, \pi_t, d_t, h_t\big).$$

6 Aggregation algorithms

This section describes the three aggregation algorithms we use in the experiments of Section 7. At a time step t, for a node $\mathrm{g} \in \mathcal{G}$, a copy $\mathcal{A}^{\mathrm{g}}$ of an aggregation algorithm $\mathcal{A}$ takes the feature vector $\mathbf{x}_t$ (generated with one of methods of the previous section) as an

input and outputs the forecast $\widehat{y}_t^{\,\mathrm{g}} = \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$. Therefore, to forecast the node g, we use $\mathbf{x}_t$, which contains the predictions of all the nodes (including that of the considering node). We remind that the features $\big(x_t^{\mathrm{g}}\big)_{\mathrm{g}\in\mathcal{G}}$ are generated independently with possibly different exogenous variables but that the observations $\big(y_t^{\mathrm{g}}\big)_{\mathrm{g}\in\mathcal{G}}$ may be strongly correlated. This is why we consider aggregation to refine some forecasts by combining the features. Our experiments demonstrate that this aggregation step improves the forecasts. Subsection 6.1 presents a trick to empirically standardize the features and the observations first. On the one hand, this preprocessing justifies boundedness assumptions (see Assumption 8) on observations and features, that ensure some theoretical guaranties of the form requested by Notation 1. On the other hand, this preprocessing simplifies hyper-parameters search (for the aggregation step) as we can choose the same for every series since they have similar statistics (scale and variance). Following Subsections 6.2 and 6.3 introduce the aggregation algorithms and some technical tricks implemented in the experiments of Section 7.

6.1 Standardization

In empirical machine learning, it is known that standardizing observations and features may significantly improve results, and sequential learning is no exception (see Gaillard et al., 2019). In addition, standardization makes the calibration of the parameters of the algorithm common to all the nodes, namely for each algorithm $\mathcal{A}^{\mathrm{g}}$, we choose the hyper-parameters $\mathbf{s}_0^{\mathrm{g}} = \check{\mathbf{s}}_0$. We can do so, because thanks to the preprocessing below, features and observations will be of the same order. Let us fix $\mathrm{g} \in \mathcal{G}$ and $t > 0$. We consider the following transformations, relying on statistics S^{g} and $\underline{\check{\mathbf{E}}}$ computed on T_0 historical time steps:

$$
\begin{aligned}
y_t^{\mathrm{g}} \to \check{y}_t^{\mathrm{g}} &\triangleq \frac{y_t^{\mathrm{g}} - x_t^{\mathrm{g}}}{S^{\mathrm{g}}} && \text{Observations tranform} \\
\mathbf{x}_t \to \check{\mathbf{x}}_t &\triangleq \underline{\check{\mathbf{E}}}\mathbf{x}_t && \text{Features transform}
\end{aligned}
$$

$$
\text{with} \quad S^{\mathrm{g}} = \max_{1-T_0 \leqslant t \leqslant 0} |y_t^{\mathrm{g}} - x_t^{\mathrm{g}}| \qquad \text{and} \qquad \underline{\check{\mathbf{E}}} \triangleq \left(\frac{1}{T_0} \sum_{t=1-T_0}^{0} \mathbf{x}_t \, \mathbf{x}_t^{\mathrm{T}} \right)^{-1/2} .
$$

We thus assume that the Gram matrix $\frac{1}{T_0}\sum_{t=1-T_0}^{0} \mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}$ is invertible, which is a reasonable assumption as soon as T_0 is large enough – otherwise, we could use the pseudo-inverse of the Gram matrix. Our standardization process differs from the usual methods (see details below) but it provides the theoretical guaranties set out below. Furthermore, it makes sense for the following reasons. Fixing $\mathrm{g} \in \mathcal{G}$, when features and observations are bounded, S^{g} is an estimation of a bound on $y_t^{\mathrm{g}} - x_t^{\mathrm{g}}$. The re-scaling of $(y_t^{\mathrm{g}} - x_t^{\mathrm{g}})$ by S^{g} should provide transformed observations lying in $[-1, 1]$ or a some neighboring range. It also reduces and homogenizes the variances for all the nodes. A simple example may illustrate this variance reduction. For deterministic features, the variance of non-transformed observations satisfy $\mathrm{Var}\big(y_t^{\mathrm{g}}\big) = \mathrm{Var}\big(y_t^{\mathrm{g}} - x_t^{\mathrm{g}}\big)$. The variance of standardized observations is then divided by $\big(S^{\mathrm{g}}\big)^2$ and we have $\mathrm{Var}\big(\check{y}_t^{\mathrm{g}}\big) = \mathrm{Var}\big(y_t^{\mathrm{g}}\big)/\big(S^{\mathrm{g}}\big)^2$. For T_0 large enough, the variance of transformed observations should be less than 1. Indeed, with high probability, the maximum of the absolute values of the random variable $(y_t^{\mathrm{g}} - x_t^{\mathrm{g}})$ on $t = 1 - T_0, \dots, 0$ (which is S^{g}), is higher than its standard deviation $\sqrt{\mathrm{Var}(y_t^{\mathrm{g}})}$ and thus $(S^{\mathrm{g}})^2 > \mathrm{Var}\big(y_t^{\mathrm{g}}\big)$. Moreover, the expectation of $(y_t^{\mathrm{g}} - x_t^{\mathrm{g}})$ should be close to 0 as soon as the features are correctly generated. Indeed, the more the benchmark forecast are relevant, the more the

observations are re-centered. Concerning the features, our standardization is classic in the case of centered features. The matrix $\breve{\underline{\mathbf{E}}}^2$ would then be an estimation of the inverse of the co-variance matrix of vectors $\mathbf{x}_t$, and the multiplication of the features by $\breve{\underline{\mathbf{E}}}$ would provide transformed features whose co-variance matrix is close to the identity matrix. Here, we do not recenter observations and features with some empirical mean as it is classically done (this would be unconvenient for our regret analysis). Anyway, Subsection 7.3 provides some experimental results which confirm that our preprocessing standardizes reasonably well observations and features. Moreover, we tested classical standardization (with re-centering) on features and obtained results similar to those presented in Section 7 (but, as hinted at above, no theoretical guaranties would be associated with this classical standardization).

We run Algorithm $\mathcal{A}^{\mathrm{g}}$ on transformed features and observations with the initialization parameter vector $\breve{\mathbf{s}}_0$ (which does not depend on g) and obtain a standardized prediction at node g, denoted by $\bar{y}_t^{\mathrm{g}}$. Then, we transform this output to get the (non-standardized) forecast

$$\widehat{y}_t^{\mathrm{g}} \triangleq S^{\mathrm{g}} \bar{y}_t^{\mathrm{g}} + x_t^{\mathrm{g}}.$$

For any vector $\breve{\mathbf{u}}^{\mathrm{g}} \in \mathbb{R}^{|\mathcal{G}|}$, we introduce the standardized regret associated with transformed observations and features, denoted by $\breve{R}_T^{\mathrm{g}}(\breve{\mathbf{u}}^{\mathrm{g}})$ as:

$$\begin{aligned}
\breve{R}_T^{\mathrm{g}}(\breve{\mathbf{u}}^{\mathrm{g}}) &\triangleq \sum_{t=1}^{T} \left(\breve{y}_t^{\mathrm{g}} - \bar{y}_t^{\mathrm{g}}\right)^2 - \sum_{t=1}^{T} \left(\breve{y}_t^{\mathrm{g}} - \breve{\mathbf{u}}^{\mathrm{g}} \cdot \breve{\mathbf{x}}_t\right)^2 \\
&= \sum_{t=1}^{T} \left(\frac{y_t^{\mathrm{g}} - x_t^{\mathrm{g}}}{S^{\mathrm{g}}} - \frac{\widehat{y}_t^{\mathrm{g}} - x_t^{\mathrm{g}}}{S^{\mathrm{g}}}\right)^2 - \sum_{t=1}^{T} \left(\frac{y_t^{\mathrm{g}} - x_t^{\mathrm{g}}}{S^{\mathrm{g}}} - \breve{\mathbf{u}}^{\mathrm{g}} \cdot \left(\breve{\underline{\mathbf{E}}} \mathbf{x}_t\right)\right)^2 \\
&= \frac{1}{\left(S^{\mathrm{g}}\right)^2} \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}}\right)^2 - \frac{1}{\left(S^{\mathrm{g}}\right)^2} \sum_{t=1}^{T} \Bigg(y_t^{\mathrm{g}} - \underbrace{\left(x_t^{\mathrm{g}} + S^{\mathrm{g}} \left(\breve{\underline{\mathbf{E}}} \breve{\mathbf{u}}^{\mathrm{g}}\right) \cdot \mathbf{x}_t\right)}_{\mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t} \Bigg)^2 .
\end{aligned}$$

In the equations above, we define $\mathbf{u}^{\mathrm{g}} \triangleq \boldsymbol{\delta}^{\mathrm{g}} + S^{\mathrm{g}} \breve{\underline{\mathbf{E}}} \breve{\mathbf{u}}^{\mathrm{g}})$ where $\boldsymbol{\delta}^{\mathrm{g}} \triangleq (\mathbf{1}_{\{i=\mathrm{g}\}})_{i \in \mathcal{G}}$ denotes the standard basis vector that points in the g direction. Equivalently, $\breve{\mathbf{u}}^{\mathrm{g}} = \breve{\underline{\mathbf{E}}}^{-1} (\mathbf{u}^{\mathrm{g}} - \boldsymbol{\delta}^{\mathrm{g}})/S^{\mathrm{g}}$, so there is a bijective correspondence between the vectors $\mathbf{u}^{\mathrm{g}}$ and $\breve{\mathbf{u}}^{\mathrm{g}}$. Therefore, by noticing that $x_t^{\mathrm{g}} = \boldsymbol{\delta}^{\mathrm{g}} \cdot \mathbf{x}_t$, the regret associated with original features and observations is related to the regret of transformed data by the following equation:

$$\breve{R}_T^{\mathrm{g}}(\breve{\mathbf{u}}^{\mathrm{g}}) = \frac{1}{\left(S^{\mathrm{g}}\right)^2} \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}}\right)^2 - \frac{1}{\left(S^{\mathrm{g}}\right)^2} \sum_{t=1}^{T} \left(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\right)^2 = \frac{R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}})}{\left(S^{\mathrm{g}}\right)^2} .$$

Furthermore, as for any $\breve{\mathbf{u}} \in \mathbb{R}^{|\mathcal{G}|}$, Notation 1 ensures

$$\breve{R}_T^{\mathrm{g}}(\breve{\mathbf{u}}^{\mathrm{g}}) = \sum_{t=1}^{T} \left(\breve{y}_t^{\mathrm{g}} - \bar{y}_t^{\mathrm{g}}\right)^2 - \sum_{t=1}^{T} \left(\breve{y}_t^{\mathrm{g}} - \breve{\mathbf{u}}^{\mathrm{g}\mathrm{T}} \breve{\mathbf{x}}_t\right)^2 \leqslant \mathrm{B}\left(\breve{\mathbf{x}}_{1:T}, \breve{y}_{1:T}^{\mathrm{g}}, \breve{\mathbf{s}}_0, \breve{\mathbf{u}}^{\mathrm{g}}\right),$$

Combining the two previous equations yields the following proposition.

Proposition 3. *For any* $\mathrm{g} \in \mathcal{G}$ *and any* $\mathbf{u}^{\mathrm{g}} \in \mathbb{R}^{|\mathcal{G}|}$*, if Notation 1 holds for Algorithm* $\mathcal{A}^{\mathrm{g}}$ *run on transformed observations and features* $\breve{y}_{1:T}^{\mathrm{g}}$ *and* $\breve{\mathbf{x}}_{1:T}$*, with the initialization parameter vector* $\breve{\mathbf{s}}_0$*, we have, for* $T > 0$*,*

$$R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}}) \leqslant \left(S^{\mathrm{g}}\right)^2 \mathrm{B}\left(\breve{\mathbf{x}}_{1:T}, \breve{y}_{1:T}^{\mathrm{g}}, \breve{\mathbf{s}}_0, \breve{\mathbf{u}}^{\mathrm{g}}\right) \quad \text{where} \quad \breve{\mathbf{u}}^{\mathrm{g}} = \breve{\underline{\mathbf{E}}}^{-1} (\mathbf{u}^{\mathrm{g}} - \boldsymbol{\delta}^{\mathrm{g}})/S^{\mathrm{g}}.$$

Throughout the section, without loss of generality and to simplify the notation, we now replace the features and observations with the standardized ones. Thus, we will write y_t^{g} for $\breve{y}_t^{\mathrm{g}}$, $\mathbf{x}_t$ for $\breve{\mathbf{x}}_t$ and so on. Moreover, we make the following assumption on the boundedness of features and observations.

Assumption 8 – *Boundedness assumptions.* For any $t > 0$ and any $\mathrm{g} \in \mathcal{G}$ we assume that there is a constant $C > 0$ such that

$$|y_t^{\mathrm{g}}| \leqslant C \quad \text{and} \quad |x_t^{\mathrm{g}}| \leqslant C.$$

Some boundedness assumptions on features and observations are frequently required to establish theoretical guaranties. Here, the constant is common to all the nodes. Practically, this assumption makes sense because of the previous transformations. As explained above, it centers and normalizes observations and features. Subsection 7.3 presents statistics on features and observation before and after standardization and indicates possible values of the constant C.

In the two next subsections, we introduce the aggregation algorithms we implemented in Section 7. We recall that, for any $\mathrm{g} \in \mathcal{G}$, at a round t, the algorithm $\mathcal{A}^{\mathrm{g}}$ provides a weight vector $\mathbf{u}_t^{\mathrm{g}}$ and thus forecasts y_t^{g} with $\mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$. In Subsection 6.2, we consider a linear aggregation algorithm: there is no restriction on the computed weight vectors. In Subsection 6.3, the two algorithms output convex combinations of features: the weight vectors are in the $|\mathcal{G}|$-simplex denoted by $\Delta_{|\mathcal{G}|}$. However, there is no reason to consider such a restriction and this is why the last paragraph of the subsection presents a trick to extend the previous algorithms to output linear combinations of features for which the weight vectors are in a L_1-ball. Thus, there are no longer restrictions on the sum or the sign of the weights.

6.2 Linear aggregation: sequential non-linear ridge regression

The first aggregation algorithm that we consider is the sequential non-linear ridge regression of Vovk [2001] and Azoury and Warmuth [2001]. So, for any $\mathrm{g} \in \mathcal{G}$, Algorithm $\mathcal{A}^{\mathrm{g}}$ refers here to Algorithm 5 run with regularization parameter $\mathbf{s}_0 = \lambda$. For any time step $t \geqslant 2$, this algorithm, chooses vectors $\mathbf{u}_t^{\mathrm{g}}$ as follow:

$$\mathbf{u}_t^{\mathrm{g}} \in \operatorname*{argmin}_{\mathbf{u}^{\mathrm{g}} \in \mathbb{R}^d} \sum_{s=1}^{t-1} \left(y_s^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_s\right)^2 + \left(\mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\right)^2 + \lambda \|\mathbf{u}^{\mathrm{g}}\|^2. \tag{6.7}$$

The solution of this minimization problem is given by:

$$\mathbf{u}_t^{\mathrm{g}} = \Big(\lambda\big(\mathbf{1}_{\{i=j\}}\big)_{(i,j)\in\mathcal{G}^2} + \sum_{s=1}^{t} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}\Big)^{\dagger} \sum_{s=1}^{t-1} y_s^{\mathrm{g}} \mathbf{x}_s,$$

where $A^{\dagger}$ denotes the pseudo-inverse of the matrix A. Algorithm 5 provides a sequential implementation of the solution of this convex minimization problem. The above non-linear ridge regression is a penalized ordinary least-squares regression. Since the features may be strongly correlated, the least squares estimator, $\mathbf{u}_t^{\mathrm{g}} \in \operatorname{argmin}_{\mathbf{u}^{\mathrm{g}} \in \mathbb{R}^d} \sum_{s=1}^{t-1} \left(y_s^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_s\right)^2$, could lead to very large prediction if a new features vector belongs to an eigenspace of the empirical gram matrix associated to a small value. The regularization term $\lambda \|\mathbf{u}^{\mathrm{g}}\|^2$ ensures

Algorithm 5 Non-Linear Sequential Ridge Regression

aim
Predict the time series $\left(y_t^{\mathrm{g}}\right)_{1\leqslant t\leqslant T}$
parameter Regularization parameter $\lambda > 0$
initialization $\underline{\mathbf{A}}_0 = \lambda\left(\mathbf{1}_{\{i=j\}}\right)_{(i,j)\in\mathcal{G}^2}$ and $\mathbf{b}_0 = (0,\ldots,0)^{\mathrm{T}}$
for $t = 1,\ldots,T$ **do**
Update matrix $\underline{\mathbf{A}}_t = \underline{\mathbf{A}}_{t-1} + \mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}$
Compute the vector $\mathbf{u}_t^{\mathrm{g}} = \underline{\mathbf{A}}_t^{-1}\mathbf{b}_{t-1}$
Output prediction $\widehat{y}_t^{\mathrm{g}} = \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$
Update vector $\mathbf{b}_t = \mathbf{b}_{t-1} + y_t^{\mathrm{g}}\mathbf{x}_t$
end for

that eigenvalues of the empirical gram matrix are not too small. We then add the regularization term $\left(\mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\right)^2$ which is the last term of the cumulative prediction error $\left(y_t^{\mathrm{g}}-\mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\right)^2$ where we have replaced unknown y_t^{g} by our best guess 0. It is known to improve the regret bound (see Vovk, 2001 and Gaillard et al., 2019). In our case (standardized targets), it particularly makes sense because it biases predictions towards 0; which, because of the standardization, biases aggregated predictions towards benchmark predictions.

Under Assumption 8, for any vector $\mathbf{u}^{\mathrm{g}} \in \mathbb{R}^{|\mathcal{G}|}$, with the algorithm $\mathcal{A}^{\mathrm{g}}$ set to the non-linear ridge regression (6.7) run with regularization parameter λ, Theorem 11.8 of the monograph *Prediction, Learning, and Games* by Cesa-Bianchi and Lugosi [2006] or Theorem 2 of Gaillard et al. [2019] provide the following theoretical guaranties:

$$R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}}) \leqslant \lambda\|\mathbf{u}^{\mathrm{g}}\|^2 + |\mathcal{G}|C^2\,\ln\left(1+\frac{C^2T}{\lambda}\right).$$

So, for any $\underline{\mathbf{U}} = \left(\mathbf{u}^1|\ldots|\mathbf{u}^{|\mathcal{G}|}\right) \in \mathcal{C}$, as $\left\|\underline{\mathbf{U}}\right\|_F^2 = \sum_{\mathrm{g}\in\mathcal{G}}\|\mathbf{u}^{\mathrm{g}}\|^2$, Theorem 9 ensures

$$R_T(\underline{\mathbf{U}}) = \sum_{\mathrm{g}\in\mathcal{G}} R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}}) \leqslant \lambda\left\|\underline{\mathbf{U}}\right\|_F^2 + |\mathcal{G}|^2C^2\,\ln\left(1+\frac{C^2T}{\lambda}\right) = \mathcal{O}\left(|\mathcal{G}|^2\ln T\right).$$

That is, since the sequential non-linear ridge regression provides a logarithmic regret bound, Meta-algorithm 7 achieves a bound of the same order.

6.3 Convex aggregation

We focus here on uniform bounds and use notation introduced in Remark 21. The following two algorithms were initially designed to compete against the best feature. Namely, for a node $\mathrm{g} \in \mathcal{G}$, the Bernstein online aggregation (BOA, see Wintenberger, 2017) and polynomially weighted average forecaster with multiple learning rates (ML-Pol, see Gaillard, 2015) provide some bound on the difference between the cumulative prediction error $L_T^{\mathrm{g}} \triangleq \sum_{t=1}^T\left(y_t^{\mathrm{g}}-\widehat{y}_t^{\mathrm{g}}\right)^2$ of the strategy and $\min_{i\in\mathcal{G}}\sum_{t=1}^T\left(y_t^{\mathrm{g}}-x_t^i\right)^2$. At each time step t, both strategies compute weight vector $\mathbf{u}_t^{\mathrm{g}} = \left(u_t^{\mathrm{g}\,i}\right)_{i\in\mathcal{G}}$ based on historical data. These vectors are in the $|\mathcal{G}|$-simplex, which we denote by $\Delta_{|\mathcal{G}|}$. For each feature $i \in \mathcal{G}$, the weight $u_t^{\mathrm{g}\,i}$ is, for BOA, an exponential function of a regularized cumulative prediction error of the feature x_t^i and, for ML-Pol, a polynomial function of the cumulative prediction error of x_t^i. However, by using gradients of prediction errors instead of the original prediction errors the average

error of these algorithms may come close to $\min_{\mathbf{u}^{\mathrm{g}}\in\Delta_{|\mathcal{G}|}} \frac{1}{T}\sum_{t=1}^{T}\big(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big)^2$. This "gradient trick" (see Cesa-Bianchi and Lugosi, 2006, Section 2.5) is presented in the next paragraph and is already integrated in the statements of the algorithms below. Moreover, for both algorithms, the computed weight vectors are in $\Delta_{|\mathcal{G}|}$. As we do not necessarily want to impose such a restriction, we use another trick, introduced by Kivinen and Warmuth [1997] and presented in the last paragraph. It extends the class of comparison from the $|\mathcal{G}|$-simplex to an L_1-ball of radius α denoted by $\mathcal{B}_\alpha \triangleq \big\{\mathbf{u}^{\mathrm{g}} \in \mathbb{R}^{|\mathcal{G}|} \,\big|\, \|\mathbf{u}\|_1 = \sum_{i\in\mathcal{G}} |u^{\mathrm{g}\,i}| \leqslant \alpha\big\}$. The aim is then to come close to the cumulative error $\min_{\mathbf{u}^{\mathrm{g}}\in\mathcal{B}_\alpha} \sum_{t=1}^{T}\big(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big)^2$.

⋆ *Gradient Trick: from the best feature to the best convex combination of features.* We consider an aggregation algorithm that takes as input, at any time step $t+1$, the previous prediction errors of each feature $(y_t^{\mathrm{g}} - x_t^i)^2$, for any $i \in \mathcal{G}$, and that of the forecast outputs at t: $(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}})^2$. Although this trick generalizes to various prediction errors, we focus here to its application in our case, namely the quadratic prediction error. We assume that the algorithm provides a bound on the quantity (see notation of Remark 21)

$$R_T^{\mathrm{g}}\big(\delta_{|\mathcal{G}|}\big) \triangleq \sum_{t=1}^{T}(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}})^2 - \min_{i\in\mathcal{G}} \sum_{t=1}^{T}(y_t^{\mathrm{g}} - x_t^i)^2\,.$$

where $\delta_{|\mathcal{G}|} \triangleq \big\{(\boldsymbol{\delta}^i)_{i\in\mathcal{G}}\big\}$ is the set of canonical basis vectors (so we have $\boldsymbol{\delta}^i\cdot\mathbf{x}_t = x_t^i$). The gradient trick consists in giving, instead of the prediction errors $(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}})^2$ and $(y_t^{\mathrm{g}} - x_t^i)^2$, for any $i \in \mathcal{G}$, the pseudo prediction errors functions defined below as input to algorithm $\mathcal{A}^{\mathrm{g}}$. This will provide a bound on the pseudo regret denoted by $\widetilde{R}_T^{\mathrm{g}}\big(\delta_{|\mathcal{G}|}\big)$. We will prove that the same bound is achieved for the minimum of $R_T^{\mathrm{g}}(\mathbf{u}^{\mathrm{g}})$ taken over $\mathbf{u}^{\mathrm{g}} \in \Delta_{|\mathcal{G}|}$ (and not only $\delta_{|\mathcal{G}|}$), namely $R_T^{\mathrm{g}}\big(\Delta_{|\mathcal{G}|}\big)$. We detail here how the trick works and gives:

$$R_T^{\mathrm{g}}\big(\Delta_{|\mathcal{G}|}\big) \leqslant \widetilde{R}_T^{\mathrm{g}}\big(\delta_{|\mathcal{G}|}\big).$$

Let us fix a vector $\mathbf{u}^{\mathrm{g}} = \big(u^{\mathrm{g}\,i}\big)_{i\in\mathcal{G}} \in \Delta_{|\mathcal{G}|}$, we have for each $t = 1, \ldots, T$

$$\begin{aligned}
\big(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}}\big)^2 - \big(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big)^2 &= \big(2y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big)\big(\mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t - \widehat{y}_t^{\mathrm{g}}\big) \\
&= 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)\big(\widehat{y}_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big) - \big(\widehat{y}_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big)^2 \\
&\leqslant 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)\big(\widehat{y}_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big)\,. \qquad (6.8)
\end{aligned}$$

By plugging this equation into the definition of the regret, we obtain

$$\begin{aligned}
R_T^{\mathrm{g}}\big(\mathbf{u}^{\mathrm{g}}\big) &\triangleq \sum_{t=1}^{T}(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}})^2 - \sum_{t=1}^{T}(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t)^2 \\
&\overset{(6.8)}{\leqslant} \sum_{t=1}^{T} 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)\big(\widehat{y}_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}}\cdot\mathbf{x}_t\big) = \sum_{t=1}^{T} 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)\widehat{y}_t^{\mathrm{g}} - \sum_{t=1}^{T}\sum_{i\in\mathcal{G}} u^{\mathrm{g}\,i} 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)x_t^i.
\end{aligned}$$

As $\mathbf{u}^{\mathrm{g}}$ belongs to the $|\mathcal{G}|$-simplex (so $\forall i \in \mathcal{G}$, $u^{\mathrm{g}\,i} \geqslant 0$ and $\sum_{i\in\mathcal{G}} u^{\mathrm{g}\,i} = 1$), we get:

$$\sum_{i\in\mathcal{G}} u^{\mathrm{g}\,i} x_t^i \;\geqslant\; \min_{j\in\mathcal{G}} x_t^j \sum_{i\in\mathcal{G}} u^{\mathrm{g}\,i} = \min_{j\in\mathcal{G}} x_t^j$$

Therefore, for any vector $\mathbf{u}^{\mathrm{g}} \in \Delta_{|\mathcal{G}|}$, the regret $R_T^{\mathrm{g}}\big(\mathbf{u}^{\mathrm{g}}\big)$ is bounded by

$$R_T^{\mathrm{g}}\big(\mathbf{u}^{\mathrm{g}}\big) \leqslant \sum_{t=1}^{T} 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)\widehat{y}_t^{\mathrm{g}} - \min_{j\in\mathcal{G}} \sum_{t=1}^{T} 2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)x_t^j \triangleq \widetilde{R}_T^{\mathrm{g}}\big(\delta_{|\mathcal{G}|}\big).$$

Thus, we now give the pseudo prediction errors associated with each feature $2\big(\widehat{y}_t^{\,\mathrm{g}} - y_t^{\mathrm{g}}\big)x_t^i$, with $i \in \mathcal{G}$, and with the output forecast $2\big(\widehat{y}_t^{\,\mathrm{g}} - y_t^{\mathrm{g}}\big)\widehat{y}_t^{\,\mathrm{g}}$ as input to algorithm $\mathcal{A}^{\mathrm{g}}$. It provides a bound on the pseudo regret defined above $\widetilde{R}_T^{\mathrm{g}}\big(\delta_{|\mathcal{G}|}\big)$; and we get the same bound on $R_T^{\mathrm{g}}\big(\Delta_{|\mathcal{G}|}\big)$. As a final note, we emphasize that Assumption 8 allows to establish that pseudo prediction errors $2(\widehat{y}_t^{\,\mathrm{g}} - y_t^{\mathrm{g}})x_t^i$ are bounded by $4C^2$. Indeed, for any $(\mathrm{g}, i) \in \mathcal{G}^2$, they ensure $|y_t^{\mathrm{g}}| \leqslant C$ and $|x_t^i| \leqslant C$. In addition, as $\mathbf{u}_t^{\mathrm{g}} \in \Delta_{\mathcal{G}}$, the output forecasts $\widehat{y}_t^{\,\mathrm{g}} = \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$ are also bounded by:

$$|\widehat{y}_t^{\,\mathrm{g}}| = \left| \sum_{j \in \mathcal{G}} u_t^{\mathrm{g}\,j} x_t^j \right| \leqslant \sum_{j \in \mathcal{G}} u_t^{\mathrm{g}\,j} |x_t^j| \leqslant \sum_{j \in \mathcal{G}} u_t^{\mathrm{g}\,j} C = C.$$

Hence, for any $i \in \mathcal{G}$, the pseudo prediction error associated with feature i satisfies

$$\big|2\big(\widehat{y}_t^{\,\mathrm{g}} - y_t^{\mathrm{g}}\big)x_t^i\big| \leqslant 4C^2. \tag{6.9}$$

⋆ *Bernstein Online Aggregation.* Wintenberger [2017] introduces an aggregation procedure

Algorithm 6 Fully adaptive Bernstein Online Aggregation (BOA) with gradient trick

aim
Predict the time series $\big(y_t^{\mathrm{g}}\big)_{1 \leqslant t \leqslant T}$
parameter Bound on pseudo prediction errors E:
for any $t = 1, \ldots, T$ and any $i \in \mathcal{G}$, $|2(\widehat{y}_t^{\,\mathrm{g}} - y_t^{\mathrm{g}})x_t^i| \leqslant E$
initialization
$\mathbf{u}_1^{\mathrm{g}} = (1/|\mathcal{G}|, \ldots, 1/|\mathcal{G}|)$
$\widehat{y}_1^{\,\mathrm{g}} = \mathbf{u}_1^{\mathrm{g}} \cdot \mathbf{x}_1$
For $i \in \mathcal{G}$, $\widetilde{R}_0^{\mathrm{g}\,i} = 0$
For $i \in \mathcal{G}$, $\eta_0^{\mathrm{g}\,i} = 0$
for $t = 1, \ldots, T-1$ **do**
For each $i \in \mathcal{G}$, update the cumulative quantity $\widetilde{Q}^{\mathrm{g}\,i}$ for feature i

$$\widetilde{Q}_t^{\mathrm{g}\,i} = \widetilde{Q}_{t-1}^{\mathrm{g}\,i} + \widetilde{r}_t^{\mathrm{g}\,i}\big(1 + \eta_{t-1}^{\mathrm{g}\,i}\widetilde{r}_t^{\mathrm{g}\,i}\big) \quad \text{where} \quad \widetilde{r}_t^{\mathrm{g}\,i} \triangleq 2(\widehat{y}_t^{\,\mathrm{g}} - y_t^{\mathrm{g}})(\widehat{y}_t^{\,\mathrm{g}} - x_t^i)$$

For each $i \in \mathcal{G}$, compute the learning rate

$$\eta_t^{\mathrm{g}\,i} = \min\left\{\frac{1}{2E}, \sqrt{\frac{\log|\mathcal{G}|}{\sum_{s=1}^t \big(\widetilde{r}_t^{\mathrm{g}\,i}\big)^2}}\right\}$$

Compute the weight vector $\mathbf{u}_{t+1}^{\mathrm{g}} = (u_{t+1}^{\mathrm{g}\,i})_{i \in \mathcal{G}}$ defined as

$$u_{t+1}^{\mathrm{g}\,i} = \frac{\exp\big(\eta_t^{\mathrm{g}\,i}\widetilde{Q}_t^{\mathrm{g}\,i}\big)}{\sum_{j \in \mathcal{G}} \exp\big(\eta_t^{\mathrm{g}\,j}\widetilde{Q}_t^{\mathrm{g}\,j}\big)}$$

Output prediction $\widehat{y}_{t+1}^{\,\mathrm{g}} = \mathbf{u}_{t+1}^{\mathrm{g}} \cdot \mathbf{x}_{t+1} = \sum_{i \in \mathcal{G}} u_{t+1}^{\mathrm{g}\,i} x_{t+1}^i$
end for

called Bernstein Online Aggregation for which weights are exponential function of the

cumulative prediction errors. Algorithm 6 describes this strategy combined with this gradient trick. Let us fix a node $\mathrm{g} \in \mathcal{G}$ and set $\mathcal{A}^{\mathrm{g}}$ to Algorithm 6 which takes as input the bound E on pseudo prediction errors ($E = 4C^2$ is a suitable choice):

$$\forall t = 1, \dots, T,\ \forall i \in \mathcal{G}, \quad 2(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}})x_t^i \leqslant E.$$

Theorem 3.4 of Wintenberger, 2017 ensures that

$$\begin{aligned} R_T^{\mathrm{g}}(\Delta_{|\mathcal{G}|}) &\leqslant \sqrt{T+1}E\left(\frac{\sqrt{2\ln|\mathcal{G}|}}{\sqrt{2}-1} + \frac{\ln(1+2^{-1}\ln T)}{\sqrt{\ln|\mathcal{G}|}}\right) + E\Big(2\ln|\mathcal{G}| + 2\ln(1+2^{-1}\ln T) + 1\Big) \\ &\lesssim \mathcal{O}\Big(\sqrt{T}\ln\ln T\Big). \end{aligned} \tag{6.10}$$

Thanks to Equation (6.9), we replace E by $4C^2$ in Equation (6.10) and we get, for each node $\mathrm{g} \in \mathcal{G}$, an upper bound on $R_T^{\mathrm{g}}(\Delta_{|\mathcal{G}|})$. By applying Theorem 9, we obtain the following uniform regret bound:

$$R_T(\Delta_{|\mathcal{G}|}) \lesssim \mathcal{O}\Big(|\mathcal{G}|\sqrt{T}\ln\ln T\Big),$$

which is of order $\sqrt{T}$ (up to poly-logarithmic terms).

⋆ *Polynomially Weighted Average Forecaster.* Gaillard et al. [2014] consider an aggregation method based on weights that are polynomial functions of the cumulative prediction errors. We use this procedure combined with the gradient trick and present it in Algorithm 7. In this description, $(\mathbf{x})_+$ denotes the vector of non-negative parts of the components of $\mathbf{x}$. With the same notation as in the previous paragraph, for any node $\mathrm{g} \in \mathcal{G}$, Theorem 5 of Gaillard et al. [2014] provides the following regret bound:

$$R_T^{\mathrm{g}}(\Delta_{|\mathcal{G}|}) \leqslant E\sqrt{|\mathcal{G}|(T+1)\big(1+\ln(1+T)\big)}. \tag{6.11}$$

With $E \leqslant 4C^2$ and by applying Theorem 9, we obtain an upper bound on the uniform regret $R_T(\Delta_{|\mathcal{G}|})$, which is also of order $\sqrt{T}$ (up to poly-logarithmic terms):

$$\begin{aligned} R_T(\Delta_{|\mathcal{G}|}) &\leqslant 4C^2|\mathcal{G}|\sqrt{|\mathcal{G}|(T+1)\big(1+\ln(1+T)\big)} \\ &\lesssim \mathcal{O}\Big(|\mathcal{G}|^{3/2}\sqrt{T\ln T}\Big). \end{aligned}$$

6.4 A scheme to extend the class of comparison from the simplex to an L_1-ball

For the previous two algorithms, we obtained an upper bound on $R_T^{\mathrm{g}}(\Delta_{|\mathcal{G}|})$. However, there is no reason for the best linear combination of features to be convex. Algorithm 8 presents a trick introduced by Kivinen and Warmuth [1997] which extends the class of comparison from the $|\mathcal{G}|$-simplex to an L_1-ball of radius $\alpha > 0$ denoted by $\mathcal{B}_\alpha$ and provides a bound on $R_T^{\mathrm{g}}(\mathcal{B}_\alpha)$. Let us fix a node $\mathrm{g} \in \mathcal{G}$. The trick consists in transforming, at each round t, the feature vector $\mathbf{x}_t$ into the $2|\mathcal{G}|$-vector $\bar{\mathbf{x}}_t = (\alpha\mathbf{x}_t | -\alpha\mathbf{x}_t)$, where $|$ is the concatenation operator between vectors. The algorithm $\mathcal{A}^{\mathrm{g}}$ is then run with these new features and it outputs the weight vector $\bar{\mathbf{u}}_t^{\mathrm{g}} \in \Delta_{2|\mathcal{G}|}$. Finally, a $|\mathcal{G}|$-vector $\mathbf{u}_t^{\mathrm{g}} \in \mathcal{B}_\alpha$ is computed from $\bar{\mathbf{u}}_t^{\mathrm{g}}$ to provide the forecast $\mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t = \bar{\mathbf{u}}_t^{\mathrm{g}} \cdot \bar{\mathbf{x}}_t$. We will actually see that we may associate any $|\mathcal{G}|$-vector $\mathbf{u} \in \mathcal{B}_\alpha$ with a vector $\bar{\mathbf{u}} \in \Delta_{2|\mathcal{G}|}$ such as $\bar{\mathbf{u}} \cdot \bar{\mathbf{x}}_t = \mathbf{u} \cdot \mathbf{x}_t$; the trick actually defines

Algorithm 7 Polynomially weighted average forecaster with Multiple Learning rates (ML-Pol) and gradient trick

aim
Predict the time series $\left(y_t^{\mathrm{g}}\right)_{1\leqslant t\leqslant T}$
parameter Bound on pseudo prediction errors E:
for any $t=1,\dots,T$ and any $i\in\mathcal{G}$, $|2(\widehat{y}_t^{\mathrm{g}}-y_t^{\mathrm{g}})x_t^i|\leqslant E$
initialization
$\mathbf{u}_1^{\mathrm{g}}=(1/|\mathcal{G}|,\dots,1/|\mathcal{G}|)$
$\widehat{y}_1^{\mathrm{g}}=\mathbf{u}_1^{\mathrm{g}}\cdot\mathbf{x}_1$
For $i\in\mathcal{G}$, $\widetilde{R}_0^{\mathrm{g}\,i}=0$
For $i\in\mathcal{G}$, $\eta_0^{\mathrm{g}\,i}=0$
for $t=1,\dots,T-1$ **do**
For each $i\in\mathcal{G}$, update the cumulative pseudo-regret of feature i

$$\widetilde{R}_t^{\mathrm{g}\,i}=\widetilde{R}_{t-1}^{\mathrm{g}\,i}+\widetilde{r}_t^{\mathrm{g}\,i}\quad\text{where}\quad\widetilde{r}_t^{\mathrm{g}\,i}\triangleq 2(\widehat{y}_t^{\mathrm{g}}-y_t^{\mathrm{g}})(\widehat{y}_t^{\mathrm{g}}-x_t^i)$$

For each $i\in\mathcal{G}$, compute the learning rate

$$\eta_t^{\mathrm{g}\,i}=\left(E+\sum_{s=1}^{t}\left(\widetilde{r}_t^{\mathrm{g}\,i}\right)^2\right)^{-1}$$

Compute the weight vector $\mathbf{u}_{t+1}^{\mathrm{g}}=(u_{t+1}^{\mathrm{g}\,i})_{i\in\mathcal{G}}$ defined as

$$u_{t+1}^{\mathrm{g}\,i}=\frac{\eta_t^{\mathrm{g}\,i}\left(\widetilde{R}_t^{\mathrm{g}\,i}\right)_+}{\sum_{j\in\mathcal{G}}\eta_t^{\mathrm{g}\,j}\left(\widetilde{R}_t^{\mathrm{g}\,j}\right)_+}$$

Output prediction $\widehat{y}_{t+1}^{\mathrm{g}}=\mathbf{u}_{t+1}^{\mathrm{g}}\cdot\mathbf{x}_{t+1}=\sum_{i\in\mathcal{G}}u_{t+1}^{\mathrm{g}\,i}x_{t+1}^i$
end for

a surjection from $\Delta_{2|\mathcal{G}|}$ to $\mathcal{B}_\alpha$. Thus, to compete against the best linear combination of features in $\mathcal{B}_\alpha$, it is enough to compete against the best convex combination of features $\bar{\mathbf{x}}_t$ in a lifted space (which we may achieve, thanks to algorithm $\mathcal{A}^{\mathrm{g}}$). We now give all the details on how this trick works and indicate its impact on the stated regret bounds. The following lemma introduces the surjection from $\Delta_{2|\mathcal{G}|}$ to $\mathcal{B}_\alpha$, which is used in Algorithm 8.

Lemma 10. *For any real $\alpha>0$, the following function ψ is a surjection from $\Delta_{2|\mathcal{G}|}$ to $\mathcal{B}_\alpha$:*

$$\psi:\begin{array}{ccc}\Delta_{2|\mathcal{G}|} & \longrightarrow & \mathcal{B}_\alpha\\ \bar{\mathbf{u}}=\left(\bar{\mathbf{u}}^+\,|\,\bar{\mathbf{u}}^-\right) & \longmapsto & \alpha(\bar{\mathbf{u}}^+-\bar{\mathbf{u}}^-),\end{array}$$

where the vector $\bar{\mathbf{u}}\in\Delta_{2|\mathcal{G}|}$ is decomposed in the two $|\mathcal{G}|$-vectors $\bar{\mathbf{u}}^+$ and $\bar{\mathbf{u}}^-$, which correspond to the $|\mathcal{G}|$ first and the $|\mathcal{G}|$ last coefficients of $\bar{\mathbf{u}}$, respectively.

By running Algorithm $\mathcal{A}^{\mathrm{g}}$ with transformed features $\bar{\mathbf{x}}_t\triangleq\left(\alpha\mathbf{x}_t\,|\,-\alpha\mathbf{x}_t\right)$ and parameter s_0^{g} (which provides weight vectors $\bar{\mathbf{u}}_t^{\mathrm{g}}$), we get the bound

$$R_T^{\mathrm{g}}(\Delta_{2|\mathcal{G}|})\triangleq\sum_{t=1}^{T}\left(y_t^{\mathrm{g}}-\bar{\mathbf{u}}_t^{\mathrm{g}}\cdot\bar{\mathbf{x}}_t\right)^2-\min_{\bar{\mathbf{u}}^{\mathrm{g}}\in\Delta_{2|\mathcal{G}|}}\sum_{t=1}^{T}\left(y_t^{\mathrm{g}}-\bar{\mathbf{u}}^{\mathrm{g}}\cdot\bar{\mathbf{x}}_t\right)^2\leqslant\mathrm{B}(\bar{\mathbf{x}}_{1:T},y_{1:T}^{\mathrm{g}},\mathbf{s}_0^{\mathrm{g}},\bar{\mathbf{u}}^{\mathrm{g}}).$$

For any time step $t = 1, \dots, T$, and for any $\bar{\mathbf{u}}^{\mathrm{g}} \in \Delta_{2|\mathcal{G}|}$, we obtain the equality of the two scalar products $\bar{\mathbf{u}}^{\mathrm{g}} \cdot \bar{\mathbf{x}}_t$ and $\psi(\mathbf{u}^{\mathrm{g}}) \cdot \mathbf{x}_t$:

$$\bar{\mathbf{u}}^{\mathrm{g}} \cdot \bar{\mathbf{x}}_t = \big(\bar{\mathbf{u}}^{\mathrm{g}+} \,|\, \bar{\mathbf{u}}^{\mathrm{g}-}\big) \cdot \big(\alpha \mathbf{x}_t \,|\, -\alpha \mathbf{x}_t\big) = \alpha\big(\bar{\mathbf{u}}^{\mathrm{g}+} - \bar{\mathbf{u}}^{\mathrm{g}-}\big) \cdot \mathbf{x}_t = \psi(\bar{\mathbf{u}}^{\mathrm{g}}) \cdot \mathbf{x}_t \,.$$

Lemma 10 implies that for any $\mathbf{u}^{\mathrm{g}} \in \mathcal{B}_\alpha$, there is at least one vector $\bar{\mathbf{u}}^{\mathrm{g}} \in \Delta_{2|\mathcal{G}|}$ such that $\psi(\bar{\mathbf{u}}^{\mathrm{g}}) = \mathbf{u}^{\mathrm{g}}$ and we get the equality:

$$\min_{\mathbf{u}^{\mathrm{g}} \in \mathcal{B}_\alpha} \sum_{t=1}^{T} \big(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\big)^2 = \min_{\bar{\mathbf{u}}^{\mathrm{g}} \in \Delta_{2|\mathcal{G}|}} \sum_{t=1}^{T} \big(y_t^{\mathrm{g}} - \psi(\bar{\mathbf{u}}^{\mathrm{g}}) \cdot \mathbf{x}_t\big)^2 = \min_{\bar{\mathbf{u}}^{\mathrm{g}} \in \Delta_{2|\mathcal{G}|}} \sum_{t=1}^{T} \big(y_t^{\mathrm{g}} - \bar{\mathbf{u}}^{\mathrm{g}} \cdot \bar{\mathbf{x}}_t\big)^2 .$$

So with, for any time step $t = 1, \dots, T$, $\mathbf{u}_t^{\mathrm{g}} \triangleq \psi(\bar{\mathbf{u}}_t^{\mathrm{g}})$, we obtain

$$R_T^{\mathrm{g}}(\mathcal{B}_\alpha) \triangleq \sum_{t=1}^{T} \big(y_t^{\mathrm{g}} - \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t\big)^2 - \min_{\mathbf{u}^{\mathrm{g}} \in \mathcal{B}_\alpha} \sum_{t=1}^{T} \big(y_t^{\mathrm{g}} - \mathbf{u}^{\mathrm{g}} \cdot \mathbf{x}_t\big)^2 = R_T^{\mathrm{g}}(\Delta_{2|\mathcal{G}|}) \,.$$

This equality provides a bound on $R_T^{\mathrm{g}}(\mathcal{B}_\alpha)$ when predictions are $\widehat{y}_t^{\mathrm{g}} = \psi^{-1}(\bar{\mathbf{u}}_t^{\mathrm{g}}) \cdot \mathbf{x}_t = \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$. With this trick, the previous bounds (6.10) and (6.11) are still true by replacing $|\mathcal{G}|$ (the dimension of the features $\mathbf{x}_t$) by $2|\mathcal{G}|$ (the dimension of the new features $\bar{\mathbf{x}}_t$) and the bound E (previously equals to $4C^2$) by $2\alpha(\alpha+1)C^2$ (the bound on the new pseudo prediction errors are calculated below):

$$\begin{aligned} R_T(\mathcal{B}_\alpha) \leqslant\; & |\mathcal{G}| \Bigg(2\,\alpha\,(\alpha+1)\,C^2\,\sqrt{T+1}\bigg(\frac{\sqrt{2\ln|\mathcal{G}|}}{\sqrt{2}-1} + \frac{\ln(1+2^{-1}\ln T)}{\sqrt{\ln|\mathcal{G}|}}\bigg) \\ & \qquad + 2\,\alpha\,(\alpha+1)\,C^2\,\big(2\ln|\mathcal{G}| + 2\ln(1+2^{-1}\ln T) + 1\big)\Bigg) \qquad \text{for BOA} \\ \leqslant\; & 2\,\alpha\,(\alpha+1)\,C^2\,|\mathcal{G}|\sqrt{|\mathcal{G}|(T+1)\big(1+\ln(1+T)\big)} \qquad \text{for ML-Poly.} \end{aligned}$$

The complete online algorithm leading to these bounds is summarized in Algorithm 8.

⋆ *Bound on new pseudo prediction errors.* Since Assumption 8 holds, the transformed features $\bar{\mathbf{x}}_t^{\mathrm{g}}$ are bounded by αC. Moreover, $\bar{\mathbf{u}}_t^{\mathrm{g}} \in \Delta_{2|\mathcal{G}|}$ implies $\big\|\bar{\mathbf{u}}_t^{\mathrm{g}}\big\|_1 = 1$, so we get

$$\big|\widehat{y}_t^{\mathrm{g}}\big| = \big|\bar{\mathbf{u}}_t^{\mathrm{g}} \cdot \bar{\mathbf{x}}_t\big| \leqslant \big\|\bar{\mathbf{u}}_t^{\mathrm{g}}\big\|_1 \big\|\bar{\mathbf{x}}_t\big\|_\infty = \alpha C \,.$$

Moreover, as the observations are still bounded by C, we have $|y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}}| \leqslant |y_t^{\mathrm{g}}| + |\widehat{y}_t^{\mathrm{g}}| \leqslant \big(\alpha+1\big)C$ and we obtain a bound on the pseudo prediction errors:

$$\big|\widetilde{\ell}_t^{\mathrm{g}}(\bar{\mathbf{x}}_t)\big| = \big\|2\big(\widehat{y}_t^{\mathrm{g}} - y_t^{\mathrm{g}}\big)\bar{\mathbf{x}}_t\big\|_\infty \leqslant 2\alpha(1+\alpha)C^2 \,.$$

Proof of Lemma 10. Denoting respectively by $(\mathbf{u})_+$ and $(\mathbf{u})_-$ the non-negative and non-positive parts of any vector $\mathbf{u}$ and by $\mathbf{1}_{|\mathcal{G}|}$ the vector of size $|\mathcal{G}|$ of which all coordinates are 1, we introduce the inverse function ψ^{-1}:

$$\psi^{-1} : \begin{array}{ccc} \mathcal{B}_\alpha & \longrightarrow & \Delta_{2|\mathcal{G}|} \\ \mathbf{u} & \longmapsto & \frac{1}{\alpha}\left(\frac{\alpha - \|\mathbf{u}\|_1}{2|\mathcal{G}|}\mathbf{1}_{|\mathcal{G}|} + (\mathbf{u})_+ \,\middle|\, \frac{\alpha - \|\mathbf{u}\|_1}{2|\mathcal{G}|}\mathbf{1}_{|\mathcal{G}|} + (\mathbf{u})_- \right). \end{array}$$

Algorithm 8 Scheme for on-line linear regression.

input Algorithm $\mathcal{A}^{\mathrm{g}}$ and bound on the weight vectors $\alpha > 0$
for $t = 1, \ldots, T$ **do**
Get the feature vector $\mathbf{x}_t$ and denote

$$\bar{\mathbf{x}}_t \triangleq \left(\alpha \mathbf{x}_t \mid -\alpha \mathbf{x}_t\right) \in \mathbb{R}^{2|\mathcal{G}|}$$

Run algorithm $\mathcal{A}^{\mathrm{g}}$ on node g with $\bar{\mathbf{x}}_t$ and get the weight vector $\bar{\mathbf{u}}_t^{\mathrm{g}} = \left(\bar{\mathbf{u}}_t^{\mathrm{g}+} \mid \bar{\mathbf{u}}_t^{\mathrm{g}-}\right)$
Output the weight vector $\mathbf{u}_t^{\mathrm{g}} = \alpha\left(\bar{\mathbf{u}}_t^{\mathrm{g}+} - \bar{\mathbf{u}}_t^{\mathrm{g}-}\right)$ and predicts $\widehat{y}_t^{\mathrm{g}} = \mathbf{u}_t^{\mathrm{g}} \cdot \mathbf{x}_t$
end for

First we will show that function images are in the right sets, meaning that for any $\mathbf{u} \in \mathcal{B}_\alpha$, $\psi^{-1}(\mathbf{u}) \in \Delta_{2|\mathcal{G}|}$ and for any $\bar{\mathbf{u}} \in \Delta_{2|\mathcal{G}|}$, $\psi(\bar{\mathbf{u}}) \in \mathcal{B}_\alpha$. Secondly, we obtain the surjectivity of ψ by proving that for any $\mathbf{u} \in \mathcal{B}_\alpha, \psi(\psi^{-1}(\mathbf{u})) = \mathbf{u}$.

⋆ *Proof that for any* $\mathbf{u} \in \mathcal{B}_\alpha$, $\psi^{-1}(\mathbf{u}) \in \Delta_{2|\mathcal{G}|}$. We set $\mathbf{u} \in \mathcal{B}_\alpha$. By definition for any $i \in \mathcal{G}$, $(u^i)_\pm \geqslant 0$ and as $\mathbf{u} \in \mathcal{B}_\alpha$, $(\alpha - \|\mathbf{u}\|_1)/(2|\mathcal{G}|) \geqslant 0$. So, all the coefficients of $\psi^{-1}(\mathbf{u})$ are non-negative. Since $\sum_{i\in\mathcal{G}} (u^i)_+ + (u^i)_- = \sum_{i\in\mathcal{G}} |u^i| = \|\mathbf{u}\|_1$, the sum of the coefficients of the vector $\psi^{-1}(\mathbf{u})$ equals 1:

$$\sum_{i\in\mathcal{G}} \left(\psi^{-1}(\mathbf{u})\right)^{i+} + \left(\psi^{-1}(\mathbf{u})\right)^{i-} = \frac{1}{\alpha} \sum_{i\in\mathcal{G}} \left(\left(u^i\right)_+ + \left(u^i\right)_- + \frac{\alpha - \|\mathbf{u}\|_1}{|\mathcal{G}|} \right) = \frac{1}{\alpha}\left(\|\mathbf{u}\|_1 + \alpha - \|\mathbf{u}\|_1\right) = 1.$$

and thus $\bar{\mathbf{u}} = \psi^{-1}(\mathbf{u}) \in \Delta_{2|\mathcal{G}|}$.

⋆ *Proof that for any* $\bar{\mathbf{u}} \in \Delta_{2|\mathcal{G}|}$, $\psi(\bar{\mathbf{u}}) \in \mathcal{B}_\alpha$. With $\bar{\mathbf{u}} = \left(\bar{\mathbf{u}}^+ \mid \bar{\mathbf{u}}^-\right) \in \Delta_{2|\mathcal{G}|}$, using that all the coefficients of $\bar{\mathbf{u}}$ are non-negative and that their sum equals 1 that is $\left\|\bar{\mathbf{u}}\right\|_1 = 1$, we get

$$\|\psi(\bar{\mathbf{u}})\|_1 \triangleq \left\|\alpha\bar{\mathbf{u}}^+ - \alpha\bar{\mathbf{u}}^-\right\|_1 \leqslant \alpha\left\|\bar{\mathbf{u}}^+\right\|_1 + \alpha\left\|\bar{\mathbf{u}}^-\right\|_1 = \alpha\left\|\bar{\mathbf{u}}\right\|_1 = \alpha.$$

⋆ *Proof that for any* $\mathbf{u} \in \mathcal{B}_\alpha$, $\psi\left(\psi^{-1}(\mathbf{u})\right) = \mathbf{u}$.

$$\psi\left(\psi^{-1}(\mathbf{u})\right) = \frac{\alpha - \|\mathbf{u}\|_1}{2|\mathcal{G}|}\mathbf{1}_{|\mathcal{G}|} + (\mathbf{u})_+ - \frac{\alpha - \|\mathbf{u}\|_1}{2|\mathcal{G}|}\mathbf{1}_{|\mathcal{G}|} - (\mathbf{u})_- = \mathbf{u}\,.$$

□

7 Experiments

Our application relies on electricity consumption data of a large number of households to which we have added meteorological data (see Subsection 7.1). Non-temporal information (sociological type, region, type of heating fuel and type of electricity contract) on the households is also provided. From these temporal and non-temporal data, we dispatch the households into clusters thanks to the methods presented in Subsection 7.2. We describe the experiments and analyze the results in Subsections 7.3 and 7.4.

7.1 The underlying real data set

The project "*Energy Demand Research Project*[1]", managed by Ofgem on behalf of the UK Government, was launched in late 2007 across Great Britain (see Raw and Ross, 2011 and Schellong, 2011). Power consumptions of approximately 18,000 households with smart-type meters were collected at half-hourly intervals for about two years. We detail below how we select only the consumption of 1,545 households over the period from April 20, 2009 to July 31, 2010 – Ben Taieb et al. [2017], who used the same data, performed similiar pre-processing in their experiments. Four non-temporal variables are associated with each household: the Region (the initial data set provides the level-4 NUTS[2] codes but we consider larger subdivisions – from 150,000 to 800,000 inhabitants – and associate each household with its level-3 code), the Acorn category value (an integer between 1 and 6 associated with an United Kingdom's population demographic type – this segmentation was developed by the company CACI Limited), the type of heating fuel ("electricity" or "electricity and gas") and the contract type ("Standard" or "Time of Use tariff" for households containing an electricity meter with a dynamic time of use tariff) for each household. In a first data cleaning step, we removed households with more than 5 missing consumption records over the period April 20, 2009 to July 31, 2010 (around $1,600$ households are thus kept) – the remaining missing consumption data points are imputed by a linear interpolation. Among the various clusterings of the households we consider in our experiments, three of them rely on three qualitative variables: "Region", "Tariff" and "Fuel + Tariff" (which is based on both the heating fuel type and the contract type). If one of the values of these qualitative variables had fewer than 20 occurrences, we have removed from the data set the households associated with that value. The final data set then contains the electrical consumption records of the 1,545 remaining households. From now on, we will denote by $\mathcal{I}$ the set of households and by $(y_{i,t})_{1-T_0 \leqslant t \leqslant T}$ the time series of the half-hourly power consumption of the $i \in \mathcal{I}$ household. Finally, we added the temperature, visibility and humidity for each region from the NOAA[3] data: we selected a weather station (with records available over the considered period) in each region and linearly interpolated the meteorological data to get H measurements per day (compared to 8 initially). Table 6.1 sums up the available variables of our data set and gives their range.

7.2 Clustering of the households

We present, in Paragraphs 7.2.1 to 7.2.3, three methods to cluster the households and we compare them in the last paragraph of this subsection. After choosing a segmentation (or two crossed segmentations), we only consider, for each cluster, the aggregated consumption of its households. Thus, for any subset $\mathrm{g} \subset \mathcal{I}$, we compute the time series $y_t^{\mathrm{g}} \triangleq \sum_{i \in \mathrm{g}} y_{i\,t}$ that we want to forecast and once clusterings are chosen, we never consider individual power consumption.

[1] https://www.ofgem.gov.uk/gas/retail-market/metering/transition-smart-meters/energy-demand-research-project

[2] *Nomenclature des Unités Territoriales Statistiques* (nomenclature of territorial units for statistics)

[3] National Oceanic and Atmospheric Administration, https://www.noaa.gov/

Variable	Description	Range / Value
Acorn	Acorn category value	From 1 to 6
Region	UK NUTS of level 3	UK- H23, -J33, -L15, -L16, -L21, -M21, or -M27
Fuel	Type of heating fuel	Electricity (E) or Electricity and Gas (EG)
Tariff	Contract type	Standard (Std) or Time of Use tariff (ToU)
Temperature	Air temperature	From -20° to 30°
Visibility	Air visibility	From 0 to 10 (integer)
Humidity	Air humidity percentage	From 0% to 100%
Date	Current time	From April 20, 2009 to July 31, 2010 (half-hourly)
Consumption	Power consumption	From 0.001 to 900 kWh
Fuel + Tariff	Cross of Fuel and Tariff variables	"E - Std", "EG - Std", "E - ToU" or "EG - ToU"
Half-hour	Half-hour of the day	From 1 to H (integer)
Day	Day of the week	From 1 (Monday) to 7 (Sunday) (integer)
Position in the year	Linear values	From 0 (Jan 1, 00:00) to 1 (Dec 31, 23:59)
Smoothed temperature	Smoothed air temperature	From -20° to 30°

Table 6.1 – Summary of the variables provided and created for each household of the data set.

7.2.1 Random clustering

We first consider the simplest way to cluster households: the segmentation is built randomly. In the experiments of Subsection 7.4, the number of clusters varies from 4 to 64. As an example here, we consider 4 clusters and we randomly assign a number between 1 and 4 to each household and obtain the weekly profiles plotted in Figure 6.5. In the following, we will call "Random (G)", a segmentation of G clusters built randomly. Naturally, the curves are almost identical and the clusters are therefore rather similar.

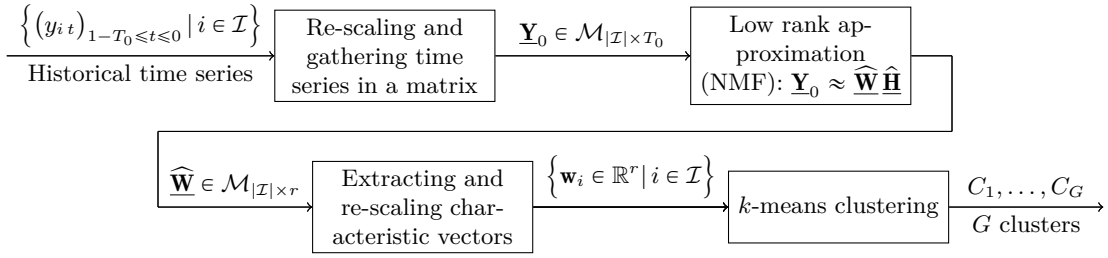
7.2.2 Segmentation based on qualitative household variables

The second approach consists in grouping households according to the provided non-temporal information. We consider the natural segmentations "Region", "Acorn" and "Fuel + Tariff" based on the corresponding qualitative variables and we plot the weekly profile of each cluster on Figures 6.6, 6.7 and 6.8. Regions have an impact on the consumption profile: the evening consumption peak time varies by location. Moreover, consumption of the Wales regions (UKL15, UKL16 and UKL21) is lower than that of the other regions (see Figure 6.6). In the Acorn classification, the lower the value, the richer the household, thus Figure 6.7 shows that wealthiest households consume the most (as expected). Finally, the type of heating fuel does not seem to have a significant impact on the weekly consumption profile (although we have observed that when the heating is partly gas, the consumption is slightly lower and in winter, it is less sensitive to the temperature drops). Similarly, it seems that the type of contract does not influence the consumption profiles. Peak consumption in the evening is however less important for a dynamic time of fuse tariff than for the standard tariff. It should be noted that since time slots of prices may change from day to day, it is difficult to quantify here the impact of the tariff, as we are only showing average consumption profiles.

7.2.3 Clustering based on non-negative matrix factorization and k-means method

The last method relies on an historical individual time series of household power consumption (April 20, 2009 to April 20, 2010). We propose a method to extract from these time

series a low number – denoted by r – of combined household characteristics and to use them to build relevant clusterings. The diagram below sums up the steps of the procedure described here quickly. We then further detail them one by one. The $|\mathcal{I}|$ historical times series $\big(y_{i\,t}\big)_{1-T_0\leqslant t\leqslant 0}$ are firstly re-scaled and gathered into a matrix $\underline{\mathbf{Y}}_0 \in \mathcal{M}_{|\mathcal{I}|\times T_0}$. We then reduce the dimension of data with a non-negative matrix factorization (NMF): we approximate $\underline{\mathbf{Y}}_0$ by $\widehat{\underline{\mathbf{W}}}\,\hat{\underline{\mathbf{H}}}$, where $\widehat{\underline{\mathbf{W}}}$ and $\hat{\underline{\mathbf{H}}}$ are $|\mathcal{I}| \times r$ and $r \times T_0$-non-negative matrices, respectively. As soon as this approximation is good enough, line i of the matrix $\widehat{\underline{\mathbf{W}}}$ is sufficient to reconstruct the historical time series of household i (with the knowledge of matrix $\hat{\underline{\mathbf{H}}}$ - which is not used for the clustering). Thus, we assign, to each household, r characteristics: the lines of $\widehat{\underline{\mathbf{W}}}$. After a re-scaling step – to give the same importance to each of those characteristics – we get the r-vectors $(\mathbf{w}_i)_{i\in\mathcal{I}}$. With this low-dimension representation of households in $\mathbb{R}^r$, we use k-means clustering algorithm in $\mathbb{R}^r$ to provide the G clusters $C_1, \ldots, C_G$ and we write "NMF (G)" for such a clustering.



⋆ *Re-scaling and Gathering Time Series in a Matrix.* For $T_0 > 0$, we consider the $|\mathcal{I}| \times T_0$ - matrix $\underline{\mathbf{Y}}_0$ which contains the re-scaled historical power consumption time series: for any $i \in \mathcal{I}$ and any $1 - T_0 \leqslant t \leqslant 0$,

$$(\underline{\mathbf{Y}}_0)_{i,t} \triangleq \frac{y_{i,t}}{\bar{y}_i}, \quad \text{with} \quad \bar{y}_i \triangleq \frac{1}{T_0}\sum_{t=1-T_0}^{0} y_{i,t}.$$

⋆ *Low Rank Approximation.* Since we are interested in power consumption, all the coefficients of $\underline{\mathbf{Y}}_0$ are non-negative - we will write $\underline{\mathbf{Y}}_0 \geqslant 0$ and say that this matrix is non-negative. To reduce dimension of non-negative matrices, Paatero and Tapper [1994] and Lee and Seung [1999] propose a factorization method whose distinguishing feature is the use of non-negativity constraints. Let us fix some integer $r \ll \min(|\mathcal{I}|, T_0)$, which will ensure a reduction of the dimension (we chose $r = 10$ in the experiments of the next subsection). The non-negative matrix factorization (NMF) approximates matrix $\underline{\mathbf{Y}}_0$ by $\underline{\mathbf{Y}}_0 \approx \underline{\mathbf{W}}^\star\underline{\mathbf{H}}^\star$, where $\underline{\mathbf{W}}^\star$ and $\underline{\mathbf{H}}^\star$ are $|\mathcal{I}| \times r$ and $r \times T_0$ non-negative matrices. They are computed by solving:

$$\big(\underline{\mathbf{W}}^\star,\, \underline{\mathbf{H}}^\star\big) \in \underset{\underline{\mathbf{W}},\,\underline{\mathbf{H}}\geqslant 0}{\operatorname{argmin}} \big\|\underline{\mathbf{Y}}_0 - \underline{\mathbf{W}}\underline{\mathbf{H}}\big\|_F^2 = \underset{\underline{\mathbf{W}},\,\underline{\mathbf{H}}\geqslant 0}{\operatorname{argmin}} \sum_{i,t}\Big(y_{i,t} - \big(\underline{\mathbf{W}}\underline{\mathbf{H}}\big)_{i,t}\Big)^2.$$

We use the function `NMF` of the `Python`-library `sklearn.decomposition` to approach a local minimum with a coordinate descent solver and denote by $\widehat{\underline{\mathbf{W}}}$ the approximation of $\underline{\mathbf{W}}^\star$. Thanks to the NMF, for any $i \in \mathcal{I}$, r characteristics (the i^{th} line of matrix $\widehat{\underline{\mathbf{W}}}$) are thus computed.

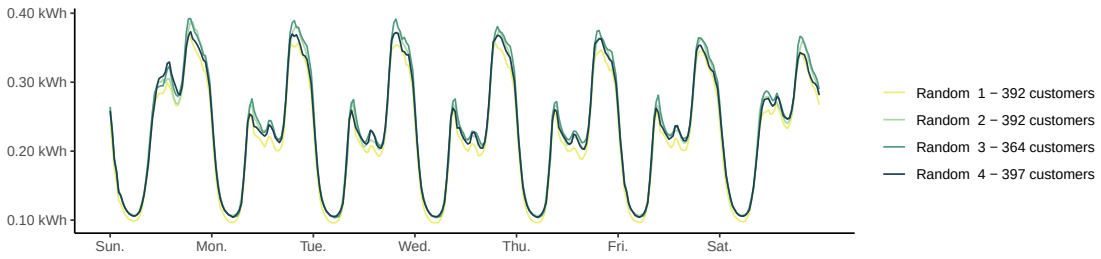


Figure 6.5 – Mean consumption per week and per cluster, with households randomly assigned to an integer from 1 to 4.

⋆ *Extracting and Re-scaling Characteristic Vectors.* To give the same impact to each of these characteristics, we re-scale the columns of $\underline{\widehat{\mathbf{W}}}$ and define, for each household i, the vector

$$\mathbf{w}_i = \left(\frac{\underline{\widehat{\mathbf{W}}}_{i\,1}}{\sum_{j\in\mathcal{I}} \underline{\widehat{\mathbf{W}}}_{j\,1}}, \ldots, \frac{\underline{\widehat{\mathbf{W}}}_{i\,r}}{\sum_{j\in\mathcal{I}} \underline{\widehat{\mathbf{W}}}_{j\,r}} \right).$$

⋆ *k-Means Clustering.* The k-means algorithm (introduced by MacQueen et al. [1967]) is then used on these r- vectors to cluster the households into a fixed number G of groups (which varies from 4 to 64 in our experiments). We recall below how this algorithm works. With $\{C_1, \ldots, C_G\}$ a G-clustering of set $\mathcal{I}$, for any $1 \leqslant \ell \leqslant G$, we define the center $\bar{\mathbf{w}}_\ell$ and the variance $\mathrm{Var}(C_\ell)$ of cluster C_ℓ by

$$\bar{\mathbf{w}}_\ell \triangleq \frac{1}{|C_\ell|} \sum_{i\in C_\ell} \mathbf{w}_i \quad \text{and} \quad \mathrm{Var}(C_\ell) \triangleq \frac{1}{C_\ell} \sum_{i\in C_\ell} \|\mathbf{w}_i - \bar{\mathbf{w}}_\ell\|^2.$$

In k-means clustering, each household belongs to the cluster with the nearest center. The best set of clusters, denoted by $\{C_1^\star, \ldots, C_G^\star\}$ – namely the best set of centers – is obtained by minimizing the following criterion:

$$\{C_1^\star, \ldots, C_G^\star\} \in \underset{\{C_1,\ldots,C_G\}}{\mathrm{argmin}} \sum_{\ell=1}^{G} \sum_{\mathbf{w}\in C_\ell} \left\|\mathbf{w} - \bar{\mathbf{w}}_\ell\right\|^2 = \underset{\{C_1,\ldots,C_G\}}{\mathrm{argmin}} \sum_{\ell=1}^{G} |C_\ell| \mathrm{Var}(C_\ell).$$

In practice, we use `KMeans` function of the `Python`-library `sklearn.cluster` to compute clusters.

⋆ *Description and Analysis of "NMF (4)".* For $G = 4$, weekly profiles are plotted in Figure 6.9. This clustering seems to detect consumption behaviors much more specific than any of the previous ones. Indeed, Clusters 3 and 4 present a peak of consumption early in the morning on working days, while the consumption of Cluster 2– which includes the largest number of households – remains almost flat throughout the morning. Moreover, the evening peak for Cluster 4 arrives earlier than for the other clusters. Finally, the consumption of Cluster 1 is generally the highest, while that of Cluster 2 is the lowest.

7.2.4 Comparison of clusterings

To measure similarity between the clusterings above, we calculate the adjusted rand index (ARI) – see Rand [1971] – for each segmentation pair and report the values thus obtained

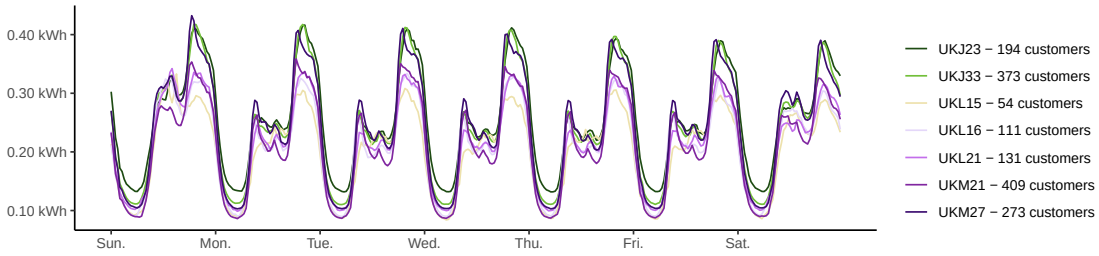


Figure 6.6 – Mean consumption per week and per region (UK NUTS of level 3).

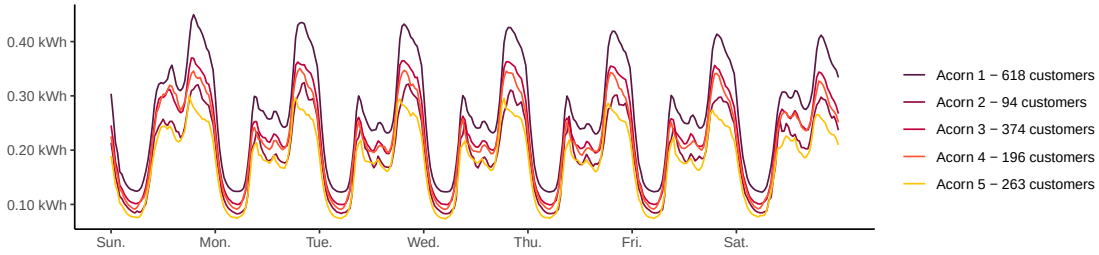


Figure 6.7 – Mean consumption per week and per Acorn category value (from 1 to 5).

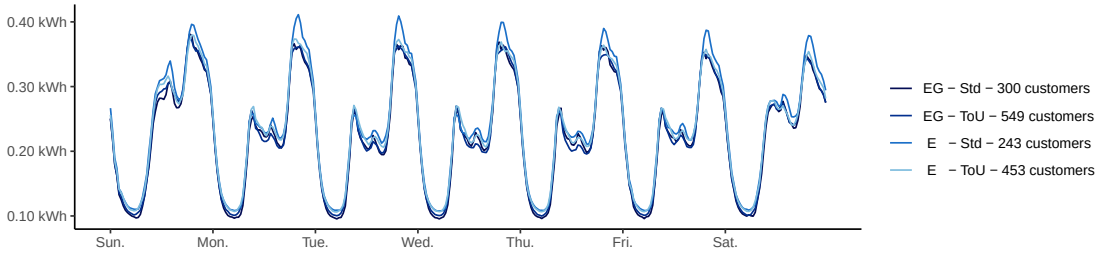


Figure 6.8 – Mean consumption per week for the households clustered according "Fuel + Tariff".

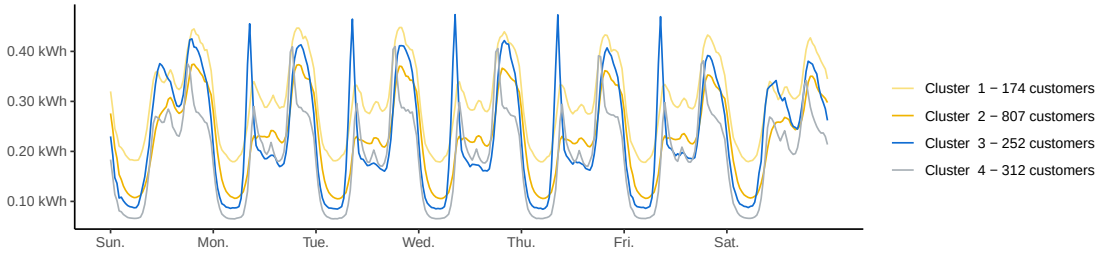


Figure 6.9 – Mean consumption per week and per cluster, with each household assigned to one of the four groups according to the NMF and k-means procedure ("NMF (4)" clustering).

in Table 6.2. We denote by $\binom{n}{k}$ = the binomial coefficient $n!/k!(n-k)!$. Given a set elements $\mathcal{I}$ and two partitions to compare, for example the segmentation "Region" $\{R_1, \dots R_N\}$ and

	Region	NMF (4)	Acorn	Fuel + Tariff
Random (4)	-0.000	0.000	0.003	-0.000
Fuel + Tariff	0.016	-0.001	0.004	
Acorn	0.043	0.018		
NMF (4)	0.011			

Table 6.2 – ARI (Adjusted Rand index) for each segmentation pair.

another clustering $\{C_1, \dots C_G\}$, the ARI is defined by

$$ARI \triangleq \frac{\sum_{\ell=1}^{G}\sum_{n=1}^{N}\binom{|C_\ell \cap R_n|}{2} - \left[\sum_{\ell=1}^{G}\binom{|C_\ell|}{2}\sum_{n=1}^{N}\binom{|R_n|}{2}\right]/\binom{|\mathcal{I}|}{2}}{\frac{1}{2}\left[\sum_{\ell=1}^{G}\binom{|C_\ell|}{2} + \sum_{n=1}^{N}\binom{|R_n|}{2} - \sum_{\ell=1}^{G}\binom{|C_\ell|}{2}\sum_{n=1}^{N}\binom{|R_n|}{2}\right]/\binom{|\mathcal{I}|}{2}}.$$

ARI lies in $[-1, 1]$ by construction, it is equal to 0 for a random matching between clusters of the two considered segmentations and to 1 for a perfect alignment. Similarity between our different household partitions is very low, only "Region" is slightly correlated with all other clusterings, and "NMF (4)" with "ACORN". But these correlations remain low and the clustering "NMF (4)" therefore seems to extract, from historical time series, some households information that are not contained in other clusterings. Its use should improve forecasts – this will be confirmed by the experiments below.

7.3 Experiment design

Thanks to the above methods, we established several partitions of the household set $\mathcal{I}$. As explained below, choosing one or two of them amounts to considering a two-level hierarchy (Example 6) or two crossed hierarchies (Example 9). We also detail the corresponding set of node $\mathcal{G}$. We then describe how we build meteorological data for each node $\mathrm{g} \in \mathcal{G}$ and generate corresponding features. Finally, we focus on standardization and online calibration of aggregation hyper-parameters. We have divided the data set into training data: one-year of historical data (from April 20, 2009 to April 19, 2010) – used for NMF clusterings, feature generation method training, and standardization – and testing data. As aggregation algorithms start from scratch, they work poorly during the first rounds. We therefore withdraw the first 10 days of testing data from the performance evaluation period. So, April 20, 2010 to April 30, 2010 is left for initializing aggregation algorithms and the hyper-parameters calibration and our methods are then tested during the last three months (from May 1, 2010 to July 31, 2010). We summarize in Table 6.3 the range of dates for each step of the procedure.

⋆ *Underlying Hierarchy.* As detailed in Section 2, we aim to forecast a set of power consumption time series $\left\{(y_t^{\mathrm{g}})_{t>0}, \mathrm{g} \in \mathcal{G}\right\}$ connected to each other by some summation constraints. These constraints are represented by one (or more) tree(s) and $\mathcal{G}$ denotes the set of its (or their) nodes. We refer to Example 6 if we consider a single segmentation and to Example 9 for two crossed clusterings. We detail below the set $\mathcal{G}$, which will contain some subsets of households set $\mathcal{I}$, for these two configurations. We recall that we denote the average power consumption of a group of households $\mathrm{g} \subset \mathcal{I}$ by $y_t^{\mathrm{g}} \triangleq \sum_{i\in\mathrm{g}} y_{i,t}$.

	Start date	End date
NMF Clusterings		
Feature Generation Model Training	April 20, 2009	April 19, 2010
Features and Observations Standardization		
Initialization of the Aggregation	April 20, 2010	April 30, 2010
Model Evaluation	May 1, 2010	July 31, 2010

Table 6.3 – Date range for the steps of the proposed method

Considering a single clustering $(C_1, \dots C_N)$ of $\mathcal{I}$, we want to forecast the consumption of each cluster C_ℓ, and also the global consumption (namely, the one for $\mathrm{g} = \mathcal{I}$). Thus, we set $\mathcal{G} = \{C_\ell\}_{1 \leqslant \ell \leqslant N} \cup \{\mathcal{I}\}$ and the associated time series respect the hierarchy of Figure 6.1 – where y^{Tot} refers to the time series associated with $\mathcal{I}$ and $y^1, y^2, \dots, y^N$ with the ones of clusters $C_1, C_2, \dots C_N$. We now consider two partitions. The first one $R_1, \dots R_N$ refers to segmentation "Region" and the second one, $C_1, \dots C_G$ to any other clustering. We would like to forecast the global consumption ($\mathrm{g} = \mathcal{I}$), the consumption associated with each region ($\mathrm{g} = R_n$, for $n = 1, \dots, N$) and with each cluster ($\mathrm{g} = C_\ell$, for $\ell = 1, \dots, G$) but also the power consumption of cluster C_1 in region R_1 ($\mathrm{g} = C_1 \cap R_1$), of cluster C_1 in region R_2 ($\mathrm{g} = C_1 \cap R_2$), and so on. Thus,we consider the set of nodes

$$\mathcal{G} = \{C_\ell \cap R_n\}_{1 \leqslant \ell \leqslant G,\, 1 \leqslant n \leqslant N} \cup \{C_\ell\}_{1 \leqslant \ell \leqslant G} \cup \{R_n\}_{1 \leqslant n \leqslant N} \cup \{\mathcal{I}\}.$$

The hierarchy associated with such crossed segmentations is represented in Figure 6.4 (with $N_1 = G$ and $N_2 = N$) – where the global consumption, associated with $\mathcal{I}$, is denoted by y^{Tot}, the one of cluster C_ℓ by $y^{\ell \cdot}$, the one of region R_n, by $y^{\cdot n}$ and where $y^{\ell n}$ refers to the local consumption of $C_\ell \cap R_n$.

⋆ *Meteorological Data of any Set of Households.* Methods presented in Section 5 for feature creation implicitly assume that meteorological data are available. We recall that we collected meteorological data for each of the N regions. Thus when $\mathrm{g} \in \mathcal{G}$ refers to one of these regions, we can directly apply the feature generation methods. However, if node g groups households from different regions, these data are not directly available and one may even wonder what they should correspond to. We take convex combinations of regional meteorological data, in proportions corresponding to the locations of the households. More precisely, for each meteorological variable (temperature, visibility or humidity), we built the meteorological variable of g as a convex combination of the N meteorological variables of the N regions. The weight associated with region n corresponds to the proportion of this region in g, in terms of contribution to the consumption – this contribution is determined from historical data.

⋆ *Feature Creation.* For each node g, we now have access to calendar and meteorological data. Considering an exponential smoothed temperature – that models the thermal inertia of buildings – is likely to improve forecasts (see among others, Taylor, 2003 and Goude et al., 2014), so we create the a-exponential smoothing of the temperature $\bar{\tau}_t^{\mathrm{g}} \triangleq a\bar{\tau}_{t-1}^{\mathrm{g}} + (1-a)\tau_t^{\mathrm{g}}$, where $a \in [0, 1]$. As in Chapter 3, to tune a, we performed an exhaustive grid search (by testing many values on the prediction models described in Section 4) and set $a = 0.999$. We then apply methods of Section 5 using available explanatory variables to generate features $\mathbf{x}_t$. Each model (auto-regressive model, generalized additive model or random forest) is trained on a year of historical data (from April

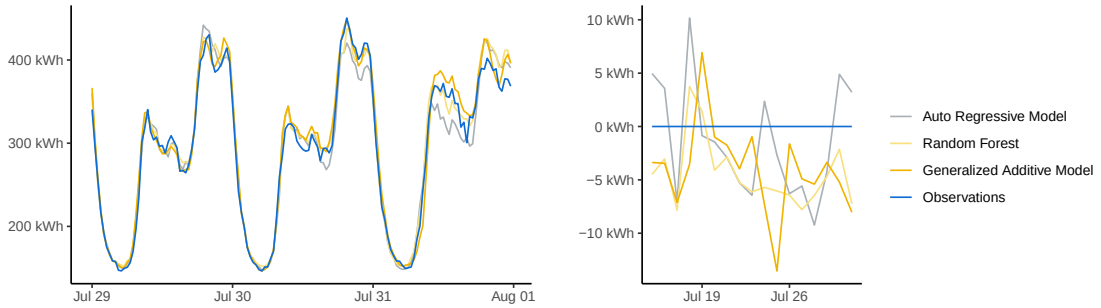


Figure 6.10 – Left picture: benchmark forecasts (auto-regressive model, generalized additive model, random forest) and observations of global consumption ($g = \mathcal{I}$) at half-hour intervals on the last three days of the test period. Right picture: corresponding daily average signed errors on the last week of the test period.

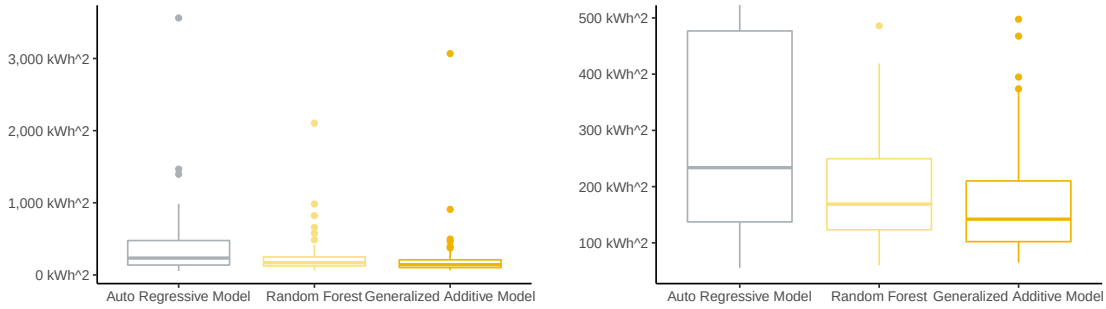


Figure 6.11 – Distribution over the test period of daily mean squared error of global consumption benchmark forecasts (auto-regressive model, generalized additive model, random forest). Left picture: original boxplots; Right picture: boxplots trimmed at 500 kWh^2.

20, 2009 to April 20, 2010). Then, forecasts are computed on the period April 20, 2010 to July 31, 2010. On the left of Figure 6.10, we represent these benchmark predictions and the observations for the global consumption (namely $g = \mathcal{I}$) over the last three days of the test period. On the right, we plot daily signed errors, $\frac{1}{H}\sum_{s=t}^{t+H}\left(y_s^g - x_s^g\right)$, for $g = \mathcal{I}$ over the last week of the test period. Finally, daily mean squared errors, $\frac{1}{H}\sum_{s=t}^{t+H}\left(y_s^g - x_s^g\right)^2$, are computed for each test period day and represented by box-plots on Figure 6.11. The generalized additive model seems to perform the best (and the auto-regressive model the worst), this will be confirmed by the numerical results of the next subsection.

⋆ *Observations and Features Standardization.* Once above features computed, they are standardized using the protocol presented in Subsection 6.1. We assess the quality of the standardization for one given configuration, namely "Region + NMF (16)", with features generated by the general additive model (this configuration, which refers to the two crossed clusterings "Region" and "NMF (16)", reaches the lower predictions errors – see Table 6.8). As there are 7 regions, the set $\mathcal{G}$ consists of $16 \times 7 + 16 + 7 + 1 = 136$ nodes, but only 129 are non-empty. For both standardized and non-standardized observations and features, we compute, for each node $g \in \mathcal{G}$, the empirical mean and empirical standard deviation over the test period. The distributions are plotted in Figures 6.12 and 6.13, respectively. Since the abscissa for non-standardized data is in logarithmic scale, the mean and standard

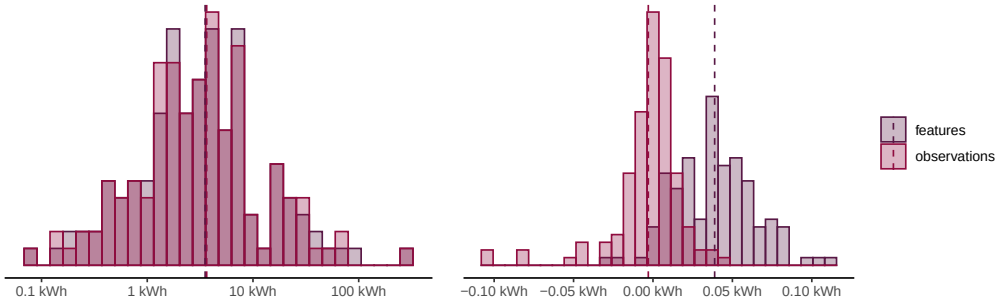


Figure 6.12 – Distribution of empirical means per cluster, for non-standardized and standardized observations and features.

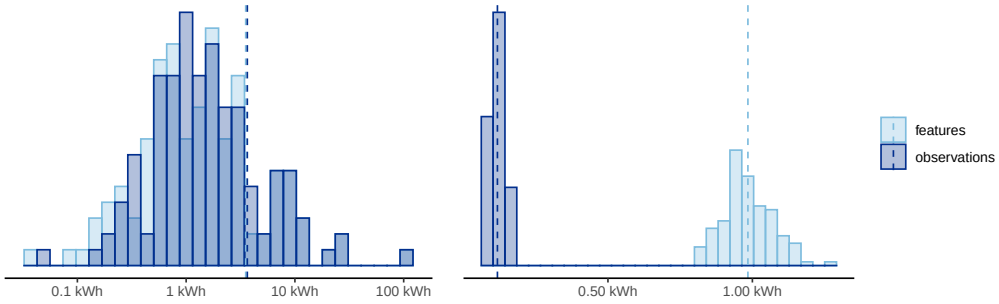


Figure 6.13 – Distribution of empirical standard deviations per cluster, for non-standardized (left) and standardized (right) observations and features.

deviation of data differ a lot from a node to another. For example, the right-hand point is the global consumption ($\mathrm{g} = \mathcal{I}$), while points on the left correspond to the consumptions of small clusters. Thus, standardization centers data and decreases standard deviations of observations, as desired. In addition, standard deviations of features are close to 1. Figure 6.14 represents correlation matrices of the $|\mathcal{G}|$-vectors $(\mathbf{x}_t)_{1\leqslant t\leqslant T}$ and $(\breve{\mathbf{x}}_t)_{1\leqslant t\leqslant T}$, that contain the non-standardized and standardized features over the test period. This shows that our standardization process is centering, re-scaling and de-correlating features. Finally, Table 6.4 gathers numerical values of the average, over $\mathrm{g} \in \mathcal{G}$, of empirical means and standard deviations (these values are indicated by dashed vertical lines on Figures 6.12 and 6.13). We also compute the maximum of the absolute value of features and observations – "Bound" column of the table. This gives an empirical approximation of the boundedness constant C – see Assumption 8.

	Mean	Bound	Standard deviation
Observations	9.53	570.02	3.65
Features	9.54	570.87	3.53
Standardized observations	-0.003	1.27	0.12
Standardized features	0.04	18.9	0.98

Table 6.4 – Mean, and maximum of absolute value and standard deviation of observations and features before and after standardization

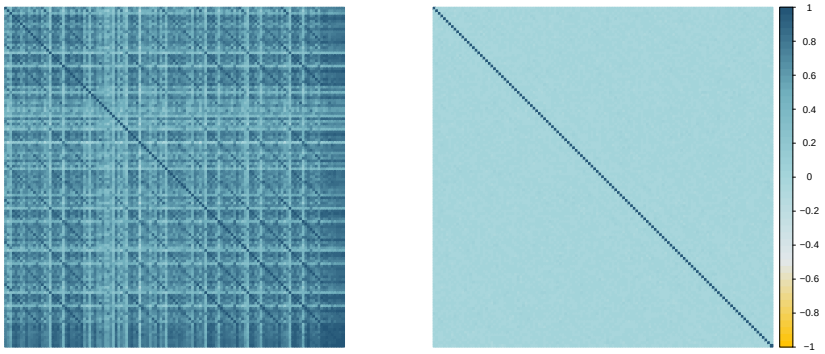


Figure 6.14 – Correlation matrix of non-standardized (left) and standardized (right) feature vectors.

⋆ *Calibration of Hyper-Parameters.* Once features and observations are standardized, we choose one of the algorithms presented in Section 6 and run it, on the $|\mathcal{G}|$ nodes, in parallel with the same hyper-parameter. For the sequential non-linear ridge regression (NL-Ridge), we have to choose the regularization parameter λ (see Equation 6.7) and for BOA and ML-Pol algorithms, we need to set α, the radius of the $L1$-ball (see Algorithm 8). Henceforth, we denote by η this hyper-parameter (which is equal to λ for NL-Ridge and to α for BOA and ML-Pol). We optimize the choice of η by grid search, which is simply an exhaustive search in a specified finite subset G of the hyper-parameter space. This optimization is performed sequentially. Indeed, for any node g and any time step $t > H$, we run $|G|$ algorithms in parallel and we chose the one – denoted by η_t – which minimizes the average prediction error on past available data. Thus, with $\widehat{y}^{\mathrm{g}}_s(\eta)$ the output, at a time step s, of algorithm $\mathcal{A}^{\mathrm{g}}$ run with η, we choose the parameter η_t as follows:

$$\eta_t \in \operatorname*{argmin}_{\eta \in G} \frac{1}{t-H} \sum_{s=1}^{t-H} \frac{1}{|\mathcal{G}|} \sum_{\mathrm{g} \in \mathcal{G}} \left(y^{\mathrm{g}}_s - \widehat{y}^{\mathrm{g}}_s(\eta)\right)^2.$$

In our experiments, to reduce the computational burden, we set $G = \left\{4^i \,|\, i = -5, -4, \ldots, 5\right\}$, so (only) 11 aggregations are run in parallel. At each new day, we check that we never reach the bounds 4^{-5} and 4^5. This kind of online calibration has shown good performance in load forecasting (see, for example, Devaine et al., 2013).

7.4 Results

In this subsection, we compare the four forecasting strategies detailed below by evaluating them on the testing period (May 1, 2010 to July 31, 2010), for each forecasting method of Section 5, for each aggregation algorithm of Section 6 and for various households clusterings. To do so, we introduce some prediction error defined below as well as a confidence bound on this error. We recall that we aim to forecast, at each time step t, a vector of time series $\mathbf{y}_t = (y^{\mathrm{g}}_t)_{\mathrm{g} \in \mathcal{G}}$. The first strategy, that we call "Benchmark", consists simply in providing the features $\mathbf{x}_t$ as forecasts. The second one considers only the projection step and thus skips the aggregation step (we will refer to it as the "Projection" strategy), the associated forecasts are thus the projected features $\Pi_{\underline{\mathbf{K}}}(\mathbf{x}_t)$. To measure the impact of the aggregation step, without projection, we also evaluate the forecasts $\hat{\mathbf{y}}_t$ (which do not necessary satisfy the hierarchical constraints) – this strategy is called "Aggregation". Finally, the strategy "Aggregation + Projection" provides the predictions $\tilde{\mathbf{y}}_t = \Pi_{\underline{\mathbf{K}}}(\hat{\mathbf{y}}_t)$. To allow

for an evaluation of the accuracy of the prediction of some time series only, we define the prediction error $E_T(\Lambda)$, for some subset of nodes $\Lambda \subset \mathcal{G}$. In the results below, this subset can be equal to $\mathcal{G}$ (to evaluate the strategies on all the nodes), to the singleton $\{\mathcal{I}\}$ (to focus on the global consumption – namely the consumption of all the households), or to the set of leaves of the tree associated with the considered segmentation(s), denoted by $\mathcal{G}_0$ (to evaluate the performance of local forecasts only). Note that $E_T(\mathcal{G})$ will correspond to $\widetilde{L}_T \times |\mathcal{G}|$ for the "Aggregation + Projection" strategy (see Equation 7.4). We now define, for any subset $\Lambda \subset \mathcal{G}$, the prediction error $E_T(\Lambda)$. First of all, for a node $\mathrm{g} \in \Lambda$ and a time step t, let us denote by $\varepsilon_t^{\mathrm{g}}$ the instantaneous squared error. It corresponds to $(y_t^{\mathrm{g}} - x_t^{\mathrm{g}})^2$ for the "Benchmark" strategy, to $\left(y_t^{\mathrm{g}} - (\Pi_{\underline{\mathbf{K}}}(\mathbf{x}_t))^{\mathrm{g}}\right)^2$ for "Projection", to $(y_t^{\mathrm{g}} - \widehat{y}_t^{\mathrm{g}})^2$ for "Aggregation", and to $(y_t^{\mathrm{g}} - \widetilde{y}_t^{\mathrm{g}})^2$ for the "Aggregation + Projection" strategy. We then consider the average (over time) squared error (which is accumulated over Λ):

$$E_T(\Lambda) \triangleq \sum_{\mathrm{g}\in\Lambda} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{\mathrm{g}}.$$

We associate with this error a confidence bound and present our results (see Tables 6.5– 6.8) in the form:

$$E_T(\Lambda) \pm \frac{\sigma_T(\Lambda)}{\sqrt{T}}, \quad \text{where} \quad \sigma_T(\Lambda)^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{\mathrm{g}\in\Lambda} \left(\varepsilon_t^{\mathrm{g}} - E_T(\Lambda)\right)^2. \tag{6.12}$$

We choose the quantity $\sigma_T(\Lambda)/\sqrt{T}$ as it is reminiscent of the error margin provided by asymptotic confidence intervals on the mean of independent and identically distributed random variables. In the next paragraph, we consider the "Region + NMF(16)" configuration and, for each of the three benchmark forecasting methods of Section 5 and for each of the three aggregation algorithms presented in Section 6, we compute these errors and confidence bounds for the four above foresting strategies. Finally, in the last paragraph, we set the benchmark forecasting method (generalized additive model) and the aggregation algorithm (ML-Pol) to test various households clusterings.

7.4.1 Impact of the benchmark forecasting methods and of the aggregation algorithms

We consider here the two crossed hierarchies "Region + NMF (16)" and we vary the benchmark forecasting approaches and the aggregation algorithms. Indeed we compute forecasts for the three methods of Section 5 – auto-regressive model, generalized additive model and random forest – and for the three algorithms of Section 6 – NL-Ridge and BOA and ML-Pol. Table 6.5 sums up $E_T(\mathcal{G}) \pm \sigma_T(\mathcal{G})/\sqrt{T}$, where $\mathcal{G}$ refers to the set of nodes associated with "Region + NMF (16)". Regarding forecasting methods, the general additive model provides the best benchmark predictions and the auto-regressive model, which is the most naive method, does not perform well. This was actually already illustrated in Figures 6.10 and 6.11. Moreover, as the theory guarantees, projection (with or without an aggregation step) always improves the forecasts. The projection step without aggregation leads to a decrease of prediction error of around 1% for the general additive and auto-regressive models and of 5% for random forest. Note that for parametric (or semi-parametric) methods, the model is assumed to be the same at all nodes. Forecasts are thus closely linked and seem to almost already satisfy the hierarchical constraints. On the contrary, for random forest methods, the forecasts seem less correlated and thus

	NL-Ridge	ML-Pol	BOA
General Additive Model			
Benchmark		455.5 ± 1.1	
Projection		450.7 ± 1.1	
Aggregation	407.6 ± 1.1	397.9 ± 1.0	406.0 ± 1.0
Aggregation + Projection	405.9 ± 1.1	396.0 ± 1.0	403.5 ± 1.0
Random Forest			
Benchmark		528.1 ± 1.0	
Projection		500.8 ± 1.0	
Aggregation	459.3 ± 1.0	467.3 ± 1.0	470.9 ± 1.0
Aggregation + Projection	451.1 ± 1.0	464.0 ± 1.0	468.1 ± 1.0
Auto-Regressive Model			
Benchmark		736.4 ± 1.6	
Projection		734.3 ± 1.6	
Aggregation	690.7 ± 1.6	690.1 ± 1.6	698.2 ± 1.6
Aggregation + Projection	689.8 ± 1.6	687.3 ± 1.6	693.1 ± 1.6

Table 6.5 – $E_T(\mathcal{G}) \pm \sigma_T(\mathcal{G})/\sqrt{T}$ (see Equation 6.12) where $\mathcal{G}$ refers to the set of nodes associated with "Region + NMF (16)" clustering, for the three benchmark forecasting methods of Section 5 (General Additive Model, Random Forest and Auto-Regressive Model), for the three aggregation algorithms of Section 6 (NL-Ridge, ML-Pol and BOA) and for the four strategies defined in Subsection 7.4 ("Benchmark", "Projection", "Aggregation" and "Aggregation + Projection"). $E_T(\mathcal{G})$ corresponds to $\tilde{L}_T \times |\mathcal{G}|$ for the "Aggregation + Projection" strategy. For strategies "Benchmark" and "Projection", the forecasts do not depend on the chosen aggregation algorithm, so the errors and the confidence bounds are the same for the three algorithms. The dark gray area corresponds to the best prediction error of the table and the light gray area to the best one, for a given benchmark forecasting method.

projection improves significantly the predictions. We point out that Wickramasuriya et al. [2019] also noted that reconciliation can improve poorly specified models, thus providing some insurance against model misspecification. The impact of aggregation step is notable: the prediction error decreases by about 10% for NL-Ridge and BOA and by about 15% for ML-Pol. Finally, our global strategy always gives the best forecasts, which, in addition, satisfy the hierarchical constraints.

Even though theoretical guarantees (see Theorem 9) are only ensured for errors summed over all nodes, we investigate the impact of our methods on global consumption predictions and on most local predictions (*i.e.*, predictions at leaves). Thus, Tables 6.6 and 6.7 contain $E_T\big(\{\mathcal{I}\}\big) \pm \sigma_T(\{\mathcal{I}\})/\sqrt{T}$ and $E_T(\mathcal{G}_0) \pm \sigma_T(\mathcal{G}_0)/\sqrt{T}$ (where $\mathcal{G}_0$ is the set of leaves), respectively. By denoting by $R_1, \ldots, R_N$, the N regions and by $C_1, \ldots C_{16}$, the 16 clusters provided by "NMF (16)", we have, in this "Region + NMF (16)" configuration, $\mathcal{G}_0 \triangleq \big\{C_\ell \cap R_n\big\}_{1 \leqslant \ell \leqslant 16,\, 1 \leqslant n \leqslant N}$. Concerning global consumption, a mere projection improves the forecasts, except in the case of auto-regressive model and, in all cases, our strategy "Aggregation + Projection" outperforms the three strategies "Benchmark", "Aggregation" and "Projection". The prediction error associated with $\mathcal{G}_0$ also decreases thanks

	NL-Ridge	ML-Pol	BOA
General Additive Model			
Benchmark		205.8 ± 9.3	
Projection		200.8 ± 9.2	
Aggregation	179.2 ± 8.9	172.0 ± 8.6	178.8 ± 8.8
Aggregation + Projection	177.6 ± 8.8	170.3 ± 8.5	176.3 ± 8.7
Random Forest			
Benchmark		231.4 ± 8.6	
Projection		228.8 ± 8.2	
Aggregation	207.1 ± 8.4	214.8 ± 8.4	218.7 ± 8.3
Aggregation + Projection	206.4 ± 8.2	212.4 ± 8.1	216.8 ± 8.2
Auto-Regressive Model			
Benchmark		380.3 ± 13.4	
Projection		380.4 ± 13.4	
Aggregation	368.6 ± 13.5	370.8 ± 13.6	376.1 ± 13.6
Aggregation + Projection	368.2 ± 13.4	369.4 ± 13.5	373.6 ± 13.5

Table 6.6 – $E_T(\{\mathcal{I}\}) \pm \sigma_T(\{\mathcal{I}\})/\sqrt{T}$ (see Equation 6.12) for "Region + NMF (16)" clustering, for the three benchmark forecasting methods of Section 5 (General Additive Model, Random Forest and Auto-Regressive Model), for the three aggregation algorithms of Section 6 (NL-Ridge, ML-Pol and BOA) and for the four strategies defined in Subsection 7.4 ("Benchmark", "Projection", "Aggregation" and "Aggregation + Projection"). The prediction error $E_T(\{\mathcal{I}\})$ corresponds to the mean squared error (over the testing period) of the global consumption. For strategies "Benchmark" and "Projection", the forecasts do not depend on the chosen aggregation algorithm, so the errors and the confidence bounds are the same for the three algorithms. The dark gray area corresponds to the best prediction error of the table and the light gray area to the best one, for a given benchmark forecasting method.

to our procedure. Therefore, our method improves the forecasting of both global and local power consumptions. Finally, Figure 6.15 represents the global power consumption on the three last day of the testing period and the daily average signed error on the last week for the four forecasts obtained with features generated with general additive model and aggregated with ML-Pol algorithm. The distributions of the daily mean squared errors for these strategies are represented in Figure 6.16. We draw the same conclusions for the daily prediction errors as for the average error on the entire test period (three months): aggregation greatly improves the forecasts, projection does too, but to a lesser extent. The box plots show that the variance of the error also decreases after the aggregation step.

7.4.2 Impact of the clustering

We now assess the impact of household segmentation on the quality of our predictions. In view of the foregoing, we set the aggregation algorithm to ML-Pol and the benchmark forecasting method to the general additive model. As clusters change from a segmentation to another, the associated sets of nodes $\mathcal{G}$ also change. Errors related to $\mathcal{G}$ or $\mathcal{G}_0$ can therefore not be compared from a segmentation to another. We thus focus here on the global consumption (namely, we compute errors related to $\{\mathcal{I}\}$). We compare our methods to a

	NL-Ridge	ML-Pol	BOA
General Additive Model			
Benchmark		66.3 ± 0.1	
Projection		66.3 ± 0.1	
Aggregation	61.6 ± 0.1	61.2 ± 0.1	61.0 ± 0.1
Aggregation + Projection	61.5 ± 0.1	61.1 ± 0.1	61.0 ± 0.1
Random Forest			
Benchmark		78.7 ± 0.1	
Projection		68.9 ± 0.1	
Aggregation	66.8 ± 0.1	65.7 ± 0.1	64.8 ± 0.1
Aggregation + Projection	63.9 ± 0.1	65.7 ± 0.1	64.8 ± 0.1
Auto-Regressive Model			
Benchmark		84.4 ± 0.1	
Projection		84.3 ± 0.1	
Aggregation	73.8 ± 0.1	72.8 ± 0.1	73.2 ± 0.1
Aggregation + Projection	73.8 ± 0.1	72.0 ± 0.1	72.0 ± 0.1

Table 6.7 – $E_T(\mathcal{G}_0) \pm \sigma_T(\mathcal{G}_0)/\sqrt{T}$ (see Equation 6.12) where $\mathcal{G}_0$ refers to the set of leaves associated with "Region + NMF (16)" clustering, for the three benchmark forecasting methods of Section 5 (General Additive Model, Random Forest and Auto-Regressive Model), for the three aggregation algorithms of Section 6 (NL-Ridge, ML-Pol and BOA) and for the four strategies defined in Subsection 7.4 ("Benchmark", "Projection", "Aggregation" and "Aggregation + Projection"). $E_T(\mathcal{G}_0)$ corresponds to a prediction errors associated with local consumptions forecasts. For strategies "Benchmark" and "Projection", the forecasts do not depend on the chosen aggregation algorithm, so the errors and the confidence bounds are the same for the three algorithms. The dark gray area corresponds to the best prediction error of the table and the light gray area to the best one, for a given benchmark forecasting method.

naive bottom-up strategy: at each time step t, we forecast the global consumption $y_t^{\{\mathcal{I}\}}$ with the sum of local consumptions $\sum_{g \in \mathcal{G}_0} x_t^g$ – instead of the benchmark predictions $x_t^{\{\mathcal{I}\}}$. Table 6.8 contains the prediction errors and the confidence bounds for the five strategies and for several household segmentations. For the "Bottom-up" strategy, the geographical clustering "Region" provides the lowest prediction error, that are much better than the one of benchmark forecasts. While when a single clustering based on household profiles or generated randomly is considered, the benchmark forecasts $x_t^{\{\mathcal{I}\}}$ are more relevant – in terms of mean squared error. Thus, taking into account regional consumptions, which depend on local meteorological variables, improves prediction. In the same way, projection significantly improves the forecasts when the regions are taken into account. Moreover, for a fixed number of clusters – for example, we compare "Fuel+Tariff", "Random (4) and"NMF (4)" – the aggregation step seems more efficient when clusters present different consumption profiles (see Figures 6.5 - 6.9). Indeed, aggregation provides much better performance for "NMF (4)" than for "Random (4)". As we had anticipated, contrary to "NMF" and "Region", clusterings "Acorn" and "Fuel + Tariff", that do not seem to detect consumption profiles, perform as well as "Random". When the number of clusters becomes too large, the performance of the strategy stagnates or even decreases. Typically

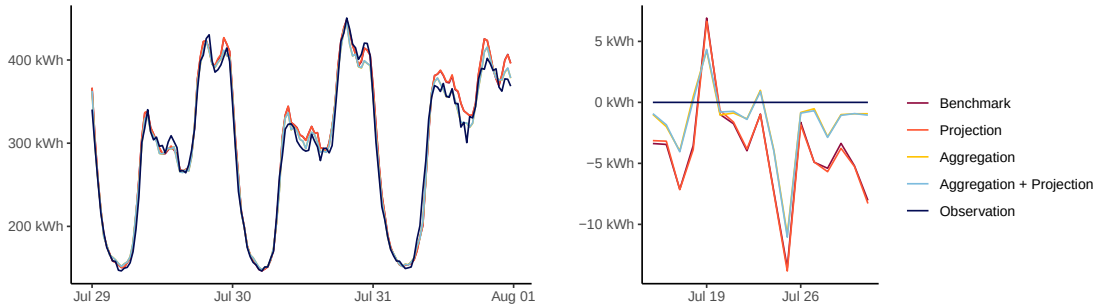


Figure 6.15 – Left picture: forecasts associated with the four strategies defined in Subsection 7.4 ("Benchmark", "Projection", "Agregation" and "Aggregation + Projection"), with benchmark forecasts generated with the generalized additive model and aggregated with ML-Pol algorithm in the "Region + NMF(16)" configuration, and observations of global consumption ($g = \mathcal{I}$) at half-hour intervals on the last three days of the test period. Right picture: corresponding daily average signed errors on the last week of the test period.

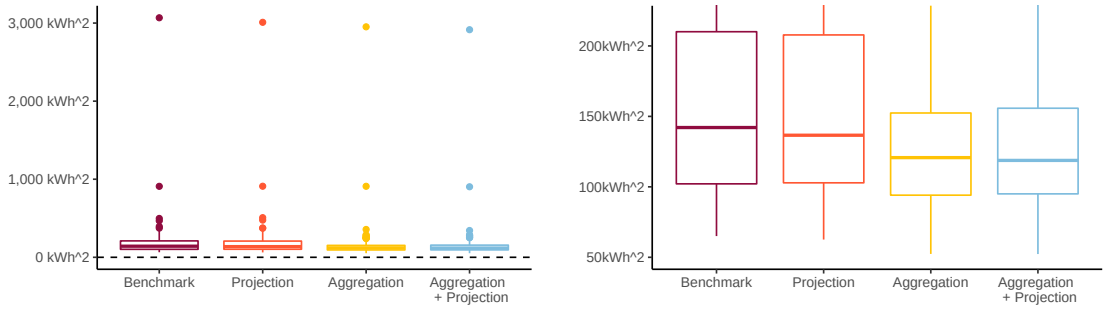
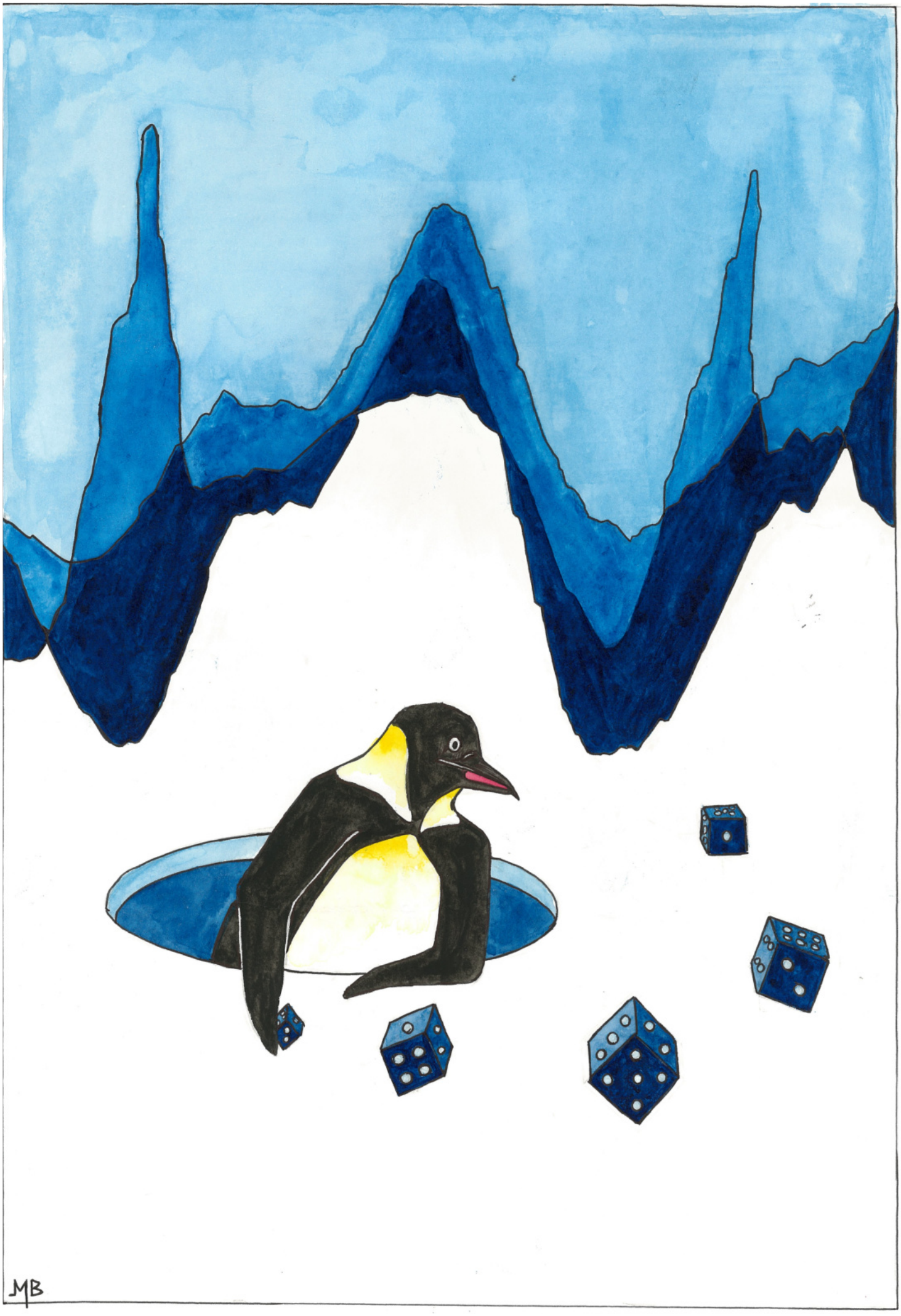


Figure 6.16 – Distribution over the test period of daily mean squared error of global consumption for the four strategies defined in Subsection 7.4 ("Benchmark", "Projection", "Agregation" and "Aggregation + Projection"), with benchmark forecasts generated with the generalized additive model and aggregated with ML-Pol algorithm in the "Region + NMF(16)" configuration. Left picture: original boxplots. Right picture: boxplots trimmed at 220 kWh^2.

for "Random" or "NMF", a number of clusters equals to 32 or 64 does not seem to improve the results compared to smaller numbers 4, 8 or 16. Another result is that aggregation and projection are robust to large number of clusters. Indeed, the performance are good for a sufficiently large number of clusters but does not decrease too much with the number of clusters – either for " Random" or "NMF" clusterings. Finally, our strategy "Aggregation + Projection" always outperforms the other four ("Bottom-up", "Benchmark", "Projection" and "Aggregation") and the "Region + NMF (16)" clustering reaches the lowest prediction error.

Clustering	Benchmark	Bottom-up	Projection	Aggregation	Aggregation + Projection
Region	205.8 ± 9.3	189.9 ± 8.3	201.3 ± 9.1	187.8 ± 8.4	186.7 ± 8.4
Region + Acorn	—	**194.2 ± 8.4**	**200.8 ± 9.2**	**182.5 ± 8.3**	**181.2 ± 8.3**
Acorn	—	205.7 ± 9.5	205.0 ± 9.3	203.3 ± 9.3	202.9 ± 9.3
Region + Fuel + Tariff	—	**199.1 ± 8.7**	**201.2 ± 9.2**	**185.4 ± 8.6**	**184.1 ± 8.6**
Fuel + Tariff	—	207.1 ± 9.7	205.5 ± 9.4	201.5 ± 9.4	201.4 ± 9.5
Region + Random (4)	—	**198.4 ± 8.7**	**201.3 ± 9.2**	**186.1 ± 8.6**	**184.6 ± 8.6**
Random (4)	—	208.0 ± 9.7	205.7 ± 9.4	199.5 ± 9.4	199.7 ± 9.4
Region + Random (8)	—	**202.3 ± 8.7**	**201.3 ± 9.2**	**182.4 ± 8.7**	**181.0 ± 8.7**
Random (8)	—	212.9 ± 9.8	205.7 ± 9.3	194.4 ± 9.1	194.4 ± 9.1
Region + Random (16)	—	**205.1 ± 8.7**	**201.3 ± 9.2**	**180.5 ± 8.7**	**178.8 ± 8.7**
Random (16)	—	218.4 ± 10.0	205.7 ± 9.3	188.6 ± 8.7	188.5 ± 8.7
Region + Random (32)	—	**205.3 ± 8.5**	**201.2 ± 9.2**	**180.4 ± 8.8**	**178.9 ± 8.7**
Random (32)	—	222.9 ± 10.1	205.6 ± 9.3	189.6 ± 8.7	189.5 ± 8.7
Random (64)	—	222.9 ± 9.8	205.6 ± 9.3	185.7 ± 8.8	185.5 ± 8.8
Region + NMF (4)	—	**196.0 ± 8.6**	**200.8 ± 9.2**	**187.4 ± 9.1**	**185.5 ± 8.9**
NMF (4)	—	205.7 ± 9.5	205.0 ± 9.3	197.0 ± 8.8	196.8 ± 8.9
Region + NMF (8)	—	**197.2 ± 8.5**	**200.7 ± 9.2**	**176.4 ± 8.9**	**174.1 ± 8.8**
NMF (8)	—	206.7 ± 9.6	205.0 ± 9.3	186.1 ± 8.9	185.7 ± 8.9
Region + NMF (16)	—	**201.0 ± 8.5**	**200.8 ± 9.2**	**172.0 ± 8.6**	**170.3 ± 8.5**
NMF (16)	—	208.4 ± 9.6	205.2 ± 9.3	179.3 ± 8.4	179.3 ± 8.4
Region + NMF (32)	—	**204.1 ± 8.5**	**201.0 ± 9.1**	**173.2 ± 8.7**	**171.5 ± 8.6**
NMF (32)	—	211.1 ± 9.6	205.4 ± 9.3	179.7 ± 8.8	179.5 ± 8.8
NMF (64)	—	214.9 ± 9.4	205.6 ± 9.3	181.9 ± 8.6	181.7 ± 8.6

Table 6.8 – $E_T(\{\mathcal{I}\}) \pm \sigma_T(\{\mathcal{I}\})/\sqrt{T}$ (see Equation 6.12) for the fives strategies defined in Subsection 7.4 ("Benchmark", "Bottom-up, "Projection", "Aggregation" and "Aggregation + Projection"), with benchmark predictions ($x_t^{\{\mathcal{I}\}}$ that are the same for all clusterings) made with General Additive Models and aggregated with ML-Pol algorithm, for many segmentations (defined in Subsection 7.2). The prediction error $E_T(\{\mathcal{I}\})$ corresponds to the mean squared error (over the testing period) of the global consumption. The dark gray area corresponds to the best prediction error of the table and the light gray area to the best one, for a given strategy.


MB

7

Simulating tariff impact in power consumption profiles with conditional variational autoencoders

The implementation of efficient demand response programs for household electricity consumption would benefit from data-driven methods capable of simulating the impact of different tariffs schemes. This chapter proposes a method based on conditional variational autoencoders (CVAE) to generate, from an electricity tariff profile combined with weather and calendar variables, daily consumption profiles of consumers segmented in different clusters. First, a large set of consumers is gathered into clusters according to their consumption behavior and price-responsiveness. The clustering method is based on a causality model that measures the effect of a specific tariff on the consumption level. Then, daily electrical energy consumption profiles are generated for each cluster with CVAE. This non-parametric approach is compared to a semi-parametric data generator based on generalized additive models. The main contribution from this new method is the capacity to reproduce rebound and side effects in the generated consumption profiles. Indeed, the application of a special electricity tariff over a time window may also affect consumption outside this time window. Another contribution is that the proposed clustering approach is capturing the reaction to a tariff change.

This chapter was written in collaboration with Ricardo J. Bessa and was accepted for publication in IEEE Access (a peer-reviewed open-access journal published by the Institute of Electrical and Electronics Engineers-IEEE).

1 Introduction

A power consumption data simulator may be very useful to study the business models of different demand response (DR) models Karlsen et al. [2020] and to conduct an *ex-ante* assessment of the DR algorithms that set tariff profiles (*i.e.*, ensure that they induce the right behavior from consumers), such as contextual bandit algorithms presented in Chapter 4. We recall that electricity demand response policies aim at modifying customers' energy consumption behavior (see Siano, 2014 for an overview) to enable higher integration levels of renewable energy sources. Most of these DR schemes rely on changes in electricity prices, which can take the form of seasonal tariffs, super-peak time-of-use, real-time pricing, critical peak pricing, etc. Dutta and Mitra [2017]. Chapter 4 proposed online learning algorithms to optimize these price incentives, considering an homogeneous population of customers. However, the responsiveness to a tariff change may change from a consumer to another. By clustering consumers according to their tariff responsiveness, an electricity supplier can send different signals depending on the cluster to which they belong, and further improve DR management. For instance, for a given temperature, day of the week, etc., the electricity supplier defines an hourly electricity tariff profile to send to some consumers clusters.

The data simulator should be able to randomly generate energy consumption profiles for different combinations of exogenous variables and tariff profiles, with consumers clustered according to their tariff responsiveness. The present chapter proposes a novel method, based on conditional variational autoencoders (CVAE), which aims to randomly simulate daily power consumption profiles conditioned by a specific electricity tariff combined with weather and calendar variables.

The remainder of this chapter is organized as follows. The next subsection conducts a literature review of for the forecasting of consumers reactions to demand response policies and data generation methods applied in the energy domain, the data set used throughout the rest of the chapter is also presented. A clustering method is first provided in

Section 3. Then, the CVAE approach used to generate energy consumption profiles is presented and discussed in Section 4. In order to evaluate the proposed method, Section 5 introduces a benchmark data generator based on semi-parametric models often used for energy consumption forecasting. Section 6 presents a comparison of the two generators and simulations that illustrate the interest of our approach.

The reproducibility of this research was ensured by applying the methodology to the open data set "SmartMeter Energy Consumption Data in London Households" from UK Power Networks presented in Chapter 3, where price incentives were sent to users via their smart meters, and by making the CVAE codes available in a GitHub repository[1].

⋆ *Literature discussion for consumers reactions to demand response policies forecasting.* Recent research developed mathematical and statistical models for modeling price responsiveness from domestic consumers. For example, Ganesan *et al.* applied a causality model to the Low Carbon London data set in order to rank consumers according to their responsiveness to tariff changes, and outperformed correlation-based metrics (see Ganesan et al., 2019 for further details). Saez-Gallego and Morales applied inverse optimization to improve the accuracy of load forecasting when aggregating a pool of price-responsive consumers and considering the effect of calendar and weather variables (see Saez-Gallego and Morales, 2017); Le Ray *et.al.* applied a clinical testing approach (based on a test and a control group) to assess whether or not loads of households participating in the EcoGrid EU DR program are price-responsive (see Le Ray et al., 2016); Mohajeryami et al. proposed an economic model to explain the consumption shift between peak and off-peak hours that maximizes customer's utility function (see Mohajeryami et al., 2016). These works are closely linked to the forecast of consumers reactions to DR policies, but, to our knowledge, were never combined with clustering techniques for consumer segmentation or used to simulate daily consumption profiles according to their price-responsiveness. We refer to the introduction of Chapter 6 for the literature review on clustering methods that do not include information about the elasticity of consumers to tariff changes.

⋆ *Data generation methods* The generation of energy consumption profiles for households is not new and it was already covered by different authors in the literature. Capasso *et al.* proposed a bottom-up approach based on the aggregation of individual appliance consumption in order to produce a household consumption profile (see Capasso et al., 1994). A Monte Carlo simulation model was proposed to combine behavioral data (home activities, availability at home from each member, etc.) and engineering functions (appliance mode of operation, technological penetration, etc.) with associated probability distributions. Park *et al.* proposed a platform, exploiting SystemC language for event-driven simulation, which simulates the behavior of individual appliances and smart plugs (see Park et al., 2010). Both works did not considered weather-dependent appliances (e.g., heating, ventilating and air conditioning - HVAC) or the effect of price signals. Physically-based models for appliances (including HVAC) are also proposed in (see Muratori et al., 2013), combined with heterogeneous Markov chain for activity patterns, to simulate households energy consumption. A similar approach was followed in Richardson et al. [2010], but using individual appliance consumption data. A set of physical models for appliances are proposed in López et al. [2019], implemented in MATLAB Simulink, and can simulate

[1] `github.com/MargauxBregere/power_consumption_simulator`

optimal on/off decisions of household appliances. Gottwalt *et al.* described a simulation engine for households with two modules: first bottom-up approach that generates consumption data for each appliance by combining statistical data about appliance use and resident presence at home; and second optimization of appliances schedule in order to find the optimal load shift according to time-based tariffs (see Gottwalt et al., 2011). Iwafune *et al.* proposed a Markov chain Monte Carlo method for simulating electric vehicle driving behaviors, which enables an evaluation of the DR potential when combined with domestic photovoltaic panels (see Iwafune et al., 2020). The aforementioned methodologies assume that information about individual appliances (usage patterns, energy consumption, etc.) and behavioral data is available, instead of just using the total household consumption collected by the smart meter. One exception is Li et al. [2019], which describes a methodology based on an elasticity coefficient (approximated by a Gaussian distribution) to estimate indices that characterize the impact of real-time prices in the consumption pattern, such as proportion of maximum load decrease, proportion of peak-valley difference of load decrease, etc. The method consisted in an empirical rule-based calculation of transferred consumption between periods, which was only applied to aggregated consumption of an electric power system and not to households.

2 Data set description and preprocessing

As a case-study for this chapter, we consider the open data set, published by UK Power Networks and containing energy consumption (in kWh per half-hour) of around 5 000 households throughout 2013. The data preprocessing is fully described in Section 2 of Chapter 3. We recall that our data set contains the power consumption records, at half-hourly intervals of 1 007 households subjected to a dynamic Time of Use (ToU) and 1 007 non-ToU customers who were on a flat rate tariff; we refer to them as Standard (Std) customers. We denote by $\mathcal{I}_{ToU}$ the set of ToU households and by $\mathcal{I}_{Std}$ the set of Std ones; for each household $i \in \mathcal{I}_{ToU} \cup \mathcal{I}_{Std}$, for any day t of year 2013, we get the H-dimensional power consumption vector $Y_t^1(i), \ldots, Y_t^H(i)$, with $H = 48$ the number of consumption readings per day. For each household, we also compute the average energy consumption, its minimum, and its maximum as well as the half-hour of the daily peak and of the daily trough, for the hot months (from April to September) and for the cold months (the others). The data set also contains half-hourly tariff and temperature profiles denoted by $p_t^1, \ldots, p_t^H$ and $\tau_t^1, \ldots, \tau_t^H$, respectively. For any day t, we also consider the smoothed temperature $\bar{\tau}_t$ – that models the thermal inertia of buildings – and two calendar variables: the type of day w_t that takes 0 on weekends and 1 on working days; and the position in the year π_t which increases linearly from 0 (on January, 1.) to 1 (on December, 31.). Therefore, the data set (presented in Table 7.2) contains, for each of the 2 014 households (half Std, half ToU), $T = 365$ observations of the energy consumption, tariff, and temperatures profiles, the smoothed temperature, the type of day, and the position in the year.

This data set is split in two sub-sets: a training set which contains about 75% of the original data – days are randomly sampled from those of 2013 – and a testing set made of the remaining data points. A perfect design of the experiments would require four data sets but the size of the original data led us to exclude this possibility. As the household clustering is a prior knowledge for the creation of the data generators (we create a generator per cluster), the entire data is used to cluster the clients. The (non-parametric

and semi-parametric) data generators are optimized on the training set. The testing set is used to calibrate CVAE-based data generators and to choose the best combination of exogenous variables to give in input. Moreover, the best CVAE among several executions of the training process (CVAEs may converges to local minima) is selected thanks to this testing set. Finally, it also permits to compare the two approaches, non-parametric and semi-parametric, in the experiments of Section 6. To simplify notation, we re-indent the observations of the original data set: observations from 1 to $T_0 = 273$ form the training set, and the ones from $T_0 + 1$ to $T = 365$ form the testing set. The dataset division and use is summarized in Table 7.1.

	Training Set	Testing Set
Households clustering	✓	✓
Semi-parametric model training	✓	
CVAE model training	✓	
CVAE hyper-parameters calibration		✓
CVAE conditional variables choice		✓
CVAE model selection		✓
Numerical experiments		✓

Table 7.1 – Summary of the use of the two data sets: the training set (75% of the original data) and the testing set (remaining data). The clustering of the households is detailed in Section 3. The training process for the CVAE-based generator is explained in Section 4; the calibration of the hyper-parameters and the selection of the best CVAE are detailed in the subsections 4.2, 4.3 and 4.4, respectively. The training process for the semi-parametric generator is in Section 5. Both data generators are compared in the experiments of Section 6.

3 Clustering of household consumers

3.1 Causality model

To measure the impact of the tariff on the energy consumption, a causality model similar to the one proposed by Ganesan *et al.* (see Ganesan et al., 2019) is considered. The finite set of available tariff is denoted by $\mathcal{P} = \{\text{Low, High, Normal}\}$ and its cardinal by $|\mathcal{P}|$. For each household and each tariff, a daily profile of the mean and the standard deviation of its energy consumption will be computed. For an household i, at an half-hour h, the random variable $Y^h(i)$ refers to the individual energy consumption of household i. It depends on the chosen tariff $p \in \mathcal{P}$ but also on the exogenous variables gathered in a vector $x^h = (\tau_t^h, \bar{\tau}_t, w_t, \pi_t)$.

Here, the aim is to estimate, for each tariff p and for each half-hour h, the expectation and the standard deviation of the random variable $Y^h(i) \,|\, P = p$. Thanks to T observations $Y_t^h(i)$, x_t^h, and p_t^h, with $t \in \{1, \ldots, T\}$, of energy consumption, tariffs, and exogenous variables, respectively, a model that gives, for the tariff p and the exogenous variables x^h, a forecast of the expected consumption at h when tariff p is picked, is trained. In the original model, the authors used kernel regression and then an approach based on bootstrapping to provide an estimation of the standard deviation (see Ganesan et al., 2019 for further details). In this work, for any exogenous variable x_t^h and tariff p_t^h, the random energy consumption $Y_t^h(i)$ is assumed to be Gaussian of mean $\mu_i(x_t^h, p_t^h)$ and

Variable	Notation
Daily energy consumption profile at half-hourly intervals	$Y_t^1(i), \dots, Y_t^H(i)$
Daily tariff profile (for ToU consumers) at half-hourly intervals	$p_t^1, \dots, p_t^H$
Daily London air temperature profile at half-hourly intervals	$\tau_t^1, \dots, \tau_t^H$
Smooth temperature (computed from past temperatures)	$\bar{\tau}_t$
Type of day (1 from Monday to Friday, 0 for week-ends)	w_t
Position in the year (0 on January, 1. and 1 on December, 31.)	π_t

Table 7.2 – Summary of the variables provided and created for each household i of the data set.

standard deviation $\sigma_i(x_t^h, p_t^h)$ and that theses mean and standard deviation depend on additive smooth predictors. They are estimated with generalized additive models (GAM), see Section 3 of Chapter 3 for further details. To do so, we train a model that gives, for the tariff p and the exogenous variables x^h, a forecast of the expected consumption at h when tariff p is selected and a forecast of the standard deviation of this consumption. For any exogenous variable x_t^h and tariff p_t^h, the random energy consumption $Y_t^h(i)$, of household i at the half hour h of the day t, is assumed Gaussian of mean $\mu_i(x_t^h, p_t^h)$ and standard deviation $\sigma_i(x_t^h, p_t^h)$. Moreover, we assume that these mean and standard deviation:

$$\mu^{i,h}(x_t^h, p_t^h) = \mathbb{E}\big[Y_t^h(i)\big] \quad \text{and} \quad \sigma^{i,h}(x_t^h, p_t^h) = \sqrt{\mathrm{Var}\left[Y_t^h(i)\right]},$$

depend on additive smooth predictors. Here, GAMs are used to estimate conjointly, for any half-hour h of a day t and any tariff $p \in \mathcal{P}$, both $\mu^{i,h}(x_t^h, p)$ and $\sigma^{i,h}(x_t^h, p)$. These approximations are denoted by $\widehat{\mu}^i, h(x_t^h, p)$ and $\widehat{\sigma}^{i,h}(p, x_t^h)$, respectively. For each half-hour h, we set the same underlying models:

$$\begin{aligned} \mu^{i,h}(x_t^h, p_t^h) &= s_\tau^{i,h}(\tau_t^h) + \xi_{\mathrm{L}}^{i,h}\mathbf{1}_{p_t^h=\mathrm{Low}} + \xi_{\mathrm{N}}^{i,h}\mathbf{1}_{p_t^h=\mathrm{Normal}} + \xi_{\mathrm{H}}^{h}\mathbf{1}_{p_t^h=\mathrm{High}} \\ \text{and} \quad \sigma^{i,h}(x_t^h, p_t^h) &= \gamma_{\mathrm{L}}^{i,h}\mathbf{1}_{p_t^h=\mathrm{Low}} + \gamma_{\mathrm{N}}^{i,h}\mathbf{1}_{p_t^h=\mathrm{Normal}} + \gamma_{\mathrm{H}}^{h}\mathbf{1}_{p_t^h=\mathrm{High}}\,. \end{aligned} \tag{7.1}$$

where $s_\tau^{i,h}$, the function catching the effect of the temperature, is approximated by a cubic spline. The `mgcv R`-package (see Wood, 2020) permits to estimate the coordinates of the spline in its basis and all the coefficients $\xi_{\mathrm{L}}^{i,h}$, $\xi_{\mathrm{N}}^{i,h}$, $\xi_{\mathrm{H}}^{i,h}$, $\gamma_{\mathrm{L}}^{i,h}$, $\gamma_{\mathrm{N}}^{i,h}$, and $\gamma_{\mathrm{H}}^{i,h}$ defined in Equation (7.1), which catch tariff effect. We highlight that models fitted on variances are linear. Both models (on mean and standard deviation) are estimated simultaneously, by setting the model family parameter of the `gam` function to the Gaussian location-scale model family. Once the function and coefficients have been estimated (we write $\widehat{s}^{i,h}$ for the estimation of $s^{i,h}$ and so on), for any tariff p, the estimations $\widehat{\mu}^{i,h}(x_t^h, p)$ and $\widehat{\sigma}^{i,h}(p, x_t^h)$ are computed:

$$\begin{aligned} \widehat{\mu}^{i,h}(x_t^h, p) &= \widehat{s}_\tau^{i,h}(\tau_t^h) + \widehat{\xi}_{\mathrm{L}}^{i,h}\mathbf{1}_{p=\mathrm{Low}} + \widehat{\xi}_{\mathrm{N}}^{i,h}\mathbf{1}_{p=\mathrm{Normal}} + \widehat{\xi}_{\mathrm{H}}^{h}\mathbf{1}_{p=\mathrm{High}} \\ \text{and} \quad \widehat{\sigma}^{i,h}(p, x_t^h) &= \widehat{\gamma}_{\mathrm{L}}^{i,h}\mathbf{1}_{p=\mathrm{Low}} + \widehat{\gamma}_{\mathrm{N}}^{i,h}\mathbf{1}_{p=\mathrm{Normal}} + \widehat{\gamma}_{\mathrm{H}}^{h}\mathbf{1}_{p=\mathrm{High}}\,. \end{aligned}$$

Therefore, for any tariff p, the trained model provides these estimations, that are denoted by $\widehat{\mu}_i(x_t^h, p)$ and $\widehat{\sigma}_i(p, x_t^h)$. Then, an approximation of the impact of a tariff change is computed with the two following quantities:

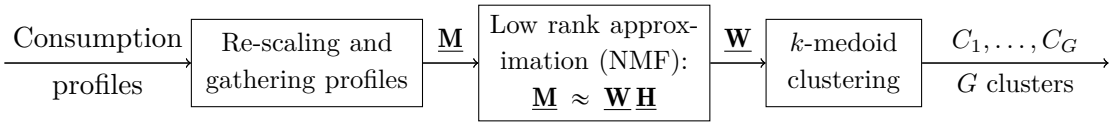
$$\mathbb{E}\big[\,Y^h(i)\,|\,P=p\,\big] \approx \frac{1}{T}\sum_{t=1}^{T}\widehat{\mu}_i\big(x_t^h, p\big) \text{ and } \sqrt{\mathrm{Var}\left[\,Y^h(i)\,|\,P=p\,\right]} \approx \frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}_i\big(x_t^h, p\big)\,. \tag{7.2}$$

For simplicity of notation, these approximations associated with an household $i \in \mathcal{I}_{ToU} \cup \mathcal{I}_{Std}$, are denoted by $\mu_i^h(p)$ and $\sigma_i^h(p)$, respectively. Vectors $\mu_i^1(p), \ldots, \mu_i^H(p)$ will be used to cluster the consumers whereas vectors $\sigma_i^1(p), \ldots, \sigma_i^H(p)$ will not be used until later, in Section 5 for the creation of the benchmark data generator. Actually, they will not be directly useful, but a similar approach will be applied to compute the standard deviation per tariff of the energy consumption of a consumer cluster, namely by replacing household i by a group of households.

3.2 Clustering method

The proposed method used to cluster the households according to their consumption profile is very similar to the one described in Section 7.2.3 of Chapter 6 – the main differences are the use of consumption profiles instead of consumption historical time series and of k-mediod algorithm instead of k-means algorithm. In this section, $\mathcal{I}$ will refer indifferently to $\mathcal{I}_{ToU}$ or to $\mathcal{I}_{Std}$. For any household, $i \in \mathcal{I}$, the causality model described in the previous section provides, for each tariff $p \in \mathcal{P}$, a daily energy consumption profile, namely H mean energy consumption $\mu_i^1(p), \ldots, \mu_i^H(p)$. As the focus is more on the shape of the profiles, rather than on the amount of consumed electricity, the profiles of an household i are first re-scaled with its average consumption associated with a base tariff, namely Normal tariff.

Then, these profiles are concatenated in a matrix $\underline{\mathbf{M}} \in \mathcal{M}_{|\mathcal{I}| \times H|\mathcal{P}|}$ that gathers all the households. The dimension of $\underline{\mathbf{M}}$ is reduced with a non-negative matrix factorization (NMF): with r a small integer, $\underline{\mathbf{M}}$ is approximated by $\underline{\mathbf{W}}\,\underline{\mathbf{H}}$, where $\underline{\mathbf{W}}$ and $\underline{\mathbf{H}}$ are $|\mathcal{I}| \times r$ and $r \times H|\mathcal{P}|$-non-negative matrices, respectively. As soon as this approximation is good enough, line i of the matrix $\underline{\mathbf{W}}$ is sufficient to reconstruct household i profiles (with the knowledge of matrix $\underline{\mathbf{H}}$ - which is not used for the clustering). Thus, for each household i, from the $H|\mathcal{P}|$-vector $(\mu_i^1(p), \ldots, \mu_i^H(p))_{p \in \mathcal{P}}$, r features are extracted: line i of $\underline{\mathbf{W}}$. With this low dimension representation of households in $\mathbb{R}^r$, k-medoids clustering algorithm provides the G clusters $C_1, \ldots, C_G$, using `KMedoid` function implemented in the `Python`-library `sklearn_extra`. The diagram below sums up the steps of the procedure described here in a summarized way and detailed below.



⋆ *Scaling and gathering profiles.* For an household $i \in \mathcal{I}$, for all $p \in \mathcal{P}$, the daily expected consumption profile $\mu_i^1(p), \ldots, \mu_i^H(p)$ is considered. We assume that there is a base tariff $p_0 \in \mathcal{P}$ that corresponds to a signal of no incentive, namely Normal tariff. We consider the quantity $\bar{\mu}_i = \frac{1}{H} \sum_{h=1}^{H} \mu_i^h(p_0)$ that is an approximation of the average daily expected consumption of household i under no DR program. Then, all the profiles of household i are rescaled by this quantity and, for each tariff $p \in \mathcal{P}$, the daily consumption profiles under tariff p of all the households $i \in \mathcal{I}$ are gathered in a matrix $\underline{\mathbf{M}}(p) \in \mathcal{M}_{|\mathcal{I}| \times H}$. Finally, the matrix $\underline{\mathbf{M}} \in \mathcal{M}_{|\mathcal{I}| \times H|\mathcal{P}|}$ is created by binding by column matrices $\underline{\mathbf{M}}(p)$, so

$$\underline{\mathbf{M}}_{i,h}(p) = \frac{\mu_i^h(p)}{\bar{\mu}_i} \quad \text{and} \quad \underline{\mathbf{M}} = \left(\underline{\mathbf{M}}(1) \;\middle|\; \ldots \;\middle|\; \underline{\mathbf{M}}(P) \right).$$

⋆ *Low rank approximation.* The dimension of the non-negative matrix of profiles $\underline{\mathbf{M}}$ is reduced with a non-negative matrix factorization (NMF) – Section 7.2.3 of Chapter 6.

The integer $r \ll \min(|\mathcal{I}|, H|\mathcal{P}|)$ that will ensure the dimension reduction is fixed (we chose $r = 5$ in our case study). The NMF approximates $\underline{\mathbf{M}}$ with $\underline{\mathbf{W}}\underline{\mathbf{H}}$, where $\underline{\mathbf{W}}$ and $\underline{\mathbf{H}}$ are non-negative matrices of size $|\mathcal{I}| \times r$ and $r \times H|\mathcal{P}|$, respectively. Function `NMF` of the `Python`-library `sklearn.decomposition` allows to approximate $\underline{\mathbf{W}}$ and $\underline{\mathbf{H}}$ with a coordinate descent solver. For simplicity of notation, $\underline{\mathbf{W}}$ is confounded with its approximation. Thus, for any household $i \in \mathcal{I}$, we get r features, namely the i^{th} line of matrix $\underline{\mathbf{W}}$, that we denote by $\underline{\mathbf{W}}_{i\cdot}$ in the following.

$\star$ *k-medoid clustering.* Now, the vectors $\underline{\mathbf{W}}_{i\cdot}$ allow to cluster households in G clusters. In k-means clustering, the center of a given cluster is simply the average between the points of this cluster. Since it can be influenced by extreme value, k-means algorithm is sensitive to outliers. Conversely, k-medoid algorithm chooses data points to represent clusters, which makes it more robust and favors a clustering where clusters have sizes of the same order. This algorithm was introduce by Kaufman and Rousseeuw [1987] with the L^1-norm. Here, we use it with the Euclidean distance and the best clustering $C_1^\star, \ldots, C_G^\star$ is the one that minimizes the following criteria:

$$\left\{C_1^\star, \ldots, C_G^\star\right\} \in \underset{\{C_1,\ldots,C_G\}}{\operatorname{argmin}} \sum_{\ell=1}^{G} \sum_{i \in C_\ell} \left\|\underline{\mathbf{W}}_{i\cdot} - \underline{\mathbf{W}}_{C_\ell\cdot}\right\|^2 \quad \text{with} \quad C_\ell \in \underset{i \in C_\ell}{\operatorname{argmin}} \sum_{j \in C_\ell} \left\|\underline{\mathbf{W}}_{i\cdot} - \underline{\mathbf{W}}_{j\cdot}\right\|^2,$$

where $\|\cdot\|$ is the Euclidean norm. The clusters are computed by using `KMedoid` function implemented in the `Python`-library `sklearn_extra`.

3.3 Evaluation of the household clustering

Three different clustering approaches of the households of $\mathcal{I}_{ToU}$ and of $\mathcal{I}_{Std}$, with $G = 4$ clusters, are compared. The first one is a random clustering: an integer between 0 and $G - 1$ is randomly assigned to each household. The second one relies on classical features used to define an households profile: the minimum, maximum, and average consumption in winter and in summer, the peek-hour, and the off-hour (average instant of maximum and minimum consumption). From these rescaled features, k-medoid algorithm is used to cluster the households. The third approach is the one proposed in this chapter and described in the previous section. For a cluster C_ℓ, and for any day t and half-hour h, we will, from now on, consider the average energy consumption $Y_t^h(C_\ell) = 1/|C_\ell| \sum_{i \in C_\ell} Y_t^h(i)$, where $Y_t^h(i)$ is the power consumption record associated with household i.

Figures 7.1 and 7.2 depict, for the three clustering approaches applied on ToU households, the weekly profile of the average power consumption of each cluster $Y_t^h(C_\ell)$ (Figure 7.1) and the normalized power consumption (Figure 7.2), namely the weekly profile of $Y_t^h(C_\ell)/\left(\frac{1}{TH}\sum_{s=1}^{T}\sum_{j=1}^{H} Y_s^j(C_\ell)\right)$. Classical features allow to discriminate households depending on the amount of electricity they consume but does not really catch daily or weekly behavior. Conversely, profile types clearly come off with the proposed method.

The Calinski-Harabasz index, (see Caliński and Harabasz, 1974) is a variance ratio criterion, that evaluate the relevance of the clustering. By denoting $Y(i)$ the vector that contains some of the consumption records associated with household i, and by $Y(C_\ell)$ the one with the average consumption records of cluster C_ℓ and by $Y(\mathcal{I})$ the average consumption records of all households, the score S_{CH} is defined as the ratio of inter-clusters

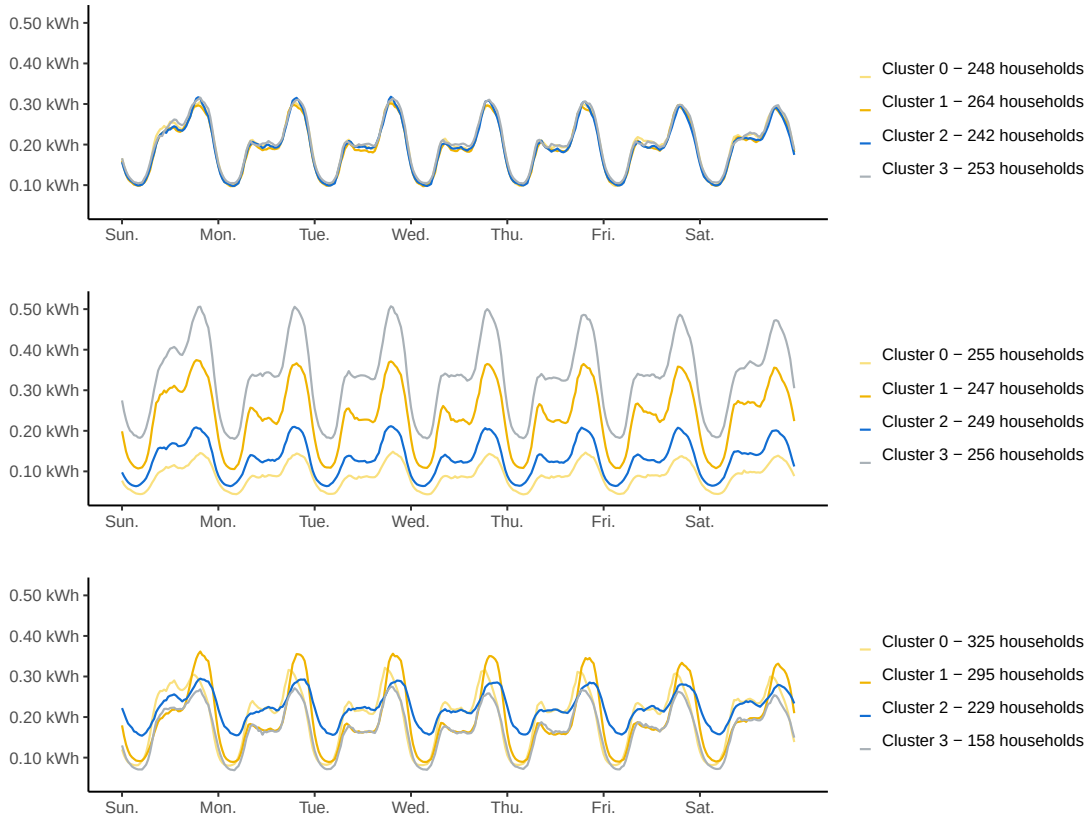


Figure 7.1 – Daily profile of ToU cluster power consumption for a random clustering (top), for a clustering based on "classical features" (middle), and the clustering method proposed in Section 3 (bottom).

variances and intra-cluster variances:

$$S_{\mathrm{CH}} = \frac{\big(|\mathcal{I}| - G\big)\,\mathrm{Var}(C_1,\ldots,C_G)}{\big(G-1\big)\sum_{\ell=1}^{G}\mathrm{Var}(C_\ell)} \quad \text{with}$$

$$\mathrm{Var}(C_1,\ldots,C_G) = \sum_{\ell=1}^{K}\big\|Y(C_\ell) - Y(\mathcal{I})\big\|^2 \text{ and } \mathrm{Var}(C_\ell) = \frac{1}{|C_\ell|}\sum_{i\in C_\ell}\big\|Y(i) - Y(C_\ell)\big\|^2. \quad (7.3)$$

where $\mathrm{Var}(C_\ell)$ is the intra-cluster variance of C_ℓ and $\mathrm{Var}(C_1,\ldots,C_G)$ is the inter-clusters variance. To compute this score, three different vectors Y are considered. First, all the records of the data set are taking into account, namely the records of the entire year 2013; therefore, in Equation (7.3), the vector $Y(i)$ is equal to $\big(Y_1^1(i), Y_1^2(i), \ldots, Y_T^H(i)\big)$. Then we look at the normalized energy consumption records, so

$$Y(i) = \Big(Y_1^1(i), Y_1^2(i), \ldots, Y_T^H(i)\Big) \,/\, \left(\frac{1}{TH}\sum_{t=1}^{T}\sum_{h=1}^{H} Y_t^h(i)\right).$$

Finally the normalized records associated with the sending of incentive signals are selected: only the normalized records associated with tariff Low or High are kept and the others are removed. The results are presented in Table 7.3, where we observe a higher score on non-normalized records for the "classical features" clustering, which is totally

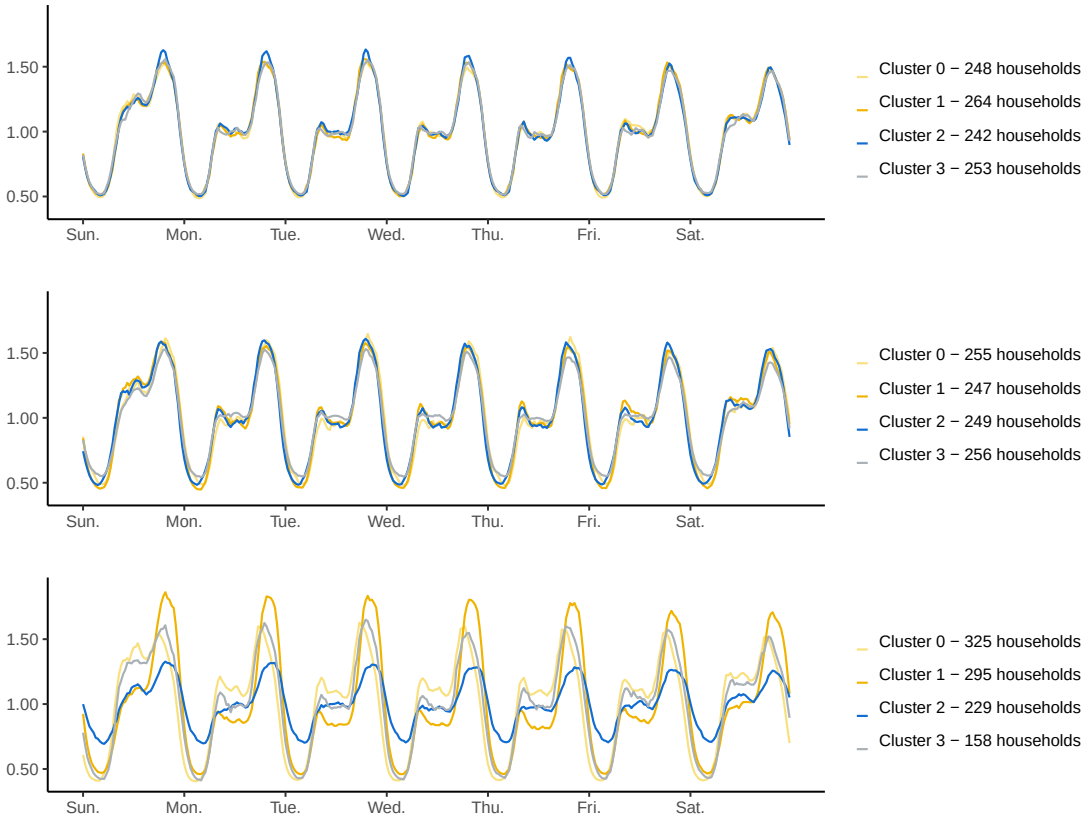


Figure 7.2 – Daily profile of ToU cluster normalized power consumption for a random clustering (top), for a clustering based on "classical features" (middle), and the clustering method proposed in Section 3 (bottom).

coherent with the curves of Figure 7.1. The proposed clustering method seems efficient for catching households behavior. Indeed it gets the higher score for normalized records and Figure 7.2 shows different shape of the power consumption profiles. Moreover, the score is even higher when we select only records associated with special tariff and this increase is more important for ToU consumers that for Std ones. This presumes that the clustering is not only catching a global behavior but also the reaction to a tariff change. It is important to mention that since we want to simulate energy consumption of quite large sub-groups of households (between one and five hundreds households), we did not investigated the optimal number of clusters G (i.e., it was fixed to 4).

In the following sections we present the two data-driven methods that simulate energy consumption profiles associated with the clusters of $\mathcal{I}_{ToU}$ obtained with the method described above. For both approaches, we will train a data generator per cluster. So from now on and for simplicity of notation, a record Y_t^h will refer to $Y_t^h(C_\ell)$, where C_ℓ designs any clusters of set $\mathcal{I}_{ToU}$.

	Rd ToU	F ToU	NMF ToU	Rd Std	F Std	NMF Std
Non-normalized	4 627	40 764	13 432	5 014	5 088	12 971
Normalized	4 950	6 870	14 088	4 834	4 741	14 934
Special tariff - N	5 151	7 033	16 070	4 904	4 694	15 460

Table 7.3 – Calinski-Harabasz score for a random clustering ("Rd"), for a clustering based on classical features ("F"), and the clustering method proposed in Section 3 ("NMF") computed for different consumption record series: all non-normalized records ("Non-rescaled"), all normalized records ("Rescaled"), and normalized records associated with tariff Low and High ("Special tariff - N").

4 Power consumption profile generation with conditional variational autoencoder

The training set made of the T_0 observations (Y_1, X_1), $(Y_2, X_2), \ldots,$ (Y_{T_0}, X_{T_0}) is considered. For a day t, $Y_t = (Y_t^1, \ldots, Y_t^H)$ is the H-dimension vector which corresponds to the daily profile of the half-hour energy consumption of a household cluster. The vector X_t gathers calendar, weather, and tariff information of day t, which will be detailed further.

4.1 Conditional variational autoencoder

4.1.1 Description

The proposed method to generate energy consumption profiles uses the conditional version of variational autoencoders (VAE), which are generative models introduced by Kingma and Welling in 2013 (see Kingma and Welling, 2014 for further details). Autoencoders were mostly used for dimensionality reduction or feature learning (see, among others Rumelhart et al., 1986 and Hinton and Zemel, 1994). They consist of two neural networks: an "encoder" E and a "decoder" D. An autoencoder learns a low dimension representation of a set of H-dimension data points by training both networks at the same time. Indeed, the encoder transforms the H-dimension vectors into d-dimension vectors (with $d \ll H$) and the decoder tries to rebuild initial vectors from the encoder outputs. Considering $Z = \mathrm{E}(Y)$ as the d-dimension output of the encoder for the H-dimensional input Y and $\mathrm{D}(Z)$ as the H-dimension output of the decoder for the d-dimension input Z, the autoencoder is trained to minimize the following "reconstruction loss"

$$L_{\mathrm{AE}} = \frac{1}{T_0} \sum_{t=1}^{T_0} \left\| Y_t - \widehat{Y}_t \right\|^2 = \frac{1}{T_0} \sum_{t=1}^{T_0} \left\| Y_t - \mathrm{D}\big(\mathrm{E}(Y_t)\big) \right\|^2,$$

where $\| \cdot \|$ is the Euclidean norm. Therefore, a data point Y can be represented in a d-dimension latent space by $\mathrm{E}(Y)$.

In the autoencoder framework, there is no constraint on this latent space and the only guarantee is that the representation $Z = \mathrm{E}(Y)$ can be decoded in the original signal $\mathrm{D}(Z) \approx Y$. Moreover, we have no idea what the decoded variable $\mathrm{D}(Z)$ would look like for a value of $Z \notin \{\mathrm{E}(Y_1), \ldots, \mathrm{E}(Y_{T_0})\}$. Thus, there is no guarantee on the shape of the latent space. Without regularization term, for any $d \geqslant 1$, by increasing the number of neurons in both the encoder and the decoder networks, we can create an autoencoder with enough degrees of freedom to fully overfit the data: for example, we could imagine

an autoencoder that encodes the first observation with 1, the second with 2 etc., and thus projects our data in the a one-dimensional space $\{1, \dots, T_0\}$. This points out the need for a regularization term. In VAEs, the introduction of a penalty on the latent space implicitly makes the strong assumption that the distribution of data points $E(Y)$ is close to a given prior distribution. This prior is often set to the standard normal distribution, which we also do in our experiments. From now on, the encoder encodes the distribution of $Z|Y$, which is wanted close to $\mathcal{N}(0, I_d)$. We consider that $Z \,|\, Y \sim \mathcal{N}(\mu(Y), \Sigma(Y))$, where $\mu(Y)$ and $\Sigma(Y)$ are the encoder outputs. The outputs $\widehat{Y}_t$ of the decoder are now $\mathrm{D}(Z_t)$, where the random variable Z_t is sampled from a d-multivariate Gaussian of mean $\mu(Y_t)$ and covariance matrix $\Sigma(Y_t)$, which are the encoder outputs. With $\mathrm{D}_{\mathrm{KL}}(P \,||\, Q)$ as the Kullback-Leibler divergence from Q to P, the VAE is trained by minimizing the following loss

$$L_{\mathrm{VAE}}(\eta) = \frac{1}{T_0} \sum_{t=1}^{T_0} \left\| Y_t - \widehat{Y}_t \right\|^2 + \eta \, \frac{1}{T_0} \sum_{t=1}^{T_0} \mathrm{D}_{\mathrm{KL}} \Big(\mathcal{N}\big(\mu(Y_t), \Sigma(Y_t)\big) \,\big|\big|\, \mathcal{N}(0, I_d) \Big) \,. \tag{7.4}$$

The first term corresponds to the reconstruction error and the second one is a regularization penalty on the latent space. The coefficient η balances these two terms. The calculations below are an adaptation of the ones proposed by Kingma and Welling [2014] to our case-study. They show how, under some assumptions on the existence of a representation of the data in a d-dimensional latent space, minimizing this loss corresponds to conjointly maximizing the likelihood of the observations with the density induced by the data generation process and minimizing an approximation error in the latent space.

The generation of the data is assumed to follow a two-steps process: firstly, a variable Z was sampled from a standard Gaussian and then, Y was sampled from the distribution $p_{\theta^\star}(\,\cdot\,|Z)$. The decoder, parametrized by θ, can model this process: with $Z \sim \mathcal{N}\big(0, I_d\big)$ as input, it generates the variable Y, conditionally to Z, by sampling it from $p_\theta(\,\cdot\,|Z)$, which is an approximation of the true distribution $p_{\theta^\star}(\,\cdot\,|Z)$. In our generation process, we will denote by $q_{_Y}(Z)$ the approximation made by the encoder of the density of $Z \,|\, Y$. The variational autoencoder is trained in a way that $q_{_Y}$ is the Gaussian of mean $\mu(Y)$ and covariance matrix $\Sigma(Y)$, where $\mu(Y)$ and $\Sigma(Y)$ are the outputs of the encoder for the input Y. For $Y \in \mathbb{R}^H$, by using Bayes' theorem and the variables Z sampled from the encoder distribution $q_{_Y}$, the log-marginal likelihood $p_\theta(Y)$ satisfies

$$\begin{aligned}
\log p_\theta(Y) &= \mathbb{E}_{Z \sim q_Y}\big[\log p_\theta(Y)\big] = \mathbb{E}_{Z \sim q_Y}\left[\log \frac{p_\theta(Y|Z)\, p_\theta(Z)}{p_\theta(Z|Y)}\right] \\
&= \mathbb{E}_{Z \sim q_Y}\left[\log \frac{q_Y(Z)}{p_\theta(Z|Y)} + \log \frac{p_\theta(Z)}{q_Y(Z)} + \log p_\theta(Y|Z)\right] \\
&= \mathrm{D}_{\mathrm{KL}}\big(q_Y(Z) \,||\, p_\theta(Z|\,Y)\big) - \mathrm{D}_{\mathrm{KL}}\big(q_Y(Z) \,||\, p_\theta(Z)\big) + \mathbb{E}_{Z \sim q_Y}\big[\log p_\theta(Y|\,Z)\big] \,.
\end{aligned}$$

The first term corresponds to the error made by approximating the distribution $\mathbf{p}_\theta(\,\cdot\,|\,Y)$ with $q_{_Y}$. Thus to conjointly maximizing the log-likelihood and minimizing this approximation error, the loss

$$\mathrm{D}_{\mathrm{KL}}\big(q_{_Y}(Z) \,||\, p(Z)\big) - \mathbb{E}_{Z \sim q_{_Y}}\big[\log p_\theta(Y|Z)\big] \,, \tag{7.5}$$

has to be minimized. The two parts of the equation above are known as the regularization term and the reconstruction term, respectively. We recall that $q_{_Y}$ is the Gaussian distribution of mean $\mu(Y)$ and of covariance matrix $\Sigma(Y)$ and that we assume $Z \sim \mathcal{N}\big(0, I_d\big)$, so

the regularization term is the Kullback–Leibler divergence between $\mathcal{N}\big(\mu(Y), \Sigma(Y)\big)$ and a standard d-multidimensional normal distribution. Moreover, we highlight that if the decoder samples $Y|Z$ from a distribution of the exponential family,

$$p_\theta(Y|Z) = a(Y)b(Z)\exp\big(\eta(Z)T(Y)\big)\,,$$

with θ gathering the functions a, b, η, and T. Then, the second term is explicit. But, for a given Z, the decoder outputs a unique vector $\mathrm{D}(Z) = \widehat{Y}$, so inferring the previous distribution is a tough task. Nevertheless, assuming that $Y \mid Z$ is a multivariate Gaussian of mean $\mathrm{D}(Z)$ and with a known covariance matrix $\sigma^2 I_d$, a very simple expression of the regularization term is obtained:

$$-\log p_\theta(Y|Z) = \frac{1}{2\sigma^2}\|Y - \mathrm{D}(Z)\|_2^2 - \log\big(2\pi^{H/2}\sigma\big)\,.$$

Therefore, the loss defined in Equation (7.5) can be re-written:

$$\frac{1}{2\sigma^2}\|Y - \widehat{Y}\|_2^2 + \mathrm{D}_{\mathrm{KL}}\big(\mathcal{N}\big(\mu(Y), \Sigma(Y)\big)\,||\,\mathcal{N}(0, I_d)\big)\,.$$

Under all the assumptions above, and given the independent observations $Y_1, \dots Y_{T_0}$, to obtain the generative process that best models the real one, we will thus consider the loss

$$L_{\mathrm{VAE}}(\eta) = \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\|Y_t - \widehat{Y}_t\|^2 + \eta\,\mathrm{D}_{\mathrm{KL}}\Big(\mathcal{N}\big(\mu(Y_i), \Sigma(Y_i)\big)\,||\,\mathcal{N}(0, I_d)\Big)\right).$$

We recall that the vectors $\widehat{Y}_t$ are the outputs of the decoder $\mathrm{D}(Z_t)$, where the random variable is sampled from a d-multivariate Gaussian of mean $\mu(Y_t)$ and covariance matrix $\Sigma(Y_t)$. This loss is conjointly maximizing the log-likelihood of the observation with the data generation process distribution: $\log p_\theta(Y_1, \dots Y_t) = \sum_{t=1}^{T_0}\log p_\theta(Y_t)$ and minimizing the approximation error $\sum_{t=1}^{T_0}\mathrm{D}_{\mathrm{KL}}\big(q_{Y_t}(Z)\,||\,p_\theta(Z|\,Y_t)\big)$. It is important to underline that the previous calculations are still valid when all the distributions are conditioned by exogenous variables.

Finally, conditional variational autoencoders (CVAE) – see Kingma et al. [2014] – are an extension of VAE where a vector of exogenous variables X is given as input to both the decoder and the encoder. Adding this conditional information may improve the reconstruction. Figure 7.3 depicts a scheme of the CVAE architecture used in the experiments. The encoder takes as input a daily energy consumption profile Y (so namely a H-vector gathering the half-hourly records of energy consumption) and an exogenous vectors X (with calendar, weather, and tariffs information) and outputs the d-dimension vectors μ and $\log\Sigma$ (it is usual to consider a log-transformation, see Marot et al. [2019]). The vector $\log\Sigma$ is also of dimension d. Indeed, only the diagonal of the covariance matrix Σ is encoded since both approaches (diagonal and full-matrix) were tested and there was no major difference on the reconstruction loss (obviously the regularization term is higher for a full covariance matrix). Since considering a full-matrix (which is symmetric) increases the dimension of encoders outputs (from $2d$ to $d(d+1)/2$) and the CVAE converges slower, we decided to keep a diagonal matrix to encode the covariance matrix Σ. The random variable Z is then sampled and given to the decoder as well as the vector of exogenous variables X. Finally, the decoder outputs $\widehat{Y}$.

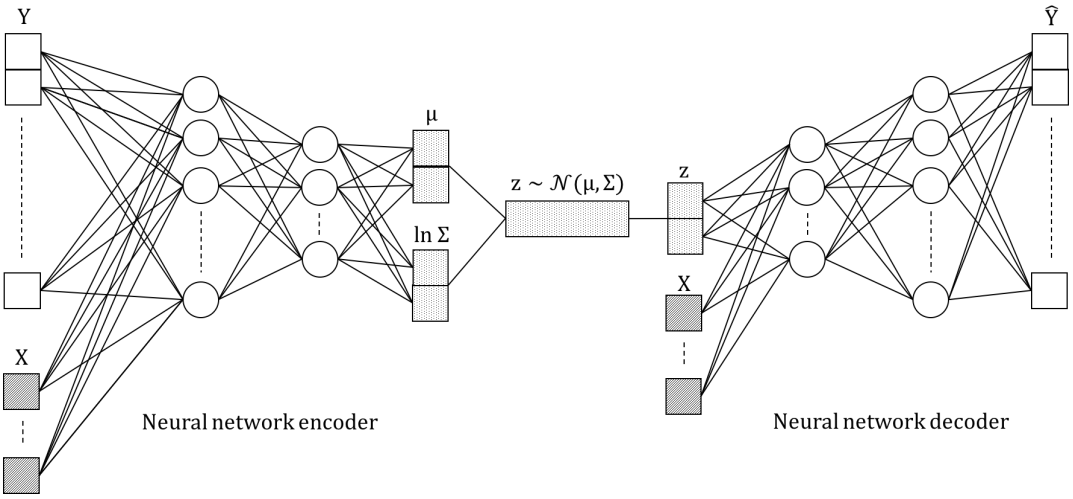


Figure 7.3 – Diagram of a conditional variational autoencoder.

Once the CVAE is trained, the decoder is isolated and used to generate data. For any day s, it is enough to sample a random variables $Z_s \sim \mathcal{N}(0, I_d)$ in the latent space and give it as input to the decoder, combined with a vector of exogenous variables X_s (that could be taken from the original data set or eventually created). Then, the decoder generates a H-vector $\widehat{Y}_s$ that corresponds to a new randomly generated daily consumption profile, for the day s and the contextual variables X_s.

4.1.2 Implementation details

The CVAE were implemented by using the software libraries `Tensorflow` and `Keras` in `Python` programming language. The architecture of a CVAE is defined by the latent dimension d as well as the number of layers and units in encoder and decoder neural networks. We use dense layers which are deeply connected neural network layers. Once the architecture of the CVAEs is set, hyperparameters are chosen: the neural activation functions, the initialization method for neural weights and the parameter η, defined in Equation (7.4), that balances the two terms of the loss. The choice of the architecture and hyper-parameters calibration is detailed in Section 4.2.

In order to optimize a CVAE, so namely to compute weights and bias for each neural of both the encoder and the decoder, the loss is minimized by using the Adam optimizer (see Kingma and Ba, 2015), an extension of stochastic gradient descent method, which is commonly used in deep learning and already implemented in `Keras`. Note that the learning rate of this optimizer is also an hyper-parameter to set before training CVAEs.

Finally, the energy consumption records are rescaled to get values between 0 and 1 by computing the maximum Y_{max} and minimum Y_{min} of the energy consumption observed on the train period. The generated value are re-scaled to get coherent profile, mostly between Y_{min} and Y_{max}.

We recall that the data described in section 7.1 was divided into two data sets: the training set contains 75% for the observations (sampled randomly from the complete data set) and is used to train the CVAE (see Table 7.1); the testing set, made with the remaining

daily observations, is used to calibrate hyper-parameters (see Section 4.2). Finally, as CVAE may converge into local minima, many CVAE are trained and the testing set is also used to select the best one (see Section 4.4).

4.2 Hyper-parameters calibration

The process described below will be applied for each of the cluster defined in Section 3, for which a half-hourly energy consumption profile for each day of 2013 is available.

4.2.1 Methodology

To perform CVAEs hyperparameter calibration we opt for a grid search approach that is simply an exhaustive searching through a manually specified subset of the hyperparameter space. This optimization is guided by the performance metric detailed below, which is simply an evaluation on a held-out validation set. For each set of parameters, namely for each point of the grid, we train a CVAE and test it according to the procedure described below. Once the CVAEs have converged, (we stop the convergence process when the loss is not decreasing any more), we compute the mean squared error (which corresponds to the reconstruction loss) on the testing set made of the observations $Y_{T_0+1}, \dots, Y_T$:

$$\text{MSE} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \left\| Y_t - \widehat{Y}_t \right\|^2$$
$$\text{where} \quad \widehat{Y}_t = D(Z_t) \quad \text{with} \quad Z_t \sim \mathcal{N}\big(\mu(Y_t), \Sigma(Y_t)\big)\,.$$

The architecture and hyperparameters of the CVAE hat reaches to lowest MSE are kept.

4.2.2 Results

We tested different values from 1 to 20 for the latent dimension d and reached a final value of 4, which is coherent with the results in Marot et al. [2019] for the daily energy consumption in France. Moreover, we also performed a principal component analysis (PCA) on the consumption data and found that 4 components were enough to explain more 80% of the variance in the data. We tested CVAEs with one or two hidden layers of 10, 15, 20 or 25 units per layer and concluded that an architecture with a hidden layer of at least 15 neurons performed much better than smaller architectures. We continued to increase the number of layers or the number of neurons per layer, but without improvement in the MSE. Moreover, the number of iterations necessary before convergence increased. So we decided to keep a single hidden layer of 15 units for both the encoder and the decoder.

Concerning the activation function of the neurons; rectified linear unit (ReLU), linear, and sigmoid functions were tested and there was no doubt that the best performance was obtained with a ReLU activation function.

For the initialization of the network weights, we compared various `Keras` initializers (Glorot uniform, HE normal, Lecun normal, Zeros, Ones) and a manual initialization with PCA (as described in Miranda et al., 2014). Classical results were observed. Indeed, when all the weights are initialized with the same weight (ones or zeros), the CVAE converges to a local minimum (all units perform the same calculations). For all the other initializers, we

observed appreciatively the same speed of convergence (a bit faster with the PCA initialization, but it requires upstream calculations) and all converged to the same minimum. Thus, we noticed it does not have a strong impact on the result and therefore the Glorot uniform initializer was selected – it is the default initialization for dense layers. For a new layer of N_{out} neurons connected to an input layer of N_{in} neurons, it draws the $N_{\text{out}}N_{\text{in}}$ weights from a uniform distribution within $\left[-\sqrt{2/(N_{\text{in}}+N_{\text{out}})}, \sqrt{2/(N_{\text{in}}+N_{\text{out}})}\right]$ (see Glorot and Bengio, 2010 for further details).

For the regularization parameter η that balances the two terms (reconstruction and regularization) in the loss function, various strategies to tune its value already exist. For example, Higgins et al. [2017] showed that a constant $\eta > 1$ may outperform classical VAE (defined with $\eta = 1$). Moreover, Liang et al. [2018] and Bowman et al. [2016] considered a moving parameter that gradually increases from 0 to 1 across iterations, linearly and according to a sigmoid, respectively. We tried the three approaches and opted for a constant regularization parameter equal to 10. Finally, we tested various learning rates for Adam optimizer but did not notice major variations in the performance, so we set it to 10^{-3}.

4.3 Conditional variables choice

We tried various combinations of the exogenous variables described in Table 7.2 and selected the one with the lowest MSE on the testing set. For a day t, the conditional vector X_t gathers the variables described below.

Without loss of generality, prices are categorical variables (Low, Normal or High), so, for an day t and an half-hour h, the prices p_t^h are encoded into two binary variables $\mathbf{1}_{p_t^h=\text{Low}}$, and $\mathbf{1}_{p_t^h=\text{High}}$ (if these two variables are null in the same time, the tariff is Normal). The position in the year $\pi_t \in [0,1]$ and the binary variable w_t for the type of day are also considered.

Taking into account the half-hourly temperature $\tau_t^1, \ldots, \tau_t^H$ significantly improves the MSE on the testing set, but the dimension of the conditional variables vector is then quite high. We tried to reduce the dimension of the temperature profile and obtained better results. A PCA was performed on the vectors made of all temperatures at day t (half-hourly records and smoothed temperature). Three components were enough to explain 98% of the variance. Therefore, we only keep the three components provided by the PCA and re-scale them between 0 and 1 to provide the variables $\widetilde{\tau}_t^1, \widetilde{\tau}_t^2, \widetilde{\tau}_t^3$. Then, they are considered as conditional variables (the daily temperature profiles $(\tau_t^1, \ldots, \tau_t^H)$ are not taking into account anymore).

Therefore, for a day t, the vector of conditional variables X_t is made of the binary variables w_t, $\mathbf{1}_{p_t^1=\text{Low}}, \ldots, \mathbf{1}_{p_t^H=\text{Low}}$, and $\mathbf{1}_{p_t^1=\text{High}}, \ldots, \mathbf{1}_{p_t^H=\text{High}}$ and of the continuous variables $\widetilde{\tau}_t^1, \widetilde{\tau}_t^2, \widetilde{\tau}_t^3$, and π_t that lie in $[0,1]$.

4.4 Simulator creation

Finally, we emphasize that CVAEs may converge into local minima. To avoid it, each CVAE is trained 50 times and the one with the lowest MSE on testing set is selected. For each of the cluster presented in Section 3, we thus get a CVAE that takes as inputs

the daily energy consumption profile $Y_t = (Y_t^1 \dots, Y_t^H)$ of the considered cluster (which is rescaled during the training process) and the conditional vector X_t described above. Then, the decoder is isolated and enables the generation of new data. Indeed, for a new vector $X_{t'}$ at a day t', which can either be created or extracted from the data test, we sample a vector $Z_{t'} \sim \mathcal{N}(0, I_d)$ and give these two vectors as inputs of the decoder, which outputs a daily energy consumption profile. The quality of the generated data is evaluated in two situations. First, samples for the conditional vectors $X_{T_0+1}, \dots, X_T$ associated with the training set are generated. Thus, we will measure the ability of the data generators to forecast energy consumption (we will see that we can deduce a foretasted density from the generated samples). Secondly, we will create new vectors X_t for which we modify the variables $\mathbf{1}_{p_t^h=\text{Low}}$, and $\mathbf{1}_{p_t^h=\text{High}}$ in order to measure the impact of tariff changes. These results are presented in Section 6 and compare them with data generated according to a semi-parametric data generator presented below.

5 Semi-parametric generator

The following semi-parametric method based on generalized additive models (GAM), see Section 3 of Chapter 3, is proposed to generate new daily consumption profile data. GAMs model electricity consumption as a sum of independent exogenous variable effects. Here, we assume that there exists a class of functions $\mathcal{F}$, such that, for a given half-hour h and a time instance t, with x_t^h a vector of exogenous variables and p_t^h the tariff, the energy consumption expectation satisfies

$$\mathbb{E}[Y_t^h] = f^h(x_t^h, p_t^h), \quad f^h \in \mathcal{F}. \tag{7.6}$$

After estimating the functions f^h (we detail further the set $\mathcal{F}$ and how GAMs may approximate these functions), we could compute the residuals and try to fit a model on them. They are centered, but a time dependence is observed, so adding a independent white noise to each forecast – as it has been done in the experiments of Chapter 4 – will not provide realistic profiles. Considering probabilistic forecasts of the energy consumption faces the same issues: the aim is to simulate trajectories. A better approach could consider multivariate probabilistic forecasts (which estimate the density of the consumption profile), but to our knowledge, GAMs do not provide such predictions. Therefore, we limit ourselves to more basic models. The simplest one could consists in fitting an autoregressive–moving-average (ARMA) process on the residuals. But residual variance depends, among others, on the tariff and on the half-hour. Thus, a profile generated with an ARMA process will, for example, present a too high variance during the night and a too low variance on peak-hours. The residuals are considered day by day (so the time dependence from a day to another is ignored – note that this is also the case with the CVAE-based model). we are aware that ignoring the day-to-day time dependence could be problematic as there are almost always long-term serial correlations present in the residuals from these models. Once generated, these noise trajectories will be added to the expected consumption forecast. Considering the daily residuals vector as multidimensional Gaussian and estimated its covariance matrix makes the generation of new samples very easy. But, with such a model, the tariff-dependence is lost. We propose an approach based on a conjoint estimation of both mean and variation of the energy consumption. Then, we tried to used Gaussian copula to create trajectories, applying the methods proposed in Pinson et al. [2009] for renewable energy scenarios (or trajectories) generation. We faced an important problem: as soon as the function f^h is not very well-estimated, the

residuals variance comes, in majority, from the estimation error. More precisely, a bad estimation of the expected consumption leads to an increase of the estimated standard deviation.

As the focus is on generating realistic a profile (and not necessary on having the best forecast in expectation), the standard deviation used to simulate data must reflect the variability observed in energy consumption data. Thanks to the causality model of Section 3.1, that is now fitted on cluster consumptions (and not on individual ones), we can estimate the standard deviation of the noise as a function of the tariff and the half-hour h. We recall that we denote by $\sigma^h(p)$ the approximation of the standard $\sqrt{\mathrm{Var}[Y^h(i) \,|\, P = p]}$ deviation associated with the half-hour h and the tariff p – see Equation (7.2). It is used to normalize the residuals, which should then be centered and of variance 1 (but not independent). Finally, we consider the standardized residual vectors and compute an estimation of their correlation matrix Σ. We can now generate new data points this way:

$$\begin{bmatrix} Y_t^1 \\ \vdots \\ Y_t^H \end{bmatrix} = \begin{bmatrix} f^1\big(x_t^1, p^1\big) \\ \vdots \\ f^H\big(x_t^H, p^H\big) \end{bmatrix} + \big(\sigma^1(p^1), \ldots, \sigma^H(p^H)\big)^{\mathrm{T}} E_t \quad \text{where} \quad E_t \sim \mathcal{N}(0, \Sigma)\,. \tag{7.7}$$

Functions $(f^h)_{1 \leqslant h \leqslant H}$ are estimated with GAMs and the exogenous vector x_t^h gathers the temperature of the time instance at the considerate half-hour τ_t^h, the smoothed temperature $\bar{\tau}_t$, the position in the year π_t, the binary variable w_t, which is equal to 1 if the day considered is a working day and 0 otherwise. For each half-hour h, we set the same underlying GAM:

$$f^h(x_t^h, p_t^h) = s_\tau^h(\tau_t^h) + s_{\bar{\tau}}^h(\bar{\tau}_t) + s_\pi^h(\pi_t) + \alpha^h w_t + \xi_{\mathrm{Low}}^h \mathbf{1}_{p_t^h = \mathrm{low}} + \xi_{\mathrm{high}}^h \mathbf{1}_{p_t^h = \mathrm{High}}\,. \tag{7.8}$$

Therefore, $\mathcal{F}$ is the set of functions that can be written this way. The s_τ^h, $s_{\bar{\tau}}^h$, and s_π^h functions are catching the effect of the temperatures and of the yearly seasonality. They are approximated by cubic splines. The `mgcv` R-package allows to estimate the coordinates of the splines in their basis and the coefficients α^h, ξ_{Low}^h, and ξ_{High}^h that catch day of the week and tariff effects. The estimation of the matrix Σ, which is used to generate profiles with correlations between temporal intervals of the same day, is computed as follows. If the model defined by Equation (7.7) was true, residuals $Y_t^h - f^h(x_t^h, p_t^h)$ should be Gaussian of mean 0 and standard deviation $\sigma^h(p_t^h)$. Thus the vector of standardized residuals $\mathrm{e}_t = (e_t^h)_{1 \leqslant h \leqslant H}$ is considered, where

$$e_t^h = \frac{Y_t^h - f^h(x_t^h, p_t^h)}{\sigma^h(p_t^h)}\,.$$

Assuming the model above, the covariance matrix Σ of vectors $\mathrm{e}_1, \ldots, \mathrm{e}_{T_0}$ should have 1 on the diagonals and all other coefficients between -1 and 1. To deal with our imperfect modeling and avoid again the problem of high standard deviation coming from the estimation error, Σ is approximated by the empirical correlation matrix of vectors $\mathrm{e}_1, \ldots, \mathrm{e}_{T_0}$. From the T_0 observations $\mathrm{e}_1, \ldots, \mathrm{e}_{T_0}$, which are assumed independent, of the H-dimensional random vector $\mathrm{e} = (e^1, \ldots, e^H)$, the coefficients of the $H \times H$-correlation matrix Σ are defined by

$$\Sigma_{i,j} = \frac{\mathrm{cov}(e^i, e^j)}{\sqrt{\mathrm{Var}(e^i)\,\mathrm{Var}(e^j)}}\,, \quad \text{where} \quad \mathrm{cov}(e^i, e^j) = \mathbb{E}(e^i e^j) - \mathbb{E}(e^i)\mathbb{E}(e^j)\,.$$

We point out that in the case of random variables $\mathrm{e}^1, \ldots, \mathrm{e}^H$ of standard deviation 1 (we assume it in the semi-parametric simulator described in Section 5), covariance and correlation matrices are equal. In there, $\Sigma_{i,j}$ is estimated by replacing covariances and variances of random variables e^i and e^j by the their empirical estimations:

$$\mathrm{cov}(e^i, e^j) \approx \frac{1}{T_0 - 1} \sum_{t=1}^{T_0} \left(e_t^i - \bar{e}^{\,i}\right)\left(e_t^j - \bar{e}^{\,j}\right)$$

$$\text{and} \quad \mathrm{Var}(e^i) \approx \frac{1}{T_0 - 1} \sum_{t=1}^{T_0} \left(e_t^i - \bar{e}^{\,i}\right)^2, \quad \text{with} \quad \bar{e}^{\,i} = \frac{1}{T_0} \sum_{t=1}^{T_0} e_t^i \,.$$

Therefore, this estimation of the correlation matrix Σ makes it possible to model the correlations between the consumption profiles of two half-hours of the same day, whereas keeping a variance of the residuals that varies according to the half-hour and the price.

Remark 22. *We could also have modeled these correlations by fitting an ARIMA model on the residuals; such an approach would have taken into account day-to-day time dependence. But on the data, it was quite difficult to estimate a significant ARIMA model. Moreover, we were not comfortable with a possible too high divergence of the residuals: after many rounds of simulation, the process may be very far from* 0.

6 Evaluation of the data generators

6.1 Evaluation metrics

By generating lots of energy consumption profiles from the simulators, an estimation of their densities can be obtained. Therefore, we use some proper scoring scores from probabilistic forecast evaluation to assess the quality of our generators. The three scores detailed below allow to evaluate the data generated on the testing period and compare both generators. For a day t of the testing set, from the vector of exogenous variables X_t, both generators output H-random vectors that are assumed to be drawn from an underlying distribution $\widehat{F}_t$. These distributions approximate the true and unknown H-dimensional distributions F_t from which the observation $(Y_t^1, \ldots, Y_t^H)$ is actually drawn. We generate $N = 200$ samples $\widehat{Y}_t^{(1)}, \ldots, \widehat{Y}_t^{(N)}$ for each generator. From these H-random vectors, we can approximate the three scores described below, that measure the adequacy between the observation vectors Y_t and the distributions $\widehat{F}_t$.

First of all, for a distribution F, and a vector of observation y, the root mean squared error is considered: $\mathrm{RMSE}\,(F, y) = \left\|\mathbb{E}[Y] - y\right\|$, where Y is a random vectors distributed according to F. The first score is thus the RMSE between the expectation of the distribution $\widehat{F}_t$ (which we approximate with empirical mean of the generated samples) and the observation Y_t:

$$\mathrm{RMSE}\,(\widehat{F}_t, Y_t) \approx \left\| \frac{1}{N} \sum_{i=1}^{N} \widehat{Y}_t^{(i)} - Y_t \right\|.$$

Here, the expectation of the distribution $\widehat{F}_t$ is actually seen as a forecast of the energy consumption Y_t. But to evaluate the quality of $\widehat{F}_t$, a criterion including the variance and shape of the densities is necessary.

The two other scores are proper scoring rules used to evaluate weather ensembles or temporal trajectories generated by a statistical method (e.g., copula model). The energy score, introduced in Gneiting and Raftery [2007], generalizes the univariate continuous ranked probability score (CRPS) and is defined as

$$\mathrm{EN}(F, y) = \mathbb{E}\Big[\|Y - y\|\Big] - \frac{1}{2}\mathbb{E}\Big[\|Y - Y'\|\Big],$$

where Y and Y' are two independent random vectors that are distributed according to F. This score is approximated by splitting the generated samples in two groups $\widehat{Y}_t^{(1)}, \ldots, \widehat{Y}_t^{(N/2)}$ and $\widehat{Y}_t^{(N/2+1)}, \ldots, \widehat{Y}_t^{(N)}$:

$$\mathrm{EN}(\widehat{F}_t, Y_t) \approx \frac{2}{N}\sum_{i=1}^{N/2}\left\|\widehat{Y}_t^{(i)} - Y_t\right\| - \frac{1}{N}\sum_{n=1}^{N/2}\left\|\widehat{Y}_t^{(i)} - \widehat{Y}_t^{(N/2+i)}\right\|.$$

Scheuerer and Hamill have shown that the ability of energy score to detect correctly correlations between the components of the multivariate distribution was limited (see Scheuerer and Hamill, 2015 for further details). To remedy, they introduced the variogram score of order p:

$$\mathrm{VG}_p(F, y) = \sum_{h,h'=1}^{H}\left(\left|y^h - y^{h'}\right|^p - \mathbb{E}\Big[\left|Y^h - Y^{h'}\right|^p\Big]\right)^2, \tag{7.9}$$

where Y is a random vectors distributed according to F. On simulated data, they compared the performance of different scores (including the energy score) with the variogram scores for various p. This score is approximated with:

$$\mathrm{VG}_p(\widehat{F}_t, Y_t) \approx \sum_{h,h'=1}^{H}\left(\left|Y_t^h - Y_t^{h'}\right|^p - \frac{1}{N}\sum_{i=1}^{N}\left|\left(\widehat{Y}_t^{(i)}\right)^h - \left(\widehat{Y}_t^{(i)}\right)^{h'}\right|^p\right)^2.$$

We emphasize that for all the scores above, the smaller the value, the better the forecast.

6.2 Numerical results

For each cluster and each day t of the testing set, we compute, for both generators (CVAE-based and GAM-based) the three scores (thanks to the 200 generated samples). Results are represented by boxplots in Figure 7.4. Moreover, for the first three days of the testing set (that are actually the first three days of 2013), 20 samples generated by the simulators for the 4 clusters, their empirical means (computed on all the samples) and the corresponding observations Y_t are plotted in Figure 7.5.

It is quite difficult to discriminate significantly both generators from these scores, but some conclusions may still be drawn. First, RMSE bloxplots and plots suggest that GAM-based generators work better than those that use CAVE when it comes to generating the average value of the original data (which is approximated by the empirical mean of the samples). However, the energy score is slightly lower for the non-parametric approach (namely for CVAE-based simulator) than for the semi-parametric one (GAM-based simulator). Thus, the method that consists in adding a noise term to a forecast in expectation may have some limits whereas CVAEs seem to catch correctly the distributions of daily energy consumption.

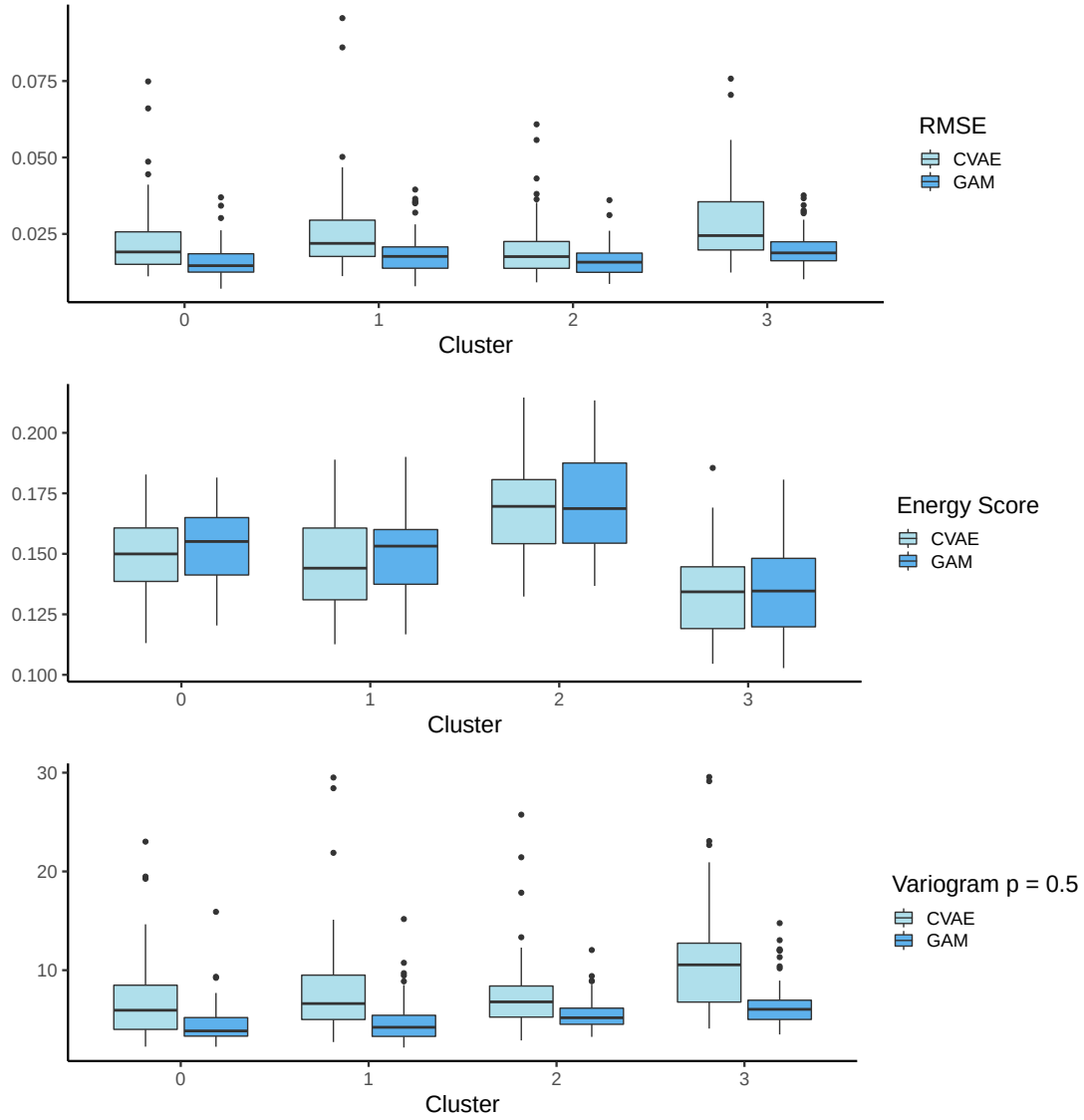


Figure 7.4 – Boxplots. From left to right Root Mean Squared Error (RMSE), Energy Score and variogram for $p = 0.5$ evaluated for each day of the data test set.

Experiments of Scheuerer and Hamill [2015] highlight that, when the estimation of the average value of the original data is incorrect (namely when the expectation of F differs from the expectation of y in Equation (7.9)), variogram scores increase. Moreover, a too low or a too high variance – when the variance of F differs from the one of y – also increases variogram. Given the variogram scores and the plots, we conclude that CVAE-based generators face an estimation of expected energy consumption worst than the semi-parametric generator but provide also samples with a too low variance (since the observations are not close to any of the generated data). Conversely, GAM-based generators provide sample with a correct variance: each of the observations is close to, at least, one of the generated data. But the trajectories are very erratic (whereas ones generated with CVAE-based generators are quite smooth); this also leads to a quite high variagram score.

Moreover, in the CVAE approach, consumption values from an half-hour to another

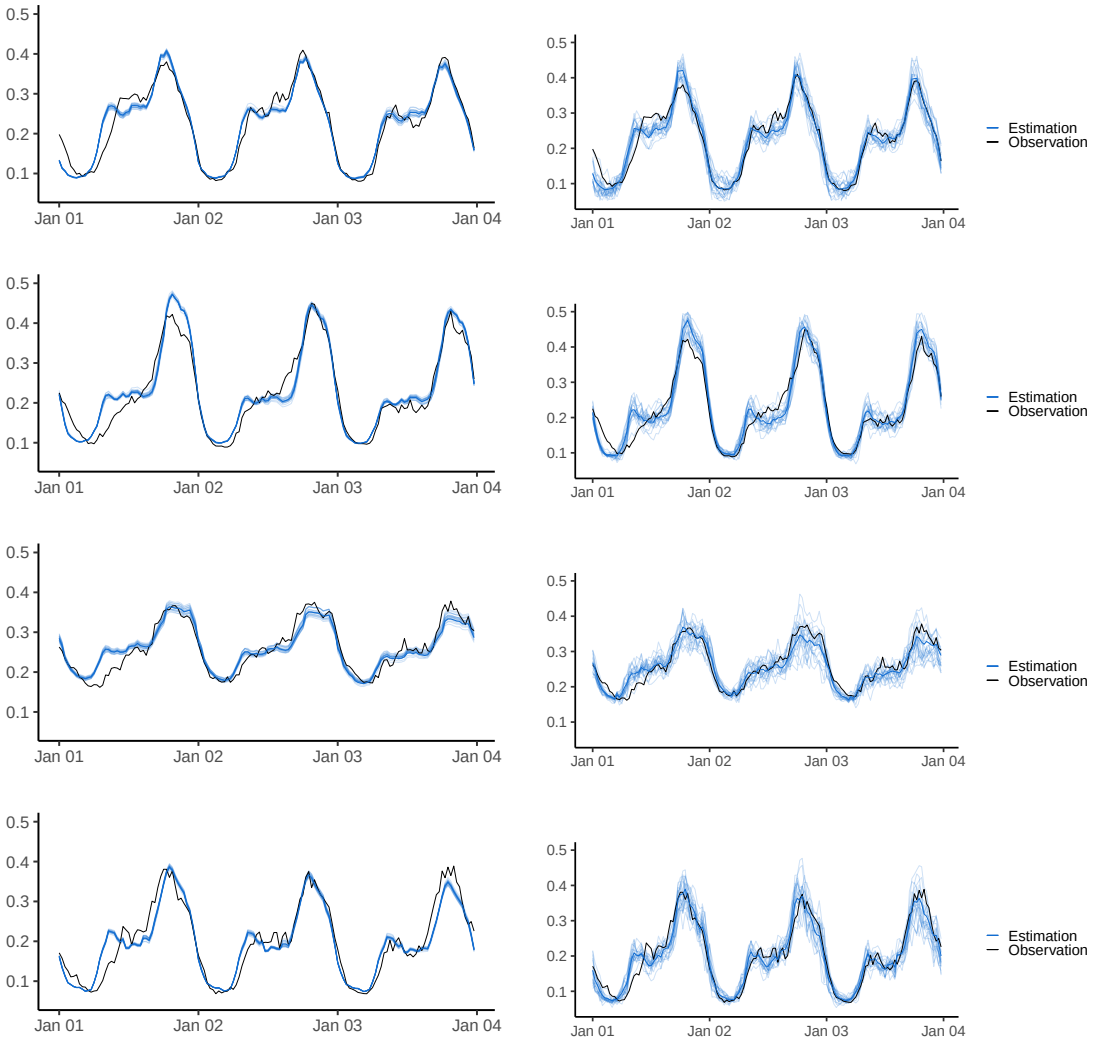


Figure 7.5 – Left: data generated with the CVAE-based generator. Right: data generated with the GAM-based generator. Blue lines: for every cluster over the first three days of the testing set, 20 energy consumption profiles and empirical mean profile, calculated on 2 200 samples (in bold), obtained by giving, to the two simulators, the exogenous variables observed over this period. Black line: real observed profiles.

are very correlated, when in the semi-parametric one, consumption profiles are more erratic. Observations suggest that the real variances and correlations lie somewhere in between. The semi-parametric method is very sensitive to the standard deviation $\sigma^h(p)$ estimations. Thus, over-estimating these variances, provide, for sure, very different samples, which may be also very erratic. Concerning CVAE-based generator, the variance of the samples could manually be increased by generating the decoder inputs according to $\mathcal{N}(0, \sigma^2 I_d)$ with $\sigma > 1$.

Finally, we emphasize that in the semi-parametric approach, the variance depends only on the tariff and on the half-hour, whereas in the CVAE, all exogenous variables are taking into account. Moreover, the next section presents some strong advantages of the CVAE generator.

6.3 Impact of the tariff

In these last experiments, for a day t of the testing set, three different conditional vectors X_t^{Normal}, X_t^{Low} and X_t^{High} are considered. The tariff is Normal for all the day long for X_t^{Normal}. For the vector X_t^{Low}, Low tariff applies from 4:30 to 9:30 a.m., and Normal one otherwise, finally, tariff is Normal expect from 7:30 to 10 p.m. where it is High for X_t^{High}. For all other components, namely for the calendar and weather variables, X_t^{Normal}, X_t^{Low}, and X_t^{High} are equal to X_t. Still for the first three days of the testing set, 20 samples generated by the generators for the 4 clusters and their empirical means (computed with all the sample) are plotted in Figure 7.6.

For both data generators, an increase of the consumption when tariff Low is applied and a decrease when the tariff is High are observed. For the GAM-based generator, the effect of the tariff is very interpretable, it is actually measured by coefficients ξ^h_{Low} and ξ^h_{High} of equation (7.8). This model makes actually this assumption that the tariff effect only depends on the half-hour. Moreover, matrix Σ models the correlations between the energy consumption at two half hours of the same day; this implicitly assumes that these correlations do not change according to the applied tariff profile. Conversely, CVAE-based generator does not have this assumption and the effect of a tariff may differ from a day to another.

Moreover, two effects that cannot be modeled by the semi-parametric approach are observed. First, the fall of the energy consumption occurs a little bit before the effective establishment of a special tariff High and continues a little after it is stopped. Thus, the effect of the High tariff exceeds the time window in which the special tariff is actually applied. This is called a side effect. Secondly, in comparison to a day of Normal tariff, when tariff Low is applied in the morning, there is a drop of the consumption in the afternoon and evening. Similarly, we observe a little increase of the consumption in the afternoon when the tariff is High during the evening. Therefore, the fall or rise in consumption shifts to another time of the day when a special tariff is applied over a time window. This is called a rebound effect. These side and rebound effects are well known behaviors of consumers and it is very valuable that the generator detects them.

The main drawback of this non-parametric generator is the generation of non-intuitive consumption profiles when the input is a tariff profile never observed in the training set, like an entire day of High tariff for example. This shows that the method has a limited gen-

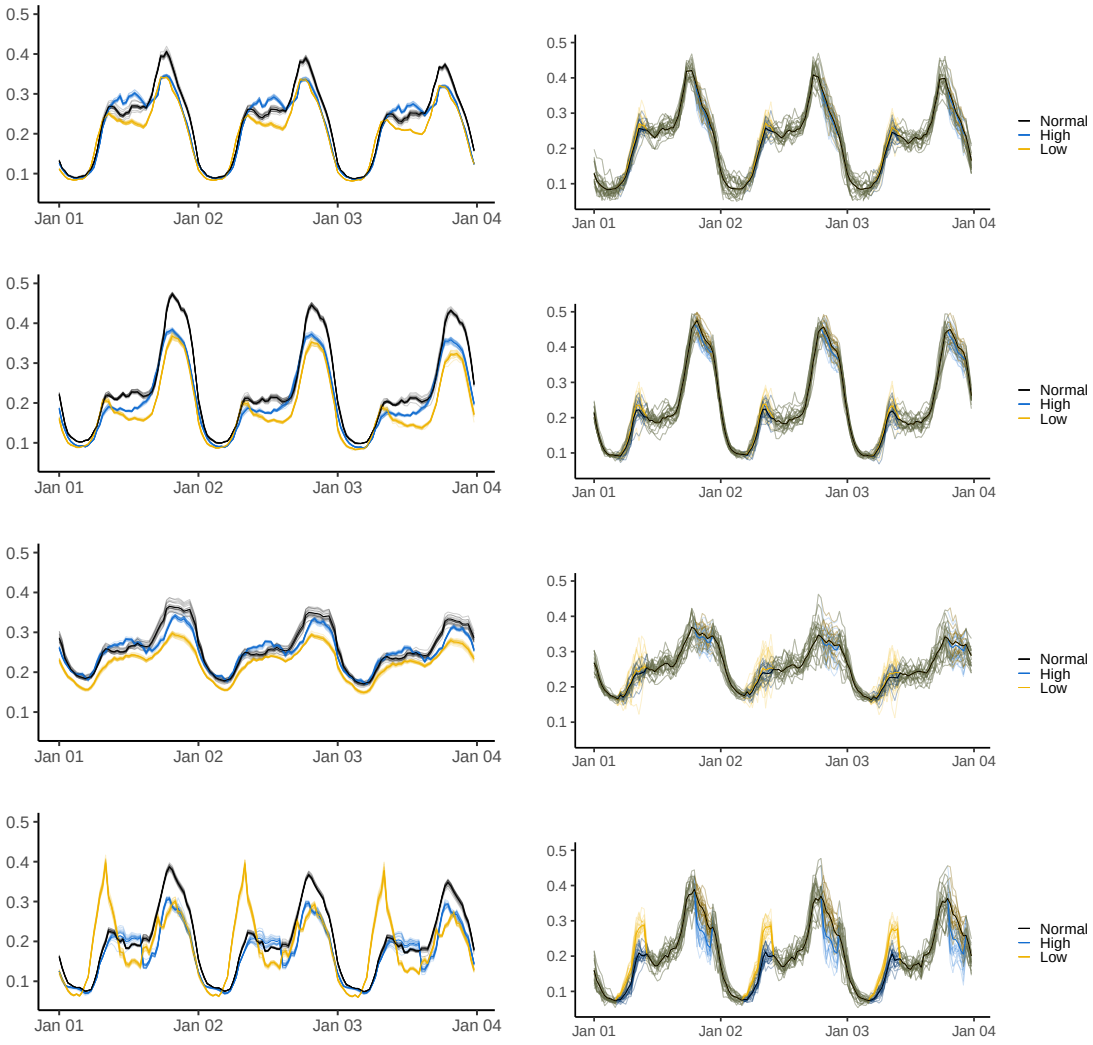


Figure 7.6 – Left: data generated with the CVAE-based generator. Right: data generated with the GAM-based generator. Black lines: for every cluster on the first three days of the testing set, 20 energy consumption profiles and empirical mean profile, computed over 200 samples (in bold), obtained by giving, to the two simulators, a Normal tariff for every half-hour and the weather and calendar variables observed over this period. Blue lines: same plots but with a High tariff in the evening and Normal tariff otherwise. Yellow lines: same plots but with a Low tariff in the early morning and Normal tariff otherwise.

eralization capacity. Enlarging the data set, especially the variety of price signals, would eliminate this limitation. To deal with the lack of variability in the sent tariff profile of the original data set, we could also imagine an online data generator: when a new tariff profile is sent, the observed consumption is integrated in the training set and the data generator is updated. The use of transfer learning methods could also improve the realism of the generated data. This machine learning field focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. Therefore, by combining data sets of consumer responsiveness to various DR programs (*i.e.* by combining diverse knowledge of electricity demand in the face of tariff changes), a data set with a higher variability in the sent tariff profiles may be obtained. On the other hand, for a full day of tariff High, the semi-parametric model generates samples with an energy consumption below the typical one for each half-hour, which is unrealistic since electricity uses cannot be delayed indefinitely.

Figure 7.6 shows that tariff-responsiveness vary from a cluster to another, i.e., rebound or side effects are not always observed and the amount of electricity over or under consumed also depends on the considered cluster. These results fully illustrate the motivation behind the use of the causality model to cluster consumers.


MB

8

Contextual bandits for personalized demand side management

This synthesis chapter generalizes the theoretical results of Chapter 4 to provide a contextual-bandit approach for personalized demand side management. The previous assumption of a homogeneous population is dropped and, by clustering of non-homogenous population into several homogenous groups (by the method of Chapter 6), a protocol for personalized demand side management, which can take into account many operational constraints, is set out. The performance of our strategies is measured in quadratic losses through a regret criterion and we offer $T^{2/3}$ upper bound on this regret (up to poly-logarithmic terms). Experiments, using the data simulator provided in Chapter 7 to test the proposed strategies, conclude the chapter.

1 Introduction

This chapter aims to manage the power consumption of some household clusters, by personalizing the tariff send according to the cluster behaviors. In Chapter 4, we proposed a contextual-bandit approach for the demand side management of an homogeneous population by offering price incentives. We recall that a target mean consumption was set at each round and that the mean consumption was modeled with a generalized additive model that took into account many covariates: some contextual variables such as the temperature, weather, and so on, as well as the distribution of prices sent. The performance of our strategies was measured in quadratic losses through a regret criterion.

As in Chapter 6, we now consider not only the aggregated power consumption of some clusters, but also the aggregated power consumption of higher aggregation levels (larger regions, entire population, etc.). The electricity consumption of each cluster is modeled a different generalized additive model and for some of the aggregation levels, we will set target consumptions. We still consider quadratic losses. We then proposed an algorithm to perform the management of this non-homogeneous population. This approach, compared to the one of Chapter 4, comes with two main refinements. First, the sent of the tariff are personalized: the population is not anymore homogeneous and the tariff chosen by the electricity provider for a sub-population depends on both its consumption behavior and its responsiveness to tariff changes. Second, global (namely for the entire population) and local (for a sub-population) target consumptions can be considered which is valuable to deal with the integration of decentralized and intermittent energies.

In Section 2 we provide a modeling of this management system. Then, an optimistic algorithm, adapted from the one provided in Chapter 4 is stated and its regret is analyzed in Section 3; we therefore show how to control the cumulative loss through a $T^{2/3}$ regret bound with respect to the best constant price allocation. Finally, experiments on the Low Carbon London data set are presented in Section 4. They rely on the clustering approach described in Chapter 7, which divides correctly the households of the Low Carbon London project according to their responsiveness to a tariff profile. The results are obtained for data simulated with the CVAE-based data generator also presented in this previous chapter, this gives an idea of the robustness of the proposed solution, initially built for data generated by generalized additives models.

2 Setting and model

2.1 Consumption modeling

We consider a large population of households of some electricity provider, constituted of G homogeneous sub-populations; the households have been previously gathered into G clusters according to their location, consumption behavior and price-responsiveness, for example by using methods introduced in Chapters 6 and 7. To manage demand the electricity provider sends some incentives signals to its customers. We assume that $K \geqslant 2$ price levels (tariffs) are available.

Remark 23. *There is no reason to consider that the same tariffs are proposed to each cluster, in particular if several types of contracts are offered, with different degrees of tariff flexibility, and if the segmentation of the households is based on contract type: we could*

consider K^1 tariffs for the cluster 1, K^2 for the cluster 2 and so on. For the ease of notation and with no loss of generality (by setting $K = \max_i K^i$), we consider the same number K of tariffs for every cluster. We point out that, pricing options can however be very different from one cluster to another.

To model the power consumption of each of the clusters, we consider Model 1 introduced in Chapter 4. Therefore, at each round t, the power consumption of a cluster $i \in \{1, \dots, G\}$ depends on the electricity prices and on some exogenous factors (temperature, wind, season, day of the week, etc.), which form a context vector $x_t^i \in \mathcal{X}^i$ (where $\mathcal{X}^i$ is some parametric space). Moreover, at round t, the electricity provider sends the tariff k to a share $p_{t,k}^i$ of the households of cluster i; we denote by p_t^i the convex vector $(p_{t,1}^i, \dots, p_{t,K}^i)$. As cluster i is rather homogeneous, it is unimportant to know to which specific household a given signal was sent; only the global proportion $p_{t,k}^i$ matter. The power consumption of cluster i, for the prices level p_t^i, is denoted by $Y_{t,p_t^i}^i$ and is assumed to be of the form:

$$Y_{t,p_t^i}^i = \varphi^i(x_t^i, p^i)^{\mathrm{T}} \theta^i + \text{noise}(i)\,,$$

where the mapping function $\varphi^i : \mathcal{X}^i \times \mathcal{P}^i \to \mathbb{R}^{d^i}$ is known and a parameter vector $\theta^i \in \mathbb{R}^{d^i}$ has to be estimated.

Remark 24. *The i-dependence in the context vector can be due to the location of the clusters: the power consumptions of clusters from different regions are affected by different weather variables, so x_t^i could contain local weather information. Moreover some of the exogenous variables may be relevant for some of the clusters and useless for the others.*

We may set some restrictions on the convex combinations $p_t^i = (p_{t,1}^i, \dots, p_{t,K}^i)$ that can be picked: we denote by $\mathcal{P}^i$ the set of legible allocations of price levels. Finally, for two clusters i and j, we assume that conditionally to the contextual vectors x_t^i and x_t^j and conditionally to the tariffs picked p_t^i and p_t^j, the power consumption of clusters i and j are independent. As in Model 1, the noise term in the power consumption depends on the price vector. Therefore, the previous assumption can be written informally

$$\mathrm{Cov}\big(\,\text{noise}(i)\,,\,\text{noise}(j)\,|\,p^i,\,p^j\,\big) = 0\,.$$

We highlight that, at a round t, price levels p_t^i and p_t^j may be correlated because the electricity provider manages both clusters in the same time. This modeling of the G power consumptions is stated with full rigor in Model 4 below.

Model 4: *G independent (conditionally to contextual variables) submodels. For a given cluster i, when the electricity provider picks the convex vector $p^i \in \mathcal{P}^i$, the consumption of the cluster obtained at round t equals*

$$Y_{t,p^i}^i = \varphi^i(x_t^i, p^i)^{\mathrm{T}} \theta^i + (p^i)^{\mathrm{T}} \varepsilon_t^i\,.$$

The noise vectors $\varepsilon_1^i, \varepsilon_2^i, \dots$ are ρ^i–sub-Gaussian[1] i.i.d. K-dimensional random variables with $\mathbb{E}[\varepsilon_1^i] = (0, \dots, 0)^{\mathrm{T}}$. Moreover, for any clusters $i \neq j$, and tariffs k, $k' \in \{1, \dots, K\}$, $\mathrm{cov}(\varepsilon_{t,k}^i, \varepsilon_{t,k'}^j) = 0$. We denote by $\Sigma^i = \mathrm{Var}(\varepsilon_1^i)$ their covariance matrix.

[1] A d–dimensional random vector ε is ρ–sub-Gaussian if for all $\nu \in \mathbb{R}^d$, one has $\mathbb{E}\big[e^{\nu^{\mathrm{T}}\varepsilon}\big] \leqslant e^{\rho^2\|\nu\|^2/2}$.

2.2 Targets and loss function

In Chapter 4, we focused on the management of the global consumption. Indeed, we aimed to influence the power consumption of the entire population of customers, which was homogeneous. The present work is a generalization of the previous one: with $G = 1$, we recover previous results related with Model 1.

We now observe the consumption of G homogeneous clusters and can send "personalized" (namely, cluster-depending) signals: at any round t, we can choose $p_t^i \neq p_t^j$ for two clusters i and j. In that follows, we will denote by $\boldsymbol{p}_t$ the matrix which contains the price vectors $(p_t^1, \dots, p_t^G)$:

$$\boldsymbol{p}_t \stackrel{\triangle}{=} \left(p_t^1 \middle| p_t^1 \middle| \dots, \middle| p_t^G \right),$$

and by $\boldsymbol{\mathcal{P}} = \mathcal{P}^1 \times \cdots \times \mathcal{P}^K \subset \Delta_K^G$, the set of legible allocations of price levels $\boldsymbol{p}_t$. We still aim to manage the global power consumption, see Example 10 below.

Example 10: *Global consumption management.* To manage demand response, for each round t, the electricity provider sets a global consumption target c_t to reach. It chooses the $K \times G$-matrix $\boldsymbol{p}_t \in \boldsymbol{\mathcal{P}}$ in order to influence the global consumption of its customers, and observes the resulting power consumption $Y_{t,\boldsymbol{p}_t} = \sum_{i=1}^G Y_{t,p_t^i}^i$. As in Chapter 4, the aim is to control the global power consumption $Y_{t,\boldsymbol{p}_t}$. The main difference comes from the segmentation: the algorithm is supposed to find flexible clusters to better manage demand. We still measure the discrepancy between the observed consumption $Y_{t,\boldsymbol{p}_t}$ and the target c_t via a quadratic loss:

$$\left(Y_{t,\boldsymbol{p}_t} - c_t\right)^2 = \left(\left(\sum_{i=1}^G Y_{t,p_t^i}^i \right) - c_t \right)^2.$$

To deal with this global management, we could directly apply the algorithm of Chapter 4, considering, at each round t, the unique observation $Y_{t,\boldsymbol{p}_t}$. But, in such a situation, we will have to estimate the power consumption associated with each set of $p_t^1, \dots, p_t^G$. Depending on how electricity consumptions and tariffs are liked (namely, depending on the underlying generalized additive model), we may pay a K^G (the dimension of $\boldsymbol{\mathcal{P}}$) factor in the regret bound. Using the observations of $Y_{t,\boldsymbol{p}_t}^1, \dots, Y_{t,\boldsymbol{p}_t}^G$, we will see below that we can reach a $K2^{G-1}G$ factor, which is better.

Electrical girds may be subjected to geographical constraints and the provider could wish a more local management of the demand. For example, we could imagine that two clusters i and j are in two different regions and that many wind farms are present in the region of cluster i, while there is none in the region of cluster j. Even if the network connects the two clusters, in windy weather, the electricity provider could want to encourage cluster i to consume "locally" by sending low prices, while maintaining normal prices for cluster j; this will minimize energy loss. To do so, it has to consider local consumption targets, see Example 11 below.

Example 11: *Local consumptions management.* The electricity provider aims to manage the clusters locally: each cluster may correspond to a region which has its own electricity production, its own grid and its own weather. It sets some local targets c_t^i for each cluster

i at each round t. Then, it chooses the matrix $\boldsymbol{p}_t$ and observes the G local consumptions. Here, the clusters' targets, tariffs and consumptions (conditionally to the exogenous variables) are completely independent, as if the G electrical grids associated with them are disjoint by design. In such a situation, the global loss is simply the sum of the local quadratic losses:

$$\sum_{i=1}^{G} \left(Y^i_{t,p^i_t} - c^i_t \right)^2 .$$

This particular example may be handled with G algorithms executed independently and in parallel (using the algorithm proposed in Chapter 4. By a union bound and by summing the regrets, we will pay a G factor in the regret bound, which is better than the bound we reach below. The interest of our approach is to manage targets common to certain groups.

We may combine the two previous examples and consider a consumption management both global and local. In a even more general modeling, we assign targets to some aggregations at variable levels of the clusters. More precisely, for a subset of clusters $\mathrm{g} \in \mathcal{P}(\{1,\dots,G\})$, we introduce, at round t, the subtarget c^{g}_t. For example, with this approach, we may introduce a global target c^{TOT}_t; two locals targets for clusters 1 and 2 denoted by c^1_t and c^2_t, respectively; and a subtarget for clusters 2 and 3 denoted by c^{23}_t; and consider the losses:

$$\left(\left(\sum_{i=1}^{G} Y^i_{t,p^i_t} \right) - c^{\mathrm{TOT}}_t \right)^2 + \left(Y^1_{t,p^2_t} - c^1_t \right)^1 + \left(Y^2_{t,p^2_t} - c^2_t \right)^2 + \left(\left(Y^2_{t,p^2_t} + Y^3_{t,p^3_t} \right) - c^{23}_t \right)^2 .$$

As it is more crucial to reach some of these subtargets than other, we also introduce some importance weights κ^{g}_t to be considered in the global loss. Therefore, at each round t, the electricity provider fixe a set of subtargets $(c^{\mathrm{g}}_t)_{\mathrm{g}\in\mathcal{G}_t}$, where $\mathcal{G}_t$ is a set of subsets of $\{1,\dots,G\}$. For example, $\mathcal{G}_t = \big\{\{1,\dots,G\}\big\}$ refers to the global consumption management of Example 10 and $\mathcal{G}_t = \big\{\{1\},\dots,\{G\}\big\}$ to the local management presented in Example 11. In addition to the subtargets, weights κ^{g}_t are given: the higher κ^{g}_t, the closer to c^{g}_t should the consumption associated with g be. Then, the provider chooses the matrix $\boldsymbol{p}_t$ to minimize the instantaneous loss

$$\ell_t \triangleq \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa^{\mathrm{g}}_t \Big(Y^{\mathrm{g}}_{t,\boldsymbol{p}} - c^{\mathrm{g}}_t \Big)^2 = \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa^{\mathrm{g}}_t \Big(\sum_{i\in\mathrm{g}} Y^i_{t,p^i_t} - c^{\mathrm{g}}_t \Big)^2 ,$$

where $\displaystyle\sum_{i\in\mathrm{g}} Y^i_{t,p^i_t}$ is the power consumption associated with subset g.

Then, it observes the G consumptions Y^i_{t,p^i_t}, for $i = 1,\dots,G$. With no loss of generality, we normalized the weights $\kappa^{\mathrm{g}}_t \in [0,1]$ for all $\mathrm{g} \in \mathcal{G}_t$ and $t \geqslant 1$. This online protocol is stated in Protocol 8.

Notation. We generally refer to a cluster with the superscript $i \in \{1,\dots,G\}$ and to a subset of clusters with the superscript $\mathrm{g} \in P(\{1,\dots,G\})$. Furthermore, we denote by $\mathcal{G}$ some set of subsets of $\{1,\dots,G\}$.

Remark 25. *For a subset* $\mathrm{g} \in P(\{1,\dots,G\})$*, we will not need to assume the consistency of the local targets* c^i_t *with respect to a possible global target* c^{g}_t*, i.e., we do not require* $\sum_{i\in\mathrm{g}} c^i_t = c^{\mathrm{g}}_t$.

Protocol 8 Target Tracking for Contextual Bandits with Subtargets

Input

Parametric context sets $\mathcal{X}^1, \ldots, \mathcal{X}^G$

Set of legible matrices of tariff allocations $\boldsymbol{\mathcal{P}} = \mathcal{P}^1 \times \ldots, \times \mathcal{P}^G$

Bounds on mean consumptions $C^1, \ldots, C^G$

Transfer functions $\varphi^i : \mathcal{X}^i \times \mathcal{P}^i \to \mathbb{R}^{d^i}$, for $i = 1, \ldots, G$

Unknown parameters

Transfer parameters $\theta^i \in \mathbb{R}^{d^i}$, for $i = 1, \ldots, G$

Covariance matrices Σ^i of size $K \times K$, for $i = 1, \ldots, G$

for $t = 1, 2, \ldots$ **do**

Observe some context $(x_t^i)_{i=1,\ldots,G}$ and some set $\mathcal{G}_t$ of subsets of $\{1, \ldots, G\}$

for $\mathrm{g} \in \mathcal{G}_t$ **do**

Observe some subtarget $c_t^{\mathrm{g}} \in [0, C^{\mathrm{g}}]$ and some weight $\kappa_t^{\mathrm{g}} \in [0, 1]$

end for

for $i = 1, 2, \ldots, G$ **do**

Choose an allocation of price levels $p_t^i \in \mathcal{P}$

Observe a resulting mean consumption

$$Y_{t,p_t}^i = \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i + (p_t^i)^{\mathrm{T}} \varepsilon_t^i$$

end for

Suffer a loss $\displaystyle\sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \Big(\sum_{i \in \mathrm{g}} Y_{t,p_t^i}^i - c_t^i \Big)^2$

end for

2.3 Expression of the regret

At a round t, for each cluster i, it may be seen, by induction, that the contextual vectors x_t^i and tariffs p_t^i, as well as the subtarget c_t^{g}, for any $\mathrm{g} \in \mathcal{G}_t$, are $\mathcal{F}_{t-1}$–measurable, where

$$\mathcal{F}_{t-1} = \sigma\big(\varepsilon_1^1, \ldots, \varepsilon_1^G, \varepsilon_2^1, \ldots, \varepsilon_2^G, \ldots, \varepsilon_{t-1}^G\big)\,.$$

Therefore, under Model 1, for any $\mathrm{g} \in \mathcal{G}_t$

$$\begin{aligned}
\mathbb{E}\Big[\big(Y_{t,\boldsymbol{p}_t}^{\mathrm{g}} - c_t^{\mathrm{g}}\big)^2 \,\big|\, \mathcal{F}_{t-1}\Big] &= \mathbb{E}\Big[\Big(\sum_{i \in \mathrm{g}} Y_{t,p_t^i}^i - c_t^{\mathrm{g}}\Big)^2 \,\big|\, \mathcal{F}_{t-1}\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{i \in \mathrm{g}} \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i + (p_t^i)^{\mathrm{T}} \varepsilon_t^i - c_t^{\mathrm{g}}\Big)^2 \,\big|\, \mathcal{F}_{t-1}\Big] \qquad \text{(Model 4)} \\
&= \Big(\sum_{i \in \mathrm{g}} \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}}\Big)^2 + \mathbb{E}\Big[\Big(\sum_{i \in \mathrm{g}} (p_t^i)^{\mathrm{T}} \varepsilon_t^i\Big)^2 \,\big|\, \mathcal{F}_{t-1}\Big] \\
&\qquad + \mathbb{E}\Big[2\Big(\sum_{i \in \mathrm{g}} \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}}\Big)\Big(\sum_{i \in \mathrm{g}} (p_t^i)^{\mathrm{T}} \varepsilon_t^i\Big) \,\big|\, \mathcal{F}_{t-1}\Big]\,.
\end{aligned}$$

In Model 1 we assume that, for all $(i, j) \in \{1, \ldots, G\}^2$ with $i \neq j$, the vectors ε_t^i and ε_t^j are independent. Furthermore, the expectations of vectors ε_t^i are null and all p_t^i, x_t^i and c_t^{g} are $\mathcal{F}_{t-1}$-measurable, so we get

$$\mathbb{E}\Big[\big(Y_{t,\boldsymbol{p}_t}^{\mathrm{g}} - c_t^{\mathrm{g}}\big)^2 \,\big|\, \mathcal{F}_{t-1}\Big] = \Big(\sum_{i \in \mathrm{g}} \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}}\Big)^2 + \sum_{i \in \mathrm{g}} (p_t^i)^{\mathrm{T}} \Sigma^i p_t^i.$$

In that follows, we aim to minimize the sum of conditionally expected losses, which equals, by summing previous equality over time and clusters,

$$\sum_{t=1}^{T} \mathbb{E}\Big[\, \ell_t \,\big|\, \mathcal{F}_{t-1} \,\Big] = \sum_{t=1}^{T} \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \bigg(\Big(\sum_{i\in\mathrm{g}} \varphi^n(x_t^i, p_t^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i\in\mathrm{g}} (p_t^i)^{\mathrm{T}} \Sigma^i p_t^i \bigg) .$$

For any matrix $\boldsymbol{p} = (p^i)_{i\in\{1,\dots,G\}}$, we introduce the notation $\ell_{t,\boldsymbol{p}}$ for the instantaneous conditionally expected loss:

$$\ell_{t,\boldsymbol{p}} \triangleq \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \bigg(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i\in\mathrm{g}} (p^i)^{\mathrm{T}} \Sigma^i p^i \bigg) . \tag{8.1}$$

To ensure the minimization of the sum of losses either in expectation or with high probability, as in Section 3.1 in Chapter 4, we compare, at each round t, our choices $\boldsymbol{p}_t = (p_t^i)_{i=1,\dots,G}$ to the choices of the best possible strategy, namely the ones which minimize the sum of conditionally expected losses. Therefore, we introduce the (conditional) regret

$$\begin{aligned}
\bar{R}_T &\triangleq \sum_{t=1}^{T} \Big(\ell_{t,\boldsymbol{p}_t} - \min_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}} \Big) \\
&= \sum_{t=1}^{T} \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \bigg(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i\in\mathrm{g}} (p_t^i)^{\mathrm{T}} \Sigma^i p_t^i \bigg) \\
&\quad - \sum_{t=1}^{T} \min_{(p^1,\dots,p^G)\in\boldsymbol{\mathcal{P}}} \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \bigg(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p_t^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i\in\mathrm{g}} (p_t^i)^{\mathrm{T}} \Sigma^i p_t^i \bigg) .
\end{aligned}$$

3 A regret bound for subtarget tracking

In this section we control the regret defined above by using arguments similar to the ones developed in Section 3 in Chapter 4. Section 3.1 presents the optimistic bandit algorithm we consider for the tracking of subtargets. The regret bound is stated in Section 3.2 and proved in Section 3.3.

As in Chapter 4, we make some assumption on the boundedness of the expected consumption: for each cluster $i \in \{1,\dots,G\}$, it is assumed to be always bounded between 0 and a known maximal value C^i. The following assumption, linked to the knowledge of C^i, indicates some normalization of the modelings.

Assumption 9 – *Boundedness assumption.* For any round t, and any value of the context vector x_t^i, we make the assumption that

$$\|\varphi^i\|_\infty \leqslant 1, \qquad \|\theta^i\|_\infty \leqslant C^i, \quad \text{and} \quad \forall p^i \in \mathcal{P}^i, \quad \varphi^i(x_t^i, p^i)^{\mathrm{T}} \theta^i \in [0, C^i] .$$

Moreover, we assume that the coefficients of the matrix Σ^i are bounded, which entails

$$\forall p^i \in \mathcal{P}^i, \quad (p^i)^{\mathrm{T}} \Sigma^i p^i \leqslant \Gamma^i,$$

where the bound Γ^i is also known. Finally, for any $\mathrm{g} \in \mathcal{G}_t$, we denote by $C^{\mathrm{g}} = \sum_{i\in\mathrm{g}} C^i$ a bound on the consumption Y_t^{g} associated with the subset g and assume that, for any

round t, if a subtarget c_t^{g} is provided for Y_t^{g}, then it is in $[0, C^{\mathrm{g}}]$.

A consequence of all these boundedness assumptions is that $L^{\mathrm{g}} = (C^{\mathrm{g}})^2 + \sum_{i \in \mathrm{g}} \Gamma^i$ upper-bounds the (conditionally) expected losses associated with the subset g: for any round t, for all values of the context vectors x_t^i with $i \in \mathrm{g}$, for any $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$,

$$\Big(\sum_{i \in \mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}} \theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i \in \mathrm{g}} (p^i)^{\mathrm{T}} \Sigma^i p^i \leqslant L^{\mathrm{g}} \quad \text{so,} \quad \ell_{t,\boldsymbol{p}} \leqslant L \triangleq \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} L^{\mathrm{g}} \,. \tag{8.2}$$

We assume, for any $i \in \{1, \ldots, G\}$, that both the transfer function φ^i and the bounds ($C^i > 0$ and $\Gamma^i > 0$) are known; consequently, for any subset g, C^{g} and L^{g} are also known. We recall that $[x]_C = \min\{\max\{x, 0\}, C\}$ is the clipped part C of a real number x (clipping between 0 and C).

3.1 Optimistic algorithm

As for previous optimistic algorithms (see Chapters 4 and 5), the key is to estimate, for each $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$, the instantaneous conditionally expected loss $\ell_{t,\boldsymbol{p}}$ and to provide a deviation bound $\alpha_{t,\boldsymbol{p}}$ for this estimator denoted by $\widehat{\ell}_{t,\boldsymbol{p}}$.

Given all the transfer functions φ^i, the definition of $\ell_{t,\boldsymbol{p}}$ (see Equation 8.1) suggests to first estimate the parameter vectors θ^i and the correlation matrices Σ^i and secondly to combine them and compute the estimator $\widehat{\ell}_{t,\boldsymbol{p}}$. For each cluster i, we modeled its power consumption exactly as the homogeneous consumers population of Chapter 4 (only the underlying models differ from one cluster to another, through the mapping functions φ^i, the parameters θ^i and the covariance matrix Σ^i). Therefore we define the estimations $\widehat{\theta}_t^i$ and $\widehat{\Sigma}_\tau^i$ exactly as we did for the estimations $\widehat{\theta}_t$ and $\widehat{\Sigma}_\tau$ of Chapter 4. At a round t, we introduce the estimators $\widehat{\theta}_t^i$ of the d^i-vectors θ^i (see Section 3.3.1 of Chapter 4), for each $i \in \{1, \ldots G\}$:

$$\widehat{\theta}_t^i \triangleq (V_t^i)^{-1} \sum_{s=1}^{t} Y_{s,p_s^i}^i \varphi^i(x_s^i, p_s^i) \,, \quad \text{where} \quad V_t^i \triangleq \lambda^i \mathrm{I}_{d^i} + \sum_{s=1}^{t} \varphi^i(x_s^i, p_s^i) \varphi^i(x_s^i, p_s^i)^{\mathrm{T}} \,.$$

After $\tau \geqslant 1$ exploration rounds, exactly as we did with a single homogeneous population in Section 3.3.2 of Chapter 4, we estimate the G covariance matrices Σ^i, for $i = 1, \ldots, G$. Here, for $\tau > 0$, these G estimators $\widehat{\Sigma}_\tau^i$ are computed in parallel, independently and cluster by cluster. Then, for $t \geqslant \tau + 1$, we can estimate the instantaneous expected loss $\ell_{t,\boldsymbol{p}}$ associated with the choices $\boldsymbol{p} = (p^i)_{i \in \{1,\ldots,G\}}$ by:

$$\widehat{\ell}_{t,\boldsymbol{p}} \triangleq \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\left(\Big[\sum_{i \in \mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}} \widehat{\theta}_{t-1}^i \Big]_{C^{\mathrm{g}}} - c_t^{\mathrm{g}} \right)^2 + \sum_{i \in \mathrm{g}} (p^i)^{\mathrm{T}} \widehat{\Sigma}_\tau^i p^i \right) .$$

With $\alpha_{t,\boldsymbol{p}}$ deviation bounds, to be set by the analysis, the optimistic algorithm picks, for $t \geqslant \tau + 1$,

$$\boldsymbol{p}_t \in \operatorname*{argmin}_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \big\{ \widehat{\ell}_{t,\boldsymbol{p}} - \alpha_{t,\boldsymbol{p}} \big\} \,. \tag{8.3}$$

For the first τ rounds, vectors $\boldsymbol{p}_t$ are picked deterministically, in order to get the deviation bounds on the estimators $\widehat{\Sigma}_\tau^i$ (see Section 3.3.2 of Chapter 4).

3.2 Statement of the regret bound

Theorem 10 below states a generalization of Theorem 4 of Chapter 4, which is proved in the next section.

Theorem 10. *Fix a risk level $\delta \in (0,1)$ and a time horizon T. Assume that Assumption 9 hold. The optimistic algorithm* (8.3) *with an initial exploration of length $\tau = \mathcal{O}(T^{2/3})$ rounds satisfies*

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

with probability at least $1-\delta$.

Remark 26. *The dependence of the regret bound on the number of clusters G is at most of order $2^{G-1}G$ (see Lemma 11 below).*

The initial exploration, namely the first τ rounds, offers a good estimation of the G matrices Σ^i, for $i = 1, \ldots, G$. From then on, the vectors θ^i are estimated online and the optimistic algorithm performs a good exploration-exploitation trade-off.

Remark 27. *If the matrices Σ^i are known and no initial exploration needs to be performed, the regret bound is in $\mathcal{O}\big(\sqrt{T \ln T \ln(T/\delta)}\big)$.*

We emphasize that if each cluster was considered independently of the others (namely if there was just a target per cluster and $\mathcal{G}_t = \big\{\{1\}, \ldots, \{G\}\big\}$, see Example 11) we could run the algorithm introduced in Chapter 4 on each cluster. Then, Theorem 4 of Chapter 4, taking $\delta = \delta'/G$, would guarantee a bound for each local regret with probability at least $1-\delta'/G$. With a union bound, as the G algorithms would be run in parallel, independently and cluster by cluster, we would obtain a regret bound with probability at least $1-\delta'$ by summing the local regrets. The main improvement here comes from the links between the clusters induced by the subtargets c_t^{g}, when g is not a singleton: all the vectors of $\boldsymbol{p}$ are linked (they cannot be chosen independently of each other if we want that the sum of the consumptions Y^i_{t,p^i_t}, for $i \in \mathrm{g}$ to be as close as possible to the target c_t^{g}) and some adaptation of the results is needed.

3.3 Analysis of the regret

The analysis of the regret is similar to the one provided in Section 3 and uses some results already proved in this previous chapter. For each round $t \geqslant \tau+1$, for each subset $\mathrm{g} \in \mathcal{G}_t$, the key is to provide a deviation bound on the instantaneous regret related to subset g. This bound will come from the deviation inequalities obtained, for each $i \in \mathrm{g}$, on $\widehat{\theta}^i_t$ and $\widehat{\Sigma}^i_\tau$ (provided in Sections 3.3.1 and 3.3.2, respectively). Then, by summing over $\mathrm{g} \in \mathcal{G}_t$ we will obtain that the instantaneous regrets $\ell_{t,\boldsymbol{p}_t} - \min_{\boldsymbol{p}\in\mathcal{P}^N} \ell_{t,\boldsymbol{p}_t}$ are bounded, with high probability, by $2\alpha_{t,\boldsymbol{p}_t}$: this is Proposition 4 below. Next, Lemma 11 shows how to control the sum over $t \geqslant \tau+1$ of these deviations bounds $\alpha_{t,\boldsymbol{p}_t}$. Finally, it will only remain to deal with the first τ rounds to conclude the proof of Theorem 10.

Proposition 4. *For a risk level $\delta \in (0,1)$ and an exploration budget $\tau \geqslant 2$, the optimistic algorithm* (8.3) *ensures, with probability at least $1-\delta$,*

$$\sum_{t=\tau+1}^{T} \ell_{t,\boldsymbol{p}_t} - \sum_{t=\tau+1}^{T} \min_{\boldsymbol{p}\in\mathcal{P}^N} \ell_{t,\boldsymbol{p}_t} \leqslant 2 \sum_{t=\tau+1}^{T} \alpha_{t,\boldsymbol{p}_t}, \tag{8.4}$$

with, for $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$,

$$\alpha_{t,\boldsymbol{p}} \triangleq \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \left[\min\left\{ L^{\mathrm{g}}, 2C^{\mathrm{g}} \sum_{i\in\mathrm{g}} B_t^i(\delta/Gt^2) \left\| (V_{t-1}^i)^{-1/2} \varphi^i(x_t^i, p^i) \right\| \right\} + \sum_{i\in\mathrm{g}} G_\tau^i(\delta/2G) \right]. \quad (8.5)$$

The quantities $B_t^i(\delta/Gt^2)$ *and* $G_\tau^i(\delta/2G)$ *are defined in Equations* (8.6) *and* (8.8) *below, respectively.*

Before proving Proposition 4, we recall the deviations inequalities obtained in 4. As the power consumption of each cluster i is modeled exactly as the homogeneous consumers population of Chapter 4, Lemmas 4 and 5, written them with a superscript i, hold for the estimations $\widehat{\theta}_t^i$ and $\widehat{\Sigma}_\tau^i$, respectively.

⋆ *Deviation inequalities on* $\widehat{\theta}_t^i$ *and* $\widehat{\Sigma}_\tau^i$ *used in the proof of Proposition 4.* For a given $i \in \{1, \dots G\}$, no matter how the provider picks p_t^i, we have, for all $t \geqslant 1$ and all $\delta \in (0,1)$ the two high probability deviation bounds below. Lemma 4 proved in Section 3.3.1 of Chapter 4 states that, with probability at least $1-\delta$,

$$\left\| (V_t^i)^{1/2} \left(\widehat{\theta}_t^i - \theta \right) \right\| \leqslant B_t^i(\delta) \triangleq \sqrt{\lambda^i d^i}\, C^i + \rho^i \sqrt{2 \ln \frac{1}{\delta} + d^i \ln\left(1 + \frac{t}{\lambda^i}\right)}. \quad (8.6)$$

Moreover, a straightforward application of Lemma 5 (see Section 3.3.2 of Chapter 4) ensures that the estimator $\widehat{\Sigma}_\tau^i$ satisfies, with probability at least $1-\delta$,

$$\sup_{p^i \in \mathcal{P}^i} \left| (p^i)^{\mathrm{T}} \left(\widehat{\Sigma}_\tau^i - \Sigma^i \right) p^i \right| \leqslant G_\tau^i(\delta) = \mathcal{O}\left(\frac{1}{\sqrt{\tau}} \ln^2(\tau/\delta) \sqrt{\ln(1/\delta)} \right), \quad (8.7)$$

where

$$G_\tau^i(\delta) \triangleq (K+8) \frac{\sqrt{\tau}}{\tau_0} \left(B_\tau^i(\delta/3) \left(C^i + \rho^i + \ln \frac{6\tau}{\delta} \right) + \left(\frac{\rho^i}{2} + \ln \frac{6\tau}{\delta} \right)^2 \sqrt{2 \ln \frac{3K^2}{\delta}} + 2 \sqrt{\exp(2\rho^i) \frac{\delta}{6}} \right) \quad (8.8)$$

with $\tau_0 = \lfloor 2\tau/(K(K+1)) \rfloor$. This application of Lemma 5 is possible because the set $\mathcal{P}^i$ is included in the simplex Δ^K.

Proof of Proposition 4. Foremost, in Step 1, we prove that, if for all $i \in \{1, \dots, G\}$ and for any $t > \tau \geqslant 2$, the inequalities

$$\| (V_{t-1}^i)^{1/2} (\widehat{\theta}_{t-1}^i - \theta^i) \| \leqslant B_t^i(\delta/Gt^2) \quad \text{and} \quad \| \widehat{\Sigma}_\tau^i - \Sigma^i \|_\infty \leqslant G_\tau^i(\delta/2G) \quad (8.9)$$

hold then, for any $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$, $| \ell_{t,\boldsymbol{p}} - \widehat{\ell}_{t,\boldsymbol{p}} | \leqslant \alpha_{t,\boldsymbol{p}}$. This induces a bound on the instantaneous regret: this is Step 2. In Step 3, we show that the $2G$ inequalities above hold for all $t \geqslant \tau + 1$ (so all the instantaneous regrets are bounded) with probability at least $1-\delta$, and we conclude the proof. This is done by a union bound over each time step but also over each cluster. This is why, in comparison to Chapter 4, a factor $1/G$ appears in the bounds $B_t^i(\delta/Gt^2)$ and $G_\tau^i(\delta/2G)$ used to define $\alpha_{t,\boldsymbol{p}}$ – see Equation (8.5). By taking $G = 1$, we recover the definition of $\alpha_{t,p}$ introduced in Chapter 4.

⋆ *Step 1: Good estimation of the losses.* For an round $t \geqslant \tau$, we assume that for all $i \in 1, \ldots, G$, the estimators $\widehat{\theta}^i_{t-1}$ and $\widehat{\Sigma}^i_\tau$ satisfy inequalities (8.9). We show below how we can then get a confidence bound on the losses. With $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$, the difference between the expected loss $\ell_{t,\boldsymbol{p}}$ and its estimation $\widehat{\ell}_{t,\boldsymbol{p}}$ is equal to

$$\begin{aligned}
\ell_{t,\boldsymbol{p}} - \widehat{\ell}_{t,\boldsymbol{p}} &= \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa^{\mathrm{g}} \Bigg[\Bigg(\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - c^{\mathrm{g}}_t \Bigg)^2 + \sum_{i\in\mathrm{g}} (p^i)^{\mathrm{T}} \Sigma^i p^i \\
&\qquad\qquad - \Bigg(\Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} - c^{\mathrm{g}}_t \Bigg)^2 - \sum_{i\in\mathrm{g}} (p^i)^{\mathrm{T}} \widehat{\Sigma}^i_\tau p^i \Bigg] \\
&= \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa^{\mathrm{g}} \Bigg[\Bigg(\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - c^{\mathrm{g}}_t \Bigg)^2 - \Bigg(\Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} - c^{\mathrm{g}}_t \Bigg)^2 \\
&\qquad\qquad + \sum_{i\in\mathrm{g}} \Big((p^i)^{\mathrm{T}} \Sigma^i p^i - (p^i)^{\mathrm{T}} \widehat{\Sigma}^i_\tau p^i \Big) \Bigg] . \qquad (8.10)
\end{aligned}$$

We will bound $\big|\ell_{t,\boldsymbol{p}} - \widehat{\ell}_{t,\boldsymbol{p}}\big|$ by bounding each term of the sum over the subsets $\mathrm{g} \in \mathcal{G}_t$. For each subset, we bound separately the two terms of the expression above: the first one deals with the estimations of θ^i and the second with the estimations of Σ^i. For any subset $\mathrm{g} \in \mathcal{G}_t$, we first have

$$\begin{aligned}
&\left| \Bigg(\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - c^{\mathrm{g}}_t \Bigg)^2 - \Bigg(\Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} - c^{\mathrm{g}}_t \Bigg)^2 \right| \\
&= \left| \Bigg(\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - \Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} \Bigg) \right. \\
&\qquad \left. \times \Bigg(\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i + \Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} - 2c^{\mathrm{g}}_t \Bigg) \right| \\
&\leqslant \left| \sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - \sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \right| \left| \sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i + \Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} - 2c^{\mathrm{g}}_t \right| .
\end{aligned}$$

We could remove the clipping in the first term of the last inequality because all $i \in \mathrm{g}$, $\varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i$ is non-negative. We now assume that for any $i \in 1, \ldots, G$, the estimator $\widehat{\theta}^i_{t-1}$ satisfies the inequality (8.6). Then, for any subset $\mathrm{g} \in \mathcal{G}_t$, using that $\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i$, $\big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1}\big]_{C^{\mathrm{g}}}$ and c^{g}_t are bounded between 0 and C^{g}, we get

$$\begin{aligned}
&\left| \Bigg(\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - c^{\mathrm{g}}_t \Bigg)^2 - \Bigg(\Big[\sum_{i\in\mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \Big]_{C^{\mathrm{g}}} - c^{\mathrm{g}}_t \Bigg)^2 \right| \\
&\qquad \leqslant 2C^{\mathrm{g}} \sum_{i\in\mathrm{g}} \Big| \varphi^i(x^i_t, p^i)^{\mathrm{T}} \big(\theta^i - \widehat{\theta}^i_{t-1} \big) \Big| \\
&\qquad \leqslant 2C^{\mathrm{g}} \sum_{i\in\mathrm{g}} \Big| \varphi^i(x^i_t, p^i)^{\mathrm{T}} (V^i_{t-1})^{-1/2} (V^i_{t-1})^{1/2} \big(\theta^i - \widehat{\theta}^i_{t-1} \big) \Big| \\
&\qquad \leqslant 2C^{\mathrm{g}} \sum_{i\in\mathrm{g}} B^i_t(\delta/Gt^2) \big\| (V^i_{t-1})^{-1/2} \varphi^i(x^i_t, p^i) \big\| .
\end{aligned}$$

The last inequality is obtain using Equation (8.9) for all $i \in \mathrm{g}$. We deal with the second part of the sum over the subsets $\mathrm{g} \in \mathcal{G}_t$ – see Equation (8.10) – by using the upper bounds of Equation (8.7)

$$\left| \sum_{i \in \mathrm{g}} \left((p^i)^{\mathrm{T}} \Sigma^i p^i - (p^i)^{\mathrm{T}} \widehat{\Sigma}^i_\tau p^i \right) \right| \leqslant \sum_{i \in \mathrm{g}} G^i_\tau(\delta/2G) \, .$$

Combining and summing over $\mathrm{g} \in \mathcal{G}_t$ the two inequalities above, we get

$$\left| \ell_{t,\boldsymbol{p}} - \widehat{\ell}_{t,\boldsymbol{p}} \right| \leqslant \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa^{\mathrm{g}} \left[2C^{\mathrm{g}} \sum_{i \in \mathrm{g}} B^i_t(\delta/Gt^2) \left\| (V^i_{t-1})^{-1/2} \varphi^i(x^i_t, p^i) \right\| + \sum_{i \in \mathrm{g}} G^i_\tau(\delta/2G) \right] .$$

Finally, Assumption 9 and the definition of clipping ensure that

$$\big(\sum_{i \in \mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \theta^i - c^{\mathrm{g}}_t \big)^2 + \sum_{i \in \mathrm{g}} (p^i)^{\mathrm{T}} \Sigma^i p^i \quad \text{and} \quad \big(\big[\sum_{i \in \mathrm{g}} \varphi^i(x^i_t, p^i)^{\mathrm{T}} \widehat{\theta}^i_{t-1} \big]_{C^{\mathrm{g}}} - c^{\mathrm{g}}_t \big)^2 + \sum_{i \in \mathrm{g}} (p^i)^{\mathrm{T}} \Sigma^i p^i$$

are in $[0, L^{\mathrm{g}}]$, for any $\mathrm{g} \in \mathcal{G}_t$ – see Equation (8.2). Therefore, we get that for any $\boldsymbol{p} \in \mathcal{P}$,

$$\begin{aligned} \left| \ell_{t,\boldsymbol{p}} - \widehat{\ell}_{t,\boldsymbol{p}} \right| &\leqslant \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa^{\mathrm{g}}_t \left[\min \left\{ L^{\mathrm{g}}, 2C^{\mathrm{g}} \sum_{i \in \mathrm{g}} B^i_t(\delta/Gt^2) \left\| (V^i_{t-1})^{-1/2} \varphi^i(x^i_t, p^i) \right\| \right\} + \sum_{i \in \mathrm{g}} G^i_\tau(\delta/2G) \right] \\ &\triangleq \alpha_{t,\boldsymbol{p}} \, . \end{aligned} \tag{8.11}$$

⋆ *Step 2: Resulting bound on the instantaneous regrets.* As in the proof of Theorem 4 of Chapter 4, thanks to the optimistic algorithm (8.3), a deviation inequality on the losses ensures a boundedness of the instantaneous regrets. Indeed, with $\boldsymbol{p}^\star_t$ an optimal matrix to be used at round t, the vectors $\boldsymbol{p}_t$ played by the optimistic algorithm satisfy:

$$\widehat{\ell}_{t,\boldsymbol{p}_t} - \alpha_{t,\boldsymbol{p}_t} \leqslant \widehat{\ell}_{t,\boldsymbol{p}^\star_t} - \alpha_{t,\boldsymbol{p}^\star_t} \, , \quad \text{where by definition,} \quad \boldsymbol{p}^\star_t \triangleq \operatorname*{argmin}_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}} \, .$$

If inequalities (8.9) hold, Equation (8.11) also holds; and with $\boldsymbol{p} = \boldsymbol{p}_t$ and $\boldsymbol{p} = \boldsymbol{p}^\star_t$, we get

$$\ell_{t,\boldsymbol{p}_t} \leqslant \widehat{\ell}_{t,\boldsymbol{p}_t} + \alpha_{t,\boldsymbol{p}_t} \quad \text{and} \quad -\ell_{t,\boldsymbol{p}^\star_t} \leqslant -\widehat{\ell}_{t,\boldsymbol{p}^\star_t} + \alpha_{t,\boldsymbol{p}^\star_t} \, .$$

By combining these three inequalities, we thus obtain

$$\ell_{t,\boldsymbol{p}_t} - \ell_{t,\boldsymbol{p}^\star_t} \leqslant \widehat{\ell}_{t,\boldsymbol{p}_t} - \widehat{\ell}_{t,\boldsymbol{p}^\star_t} + \alpha_{t,\boldsymbol{p}_t} + \alpha_{t,\boldsymbol{p}^\star_t} \leqslant 2\alpha_{t,\boldsymbol{p}_t} \, .$$

Finally, by summing over t, we get Equation (8.4).

⋆ *Step 3: Special cases.* We conclude the proof by dealing with the time steps when at least one of the events (8.9) does not hold. For a given $t > \tau$ and for $i \in \{1, \dots, G\}$, Equations (8.6) and (8.7) ensure that

$$\mathbb{P}\Big(\|(V^i_{t-1})^{1/2} (\widehat{\theta}^i_{t-1} - \theta^i)\| \geqslant B^i_t(\delta/Gt^2) \Big) \leqslant \frac{\delta}{Gt^2} \quad \text{and} \quad \mathbb{P}\Big(\|\widehat{\Sigma}^i_\tau - \Sigma^i\|_\infty \geqslant G^i_\tau(\delta/2G) \Big) \leqslant \frac{\delta}{2G} \, .$$

Thus, by union bound, at least one of the events (8.9) does not hold for some $t > \tau$ and some $i \in \{1, \dots, G\}$, with probability smaller than

$$\sum_{i=1}^{G} \sum_{t \geqslant \tau+1} \frac{\delta}{Gt^2} + \sum_{i=1}^{G} \frac{\delta}{2G} \leqslant \delta \int_2^\infty \frac{1}{t^2} \mathrm{d}t + \frac{\delta}{2} = \delta \, .$$

□

Exactly as in the analysis of the regret stated in Chapter 4, we are now left with proving the following lemma to conclude the analysis. This result is based on Lemma 7 of Chapter 4.

Lemma 11. *By denoting by $\mathcal{G}$ the set of the subsets of $\{1,\dots,G\}$, no matter how the environment and provider pick the $(x_t^i)_{i\in\{1,\dots G\}}$ and $\boldsymbol{p}_t$,*

$$\begin{aligned}\sum_{t=\tau+1}^{T}\alpha_{t,\boldsymbol{p}_t} &\leqslant \sum_{\mathrm{g}\in\mathcal{G}}\sum_{i\in\mathrm{g}}\left(\sqrt{\left(2C^{\mathrm{g}}B_T^i(\delta/T^2G)\right)^2+\frac{(L^{\mathrm{g}})^2}{2}}\sqrt{d^iT\ln\frac{\lambda^i+T}{\lambda^i}}+TG_\tau^i(\delta/2G)\right)\\ &\leqslant \mathcal{O}\left(2^{G-1}G\times\left(\sqrt{T\ln T\ln(T/G\delta)}+\frac{T}{\sqrt{\tau}}\ln^2(\tau/G\delta)\sqrt{\ln(1/G\delta)}\right)\right).\end{aligned}$$

Proof of Lemma 11. For any set $\mathrm{g}\in\mathcal{G}_t$, Lemma 12, proved at the end of the section, states that for any $f:\mathrm{g}\to\mathbb{R}^+$, $\min\left\{L,\sum_{i\in\mathrm{g}}f(i)\right\}\leqslant\sum_{i\in\mathrm{g}}\min\{L,f(i)\}$. Considering the function

$$\begin{array}{rcccl} f & : & g & \to & \mathbb{R}^+\\ & & i & \mapsto & B_t^i(\delta/Gt^2)\left\|(V_{t-1}^i)^{-1/2}\varphi^i(x_t^i,p_t^i)\right\|,\end{array}$$

we obtain that, for any round $t\geqslant\tau+1$,

$$\begin{aligned}&\min\left\{L^{\mathrm{g}},\,2C^{\mathrm{g}}\sum_{i\in\mathrm{g}}B_t^i(\delta/Gt^2)\left\|(V_{t-1}^i)^{-1/2}\varphi^i(x_t^i,p_t^i)\right\|\right\}\\ &\qquad\leqslant\sum_{i\in\mathrm{g}}\underbrace{\min\left\{L^{\mathrm{g}},2C^{\mathrm{g}}B_t^i(\delta/Gt^2)\left\|(V_{t-1}^i)^{-1/2}\varphi^i(x_t^i,p_t^i)\right\|\right\}}_{\triangleq a^i_{t,p_t^i}(L^{\mathrm{g}},C^g)}.\end{aligned}$$

Therefore, by summing over $\mathrm{g}\in\mathcal{G}_t$, we get the bound

$$\alpha_{t,\boldsymbol{p}_t}\leqslant\sum_{\mathrm{g}\in\mathcal{G}_t}\kappa_t^{\mathrm{g}}\sum_{i\in\mathrm{g}}\left(a^i_{t,p_t^i}(L^{\mathrm{g}},C^g)+G_\tau^i(\delta/2G)\right).$$

Now, we artificially enlarge the sum over $\mathrm{g}\in\mathcal{G}_t$ into a sum over $\mathrm{g}\in\mathcal{G}\triangleq\mathcal{P}(\{1,\dots,G\})$. This rough upperbound can easily be improved if the subset sets $\mathcal{G}_t$ are known in advance, by setting $\mathcal{G}=\bigcup_{t\in\{1,\dots,T\}}\mathcal{G}_t$. As we assume all κ_t^{g} lie in $[0,1]$, we obtain

$$\alpha_{t,\boldsymbol{p}_t}\leqslant\sum_{\mathrm{g}\in\mathcal{G}}\sum_{i\in\mathrm{g}}\left(a^i_{t,p_t^i}(L^{\mathrm{g}},C^{\mathrm{g}})+G_\tau^i(\delta/2G)\right).$$

We recall the result stated by Lemma 7 in Chapter 4: for a given $i\in\{1,\dots G\}$, no matter how the environment and provider pick the x_t^i and p_t^i,

$$\begin{aligned}&\sum_{t=\tau+1}^{T}\min\left\{L^{\mathrm{g}},2C\,B_{t-1}^i(\delta t^{-2}/G)\left\|(V_{t-1}^i)^{-1/2}\varphi^i(x_t^i,p_t^i)\right\|\right\}\\ &\qquad\leqslant\sqrt{\left(2CB_T^i(\delta/T^2G)\right)^2+\frac{(L^{\mathrm{g}})^2}{2}}\sqrt{d^iT\ln\frac{\lambda^i+T}{\lambda^i}}=\mathcal{O}\left(\sqrt{T\ln T\ln(T/G\delta)}\right).\end{aligned}$$

By summing over t, we conclude the proof:

$$\begin{aligned}\sum_{t=\tau+1}^{T}\alpha_{t,\boldsymbol{p}_t} &\leqslant \sum_{\mathrm{g}\in\mathcal{G}}\sum_{i\in\mathrm{g}}\sum_{t=\tau+1}^{T}\left(a^i_{t,p^i_t}(L^{\mathrm{g}},C^{\mathrm{g}})+G^i_\tau(\delta/2G)\right)\\ &\leqslant \sum_{\mathrm{g}\in\mathcal{G}}\sum_{i\in\mathrm{g}}\left(\sqrt{\left(2C^{\mathrm{g}}B^i_T(\delta/T^2G)\right)^2+\frac{(L^{\mathrm{g}})^2}{2}}\sqrt{d^iT\ln\frac{\lambda^i+T}{\lambda^i}}+TG^i_\tau(\delta/2G)\right)\\ &= \mathcal{O}\left(\sum_{\mathrm{g}\in\mathcal{G}}\sum_{i\in\mathrm{g}}\left(\sqrt{T\ln T\ln(T/G\delta)}+\frac{T}{\sqrt{\tau}}\ln^2(\tau/G\delta)\sqrt{\ln(1/G\delta)}\right)\right).\end{aligned}$$

As the sum of the cardinals of all the subsets of $\{1,\dots,G\}$ equals to $G2^{G-1}$ (see Lemma 13 below) we obtain the dependency on G. □

To conclude the proof of Theorem 10, it only remains to bound the instantaneous expected regret $\left(\ell_{t,\boldsymbol{p}_t}-\min_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}}\ell_{t,\boldsymbol{p}}\right)$ by $L=\sum_{g\in\mathcal{G}}L^{\mathrm{g}}$ for the first τ rounds. Thus, using Proposition 3 and Lemma 11, we get

$$\begin{aligned}\bar{R}_T &\leqslant \tau\sum_{g\in\mathcal{G}}L^{\mathrm{g}}+\sum_{t=\tau+1}^{T}\alpha_{t,\boldsymbol{p}_t}\\ &= \mathcal{O}\left(\tau L+2^{G-1}G\left(\sqrt{T\ln T\ln(T/G\delta)}+\frac{T}{\sqrt{\tau}}\ln^2(\tau/G\delta)\sqrt{\ln(1/G\delta)}\right)\right).\end{aligned}$$

Picking τ of order $T^{2/3}$ ensures the result claimed in Theorem 10.

We now state and prove the two tricks we used in the proof of Lemma 11.

Lemma 12 (Proof of Lemma 12)**.** *Given a finite set* g*, a function* $f:\mathrm{g}\to\mathbb{R}^+$ *and a constant* $L\geqslant 0$*, we have*

$$\min\left\{L,\sum_{i\in\mathrm{g}}f(i)\right\}\leqslant\sum_{i\in\mathrm{g}}\min\{L,f(i)\}.$$

Proof. This lemma only relies on the subadditivity of the function $\min\{L,\cdot\}$, having the nonnegative real numbers as domain. Therefore, we just have to show that

$$\forall\, u,\, v\in\mathbb{R}^+\qquad \min\{L,u+v\}\leqslant\min\{L,u\}+\min\{L,v\},$$

to conclude the proof. Let $0\leqslant u\leqslant v$ be two nonnegative real numbers. If $L\geqslant u+v$, the inequality above is an equality. Otherwise, we distinguish the two following cases:

- if $L\geqslant v$, then $\min\{L,u\}+\min\{L,v\}=u+v\geqslant L=\min\{L,u+v\}$;
- else, $L<v$ and $\min\{L,u\}+\min\{L,v\}=\min\{L,u\}+L\geqslant L=\min\{L,u+v\}$.

□

Lemma 13. *With an integer* $G\geqslant 1$*, the sum of the cardinals of all the subsets of* $\{1,\dots,G\}$ *is equal to* $G2^{G-1}$*, so*

$$\sum_{\mathrm{g}\in\mathcal{P}(\{1,\dots,G\})}\sum_{i\in\mathrm{g}}1=G2^{G-1}.$$

Proof of Lemma 13. For any integer $g \in \{1, \dots, G\}$, we have

$$g\binom{G}{g} = g\frac{G!}{(G-g)!g!} = G\frac{(G-1)!}{\big((G-1)-(g-1)\big)!(g-1)!} = G\binom{G-1}{g}.$$

As there are $\binom{G}{g}$ subsets of $\{1, \dots, G\}$ of cardinal g, by using the equation above and the binomial theorem, we obtain

$$\sum_{\mathrm{g}\in\mathcal{P}(\{1,\dots,G\})} |\mathrm{g}| = \sum_{g=1}^{G} g\binom{G}{g} = G\sum_{g=1}^{G-1}\binom{G-1}{g} = G2^{G-1}.$$

□

4 Application to the Low Carbon London data set

This section provides an illustration of the algorithm presented above. We consider the Low Carbon London data set presented in Chapter 3 and the data generator based on conditional variational auto-encoders (CVAE) built in Chapter 7. The data are not anymore simulated with generalized additive models and we will thus be able to assess the robustness of our solutions. The algorithm implemented takes into account rebound and side effects, namely, a round t refers to a day and the algorithm choices a daily tariff profile (that is $H = 48$ price allocations) instead of a single price allocation (when t refereed to half-hour time steps). To do so, we adapt algorithm (8.3). In our experiments of personalized daily demand side management, we consider some restrictions on the set of price allocations $\boldsymbol{\mathcal{P}}$ and strong assumptions on the underlying electricity consumption model. This is detailed in Subsection 4.1 below. Results are presented in Subsection 4.2 and in Subsection 4.3, we discuss several points that may improve this experiment.

4.1 Experiment design

We recall that the Low Carbon London data set is made of electricity consumption records (in kWh) at half hourly intervals of a thousand customers subjected to dynamic tariffs, through 2013. Here, we consider $G = 4$ sub-populations that were obtained with the procedure described in Chapter 7: for each household and each tariff (among Low, Normal and Low), a daily consumption profile is built from the individual consumption time series (using a causality model) and thanks to a dimension reduction technique (non-negative matrix factorization), these profiles are reduced into a few features that are used to cluster the households (using k-medoid algorithm). Once these clusters have been set, using the data generator described in Chapter 7, a full information data set can be obtained. We swiftly recall and illustrate the generation of electricity consumption data in the next paragraph, then we set some targets electricity consumption and some operational constraints. At the end of the section, we focus on the assumptions we made on the electricity consumption modeling and how we adapted the algorithm accordingly.

4.1.1 Data generator

The daily electricity consumption profiles are generated with the CVAE-based data simulator of Chapter 7. Given exogenous weather, calendar and tariff variables, it outputs

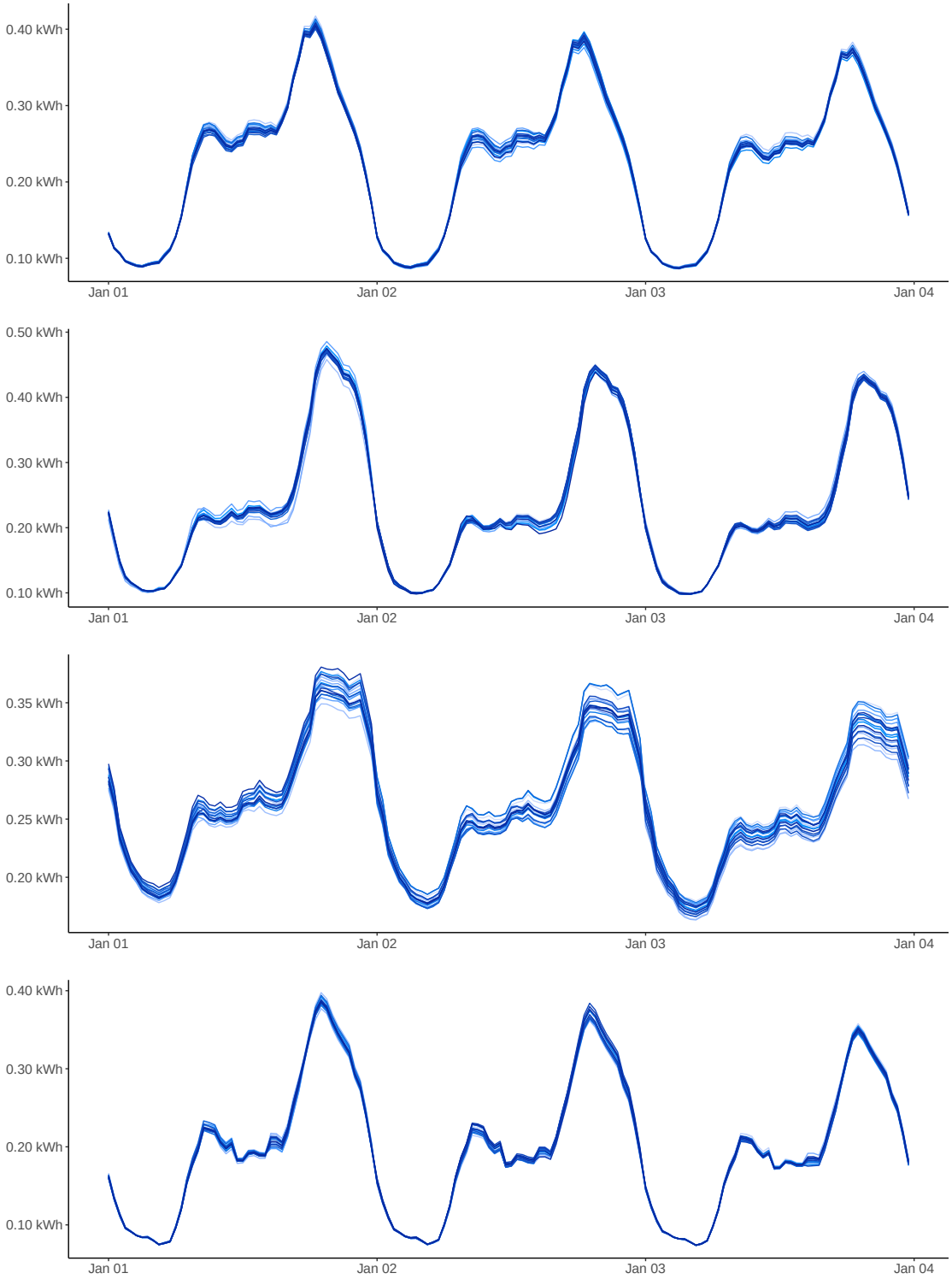


Figure 8.1 – 20 simulated electricity consumption (in kWh), for the first three days of 2013, for Normal tariff days and for the four household clusters (from cluster 1 at the top to cluster 4 at the bottom).

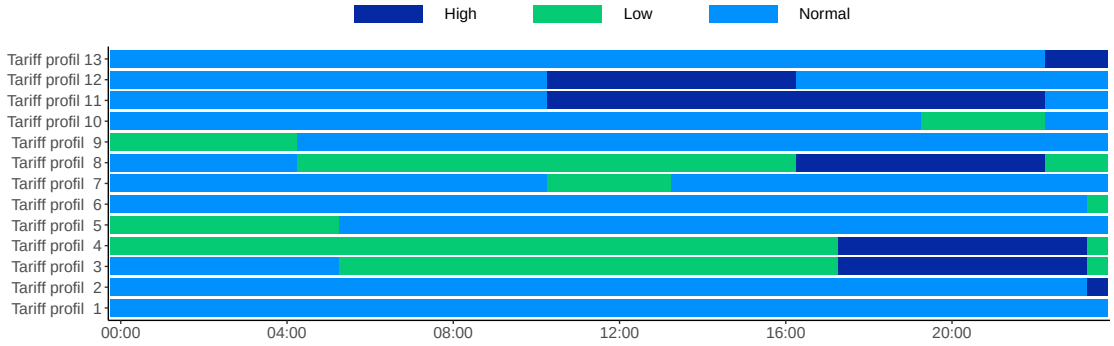


Figure 8.2 – Representation of the 13 daily electricity price profiles available: tariffs (High in navy, Low in green and Normal in blue) according the half-hour of the day.

random profiles. In Figure 8.1, 20 simulated profiles are plotted for the four household clusters. These electricity consumption records were generated for three days of "non-demand side management": Normal tariff was sent every half hour.

4.1.2 Targets and operational constraints

Restriction of the set of price allocations. In all this experiment, we assume that, for a given cluster i, the electricity provider must send the same tariff profile to all the households, namely, at each half-hour, the price allocation p^i is a Dirac mass. From now on, we denote by $j(i)$ the tariff sent to cluster i. We highlight that for each half-hour, the electricity provider may choose among $K = 3$ tariffs, and that it can give different tariff profiles to each of the $G = 4$ groups. Therefore, there exist $(K^H)^G = 3^{192} \approx 4.10^{91}$ possible configurations which makes the implementation of our algorithm tough – even if some of these configuration are totally inconceivable (e.g., a profile that changes from Low to High tariff and vice versa at every half hour). We emphasized in Chapter 7 that the CVAE-based data simulator generates non-intuitive consumption profiles when the input is a tariff profile never observed in the training set. Fortunately, the tariff profiles sent to households which took part in the Low Carbon London experiment were especially designed to test the potentiality of demand side management strategies (see the report Schofield et al., 2014 for further details). For all these reasons, we restrict the number of tariff profiles and consider only the ones that have been sent at least twice (so for which we have two days of electricity consumption records) during the Low Carbon London experiment. The 13 tariff profiles obtained are represented in Figure 8.2. Let us number them with $J = 1, 2, \dots \bar{J}$ and $\bar{J} = 13$. These profiles are associated with H-vectors made of the tariffs (among Low, High and Normal) associated with each half-hour of the day. With this notation, for a day t and for each cluster i, the electricity provider chooses a profile $J_t(i)$, that is, it picks

$$\boldsymbol{p}_t = \big(J_t(1), \dots, J_t(G)\big) \in \boldsymbol{\mathcal{P}} \quad \text{where} \quad \boldsymbol{\mathcal{P}} = \Big\{1, \dots, \bar{J}\Big\}^G.$$

The cardinal of the set of price allocations is then $\bar{J}^G = 13^4$ and our application fits into the discrete bandit framework. Throughout this experiment, we also force the electricity provider to pick same tariff profiles for both clusters 1 and 2 (this decision is arbitrary). Therefore, we merge these two clusters (we will consider a unique model, namely a unique parameter vector, to estimate the sum of the two electricity consumptions). This will

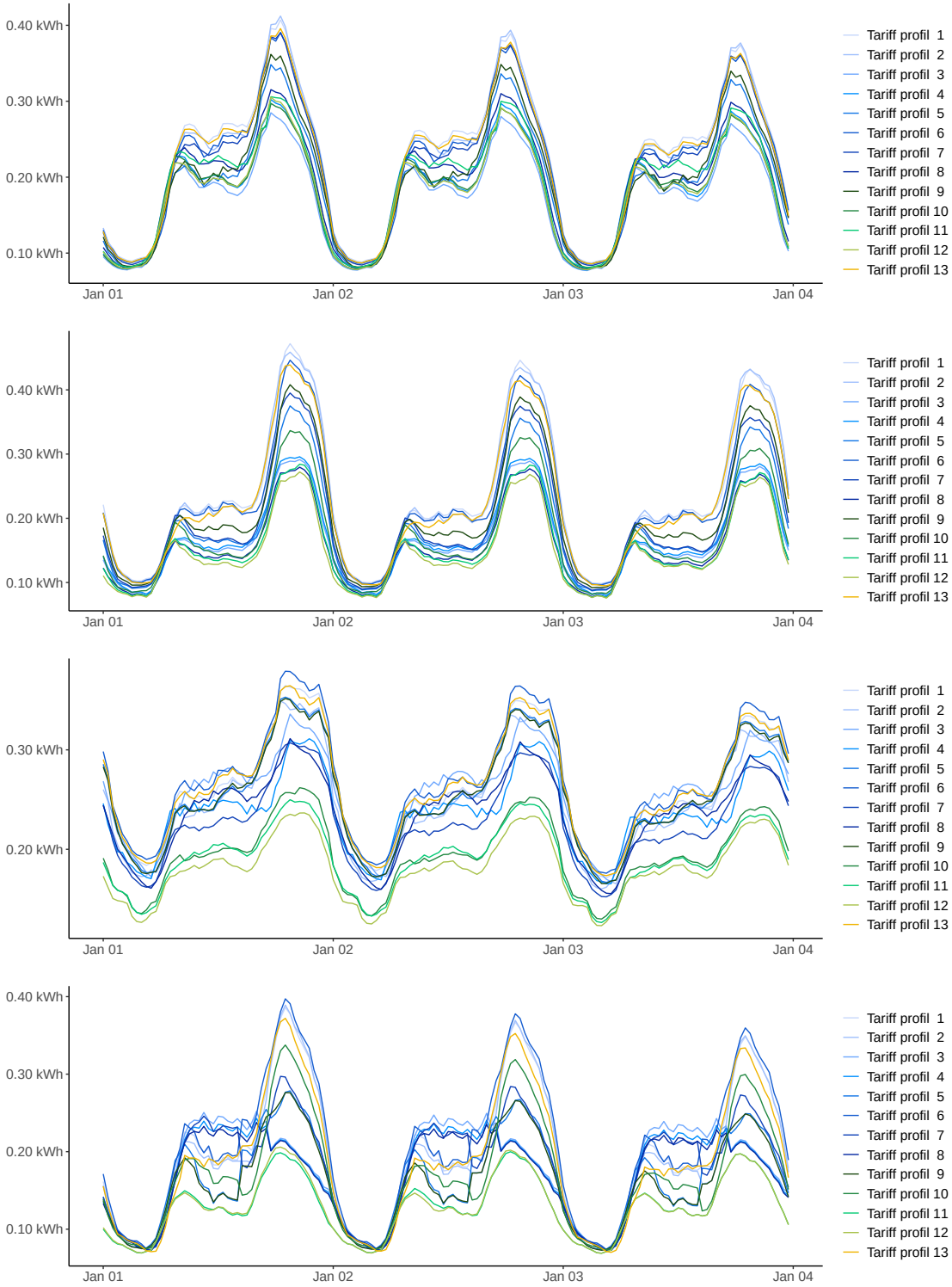


Figure 8.3 – Expected average consumption in kWh (computed over 1,000 simulations) depending on the tariff profile sent for each cluster (from cluster 1 at the top to cluster 4 at the bottom) for the first three days of 2013.

thus reduce the dimension of $\boldsymbol{\mathcal{P}}$: at each day t, the electricity provider has now to choose among $13^3 = 2,197$ configurations. Once these $\bar{J}$ tariff profiles have been selected, for each cluster i, for each $J(i) = 1, \ldots, \bar{J}$, and each day t of 2013, by keeping the weather variables observed in the Low Carbon London data set, we simulated the H-dimensional consumption profile $Y^i_{t,J(i)}$. In Figure 8.3, we plotted, for each cluster, the expected average power consumption (computed over 1,000 simulations) associated with each tariff profile.

Standardization of the electricity consumptions. For a cluster i and a day t, the vector $Y^i_{t,J(i)}$ is the electricity consumption profile of cluster i receiving profile $J(i)$, averaged over the households which are in the cluster. In all what follows, for simplicity, we assume that each cluster is made of the same number N of households. Therefore, it is enough to consider consumptions of the form

$$Y^{\mathrm{g}}_t = \sum_{i \in \mathrm{g}} Y^i_{t,J(i)}\,,$$

and by multiplying all the consumptions by N, we can get the non-averaged consumptions.

Creation of the targets. We recall that, as turning on and off power plants is not instantaneous and requires scheduling, electricity providers generally want to smooth as much as possible the demand curve. Therefore, we build a smooth global target for the global consumption

$$Y^{\mathrm{TOT}}_t = \sum_{i=1}^{G} Y^i_t\,.$$

The ideal would be to obtain a curve close to a constant curve but which remains realistic, i.e. which is attainable (or almost attainable) so it can be approached by applying one of the available tariff profiles. To do so, we compute the mean consumption $\bar{c}^{\mathrm{TOT}}$ of the entire households population over the year 2013. Then, for each day t, we look at the 13 expected average global electricity consumption profiles (computed over 1,000 simulations) associated with the configurations for which the entire population received the same tariff profile. For each half-hour h, we thus obtained a minimal and a maximal average consumption, denoted by $Y^h_{t,\min}$ and $Y^h_{t,\max}$, respectively. Then, we clipped the target $\bar{c}^{\mathrm{TOT}}$ between these two values. To obtain a realistic and rather flat target, we smooth it using the `R`-function `loess`; that is

$$c^{\mathrm{TOT},h}_t = \texttt{loess}\left(\min\left\{ \max\left\{\bar{c}^{\mathrm{TOT}}, Y^h_{t,\min}\right\}, Y^h_{t,\max} \right\}\right).$$

On the top of Figure 8.4 , we represent, for the first three days of 2013, the minimal and maximal average consumptions, denoted by $Y^h_{t,\min}$ and $Y^h_{t,\max}$, in blue and navy, respectively. In red dashed line, we plot the quantities used to computed the target: $\min\{\max\{\bar{c}^{\mathrm{TOT}}, Y^h_{t,\min}\}, Y^h_{t,\max}\}$. Finally the target consumption is in red solid line.

We also emphasize that demand side management strategies are developed to deal with intermittent and decentralized energies, like solar or wind power. To test the viability of our solution in such a situation, we may consider that one of the clusters (we arbitrary pick cluster 4 in our experiments) is located near a solar farm, and that between 10 a.m. and 5 p.m., this farm provides some electricity that has to be be consumed. To do so, it is

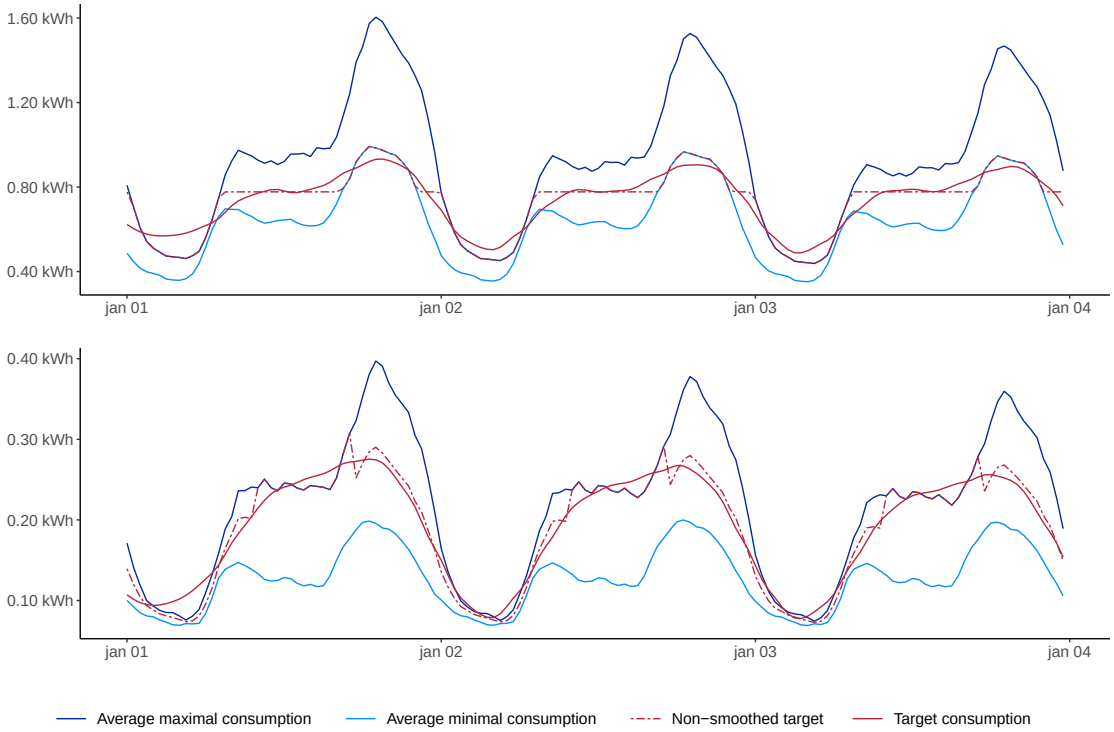


Figure 8.4 – Target consumption (red solid lines), and non-smoothed target before smoothing (red dashed lines) and minimal and a maximal average consumptions (blue and navy lines, respectively) for the global electricity consumption (at the top) and for cluster 4 (at the bottom) for the first three days of 2013.

enough to set the local target $c_t^{4,h}$ to a high value (namely, at $Y_{t,\max}^{4,h}$, the maximal average consumption of cluster 4 for the day t and the half-hour h) between 10 a.m. and 5 p.m.; and to the mean (over the tariff profiles) expected value otherwise. We also smooth the local target $c_t^{4,h}$ with the `loess`-function. This target is represented at the bottom of Figure 8.4 in red solid line. The red dashed line is the non-smoothed target and the minimal and a maximal average consumptions for cluster 4 are in blue and navy, respectively.

We highlight that contrary to the experiences of Chapter 3, these targets are not necessarily attainable. On the one hand because the smoothing introduces half-hours for which the target is higher than the maximum expected consumption or lower than the minimum expected consumption; and on the other hand, because sending of Dirac masses makes the notion of attainability obsolete. Indeed, by considering, for example, a global consumption target, as the global electricity consumption is $Y^{\mathrm{TOT}} = \sum_{i=1}^{G} Y_{t,J(i)}^{i}$, the targets which are attainable (in the way it was defined in Chapter 4) are that quantities

$$\mathbb{E}\Big[\sum_{i=1}^{G} Y_{t,J(i)}^{i}\Big], \quad \text{for}\big(J(1),\dots,J(G)\big) \in G^{\bar{J}}.$$

Practically, this values are unknown. Moreover, as soon as we consider several targets, the attainability assumption becomes even more complex. This assumption was useful in Section 4 of Chapter 4 to obtain a reduction of the order of magnitude of the regret bound to a poly-logarithmic rate; but the regret bound stated in Section 3 does not required any

attainability assumption.

4.1.3 Assumptions and optimistic algorithm

We recall that the algorithm stated in Equation (8.3) picks tariff allocations by solving:

$$\boldsymbol{p}_t \in \operatorname*{argmin}_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \left\{ \widehat{\ell}_{t,\boldsymbol{p}} - \alpha_{t,\boldsymbol{p}} \right\}.$$

Computationally, this minimization problem may be difficult to solve. We explain below the assumptions we made on the electricity consumption and how we practically adapted this algorithm.

Covariance matrices. We highlight that if, at a day t and a half-hour h, all price allocations p^i, which are sent to clusters i, are Dirac masses in $j(i)$, the expected loss becomes

$$\begin{aligned}\ell_{t,\boldsymbol{p}} &= \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}}\theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i\in\mathrm{g}} (p^i)^{\mathrm{T}} \Sigma^i p^i \right) \\ &= \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}}\theta^i - c_t^{\mathrm{g}} \Big)^2 + \sum_{i\in\mathrm{g}} \Sigma^i_{j(i)j(i)} \right),\end{aligned}$$

without any h-indexation (for the ease of notation). By making the strong assumption that all tariffs have the same variance, namely that, for any j, $\Sigma^i_{jj} = \sigma^2$, we get that

$$\ell_{t,\boldsymbol{p}} = \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}}\theta^i - c_t^{\mathrm{g}} \Big)^2 + |\mathrm{g}|\sigma^2 \right).$$

Thus, when we look at the instantaneous regrets $\ell_{t,\boldsymbol{p}_t} - \min_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \ell_{t,\boldsymbol{p}}$, the terms $\sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} |\mathrm{g}|\sigma^2$ cancel each other out. Moreover, as

$$\operatorname*{argmin}_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \big\{ \widehat{\ell}_{t,\boldsymbol{p}} - \alpha_{t,\boldsymbol{p}} \big\} = \operatorname*{argmin}_{\boldsymbol{p}\in\boldsymbol{\mathcal{P}}} \Big\{ \widehat{\ell}_{t,\boldsymbol{p}} - \alpha_{t,\boldsymbol{p}} - \sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} |\mathrm{g}|\sigma^2 \Big\},$$

there is no need to estimate σ^2 and it is enough to estimate

$$\sum_{\mathrm{g}\in\mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\Big(\sum_{i\in\mathrm{g}} \varphi^i(x_t^i, p^i)^{\mathrm{T}}\theta^i - c_t^{\mathrm{g}} \Big)^2 \right).$$

Underlying additive models. We recall that the available tariff profiles are numbered from 1 to $\bar{J} = 13$. For a day t and a half-hour h, the exogenous variables τ_t^h, $\bar{\tau}_t^h$, π_t^h and w_t refer to the temperature, the smoothed temperature, the position in the year and the type of day, respectively (see Section 2 of Chapter 3 for further details on these variables). For each $i = 1, \ldots, G$, we consider that the power consumption of cluster i receiving tariff profile J can be modeled with:

$$\begin{aligned}Y_{t,J}^{i,h} &= s_\tau^{i,h}(\tau_t^h) + s_{\bar{\tau}}^{i,h}(\bar{\tau}_t^h) + s_\pi^{i,h}(\pi_t^h) + \sum_{w=0}^{6} \zeta_w^h \mathbb{1}_{\{w_t = w\}} + \xi_J^{i,h} + \text{noise} \\ &= f^{i,h}(x_t) + \xi_J^{i,h} + \text{noise}\,, \end{aligned} \tag{8.12}$$

where the functions $s_\tau^{i,h}$, $s_{\bar{\tau}}$ and $s_\pi^{i,h}$ are cubic splines; $\xi_J^{i,h}$ is this effect of profile J at the half-hour h; and the noises are centered, Gaussian, and of variance σ^2. Therefore, we consider that the effect of the tariff profile J does not depend on the exogenous variables τ_t^h, $\bar{\tau}_t^h$, etc.; but only on the half-hour h and on the cluster i. Moreover, we assume that the functions $f^{i,h}$ are known. In the experiments, we used a year of power consumption data, simulated only for "non-demand side management" days (namely, with a tariff which was always equal to Normal), and estimated these functions using the `mgcv`-function (see Wood, 2020).

Expected losses. For a day t, the loss suffered by the electricity provider is the sum of the losses suffered at each half-hour (we refer to the Section 6 of Chapter 4 for further details). Therefore, for each day t and any $(p^1, \ldots, p^G)$, we need to estimate

$$\sum_{h=1}^{H} \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\left(\sum_{i \in \mathrm{g}} \varphi^{i,h}(x_t^i, p^i)^{\mathrm{T}} \theta^{i,h} - c_t^{\mathrm{g},h} \right)^2 \right).$$

We recall that in our experiments, price allocations p_i are Dirac masses and that for each day t, the electricity provider picks, for each cluster i, a profile among $\{1, \ldots, \bar{J}\}$; therefore, $\boldsymbol{p}$ is from now on a set of profiles $\{J(1), \ldots, J(G)\}$ – when $\boldsymbol{p}$ is picked, each cluster i receives profile $J(i)$. If the electricity provider chooses the profiles $\boldsymbol{p} = \{J(1), \ldots, J(G)\}$, by replacing the local expected electricity consumptions $\varphi^i(x_t^i, p^i)^{\mathrm{T}} \theta^i$ by $f^{i,h}(x_t) + \xi_J^{i,h}$ in the above expression, we have to estimate

$$\sum_{h=1}^{H} \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\sum_{i \in \mathrm{g}} f^{i,h}(x_t) + \xi_{J(i)}^{i,h} - c_t^{\mathrm{g},h} \right)^2.$$

We highlight that we could have weighed the half-hourly losses. To estimate these losses, it is enough to estimate the tariff effects $\xi_{J(i)}^{i,h}$. We emphasize that there are $\bar{J} \times H \times G = 13 \times 48 \times 4 = 2,496$ of them (but as clusters 1 and 2 received the same tariff profiles and as there is no local target on these clusters, we can consider them as a single cluster and therefore, there are $13 \times 48 \times 3 = 1,872$ coefficients to estimate). As the set $\boldsymbol{\mathcal{P}}$ is finite and because the effects of the tariffs do not depend on the contextual variables, we are almost as in a classical multi-armed bandit problem (and not any more in a linear bandit setting); there is thus no need to consider Ridge regression to estimate the tariff effects coefficients $\xi_{J(i)}^{i,h}$ (that form the parameter vectors θ^i). As in classical bandit, this can be done by taking the empirical mean, and for a cluster i, a tariff profile J and a half-hour h, we consider the estimators:

$$\widehat{\xi}_{J,t}^{i,h} = \frac{1}{N_{J,t}^i} \sum_{s=1}^{t} \left(Y_s^{i,h} - f^{i,h}(x_s) \right) \mathbb{1}_{\{J_s(i)=J\}}, \quad \text{with} \quad N_{J,t}^i = \sum_{s=1}^{t} \mathbb{1}_{\{J_s(i)=J\}},$$

where $J_s(i)$ is the tariff profile sent to cluster i at day s and $N_{J,t}^i$ the number of times profile J has been picked for cluster i. Then, for each set of tariff profiles $\boldsymbol{p} = (J(1), \ldots, J(G))$, we consider the expected losses

$$\widetilde{\ell}_{t,\boldsymbol{p}} = \sum_{h=1}^{H} \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \left(\sum_{i \in \mathrm{g}} f^{i,h}(x_t) + \widehat{\xi}_{J(i),t-1}^{i,h} - c_t^{\mathrm{g},h} \right)^2.$$

Exploration term. It only remains to adapt the exploration terms $\alpha_{t,\boldsymbol{p}}$ defined in Proposition 4. As we did not take into account the variance terms, $\alpha_{t,\boldsymbol{p}}$ only depends on how well the coefficients $\xi^{i,h}_{J(i)}$ have been estimated. Moreover, as we estimated these coefficients with empirical means (as in the classical UCB algorithm) we also considered exploration terms of the form of the ones of the UCB algorithm (see Chapter 2 for further details) and for each $\boldsymbol{p} = (J(1), \dots, J(G))$ we define

$$\widetilde{\alpha}_{t,\boldsymbol{p}} = \sum_{h=1}^{H} \sum_{\mathrm{g} \in \mathcal{G}_t} 2\sigma C^{\mathrm{g}} \kappa_t^{\mathrm{g}} \sum_{i \in \mathrm{g}} \sqrt{\frac{8 \ln t}{N^i_{J(i),t-1}}} = \sum_{\mathrm{g} \in \mathcal{G}_t} 2H\sigma C^{\mathrm{g}} \kappa_t^{\mathrm{g}} \sum_{i \in \mathrm{g}} \sqrt{\frac{8 \ln t}{N^i_{J(i),t-1}}} \,.$$

The standard deviation σ is set to 2.10^{-3} after estimating it on the simulated data. This value is quite low, and this is due to CVAEs, which produce samples with low variability (see Section 6 of Chapter 7). However, we kept the value to favor exploitation (the confidence levels provided by theoretical results often lead to high exploration terms – we tried different values and this ones gives a correct trade-off between exploration and exploitation).

Remark 28. *We highlight that these confidence terms are not so far from the one of Proposition 4. For the ease for notation, we do not index the variables by the half-hour and we recall that we had*

$$\alpha_{t,\boldsymbol{p}} \triangleq \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \left[\min \left\{ L^{\mathrm{g}}, 2C^{\mathrm{g}} \sum_{i \in \mathrm{g}} B^i_t(\delta/Gt^2) \left\| (V^i_{t-1})^{-1/2} \varphi^i(x^i_t, p^i) \right\| \right\} + \sum_{i \in \mathrm{g}} G^i_\tau(\delta/2G) \right]$$

$$\text{with} \quad V^i_t \triangleq \lambda^i \mathrm{I}_{d^i} + \sum_{s=1}^{t} \varphi^i(x^i_s, p^i_s) \varphi^i(x^i_s, p^i_s)^{\mathrm{T}} \,,$$

As we do not take into account the variance term, $\sum_{i \in \mathrm{g}} G^i_\tau(\delta/2G)$ can be considered null. Moreover, in our experiment setting, the parameter θ^i is made of the coefficients ξ^i_J (so $d(i) = \bar{J}$) and the mapping function is simply $\varphi^i(x^i, p^i) = (\mathbb{1}_{\{J(i)=J\}})_{J=1,\dots,\bar{J}}$. Therefore the matrix V^i_t is diagonal, and we have

$$\sum_{s=1}^{t} p^i_s (p^i_s)^{\mathrm{T}} = \sum_{s=1}^{t} \left(\mathbb{1}_{\{J(i)_s=J\}}\right)_{J=1,\dots,\bar{J}} \left(\mathbb{1}_{\{J(i)_s=J\}}\right)^{\mathrm{T}}_{J=1,\dots,\bar{J}} \begin{bmatrix} N^i_{1,t} & 0 & \dots \\ 0 & \ddots & 0 \\ \dots & 0 & N^i_{\bar{J},t} \end{bmatrix} .$$

With $\lambda = 0$, for a tariff profile J, we get that

$$\left\| (V^i_{t-1})^{-1/2} \left(\mathbb{1}_{\{J(i)=J\}}\right)_{J=1,\dots,\bar{J}} \right\| = \frac{1}{\sqrt{N^i_{J,t}}} \,.$$

Finally we recall that $B^i_t(\delta/Gt^2)$ is of order $\rho \ln t$, where ρ is the sub-Gaussian coefficient of the noise (which is equal to σ is our experiments). We point out that we can not take $\lambda = 0$ in $B^i_t(\delta/Gt^2)$ definition. Finally, we highlight that $B^i_t(\delta/Gt^2)$ also depends on δ but that we consider an algorithm very similar to the UCB algorithm (and thus which does not depend on any risk level).

Optimistic algorithm. After $\bar{J} = 13$ initial exploration round ($J_t(i) = t$ for all clusters), at a day t, the algorithm of the experiment picks the tariff profile

$$\boldsymbol{p}_t = \Big(J_t(1), \dots, J_t(G) \Big) \in \operatorname*{argmin}_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \left\{ \widetilde{\ell}_{t,\boldsymbol{p}} - \widetilde{\alpha}_{t,\boldsymbol{p}} \right\} .$$

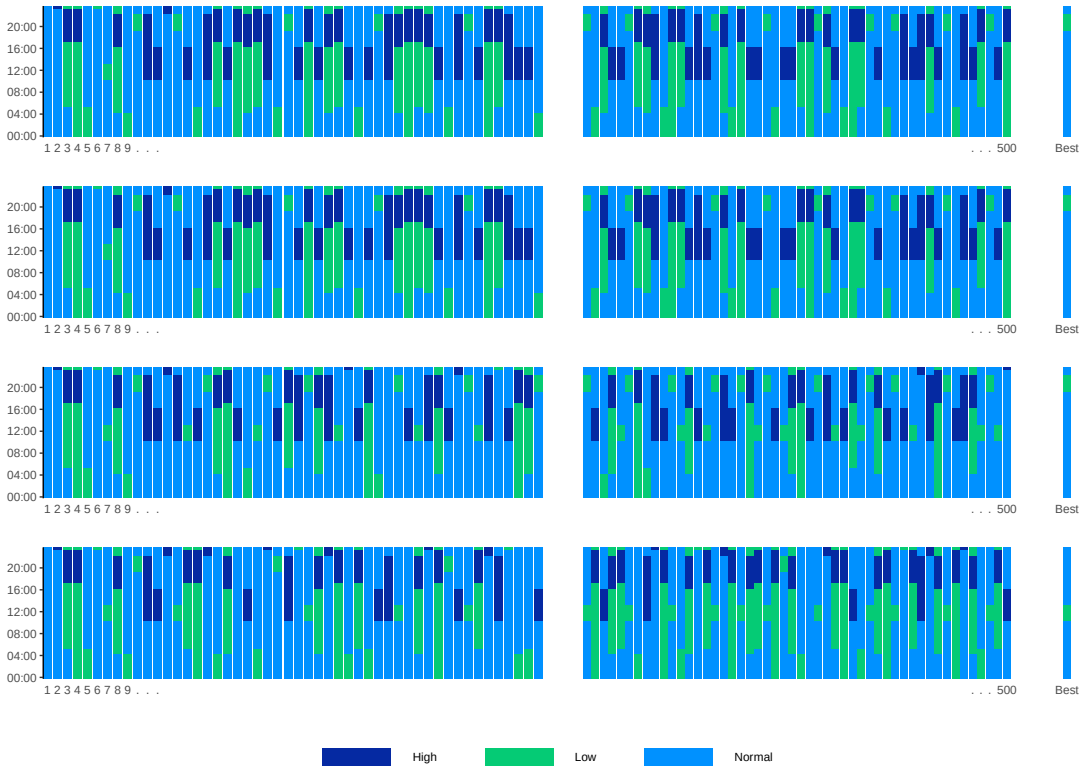


Figure 8.5 – Tariff profiles picked for each cluster (from cluster 1 at the top to cluster 4 at the bottom) for the 50 first (left) and 50 last (right) iterations of the algorithm, with $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$.

4.2 Results

We consider two experiments: in the first one, to see if the algorithm is both learning and optimizing the tariff effects, we set the targets and exogenous variables to the ones of January, 1st (that is, except the tariff picked, all variables are constant), while in the second experiments, targets and exogenous variables vary across 2013. For both experiments, we consider the targets of Figure 8.4 and, by denoting by κ^{TOT} and κ^4 the weights associated with the global electricity consumption and the one of cluster 4, respectively, we compare the results for $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$; and for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$, thus, for the second configuration, only a global target is considered. We run the algorithm for $T = 500$ iterations for the first experiment and for T =1,500 for the second one (after 365, 730, 1,095 etc. iterations, we start again the year 2013, with the same weather variables).

4.2.1 Constant exogenous variables

For both configurations, that is for $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$ and for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$, we represent in Figures 8.5 and 8.6, respectively, the tariff profiles sent for the 50 first and the 50 last iterations to each cluster (as we impose it, clusters 1 and 2 received the same profiles); on the far right of the figures, the best tariff (in expectation) profile is also plotted. First, and fortunately, by looking at the far right of the figures, we can see that the best strategy (namely the one which consist in picking, for each day, the "best" profiles) varies depending on whether we consider a local target or not: the best tariff

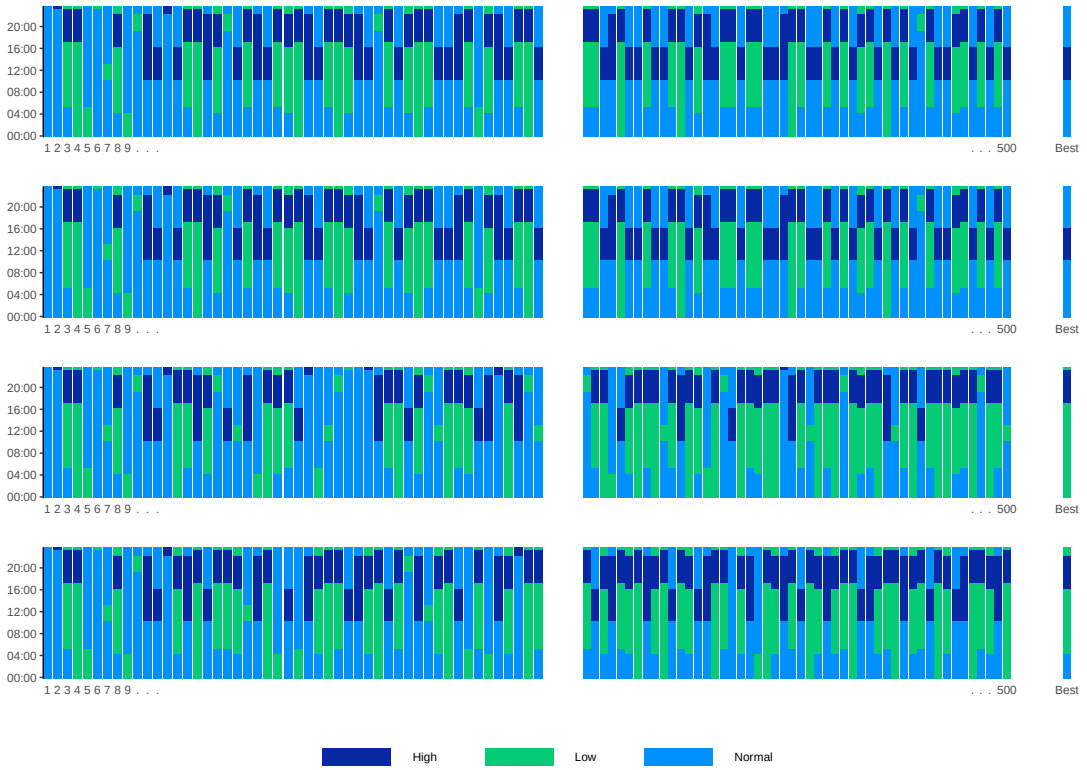


Figure 8.6 – Tariff profiles picked for each cluster (from cluster 1 at the top to cluster 4 at the bottom) for the 50 first (left) and 50 last (right) iterations of the algorithm, with $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$.

profile to pick has changed for cluster 4, as well as for all the others. Let us focus on cluster 4. When a local target is imposed on it, the best tariff profile to pick is profile 7 (see Figure 8.2). Figure 8.5 shows that at the beginning of the algorithm, after the 13 initial exploration rounds, the algorithm is still exploring, while at the end it picks many times profile 7, but also profiles 8 and 4. We highlight that when we look at the expected electricity consumption of cluster 4, for profiles 8 and 7, the curves are very closed together (see Figure 8.1). But even after 450 iterations, the algorithm is still exploring: each tariff profile has been picked at least once between the iterations 450 and 500. For the second configuration, when $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$, the algorithm seems to converge more quickly and at the end of the execution, it hesitates between few profiles. These differences in the speed of "convergence" may be explained by looking at the expected losses and exploration terms formulas: both depend linearly on the coefficients κ^{g} but the differences between the expected consumptions and the targets are generally lower for a single cluster than for the global household population (and in both cases lower than 1), raising them to the power of two accentuates this differences that are not enough compensated by the coefficients κ^{g}. As smaller expected losses favor exploration, the convergence of the first configuration is slower.

In Figures 8.7 and 8.8, we plotted, for cluster 4 and for the global household population, the target and the average consumption over the 25 first (on the left of the figure) and over the 25 last (on the right) iterations associated with the tariff picked in blue. The

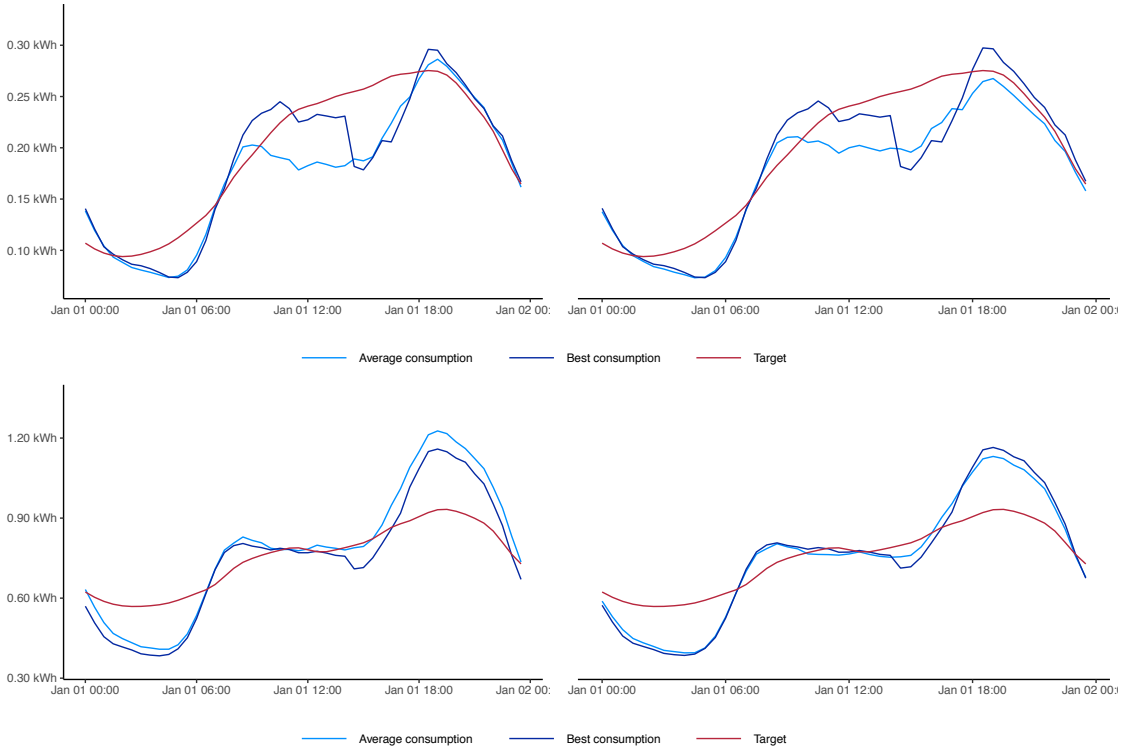


Figure 8.7 – Average, over the 25 first (on the left of the figure) and over the 25 last (on the right) iterations, consumption (in kWh) for tariffs profile picked by the algorithm (in blue), target (in red) and consumption for tariffs profile for the best tariff to play for $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$, for cluster 4 (on the top) and for the global households population (on the bottom).

consumption associated with best profile to pick is also plotted in navy ("best" according to our modeling). Over the iterations, the algorithm approaches the targets and reach the best strategies (the two blue curves overlap). In Figure 8.7, the consumption of cluster 4 seems closer to the target than the best strategy: the best profile to choose is not necessarily the one associated with the consumption closest to the target because of the term relative to c_t^{TOT} in the loss. The algorithm has to find the right compromise between the two targets and the observation of consumptions close to the target at the beginning of the execution of the algorithm is probably due to chance. To support these results, the instantaneous regrets are plotted in 8.9. We highlight that to compute this quantities we had to estimate the effects $\xi_j^{i,h}$, and we did it using 1,000 samples of power consumption records provided by the CVAE-based generator. The instantaneous regret is very high at the beginning of the execution (during the exploration rounds) and then it totally vanishes and oscillates between very low values. The algorithm seems to discriminate strongly some of the tariff profiles configurations. For others, the difference between the associated expected losses are not so important and it is difficult to choose between them. To make a comparison with classical bandits, the situation may be similar to a multi-armed bandit problem with sub-optimal arms gaps which are either very large or very small (we recall that the gap of an arm is the differences between the expected reward of the optimal arm and the one of the considered arm). We point out that, by cleverly allocating the tariff profiles to the clusters, the algorithm can continue to explore the various profiles, it just never picks some allocations which lead to large expected losses (e.g., profiles leading

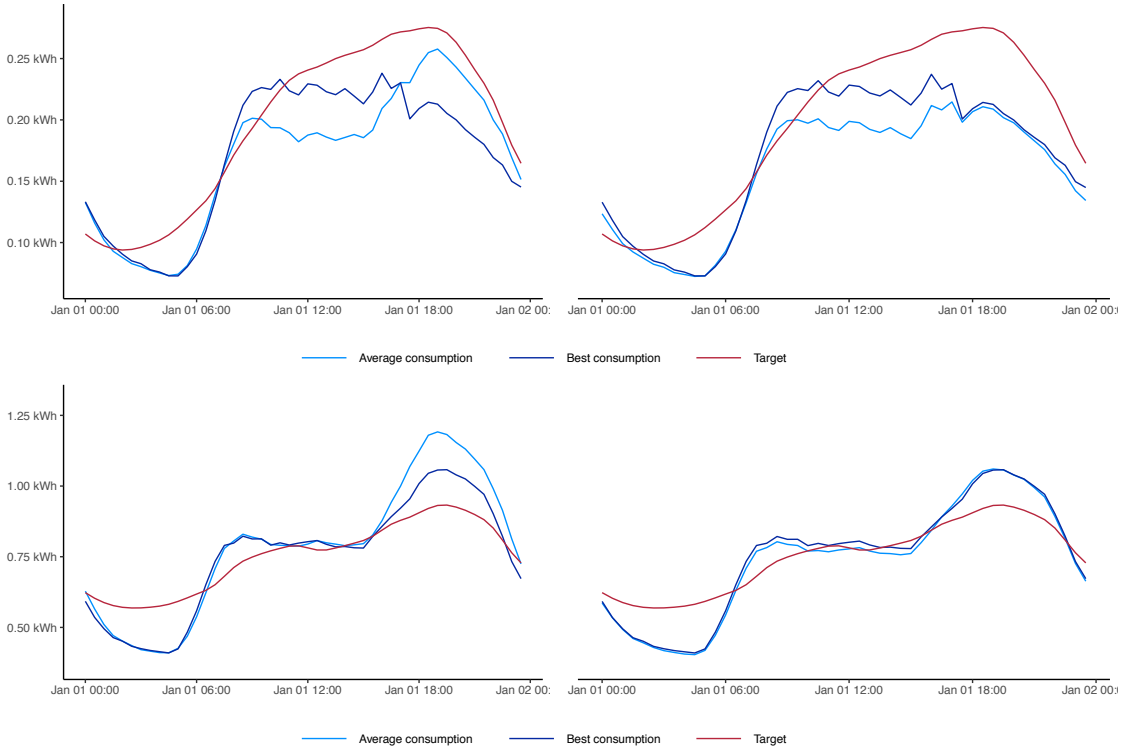


Figure 8.8 – Average, over the 25 first (on the left of the figure) and over the 25 last (on the right) iterations, consumption (in kWh) for tariffs profile picked by the algorithm (in blue), target (in red) and consumption for tariffs profile for the best tariff to play for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$, for cluster 4 (on the top) and for the global households population (on the bottom).

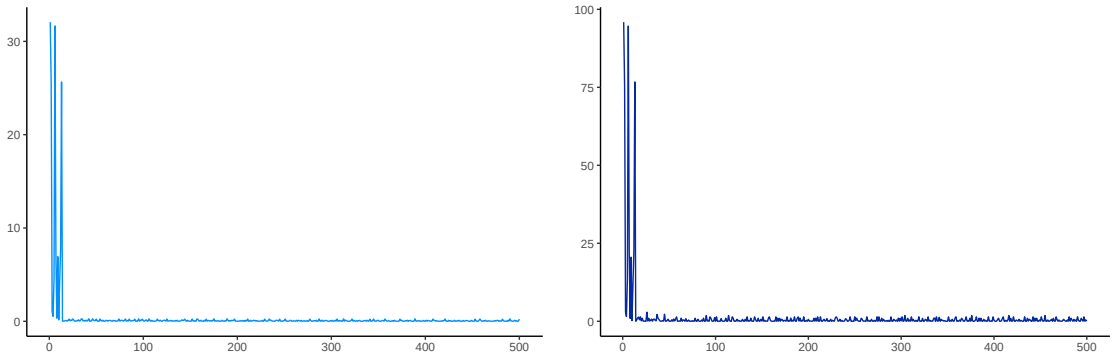


Figure 8.9 – Instantaneous regrets (in kWh^2) for the configuration $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$ (blue, to the left) and for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$ (navy, to the right).

to high evening consumption peaks for all clusters).

To be sure that the algorithm is learning thought the iterations, in Table 8.1, we computed the root average expected losses

$$\sqrt{\frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \ell_{t,\boldsymbol{p}}}\,,$$

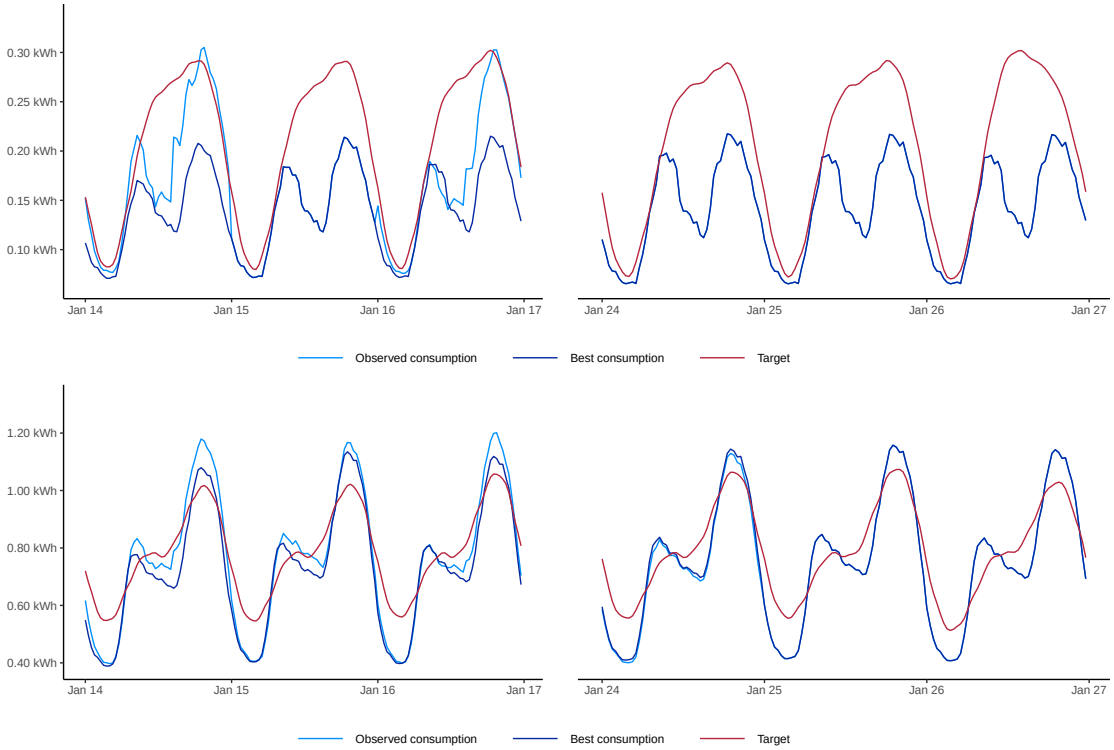


Figure 8.10 – Target consumption (in red) and observed electricity consumption (in kWh) for the tariff profiles picked by the algorithm (in blue) and for the best tariff profiles to pick (in navy) for $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$ and for the cluster 4 (on the top) and the global household population (on the bottom) for the iterations 14 to 16, namely just after the initial exploration rounds, (on the left); and for the iterations 379 to 381, namely after a year of learning, (on the right).

for the 100 iterations just after the exploration rounds and for the 100 last ones, and observe a slight improvement in the results.

Root average expected loss (in kWh)	Start	End
$\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$	0.36	0.31
$\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$	0.65	0.62

Table 8.1 – Root average expected loss (in kWh), computed over the 100 iterations ("Start") after initial exploration rounds, namely iterations from 14 to 114; and over the last 100 iterations ("End"), namely iterations from 400 to 500.

4.2.2 Non-constant exogenous variables

For this second experiment, we consider that weather and calendar variables are changing from an iteration to another. In Figures 8.10 and 8.11, we plotted, the consumption (in blue) observed for the tariff profiles picked by the algorithm and the consumption associated with the "best" strategy (in navy) from January 14. to 17., after the explorations rounds (that is, at the beginning of the execution) and after a year of learning – so we compare consumptions associated with the same exogenous and calendar variables. It seems clear (even it is only over three iterations) that the algorithm can learn: after a year of learning, it chooses correctly (according to our modeling) the tariff profiles and

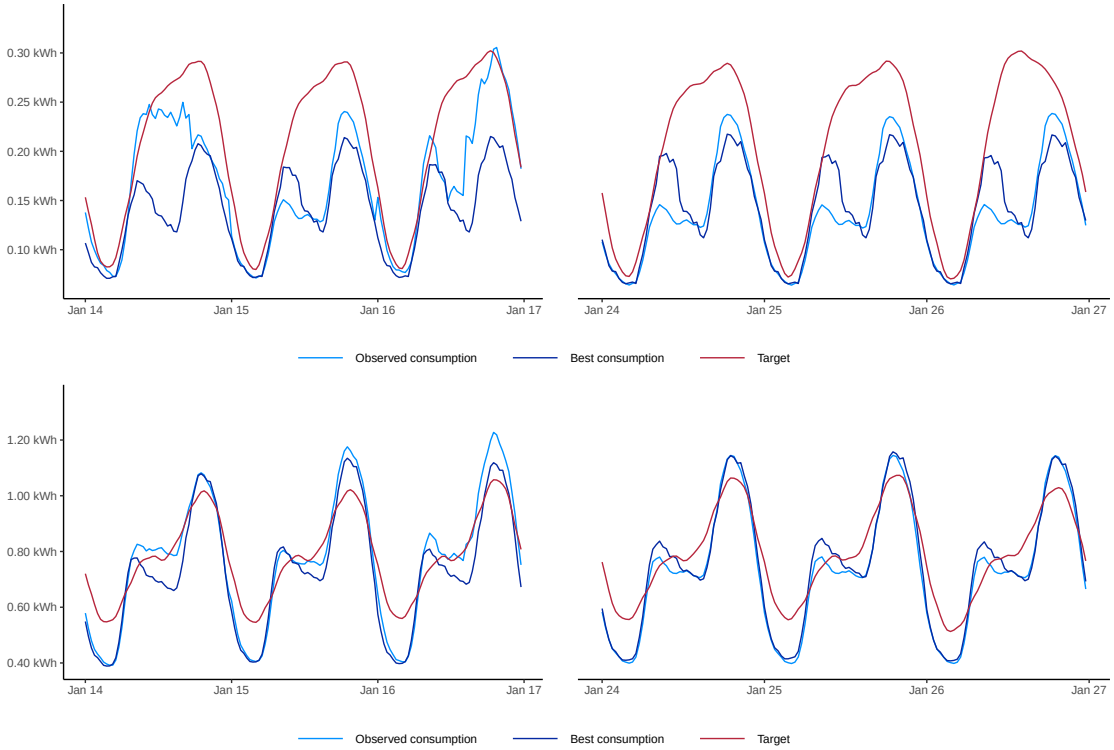


Figure 8.11 – Target consumption (in red) and observed electricity consumption (in kWh) for the tariff profiles picked by the algorithm (in blue) and for the best tariff profiles to picked (in navy) for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$ and for the cluster 4 (on the top) and the global household population (on the bottom) for the iterations 14 to 16, namely just after the initial exploration rounds, (on the left); and for the iterations 379 to 381, namely after a year of learning, (on the right).

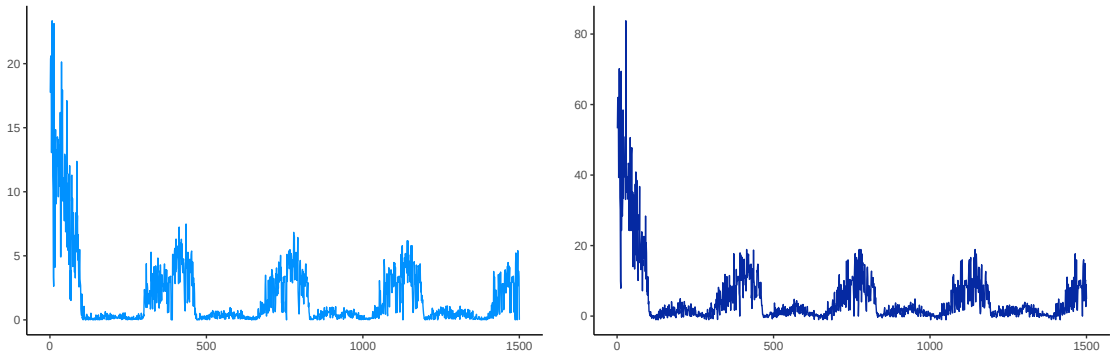


Figure 8.12 – Instantaneous regrets (in kWh^2) for the configuration $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$ (blue, to the left) and for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$ (navy, to the right).

both navy and blue curves overlap. The root average expected losses computed just after the exploration rounds and after a year of learning provided in Table 8.2 confirm these results.

In Figure 8.12, we plot the instantaneous regret and obtain surprising results: after a hundred iterations, the regret decreases rapidly towards 0, but it increases again around the iteration 300, before decreasing again. This process repeats each 365 iterations and

Root average expected loss (in kWh)	Start	End
$\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$	3.31	2.71
$\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$	5.86	4.64

Table 8.2 – Root average expected loss (in kWh), computed over the 100 iterations ("Start") after initial exploration rounds, namely iterations from 14 to 114; and after a year of learning ("End"), namely iterations from 379 to 479.

Root average loss (in kWh)	Start	End
$\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$	1.28	0.98
$\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$	2.20	1.68

Table 8.3 – Root average loss (in kWh), computed over the 100 iterations("Start") after initial exploration rounds, namely iterations from 14 to 114; and after a year of learning ("End"), namely iterations from 379 to 479.

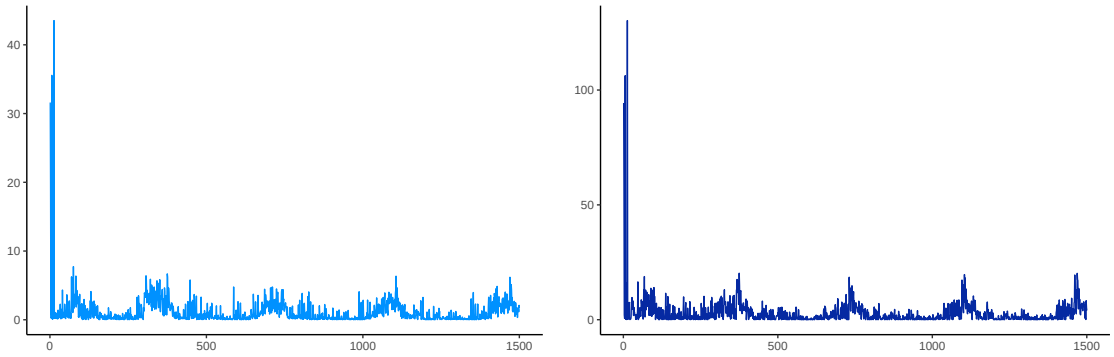


Figure 8.13 – Instantaneous losses (in kWh^2) for the configuration $\kappa^{\text{TOT}} = 1/3$ and $\kappa^4 = 1$ (blue, to the left) and for $\kappa^{\text{TOT}} = 1$ and $\kappa^4 = 0$ (navy, to the right).

an annual pattern appears (that is why we plot the regret over 1,500 iterations). This rebound may be explained by a seasonal effect. It seems that it always appears in winter and that it is a little less important from year to year. The variance of consumption, high in winter, could explain this phenomenon. Moreover, we recall that the instantaneous regret is the difference between the expected loss (given a model) and the best possible expected loss (given the same model). Here, the expected losses are approximated: it is clear that the power consumption profiles generated by the CVAE-based generator do not satisfy the model of Equation 8.12. These results on the instantaneous regret suggest that for some iterations (in winter), the estimations of the tariff effects $\xi_j^{i,h}$ are not satisfying; but it is not so surprising since the observations do not follow the model. To see if our algorithm is still robust, we compute the average (over 100 iterations) true loss (namely the losses suffered by the electricity provider):

$$\sqrt{\frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \sum_{h=1}^{H} \sum_{\mathrm{g} \in \mathcal{G}_t} \kappa_t^{\mathrm{g}} \Big(\sum_{i \in \mathrm{g}} Y_t^{i,h} - c_t^{\mathrm{g},h} \Big)^2}.$$

Results are in Table 8.3. Moreover, the instantaneous (true) losses are plotted in Figure 8.13. They decrease throughout the iterations, which suggests that our algorithm will be somewhat robust to observations that do not follow the model of Equation 8.12. But in

the same way as for instantaneous regrets, we observe higher losses in winter than for the rest of the year. We should also note that the target changes every day and that the latter can be more difficult to reach for certain days (in winter) than for others – and therefore be linked to high losses.

4.3 Perspectives

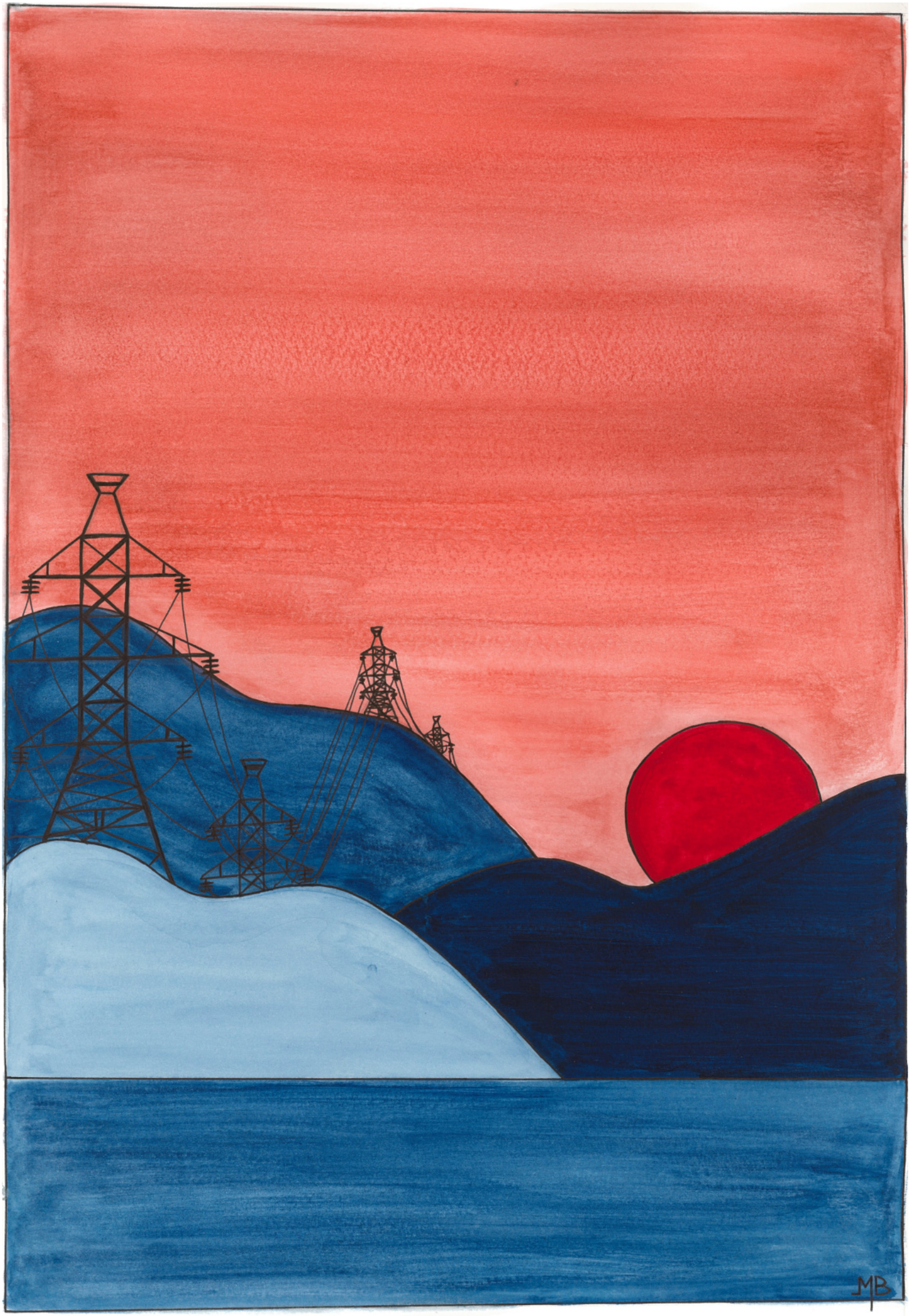
These experiments show that the algorithm quickly finds the best strategy (among the 2,197 configurations) and performs a personalized demand side management whether the exogenous variables are constant or not: the regrets and true losses are decreasing thought the iterations. It can do it on data that have been simulated with a data-driven approach and do not satisfy any generalize additive model
a priori. This argues for the robustness of our strategies. We also emphasize that the algorithm is able to find the best trade-off when several targets are given (with possibly different weights). However, these results are subject to improvement and open up many perspectives.

First, we highlight that we could have introduced some weights in the losses to give some importance to half-hours of the evening, that generally come with a peak in the electricity consumption and which are therefore crucial in demand side management.

We recall that there is no doubt on the efficiency of generalized additive models to model and forecast the electricity power consumption. In the experiments presented above, we consider a very basic model. Even if the results are promising, it is clear that under a changing environment we have to consider a better model and therefore to adapt the algorithm. Results should then be considerably improved. But such a modeling comes with computational difficulties: the dimension of the coefficients to estimate will increase, as well as the exploration terms. It is possible that an adaptation of the balance between exploration and exploitation is then necessary. We also point out that we completely dropped the variance matrices in these experiments, while they play an important role in the electricity consumption modeling.

Finally, a significant improvement would be to enlarge the set of price allocations. We highlight that other data sets would enrich these experiments: other tariff profiles could be considered. We also would like to be able to split the clusters and send different tariff profiles to each sub-cluster (as in the theory). Therefore, we could attain, in expectation, the targets (as in Chapter 4). However, by enlarging the set of possible price allocations, the minimization problem may then have several solutions (different ways of splitting the clusters can lead to the same expected loss) and that it is not trivial to solve it – because it is not convex.

These results are therefore encouraging and offer many ways of improvement. Challenges will be technical, because of the implementation of algorithms and models, as well as theoretical, with the resolution of minimization problems.


MB

Bibliography

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems (NIPS'11)*, pages 2312–2320, 2011.

Naoki Abe and Philip M Long. Associative reinforcement learning using linear probabilistic concepts. In *Proceedings of the 16th International Conference on Machine Learning (ICML'99)*, pages 3–11, 1999.

Ajith Abraham and Baikunth Nath. A neuro-fuzzy approach for modelling electricity demand in victoria. *Applied Soft Computing*, 1(2):127–138, 2001.

Mohamed H Albadi and Ehab F El-Saadany. Demand response in electricity markets: An overview. In *2007 IEEE power engineering society general meeting*, pages 1–5. IEEE, 2007.

Christophe Amat, Tomasz Michalski, and Gilles Stoltz. Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance*, 88:1–24, 2018.

Anestis Antoniadis, Efstathios Paparoditis, and Theofanis Sapatinas. A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):837–857, 2006.

A-U Asar and James R Mcdonald. A specification of neural network applications in the load forecasting problem. *IEEE transactions on control systems technology*, 2(2): 135–141, 1994.

Benjamin Auder, Jairo Cugliari, Yannig Goude, and Jean-Michel Poggi. Scalable clustering of individual electrical curves for profiling and bottom-up forecasting. *Energies*, 11(7): 1893, 2018.

Jean-Yves Audibert and Sébastien Bubeck. Best Arm Identification in Multi-Armed Bandits. In *COLT - 23th Conference on Learning Theory - 2010*, page 13 p., Haifa, Israel, June 2010. URL `https://hal-enpc.archives-ouvertes.fr/hal-00654404`.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002a.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The non-stochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002b.

Maryam Aziz, Emilie Kaufmann, and Marie-Karelle Riviere. On Multi-Armed Bandit Designs for Dose-Finding Trials. April 2020. URL `https://hal.archives-ouvertes.fr/hal-02533297`. working paper or preprint.

Katy S. Azoury and Manfred K. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, 43(3):211–246, 2001.

Amadou Ba, Mathieu Sinn, Yannig Goude, and Pascal Pompey. Adaptive learning of smoothing functions: Application to electricity load forecasting. In *Advances in neural information processing systems*, pages 2510–2518, 2012.

Hamsa Bastani and Mohsen Bayati. Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294, 2020.

Souhaib Ben Taieb and Rob J Hyndman. A gradient boosting approach to the kaggle load forecasting competition. *International journal of forecasting*, 30(2):382–394, 2014.

Souhaib Ben Taieb, James W Taylor, and Rob J Hyndman. Coherent probabilistic forecasts for hierarchical time series. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3348–3357. JMLR. org, 2017.

Souhaib Ben Taieb, James W Taylor, and Rob J Hyndman. Hierarchical probabilistic forecasting of electricity demand with smart meter data. *Journal of the American Statistical Association*, pages 1–17, 2020.

Alina Beygelzimer, John Langford, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS'11)*, pages 19–26, 2011.

Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, 2016.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Leo Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, 1984. ISBN 0-534-98053-8.

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.

D Bunn and E D Farmer. Comparative models for electrical load forecasting. 1 1985.

Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

Alfonso Capasso, W Grattieri, R Lamedica, and A Prudenzi. A bottom-up approach to residential load modeling. *IEEE transactions on power systems*, 9(2):957–964, 1994.

Raymond J Carroll, Jianqing Fan, Irene Gijbels, and Matt P Wand. Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438): 477–489, 1997.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Bo-Juen Chen, Ming-Wei Chang, et al. Load forecasting using support vector machines: A study on eunite competition 2001. *IEEE transactions on power systems*, 19(4): 1821–1830, 2004.

Xin Chen, Yutong Nie, and Na Li. Online residential demand response via contextual multi-armed bandits. *arXiv preprint arXiv:2003.03627*, 2020.

Gianfranco Chicco, Roberto Napoli, and Federico Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21(2):933–940, 2006.

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.

Thomas M. Cover. Universal Portfolios. *Mathematical Finance*, 1(1):1–29, 1991. ISSN 14679965. doi: 10.1111/j.1467-9965.1991.tb00002.x.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT'08)*, 2008.

Raphaël Deswarte, Véronique Gervais, Gilles Stoltz, and Sébastien da Veiga. Sequential model aggregation for production forecasting. 2019.

Marie Devaine, Pierre Gaillard, Yannig Goude, and Gilles Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90(2):231–260, 2013.

Joseph Leo Doob. *Stochastic processes*, volume 101. New York Wiley, 1953.

Grzegorz Dudek. Short-term load forecasting using random forests. In *Intelligent Systems' 2014*, pages 821–828. Springer, 2015.

D. M. Dunn, W. H. Williams, and T. L. DeChaine. Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association*, 71(353):68–71, 1976. ISSN 01621459.

Goutam Dutta and Krishnendranath Mitra. A literature review on dynamic pricing of electricity. *Journal of the Operational Research Society*, 68(10):1131–1145, 2017.

Robert F Engle, Clive WJ Granger, John Rice, and Andrew Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394):310–320, 1986.

M Ernoult, R Mattatia, F Meslier, and P Rabut. Estimation of the sensitivity of the electrical energy demand to variations in meteorological conditions: History of methods and development of new approaches at edf. *International Journal of Electrical Power & Energy Systems*, 5(3):189–198, 1983.

Shu Fan and Rob J Hyndman. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1):134–141, 2012.

Matteo Fasiolo, Simon N Wood, Margaux Zaffran, Raphaël Nedellec, and Yannig Goude. Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, pages 1–11, 2020.

José Nuno Fidalgo, Manuel António Matos, and Luis Ribeiro. A new clustering algorithm for load profiling based on billing data. *Electric Power Systems Research*, 82(1):27–33, 2012.

Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems (NIPS'10)*, pages 586–594, 2010.

David Fischer, Johannes Scherer, Alexander Flunk, Niklas Kreifels, Karen Byskov-Lindberg, and Bernhard Wille-Haussmann. Impact of hp, chp, pv and evs on households' electric load profiles. In *2015 IEEE Eindhoven PowerTech*, pages 1–6. IEEE, 2015.

Dylan J Foster, Alekh Agarwal, Miroslav Dudík, Haipeng Luo, and Robert E Schapire. Practical contextual bandits with regression oracles. *arXiv preprint arXiv:1803.01088*, 2018.

Pierre Gaillard. *Contributions to online robust aggregation : work on the approximation error and on probabilistic forecasting. Applications to forecasting for energy markets.* Theses, Université Paris Sud - Paris XI. `https://tel.archives-ouvertes.fr/tel-01250027`, July 2015. URL `https://tel.archives-ouvertes.fr/tel-01250027`.

Pierre Gaillard, Gilles Stoltz, and Tim van Erven. A second-order bound with excess losses. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of the 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 176–196, Barcelona, Spain, June 2014. PMLR.

Pierre Gaillard, Yannig Goude, and Raphaël Nedellec. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, 32(3):1038–1050, 2016. doi: 10.1016/j.ijforecast.2015.12.001.

Pierre Gaillard, Sébastien Gerchinovitz, Malo Huard, and Gilles Stoltz. Uniform regret bounds over $\mathbb{R}^d$ for the sequential linear regression problem with the square loss. In Aurélien Garivier and Satyen Kale, editors, *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, volume 98 of *Proceedings of Machine Learning Research*, pages 404–432, Chicago, Illinois, March 2019. PMLR.

Kamalanathan Ganesan, João Tomé Saraiva, and Ricardo J Bessa. On the use of causality inference in designing tariffs to implement more effective behavioral demand response programs. *Energies*, 12(14):2666, 2019.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 249–256, 2010.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Benjamin Goehry, Yannig Goude, Pascal Massart, and Jean-Michel Poggi. Aggregation of multi-scale experts for bottom-up load forecasting. *IEEE Transactions on Smart Grid*, 11(3):1895–1904, 2019.

Aditya Gopalan, Shie Mannor, and Yishay Mansour. Thompson sampling for complex online problems. In *Proceedings of the 31st International Conference on Machine Learning (ICML'14)*, pages 100–108, 2014.

Sebastian Gottwalt, Wolfgang Ketter, Carsten Block, John Collins, and Christof Weinhardt. Demand side management–A simulation of household behavior undervariable prices. *Energy Policy*, 39:8163–8174, 2011.

Yannig Goude. *Mélange de prédicteurs et application à la prévision de consommation électrique*. PhD thesis, Thèse de doctorat, Université Paris-Sud, 2008.

Yannig Goude and Pierre Gaillard. Opera: a r package for online prediction by expert aggregation. 06 2016.

Yannig Goude, Raphaël 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.

Charles W. Gross and Jeffrey E. Sohl. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, 9(3):233–254, 1990. doi: 10.1002/for.3980090304.

Chong Gu. *Smoothing Spline ANOVA Models*. New York: Springer, 2002.

Takeshi Haida and Shoichi Muto. Regression based peak load forecasting using a transformation technique. *IEEE Transactions on Power Systems*, 9(4):1788–1794, 1994.

Andrew Harvey and Siem Jan Koopman. Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88(424):1228–1236, 1993.

Trevor Hastie and Robert Tibshirani. Generalized linear models. *Statistical Science*, 1(3): 297–318, 1986.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. *5th International Conference on Learning Representations (ICLR)*, 2(5):6, 2017.

Geoffrey E Hinton and Richard S Zemel. Autoencoders, minimum description length and helmholtz free energy. In *NIPS'93: Proceedings of the 6th International Conference on Neural Information Processing Systems*, pages 3–10, 1994.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. pages 409–426, 1994.

Arthur E Hoerl, Arthur E Hoerl, AE HOERL, and C Hoerl. Application of ridge analysis to regression problems. *Chem. Eng. Prog.*, 1962.

Tao Hong, Jingrui Xie, and Jonathan Black. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. *International Journal of Forecasting*, 35(4): 1389–1399, 2019.

Rob J Hyndman and Shu Fan. Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, 25(2):1142–1153, 2010.

Rob J Hyndman, Anne B Koehler, Ralph D Snyder, and Simone Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, 18(3):439–454, 2002.

Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

Yumiko Iwafune, Kazuhiko Ogimoto, Yuki Kobayashi, and Kensuke Murai. Driving simulator for electric vehicles using the Markov chain monte carlo method and evaluation of the demand response effect in residential houses. *IEEE Access*, 8:47654–47663, 2020.

Pooria Joulani, Andras Gyorgy, and Csaba Szepesvári. Online learning under delayed feedback. In *International Conference on Machine Learning*, pages 1453–1461, 2013.

Sophie Schönfeldt Karlsen, Mohamed Hamdy, and Shady Attia. Methodology to assess business models of dynamic pricing tariffs in all-electric houses. *Energy and Buildings*, 207:109586, 2020.

Leonard Kaufman and Peter J Rousseeuw. Clustering by means of medoids. statistical data analysis based on the l1 norm. *Y. Dodge, Ed*, pages 405–416, 1987.

Hiroshi Kikusato, Kohei Mori, Shinya Yoshizawa, Yu Fujimoto, Hiroshi Asano, Yasuhiro Hayashi, Akihiko Kawashima, Shinkichi Inagaki, and Tatsuya Suzuki. Electric vehicle charge–discharge management for utilization of photovoltaic by coordination between home and grid energy management systems. *IEEE Transactions on Smart Grid*, 10(3): 3186–3197, 2018.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations (ICLR2015)*, 2015.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2014.

Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *28th Conference on Neural Information Processing Systems (NIPS)*, pages 3581–3589, 2014.

Jyrki Kivinen and Manfred K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, January 1997. ISSN 0890-5401. doi: 10.1006/inco.1996.2612.

Weicong Kong, Zhao Yang Dong, David J Hill, Fengji Luo, and Yan Xu. Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems*, 33(1):1087–1088, 2017.

Katarina Kostková, L' Omelina, P Kyčina, and Peter Jamrich. An introduction to load management. *Electric Power Systems Research*, 95:184–191, 2013.

Tze Leung Lai. Adaptive treatment allocation and the multi-armed bandit problem. *The Annals of Statistics*, pages 1091–1114, 1987.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Peter Lancaster and Kęstutis Šalkauskas. Curve and surface fitting: an introduction. *csfa*, 1986.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Guillaume Le Ray and Pierre Pinson. Online adaptive clustering algorithm for load profiling. *Sustainable Energy, Grids and Networks*, 17:100181, 2019.

Guillaume Le Ray, Emil Mahler Larsen, and Pierre Pinson. Evaluating price-based demand response in practice?with application to the ecogrid eu experiment. *IEEE Transactions on Smart Grid*, 9(3):2304–2313, 2016.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

B Lescoeur and JB Galland. Tariffs and load management: The french experience. *IEEE transactions on power systems*, 2(2):458–464, 1987.

Jinghua Li, Yongsheng Lei, Qian Huang, Zhijun Qin, and Bo Chen. Feature analysis of generalized load patterns considering active load response to real-time pricing. *IEEE Access*, 7:119443–119453, 2019.

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.

Yingying Li, Qinran Hu, and Na Li. A reliability-aware multi-armed bandit approach to learn and select users in demand response. *arXiv preprint arXiv:2003.09505*, 2020.

Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 World Wide Web Conference*, pages 689–698, 2018.

Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.

Juan Miguel Gonzalez López, Edris Pouresmaeil, Claudio A. Cañizares, Kankar Bhattacharya, Abolfazl Mosaddegh, and Bharatkumar V. Solanki. Smart residential load simulator for energy management in smart grids. *IEEE Transactions on Industrial Electronics*, 66(2):1443–1452, 2019.

James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

Pierre Mallet, Per-Olof Granstrom, Per Hallberg, Gunnar Lorenz, and Pavla Mandatova. Power to the people!: European perspectives on the future of electric distribution. *IEEE Power and Energy magazine*, 12(2):51–64, 2014.

Vivien Mallet, Gilles Stoltz, and Boris Mauricette. Ozone ensemble forecast with machine learning algorithms. *Journal of Geophysical Research: Atmospheres*, 114(D5), 2009.

Shie Mannor. Misspecified and complex bandits problems, 2018. Talk at "50$^{\text{èmes}}$ Journées de Statistique", EDF Lab Paris Saclay, May 31st, 2018.

Antoine Marot, Antoine Rosin, Laure Crochepierre, Benjamin Donnot, Pierre Pinson, and Lydia Boudjeloud-Assala. Interpreting atypical conditions in systems with deep conditional autoencoders: the case of electrical consumption. In *ECML PKDD 2019 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2019.

Jérémie Mary, Romaric Gaudel, and Philippe Preux. Bandits and recommender systems. In *International Workshop on Machine Learning, Optimization and Big Data*, pages 325–336. Springer, 2015.

H Brendan McMahan and Matthew J Streeter. Tighter bounds for multi-armed bandits with expert advice. In *Proceedings of the 22nd Conference on Learning Theory (COLT'09)*, 2009.

Jiali Mei, Yohann De Castro, Yannig Goude, and Georges Hébrail. Nonnegative matrix factorization for time series recovery from a few temporal aggregates. In *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*, pages 2382–2390, 2017.

Vladimiro Miranda, Joana da Hora Martins, and Vera Palma. Optimizing large scale problems with metaheuristics in a reduced space mapped by autoencoders–Application to the wind-hydro coordination. *IEEE Transactions on Power Systems*, 29(6):3078–3085, 2014.

Saeed Mohajeryami, Iman N. Moghaddam, Milad Doostan, Behdad Vatani, and Peter Schwarz. A novel economic model for price-based demand response. *Electric Power Systems Research*, 135:1–9, 2016.

Ahmadreza Moradipari, Cody Silva, and Mahnoosh Alizadeh. Learning to dynamically price electricity demand based on multi-armed bandits. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 917–921. IEEE, 2018.

Matteo Muratori, Matthew C. Roberts, Ramteen Sioshansi, Vincenzo Marano, and Giorgio Rizzoni. A highly resolved modeling technique to simulate residential power demand. *Applied Energy*, 107:465–473, 2013.

John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

Pentti Paatero and Unto Tapper. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2): 111–126, 1994.

P. Palensky and D. Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, 7(3): 381–388, 2011.

Anastasios Panagiotelis, George Athanasopoulos, Puwasala Gamakumara, and Rob J Hyndman. Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, 2020.

Seunghyun Park, Hanjoo Kim, Hichan Moon, Jun Heo, and Sungroh Yoon. Concurrent simulation platform for energy-aware smart metering systems. *IEEE Transactions on Consumer Electronics*, 56(3):1918–1926, 2010.

Vianney Perchet and Philippe Rigollet. The multi-armed bandit problem with covariates. *The Annals of Statistics*, pages 693–721, 2013.

Amandine Pierrot and Yannig Goude. Short-term electricity load forecasting with generalized additive models. *Proceedings of ISAP power*, 2011, 2011.

Pierre Pinson, Henrik Madsen, Henrik Aa Nielsen, George Papaefthymiou, and Bernd Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.

Ramu Ramanathan, Robert Engle, Clive WJ Granger, Farshid Vahid-Araghi, and Casey Brace. Short-run forecasts of electricity loads and peaks. *International journal of forecasting*, 13(2):161–174, 1997.

William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

Gary Raw and David Ross. Energy demand research project: Final analysis. 2011.

Ian Richardson, Murray Thomson, David Infield, and Conor Clifford. Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, 42(10):1878–1887, 2010.

John V Ringwood, D Bofelli, and Fiona T Murray. Forecasting electricity demand on short, medium and long time scales using neural networks. *Journal of Intelligent and Robotic Systems*, 31(1-3):129–147, 2001.

Pedro Pereira Rodrigues, João Gama, and Joao Pedroso. Hierarchical clustering of time-series data streams. *IEEE Transactions on Knowledge and Data Engineering*, 20(5): 615–627, 2008.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

Paat Rusmevichientong and John N Tsitsiklis. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

Javier Saez-Gallego and Juan M Morales. Short-term forecasting of price-responsive loads using inverse optimization. *IEEE Transactions on Smart Grid*, 9(5):4805–4814, 2017.

Wolfgang Schellong. Energy demand analysis and forecast. *Energy Management Systems*, pages 101–120, 2011.

Michael Scheuerer and Thomas M Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334, 2015.

James Schofield, Richard Carmichael, Simon Tindemans, Matt Woolf, Mark Bilton, and Goran Strbac. Residential consumer responsiveness to time-varying pricing, 2014. Technical report.

Raffi Sevlian and Ram Rajagopal. A scaling law for short term load forecasting on varying levels of aggregation. *International Journal of Electrical Power & Energy Systems*, 98: 350–361, June 2018.

Hussain Shareef, Maytham S Ahmed, Azah Mohamed, and Eslam Al Hassan. Review on home energy management system considering demand responses, smart technologies, and intelligent controllers. *IEEE Access*, 6:24498–24509, 2018.

Weiwei Shen, Jun Wang, Yu-Gang Jiang, and Hongyuan Zha. Portfolio choices with orthogonal bandit learning. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

Eli Shlifer and Ronald W. Wolff. Aggregation and proration in forecasting. *Management Science*, 25(6):594–603, June 1979. ISSN 0025-1909. doi: 10.1287/mnsc.25.6.594.

Pierluigi Siano. Demand response and smart grids? A survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

Ingo Steinwart, Andreas Christmann, et al. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17(1):211–225, 2011.

Mingyang Sun, Ioannis Konstantelos, and Goran Strbac. C-vine copula mixture model for clustering of residential electrical load pattern data. *IEEE Transactions on Power Systems*, 32(3):2382–2393, 2017.

James W Taylor. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805, 2003.

James W Taylor and Roberto Buizza. Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power systems*, 17(3):626–632, 2002.

James W Taylor and Roberto Buizza. Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19(1):57–70, 2003.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits, 2013. arXiv preprint arXiv:1309.6869.

Tim Van Erven and Jairo Cugliari. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, pages 297–317. Springer, 2015.

Claire Vernade, Olivier Cappé, and Vianney Perchet. Stochastic bandit models for delayed conversions. *arXiv preprint arXiv:1706.09186*, 2017.

Vladimir Vovk. Competitive on-line statistics. *International Statistical Review*, 69(2): 213–248, 2001.

Volodimir G. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, COLT '90, pages 371–386, San Francisco, CA, USA, 1990. Morgan Kaufmann Publishers Inc. ISBN 1-55860-146-5.

Grace Wahba. *Spline models for observational data*, volume 59. Siam, 1990.

Chih-Chun Wang, Sanjeev R Kulkarni, and H Vincent Poor. Arbitrary side observations in bandit problems. *Advances in Applied Mathematics*, 34(4):903–938, 2005a.

Chih-Chun Wang, Sanjeev R Kulkarni, and H Vincent Poor. Bandit problems with side observations. *IEEE Transactions on Automatic Control*, 50(3):338–355, 2005b.

Qingsi Wang, Mingyan Liu, and Johanna L Mathieu. Adaptive demand response: Online learning of restless and controlled bandits. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 752–757. IEEE, 2014.

Yi Wang, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, and Yun Zhao. Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, 20(2):117–129, 2015.

Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526):804–819, 2019.

Olivier Wintenberger. Optimal learning with Bernstein online aggregation. *Machine Learning*, 106(1):119–141, 2017.

Simon Wood. *Generalized Additive Models: An Introduction with R*. CRC Press, 2006.

Simon Wood. Mgcv: a r package for mixed gam computation vehicle with automatic smoothness estimation. 08 2020.

Simon N Wood, Yannig Goude, and Simon Shaw. Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1): 139–155, 2015.

Ye Yan, Yi Qian, Hamid Sharif, and David Tipper. A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE communications surveys & tutorials*, 15(1):5–20, 2012.

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre : Algorithmes de bandits stochastiques pour la gestion de la demande électrique

Mots clés : bandits stochastiques, apprentissage séquentiel, gestion de flexibilités électriques, prévision de la demande électrique, auto-encodeur variationnel conditionnel, segmentation

Résumé : L'électricité se stockant difficilement à grande échelle, l'équilibre entre la production et la consommation doit être rigoureusement maintenu. Une gestion par anticipation de la demande se complexifie avec l'intégration au mix de production des énergies renouvelables intermittentes. Parallèlement, le déploiement des compteurs communicants permet d'envisager un pilotage dynamique de la consommation électrique. Plus concrètement, l'envoi de signaux - tels que des changements du prix de l'électricité - permettrait d'inciter les usagers à moduler leur consommation afin qu'elle s'ajuste au mieux à la production d'électricité. Les algorithmes choisissant ces signaux devront apprendre la réaction des consommateurs face aux envois tout en les optimisant (compromis exploration-exploitation). Notre approche, fondée sur la théorie des bandits, a permis de formaliser ce problème d'apprentissage séquentiel et de proposer un premier algorithme pour piloter la demande électrique d'une population homogène de consommateurs. Une borne supérieure d'ordre $T^{2/3}$ a été obtenue sur le regret de cet algorithme. Des expériences réalisées sur des données de consommation de foyers soumis à des changements dynamiques du prix de l'électricité illustrent ce résultat théorique. Un jeu de données en « information complète » étant nécessaire pour tester un algorithme de bandits, un simulateur de données de consommation fondé sur les auto-encodeurs variationnels a ensuite été construit. Afin de s'affranchir de l'hypothèse d'homogénéité de la population, une approche pour segmenter les foyers en fonction de leurs habitudes de consommation est aussi proposée. Ces différents travaux sont finalement combinés pour proposer et tester des algorithmes de bandits pour un pilotage personnalisé de la consommation électrique.

Title: Stochastic Bandit Algorithms for Demand Side Management

Keywords: stochastic bandit, online learning, demand response, load forecasting, conditional variational auto-encoder, clustering

Abstract: As electricity is hard to store, the balance between production and consumption must be strictly maintained. With the integration of intermittent renewable energies into the production mix, the management of the balance becomes complex. At the same time, the deployment of smart meters suggests demand response. More precisely, sending signals - such as changes in the price of electricity - would encourage users to modulate their consumption according to the production of electricity. The algorithms used to choose these signals have to learn consumer reactions and, in the same time, to optimize them (exploration-exploration trade-off). Our approach is based on bandit theory and formalizes this sequential learning problem. We propose a first algorithm to control the electrical demand of a homogeneous population of consumers and offer $T^{2/3}$ upper bound on its regret. Experiments on a real data set in which price incentives were offered illustrate these theoretical results. As a "full information" dataset is required to test bandit algorithms, a consumption data generator based on variational autoencoders is built. In order to drop the assumption of the population homogeneity, we propose an approach to cluster households according to their consumption profile. These different works are finally combined to propose and test a bandit algorithm for personalized demand side management.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France