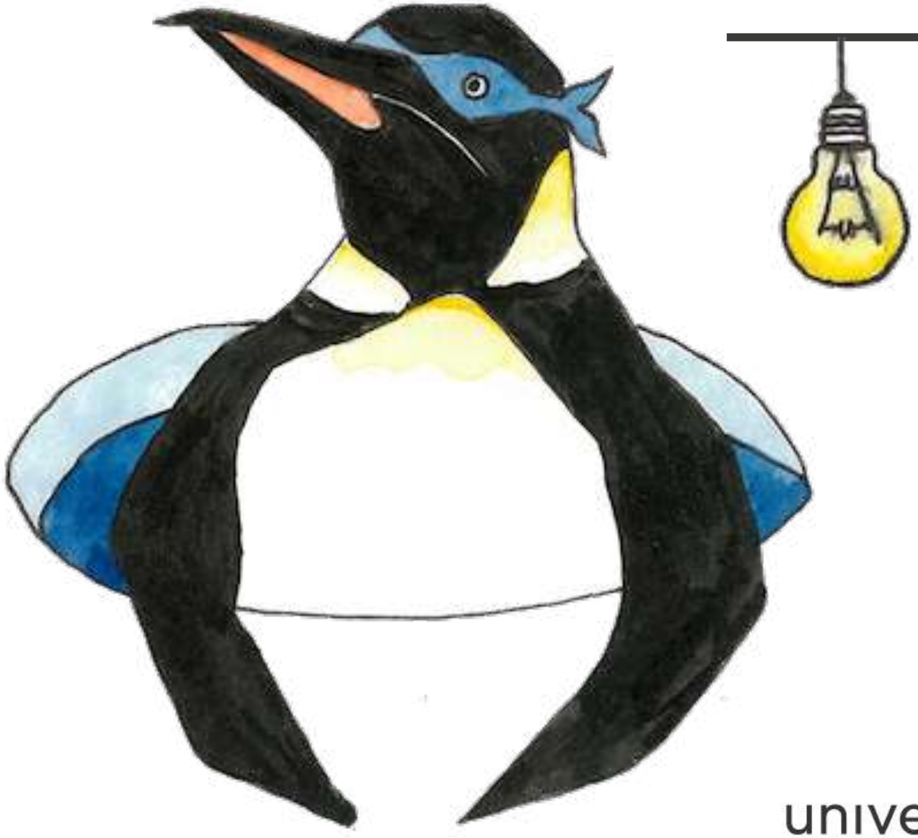
Stochastic Bandit Algorithms for Demand Side Management



Margaux Brégère – Dec. 10, 20

Under the supervision of Gilles Stoltz,
Yannig Goude and Pierre Gaillard

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

université PARIS-SACLAY
ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)
Inria INVENTEURS DU MONDE NUMÉRIQUE
eDF

Introduction - Motivation

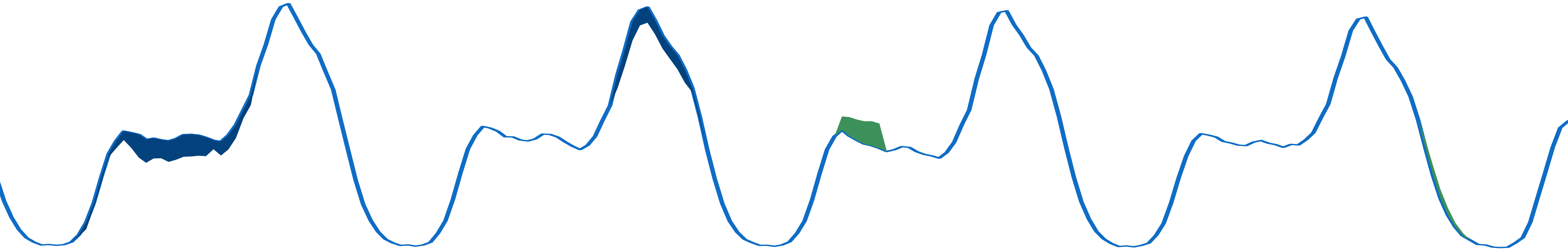
As electricity is hard to store, balance between production and demand must be strictly maintained

Current solution: forecast demand and adapt production accordingly

- With the development of renewable energies, production becomes harder to adjust
- New (smart) meters provide access to data and instantaneous communication

Prospective solution: send incentive signals (electricity tariff variations) to manage demand response

Introduction - Motivation



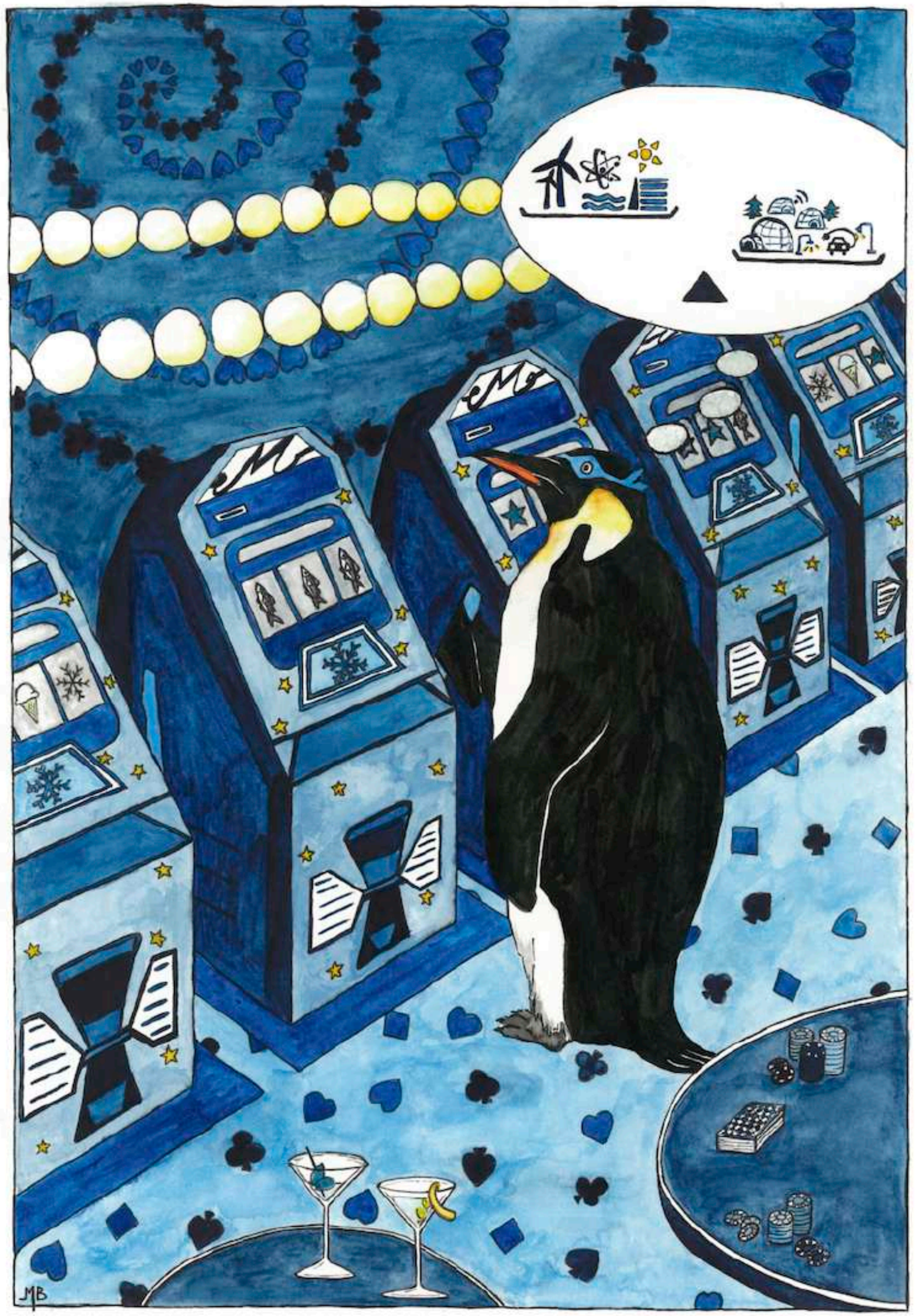
How to develop automatic solutions
to chose incentive signals dynamically?

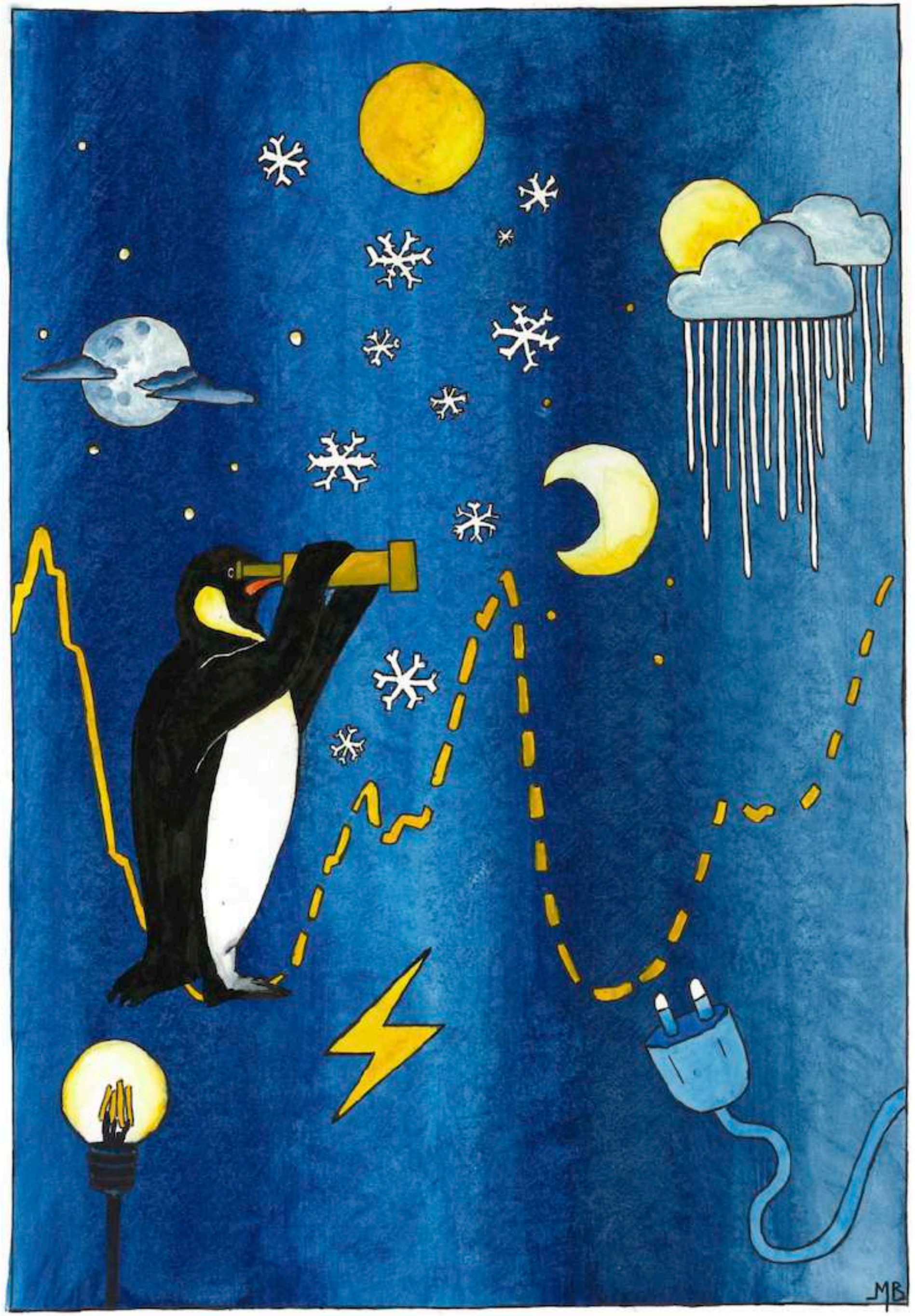
Exploration: learn consumer behavior

Exploitation: optimize signal sending

PhD topic

Apply mathematical bandit theory to the sequential learning problem of demand side management





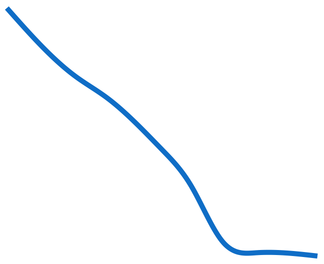
First of all: modeling

How to model electricity demand?
- Using classical (for EDF) power consumption forecasting methods

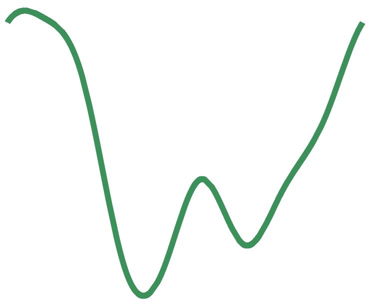
How to formalize the sequential learning?
- Defining a protocol (under some assumptions)

Generalized additive models for electricity demand

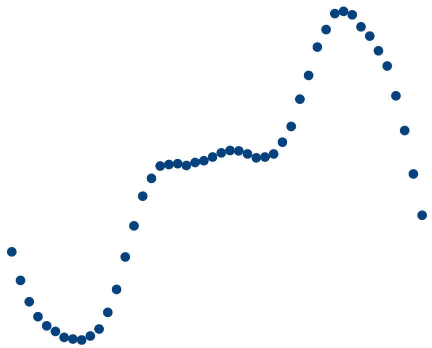
$$Y_t = f_1(\text{temperature}) + f_2(\text{position in the year}) + f_3(\text{hour}) + f_4(\text{tariff}) + \cdots + \text{noise}$$



+



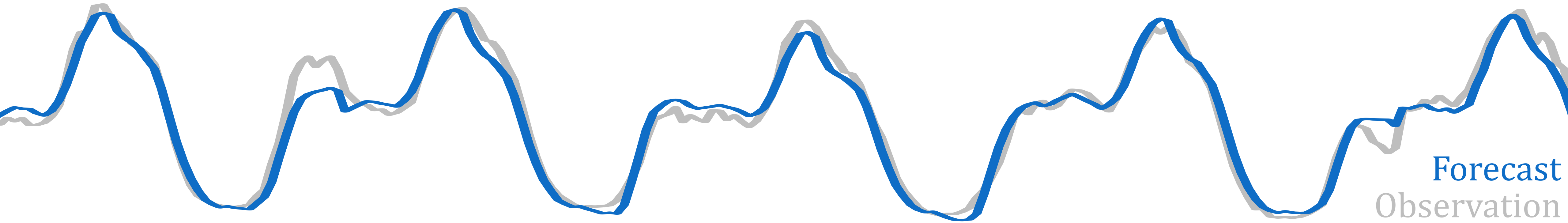
+



+ ...

- There is a known transfer function ϕ and an unknown parameter θ such that

$$Y_t = \phi(\text{temperature}, \text{position in the year}, \text{hour}, \text{tariff} \ldots)^T \theta + \text{noise}$$



Electricity demand modeling

Assumption:

- K tariffs
- Homogenous population

At each round $t = 1, \ldots$

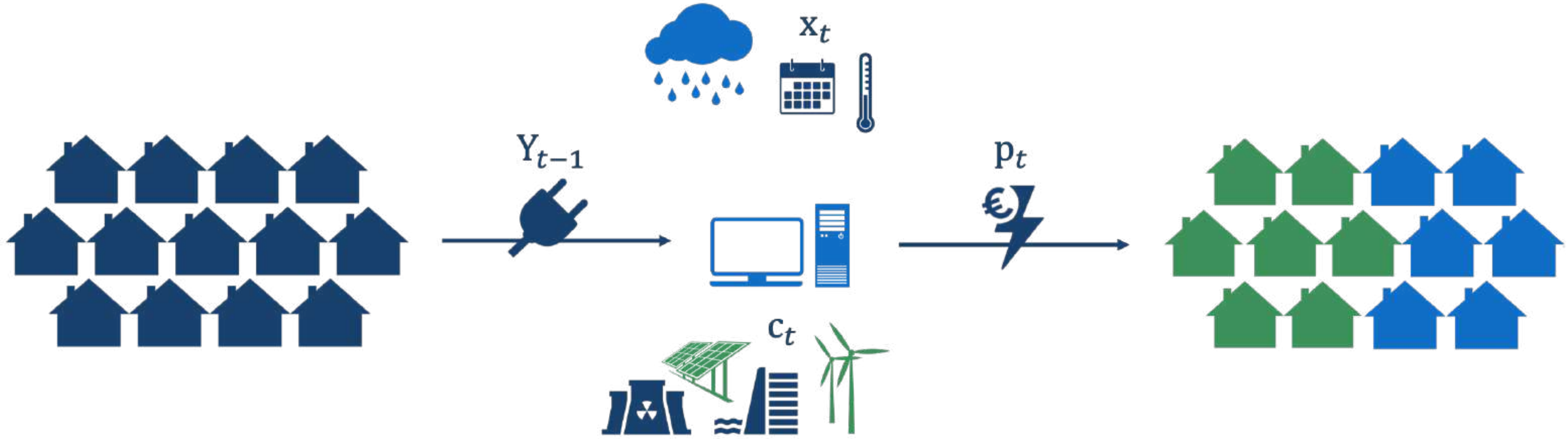
- Observe a context x_t
- Choose price levels p_t
- Observe the electricity demand $Y_t = \phi(x_t, p_t)^T\theta + p_t^T \varepsilon_t$

with $\mathbb{E}[\varepsilon_t] = (0, \ldots 0)^T$ and $\mathbb{V}[\varepsilon_t] = \Sigma \in \mathcal{M}_K(\mathbb{R})$

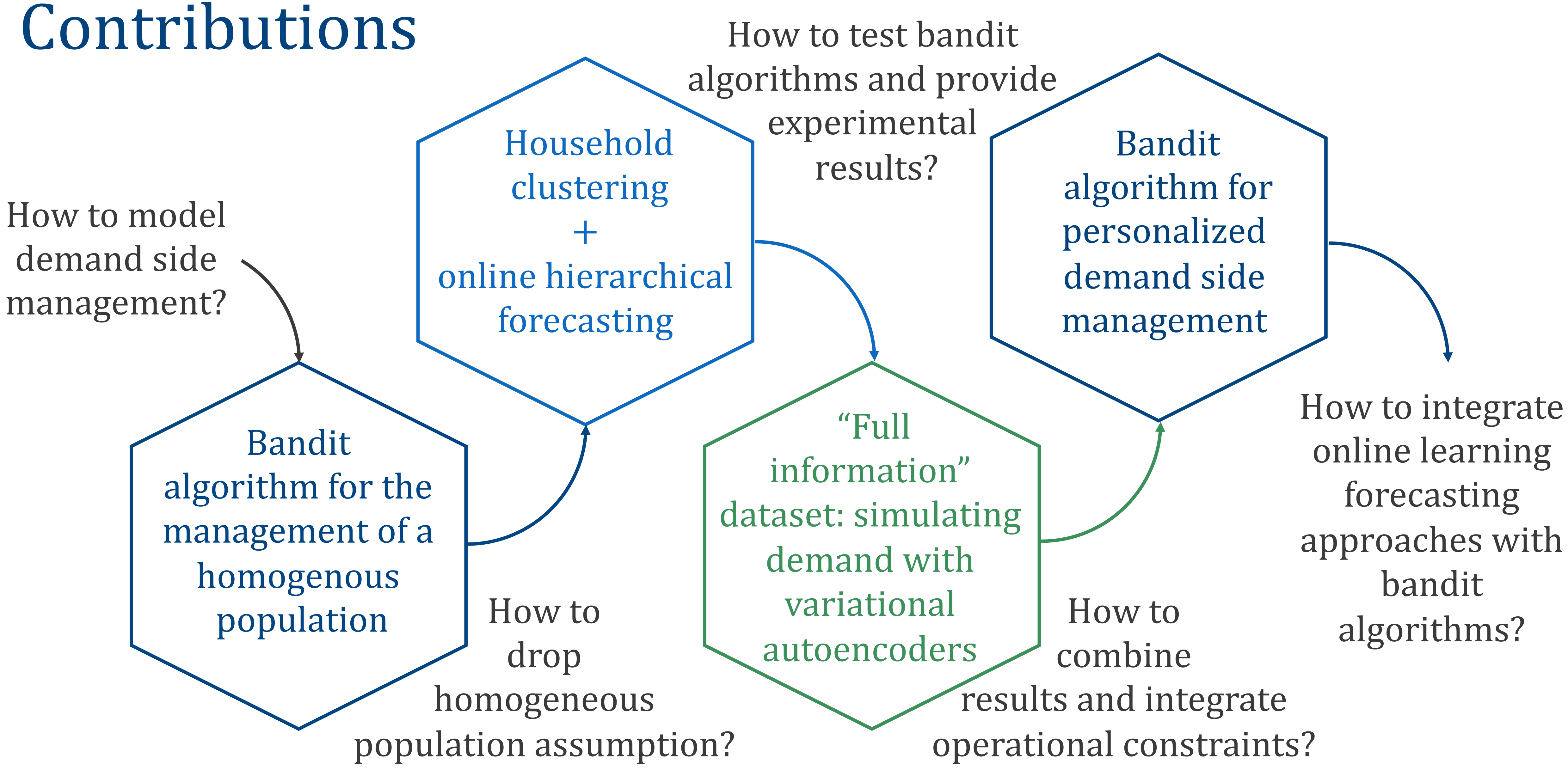
Protocol for target tracking

At each round $t = 1, \dots, T$

- Observe a context x_t and a target c_t
- Choose price levels p_t
- Observe the resulting demand Y_t and suffer a loss $(Y_t - c_t)^2$



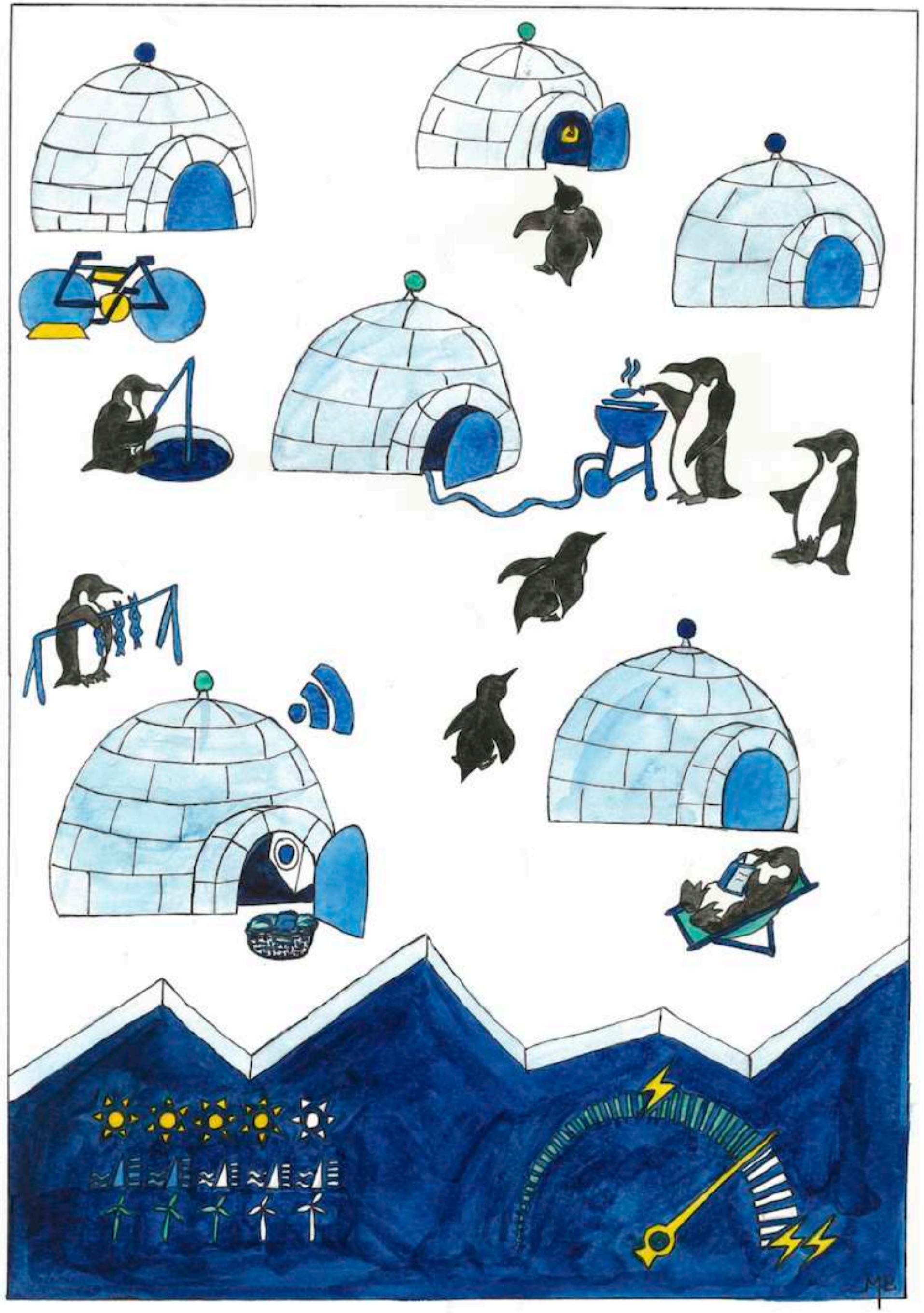
Contributions



Bandit algorithm for the management of a homogenous population

How to evaluate a target tracking algorithm?
▸ Defining a regret criterion

How to adapt existing bandit theory?
▸ Adapting LinUCB algorithm (Li et al. 2010)



Joint work with Pierre Gaillard, Yannig Goude and Gilles Stoltz, International Conference on Machine Learning, 2019

Protocol: target tracking for contextual bandits

At each round $t = 1, \ldots, T$

- Observe a context x_t and a target c_t
- Choose price levels p_t
- Observe a resulting demand $Y_t = \phi(x_t, p_t)^T \theta + p_t^T \varepsilon_t$ with $\mathbb{V}[\varepsilon_t] = \Sigma$
- Suffer a loss $(Y_t - c_t)^2$ such that

$$\mathbb{E}[\,(Y_t - c_t)^2 \mid \text{past}, x_t, p_t] = \left(\phi(x_t, p_t)^T \theta - c_t\right)^2 + p_t^T \Sigma p_t$$

Aim: minimize the pseudo-regret

$$R_T = \sum_{t=1}^{T} \left(\phi(x_t, p_t)^T \theta - c_t\right)^2 + p_t^T \Sigma p_t - \sum_{t=1}^{T} \min_p \left(\phi(x_t, p)^T \theta - c_t\right)^2 + p^T \Sigma p$$

- Estimate parameters θ and Σ to estimate losses to reach a bias-variance trade-off

Optimistic algorithm

Inspired from Lin-UCB (Li et al. 2010)

For $t = 1,2, \dots, \tau$

- Select price levels deterministically to estimate Σ offline with $\hat{\Sigma}_\tau$

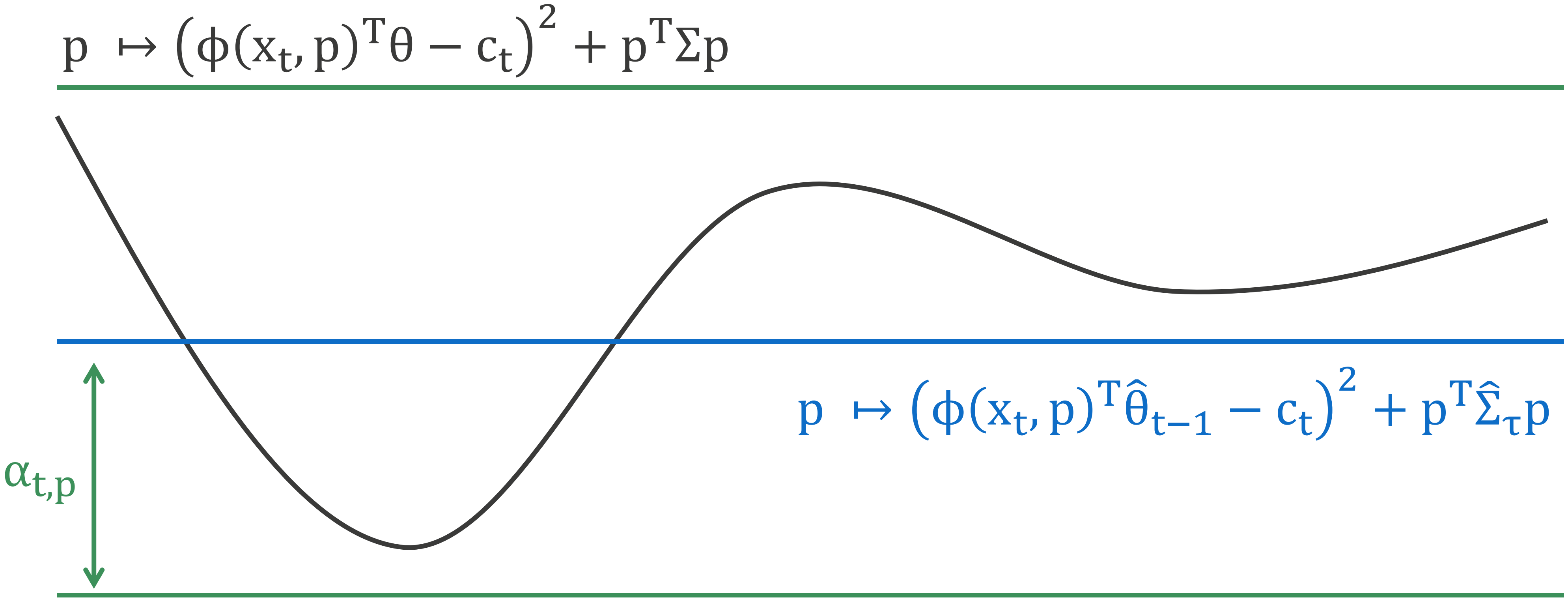
For $t = \tau, \dots, T$

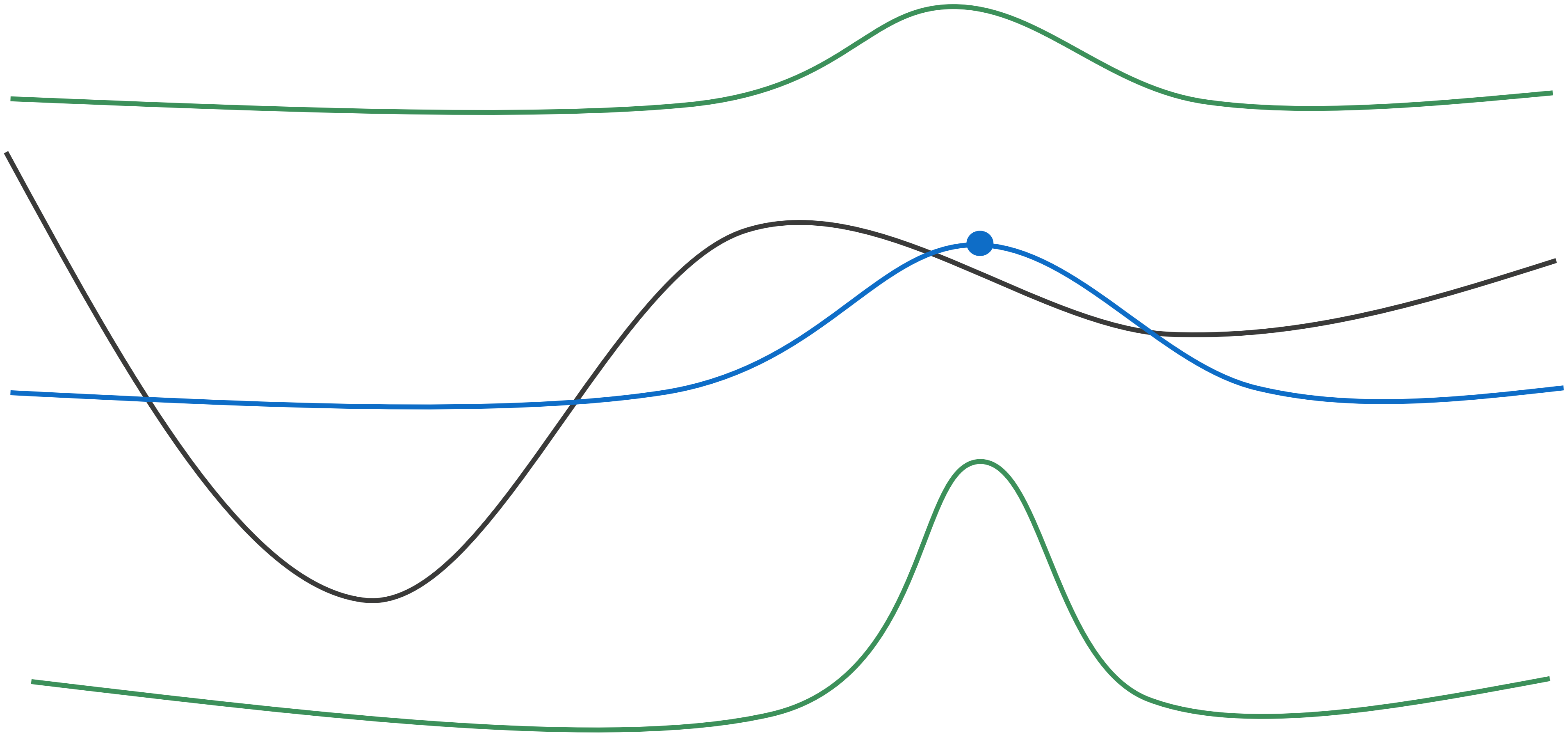
- Estimate θ based on past observations with $\hat{\theta}_{t-1}$ (Ridge regression)
- Estimate the future expected loss for each p: $\left(\phi(x_t, p)^T \hat{\theta}_{t-1} - c_t\right)^2 + p^T \hat{\Sigma}_\tau p$
- Get a confidence bound for each p

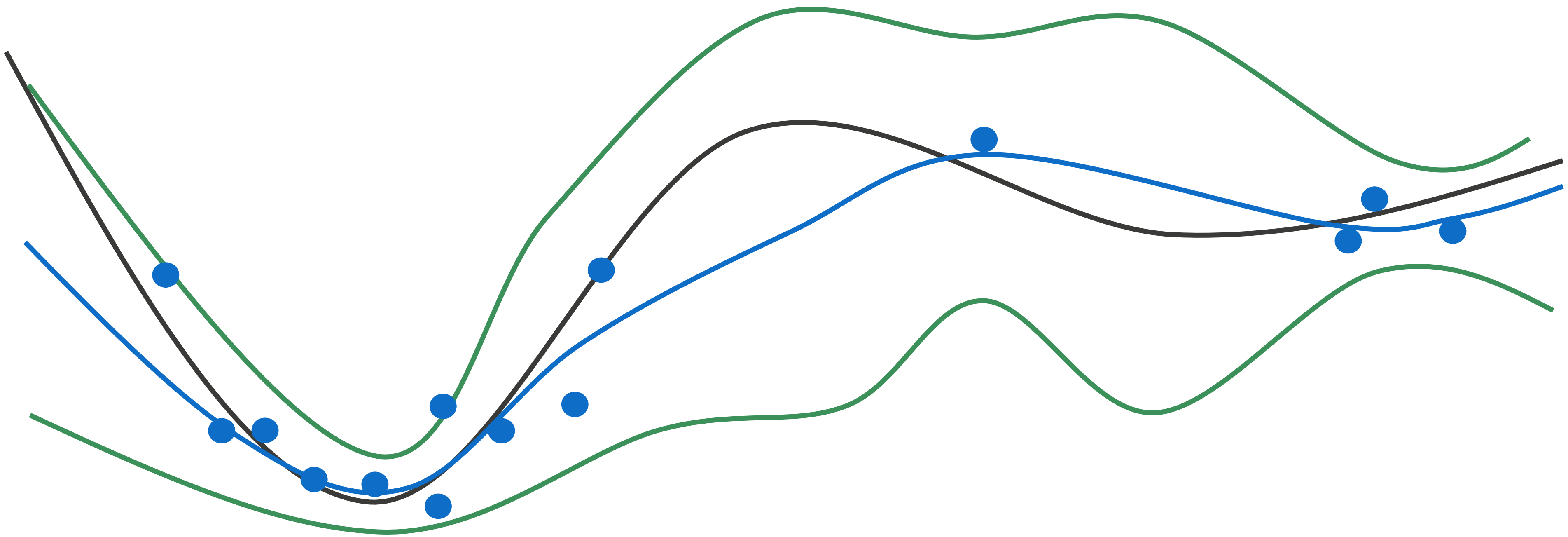
$$\left\| \left(\phi(x_t, p)^T \hat{\theta}_{t-1} - c_t\right)^2 + p^T \hat{\Sigma}_\tau p - \left(\phi(x_t, p)^T \theta - c_t\right)^2 + p^T \Sigma p \right\| \leq \alpha_{t,p}$$

- Select price levels optimistically

$$p_t \in \arg\min_p \left\{ \left(\phi(x_t, p)^T \hat{\theta}_{t-1} - c_t\right)^2 + p^T \hat{\Sigma}_\tau p - \alpha_{t,p} \right\}$$


$p \mapsto \left(\phi(x_t, p)^T\theta - c_t\right)^2 + p^T\Sigma p$
$p \mapsto \left(\phi(x_t, p)^T\hat{\theta}_{t-1} - c_t\right)^2 + p^T\hat{\Sigma}_\tau p$
$\alpha_{t,p}$





The problem is a bit more complex: curves vary with time t

Regret bound

Theorem

For proper choices of confidence levels $\alpha_{t,p}$ and number of exploration rounds τ, with high probability

$$R_T = \sum_{t=1}^{T} \left(\phi(x_t, p_t)^T \theta - c_t\right)^2 + p_t^T \Sigma p_t - \sum_{t=1}^{T} \min_{p \in \mathcal{P}} \left\{\left(\phi(x_t, p)^T \theta - c_t\right)^2 + p^T \Sigma p\right\} \leq \mathcal{O}\left(T^{2/3}\right)$$

Remark $R_T \leq \mathcal{O}\left(\sqrt{T} \ln T\right)$ if Σ is known

Elements of proof

- Deviation inequalities on $\hat{\theta}_t$ [1] and on $\hat{\Sigma}_\tau$
- Inspired from LinUCB regret bound analysis [2]

[1] Laplace's method on supermartingales: Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits, 2011

[2] Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions, 2011

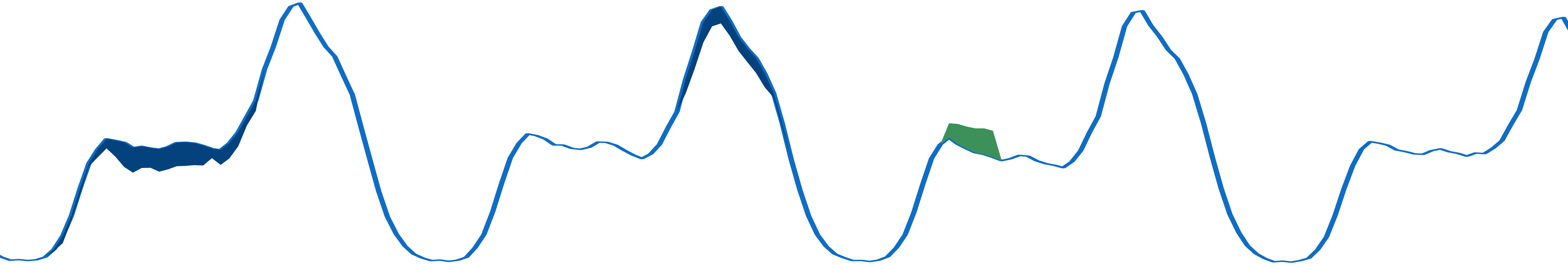
Smart Meter Energy Consumption Data

"Smart Meter Energy Consumption Data in London Households"
Public dataset - UK Power Networks

Individual electricity demand at half-an-hour intervals throughout 2013 of

~1 000 clients subjected to Dynamic Time of Use energy prices

Three tariffs: Low, Normal, High



Design of the experiment

- Alternative policies cannot be tested on historical data… How to test bandit algorithms?
  - Simulating data with $Y_t = f(x_t) + p_t^T \begin{bmatrix} \xi_{Low} \\ \xi_{Normal} \\ \xi_{High} \end{bmatrix} + p_t^T \varepsilon_t$ and $\mathbb{V}[\varepsilon_t] = \Sigma$

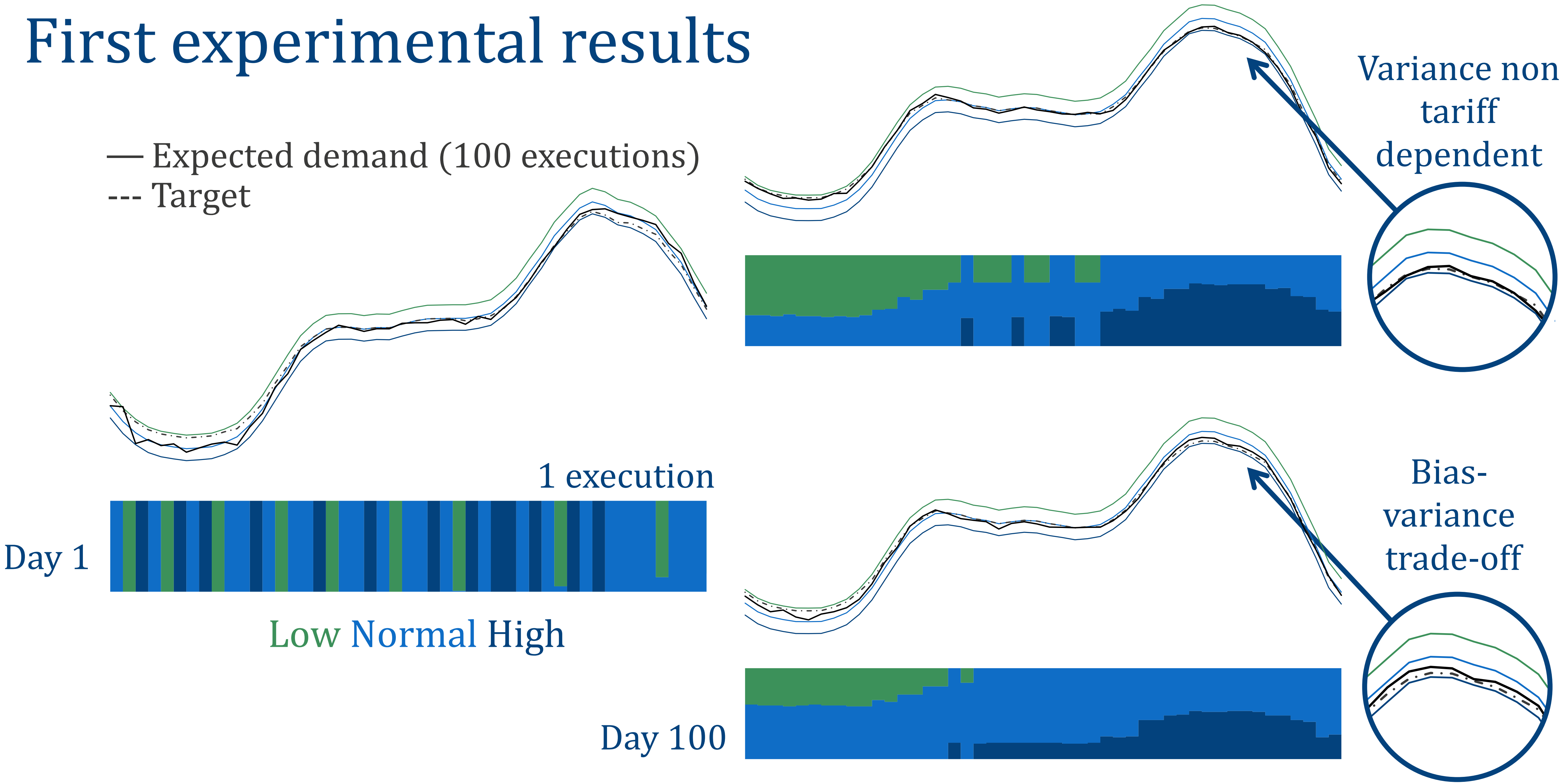
$$\text{where } \Sigma = \begin{pmatrix} \sigma_{Low} & 0 & 0 \\ 0 & \sigma_{Normal} & 0 \\ 0 & 0 & \sigma_{High} \end{pmatrix}$$

Experiment 1: $\sigma_{Low} = \sigma_{Normal} = \sigma_{High}$

Experiment 2: $\sigma_{Low} > \sigma_{High} > \sigma_{Normal}$

- Which target to choose?
  - Close to average High demand during the evening
  - Close to average Low demand during the night
- Which context to choose?
  - Algorithm executed on historical context
- Operational constraints on legible allocations of price levels:
  - Impossible to send Low and High tariffs at the same time
  - Population split in 100 equal subsets

First experimental results



First extensions and prospects

Extensions

- Generalization to general loss functions (polynomial approximation)
- Considering "rebound" effect: Y_t also depends on $\dots, p_{t-2}, p_{t-1}$ and $p_{t+1}, p_{t+2} \dots$

→ a daily profile management, for each day d:

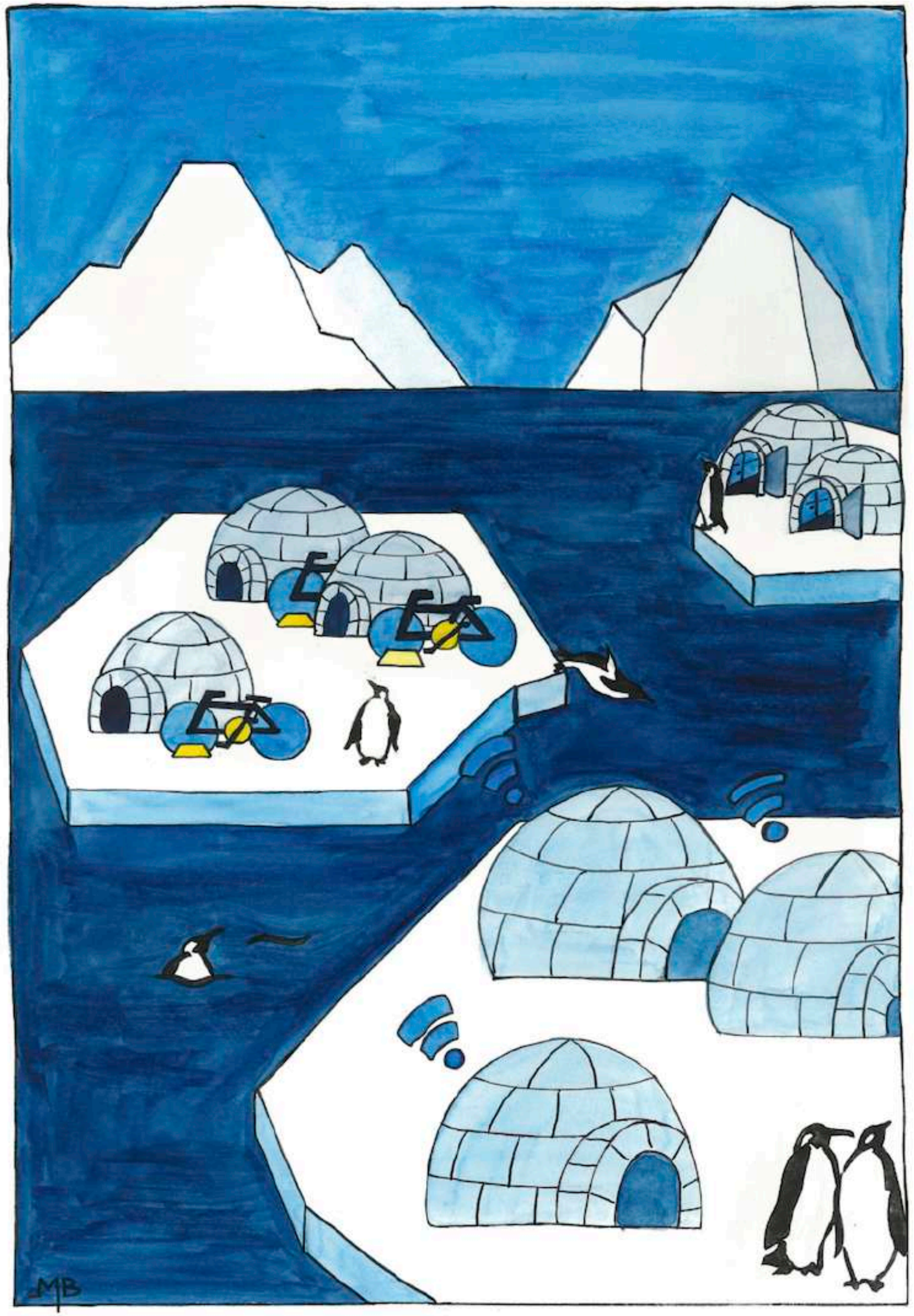
$$\begin{bmatrix} Y_d^1 \\ \vdots \\ Y_d^H \end{bmatrix} = \begin{bmatrix} \phi(x_d, p_d^1)^T \theta^1 + p_d^{1^T} \varepsilon_d^1 \\ \vdots \\ \phi(x_d, p_d^H)^T \theta^H + p_d^{H^T} \varepsilon_d^H \end{bmatrix}$$

Adaptation of the optimistic algorithm:

$$(p_d^1, \dots, p_d^H) \in \underset{(p^1,\dots,p^H)}{\arg\min} \left\{ \sum_{h=1}^{H} \left(\left(\phi(x_d, p^h)^T \hat{\theta}_{d-1}^h - c_d^h \right)^2 + p^{h^T} \hat{\Sigma}_\tau^h p^h \right) - \alpha_{d,p^1,\dots,p^H} \right\}$$

Prospects

- Algorithm optimality study (Lower bounds)
- Nonconvex optimization problem resolution



Towards the application of theoretical results

How to drop
the homogenous population assumption?

- Clustering households (or igloos)

How to test bandit algorithms?

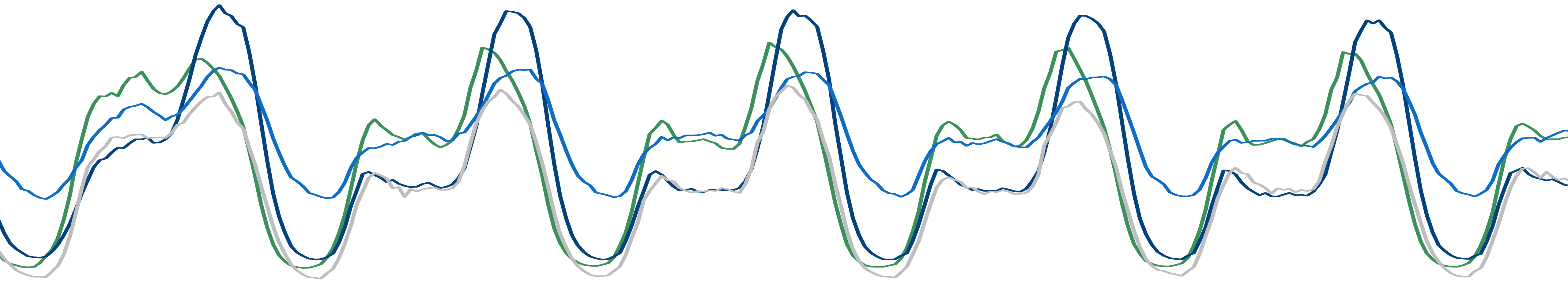
- Using a data simulator

Dropping the homogeneous population assumption

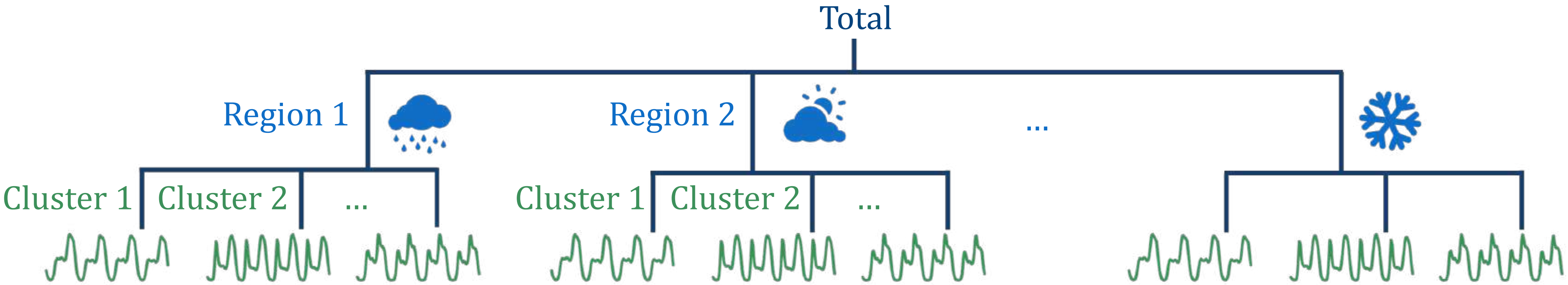
Double segmentation:

- geographical, based on region information
- behavioral:

Historical individual time series → Feature creation using low rank approximation → Clustering algorithm (k-medoid) → k clusters



Online hierarchical forecasting



Historical data and contextual variables → **Generate** base forecasts → Linearly combine the forecasts with an **aggregation** algorithm → **Project** the forecasts onto a coherent subspace → Forecasts fulfilling the hierarchical constraints

Joint work with Malo Huard, paper submitted

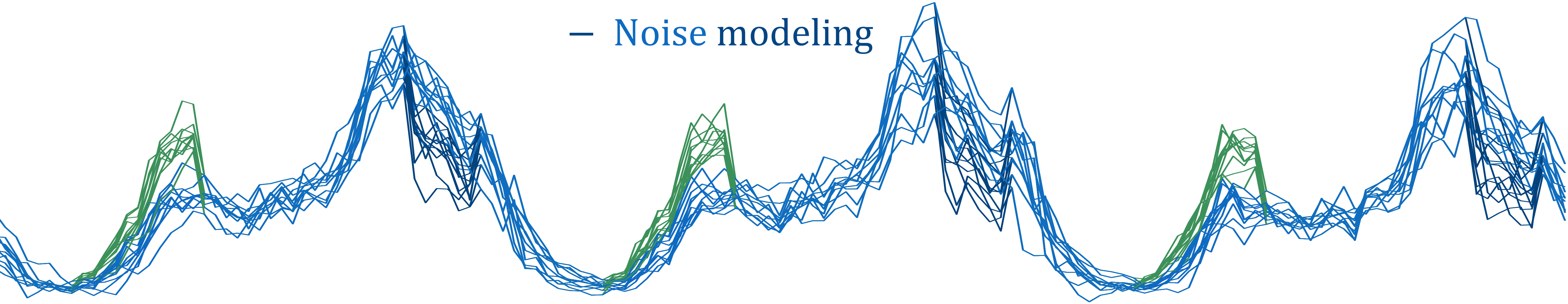
Simulating electricity demand

- A semi-parametric approach with "generalized additive models + noise"
  - Illustrate the theory
- A black-box approach with conditional variational auto-encoders
  - Test the algorithm robustness

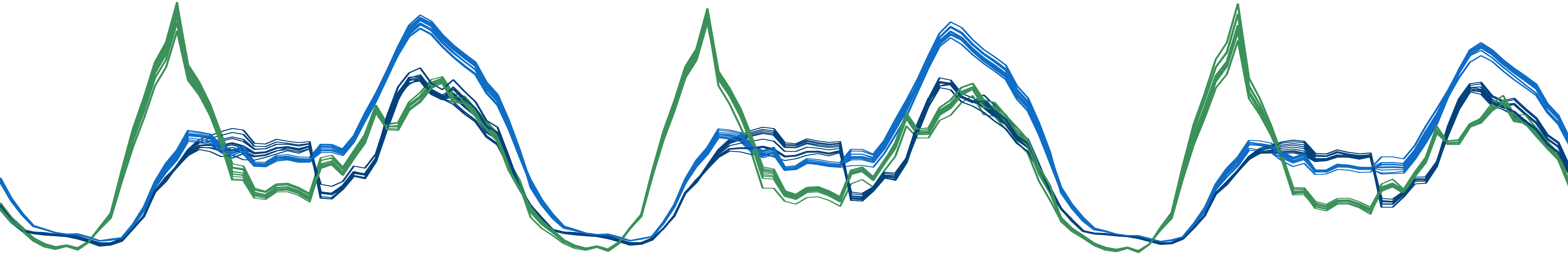
Joint work with Ricardo Jorge Bessa, IEEE access, 2020

Demand generated for different tariff signals

- Semi-parametric generator: + Interpretable
  – Noise modeling



- Black-box generator: + Rebound effect
  – Limited generalization capacity → transfer learning

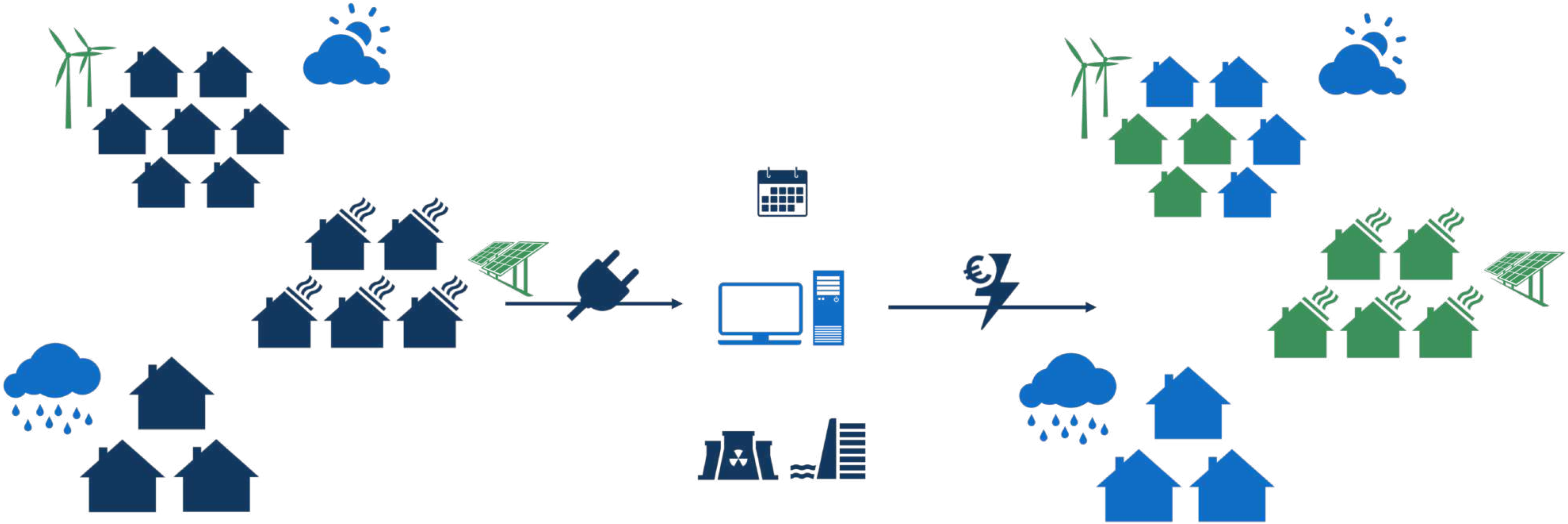


Synthesis - Operational demand side management

- Personalizing incentive signals according to
  - Local meteorological condition
  - Consumption behavior

- Taking into account operational
  - Network constraints (renewable energies integration)
  - Commercial constraints (electricity supply contract)

Personalized demand side management



Protocol

At each round $t = 1, \dots, T$

- Observe G contexts $(x_t^i)_{i=1,\dots,G}$
- Observe some sub-targets, which may correspond to renewable energy production, c_t^g, with $g \in \mathcal{P}(1, \dots, G)$ and some weights κ_t^g
- For $i = 1, \dots, G$
  - Choose price levels $p_t^i \in$ {prices allowed by the electricity contract at t}
  - Observe the resulting demand $Y_t^i = \phi^i(x_t^i, p_t^i)^T \theta^i + p_t^{iT} \varepsilon_t^i$, with $\mathbb{V}[\varepsilon_t^i] = \Sigma^i$
- Suffer a loss

$$\sum_g \kappa_t^g \left(\sum_{i \in g} Y_t^i - c_t^g \right)^2$$

Optimistic algorithm with a $T^{2/3}$ upper regret bound

Initialization: for $t = 1,2,\dots,\tau$, estimate $\Sigma^1,\dots,\Sigma^G$ offline and build confidence sets

For $t = \tau, \dots, T$

- For $i = 1, \dots, G$
  - Estimate parameters θ^i

- Estimate the future expected loss for each price level combination:

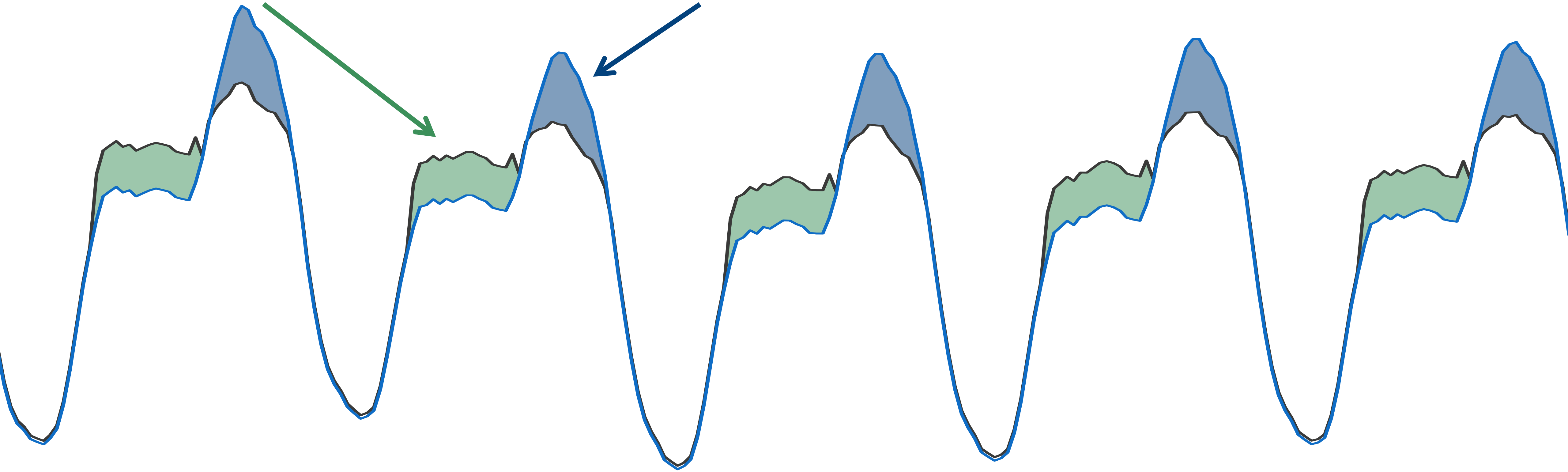
$$\sum_g \kappa_t^g \left(\left(\sum_{i \in g} \phi^i(x_t^i, p_t^i)^T \hat{\theta}_{t-1}^i - c_t^g \right)^2 + \sum_{i \in g} p^{i^T} \hat{\Sigma}_\tau^i p^i \right)$$

- Get a confidence bound for each $(p^1, \dots, p^G)$
- Select price levels optimistically

Experiments – Target creation

Renewable energies integration

Peak demand reduction (10%)



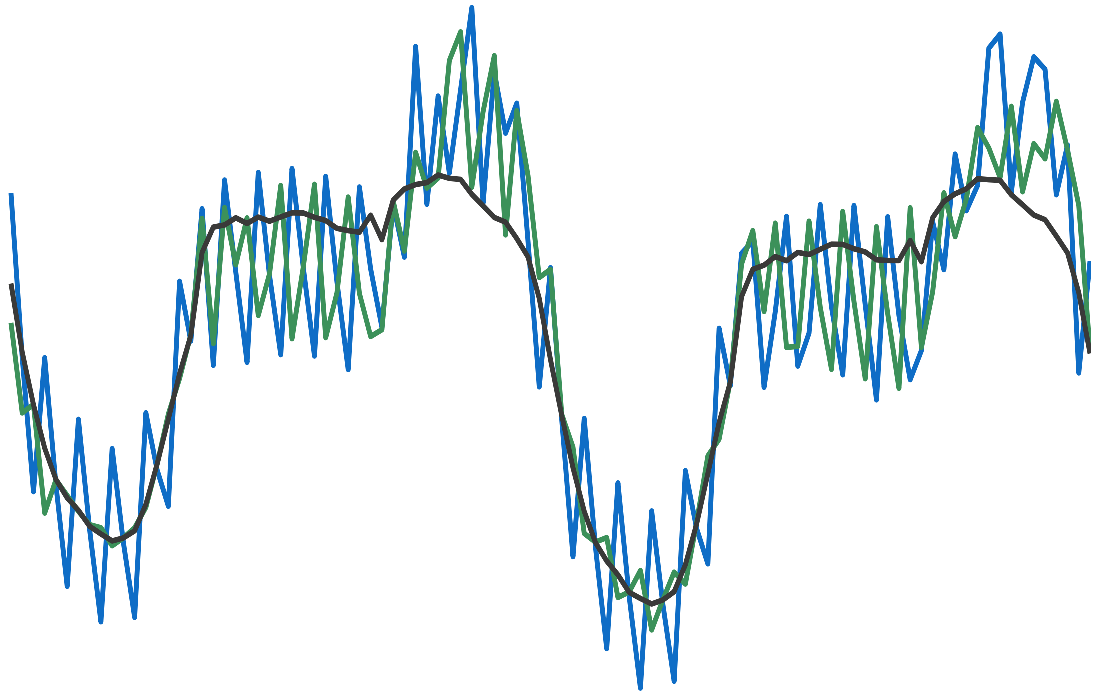
Both curves have the same integral (daily demand remains constant)

Experiments – Global vs personalized management

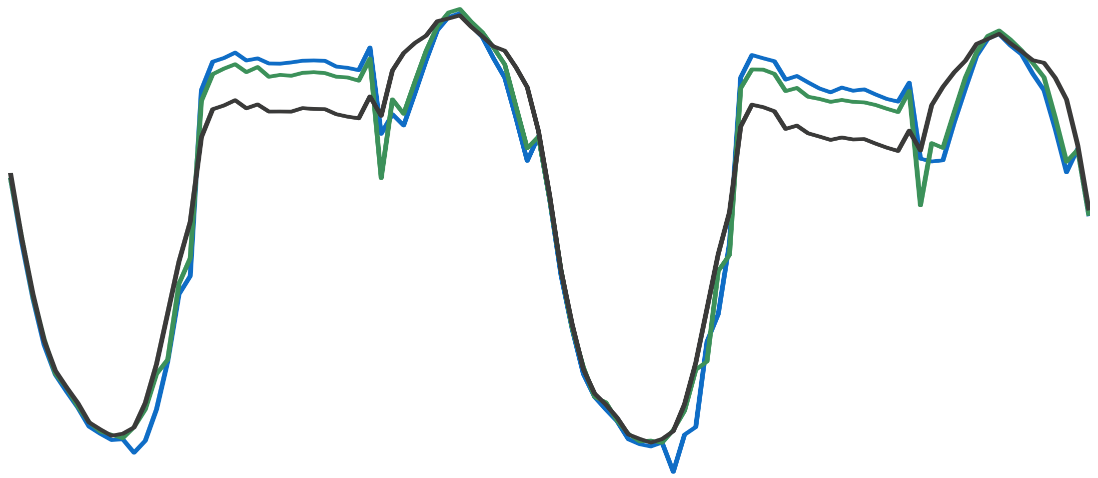
A single tariff per cluster

2 clusters: $Y_t^i = f^i(x_t) + \xi^i_{tariff_t} + \varepsilon^i_{tariff_t,t}$, $i = 1,2$

Decrease in loss of 30%!



Days 1 and 2



Days 99 and 100

Target

Expected total demand (100 executions) with a global and a personalized management

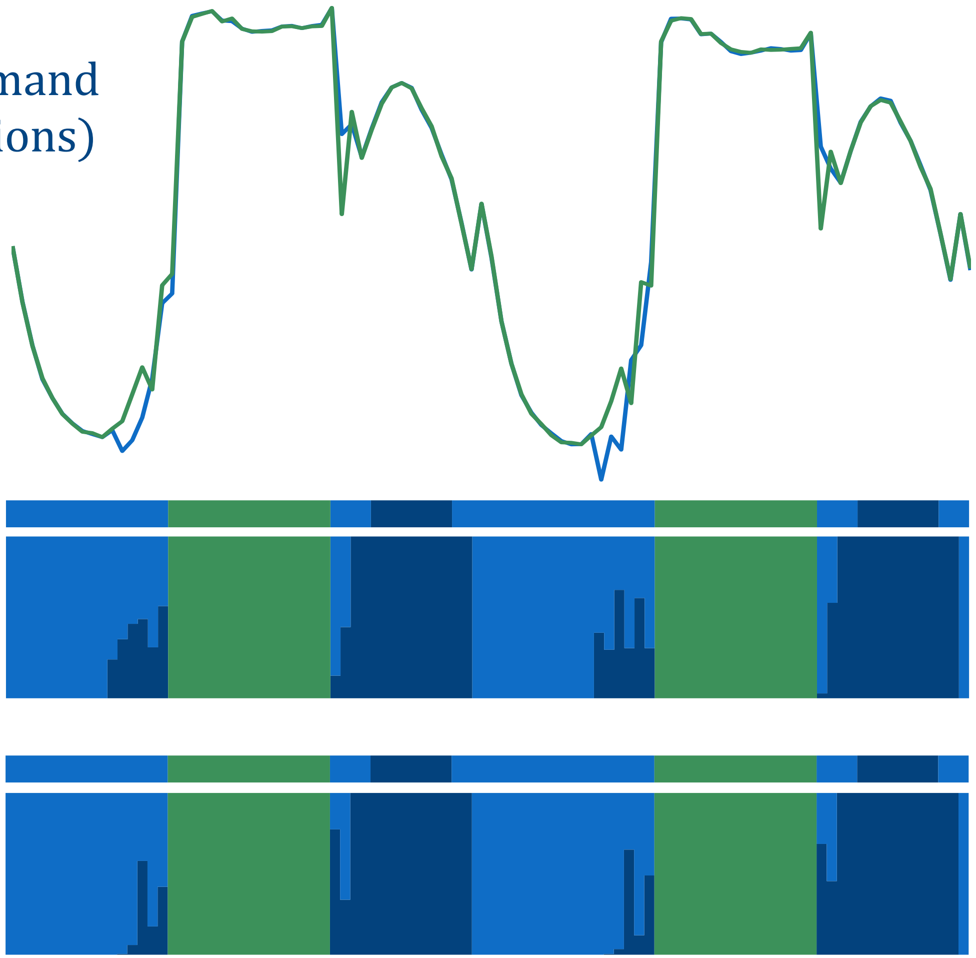
Experiments – Global vs personalized management

Cluster 1

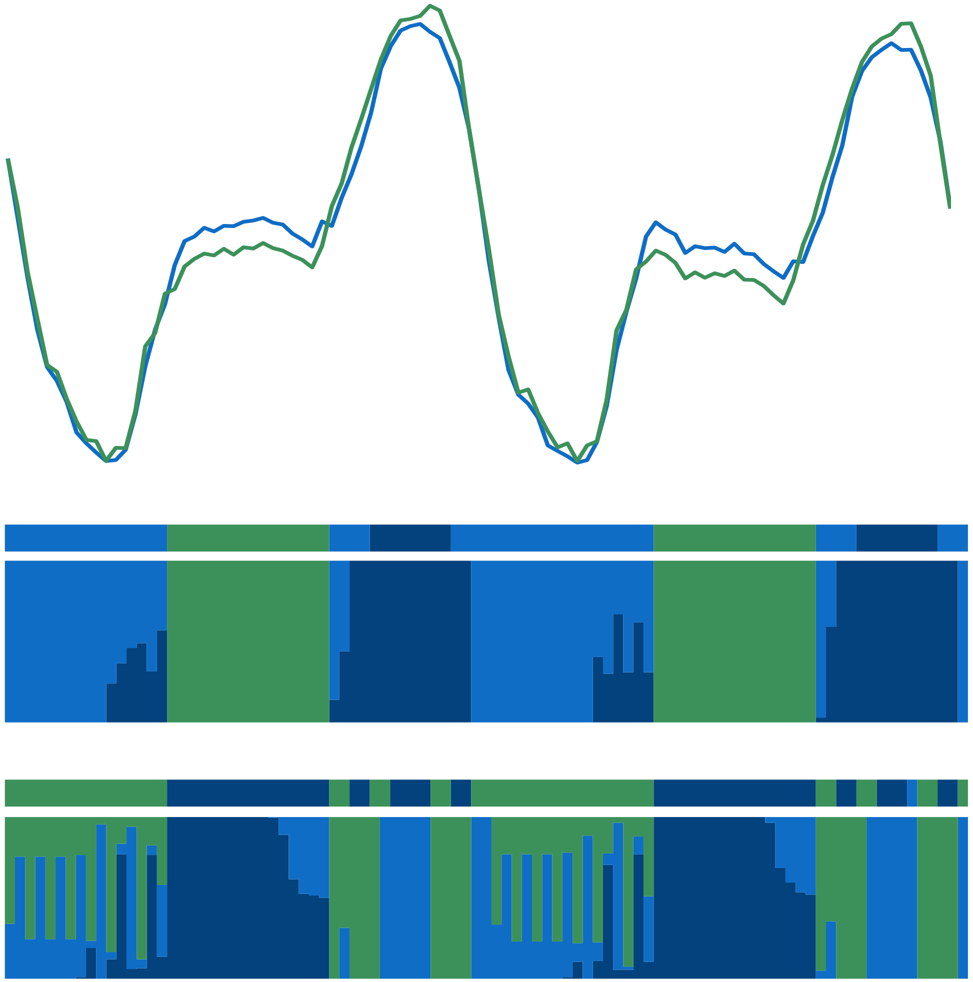
Expected demand
(100 executions)

Optimal tariff

Tariffs
sent on 100
executions



Cluster 2



Global
management

Personalized
management

Conclusions and prospects

Contributions

- A bandit algorithm for personalized demand side management which can take into account operational constraints
- Household segmentation and online hierarchical forecasting approaches
- A data generator using deep-learning

Prospects

- Improving experiments (by integrating operational constraints, splitting clusters to send several tariffs, testing with various data generators…)
- Integrating online hierarchical forecasting to personalized demand side management bandit algorithm

Un immense merci à

Gilles Stoltz, Yannig Goude et Pierre Gaillard, pour cette aventure, pour leur encadrement et pour leur complémentarité,

Rob Hyndman and Odalric-Ambrym Maillard, for their careful reading and relevant remarks which helped me to improve the manuscript,

Christophe Giraud, Émilie Kaufmann et Nadia Oudjane, pour avoir accepté de constituer le jury dont je rêvais,

Malo Huard, pour une collaboration des plus amicales,

Ricardo Jorge Bessa, pela calorosa recepção no Porto,

Clotilde d'Épenoux et Stéphane Nonnenmacher de l'EDMH pour leur aide précieuse dans l'organisation de cette soutenance *Covid-free,*

L'équipe Proba-Stat du Laboratoire de Mathématiques d'Orsay, au département Osiris et au groupe R39 à EDF R&D et aux équipes Sierra et Willow à l'Inria Paris pour trois merveilleuses années à leurs côtés.