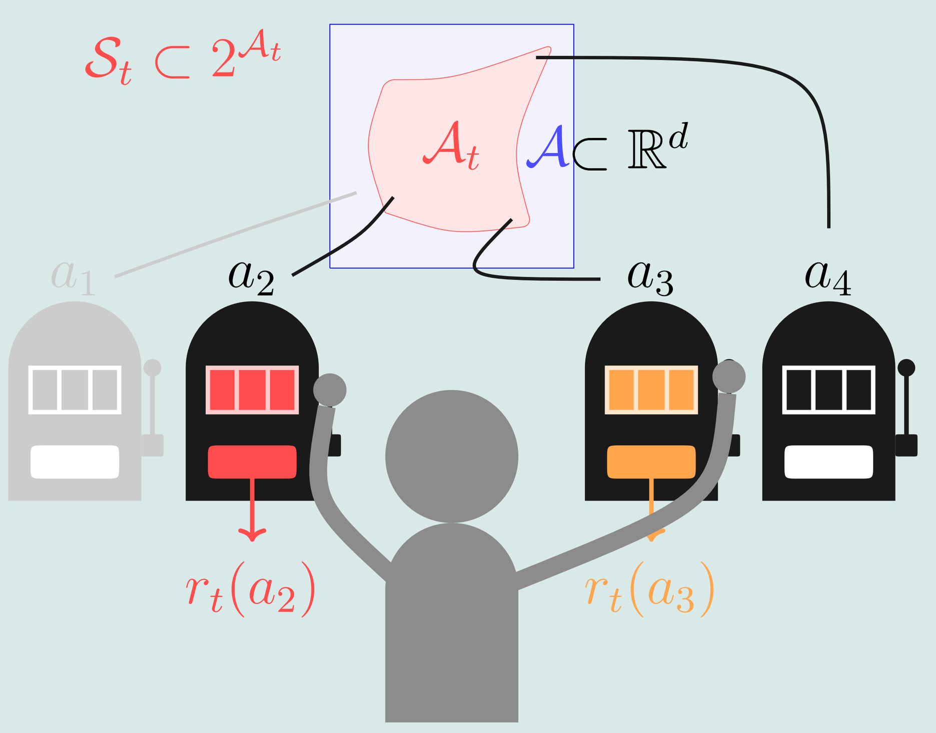
BAYESIAN ANALYSIS OF COMBINATORIAL GAUSSIAN PROCESS BANDITS

WITH APPLICATIONS TO ENERGY-EFFICIENT NAVIGATION

Jack Sandberg[1] Niklas Åkerblom[1,2] Morteza Haghir Chehreghani[1]

[1]Chalmers University of Technology and University of Gothenburg [2]Volvo Car Corporation

1. Combinatorial volatile Gaussian process semi-bandit problem (CVGP-MAB)



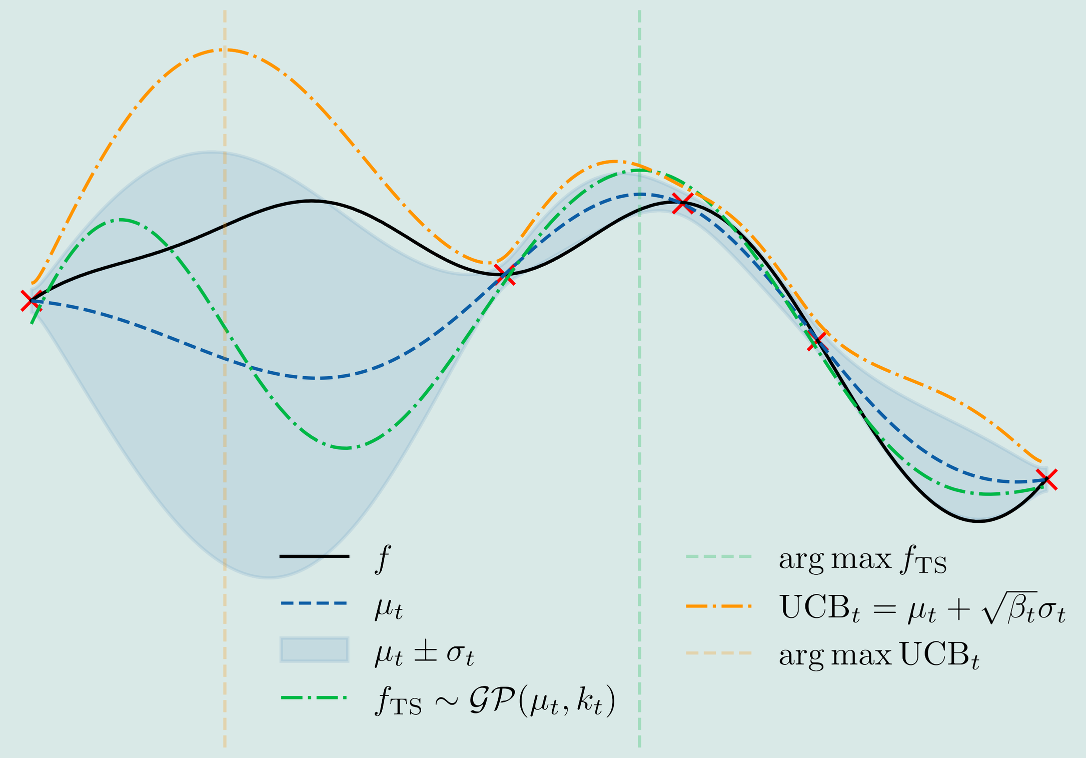
$\mathbb{E}[R_t] = f(a_2) + f(a_3), f \sim \mathcal{GP}(\mu, k)$

- Select superarm $\mathbf{a}_t \in \mathcal{S}_t$ s.t. $|\mathbf{a}_t| \leq K$.
- **Volatile (≡ contextual): Available arms vary with time.**
- Observe rewards $r_t(a)\ \forall a \in \mathbf{a}_t$.
- Goal: Minimize Bayesian regret

$$\mathrm{BR}(T) = \sum_{t=1}^{T} \mathbb{E}\left[f(\mathbf{a}_t^*) - f(\mathbf{a}_t)\right]$$

where $\mathbf{a}_t^* = \arg\max_{\mathbf{a}\in\mathcal{S}_t} f(\mathbf{a})$ and $f(\mathbf{a}) = \sum_{a\in\mathbf{a}} f(a)$.

2. Algorithms and theoretical results

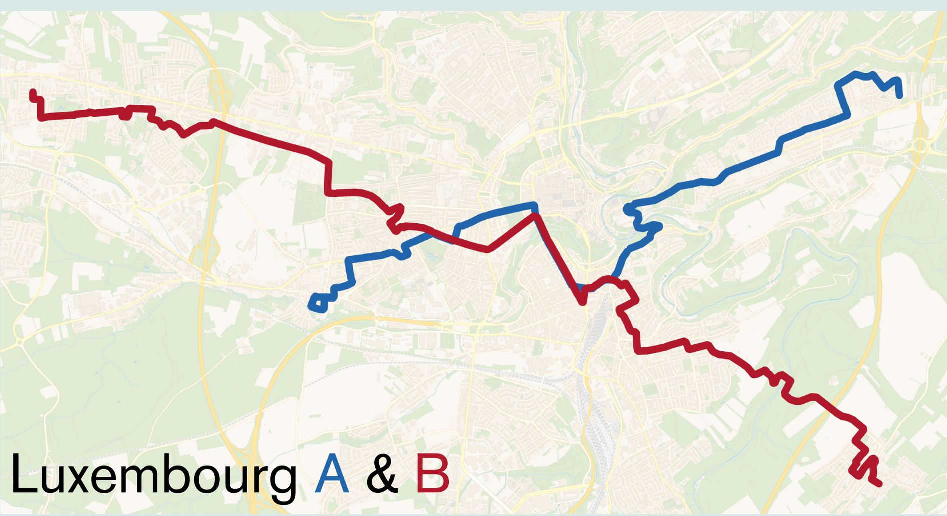


Acquisition functions:

GP-UCB	$\mu_t + \sqrt{\beta_t}\sigma_{t-1}(a)$
GP-BayesUCB[1,2]	$Q(1-\eta_t, \mathcal{N}(\mu_t, \sigma_t^2(a)))$
GP-TS	$\tilde{f}_t(a) \sim \mathcal{GP}(\mu_t, k_t)$

- GP-UCB and -BayesUCB differ in parametrization of β_t.
- **GP-UCB and GP-TS: First regret bound for infinite and volatile setting (and combinatorial).**
- **First regret bound for GP-BayesUCB.**
- Bayesian regret of $\tilde{\mathcal{O}}(K\sqrt{T\gamma_{TK}})$, where γ_{TK} is the maximum information gain of TK arms.

3. Application to online energy efficient navigation



- Find most efficient route from A to B using minimal energy.
- Random energy cost with unknown means.
- Minimize total energy used over T steps.
- Modelled as combinatorial MAB by Åkerblom et al. [3].
- **We model as CVGP-MAB and use the features of the road segments to learn faster.**

References

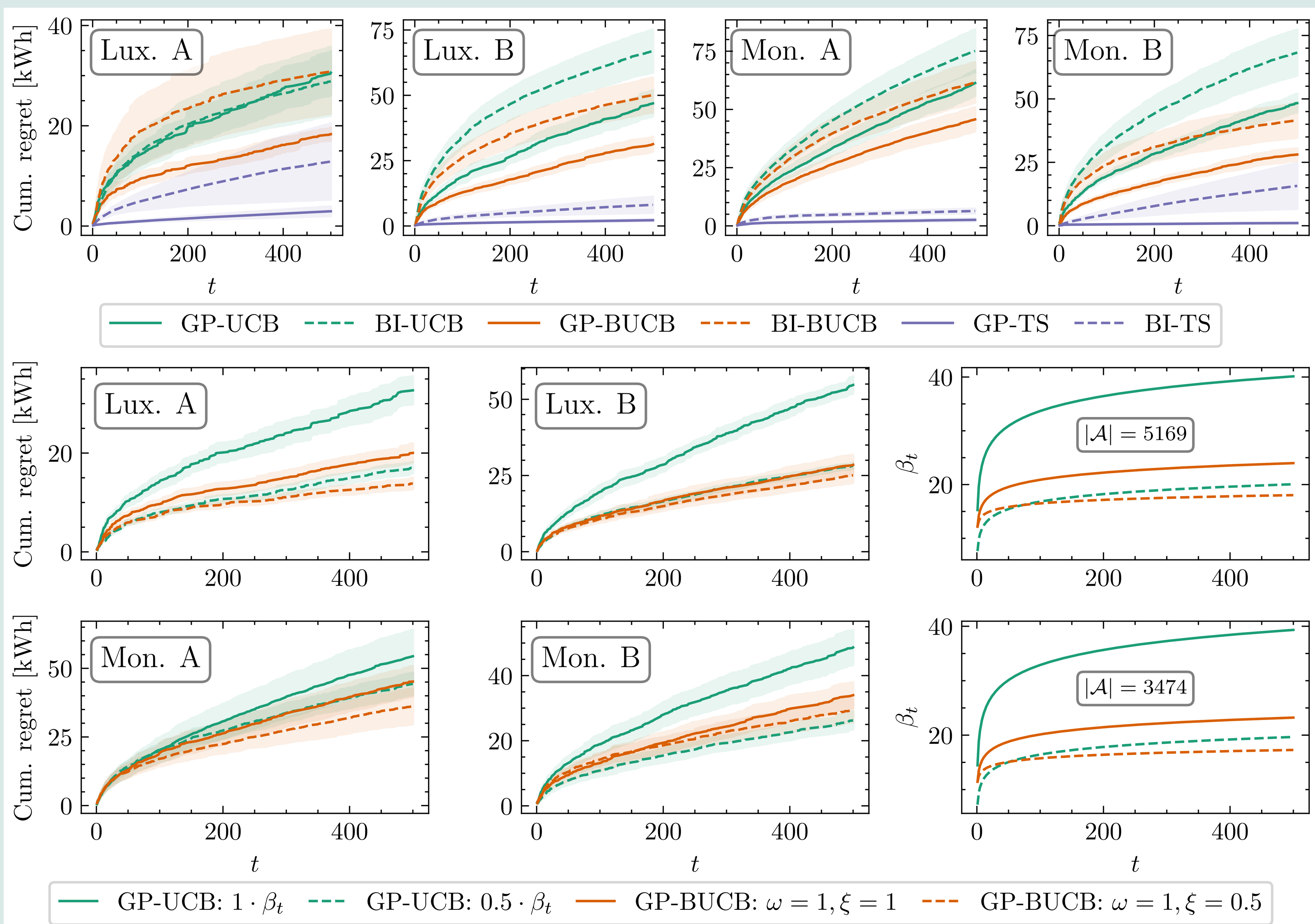
[1] Kaufmann et al. On Bayesian Upper Confidence Bounds for Bandit Problems. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, 2012

[2] Nuara et al. A Combinatorial-Bandit Algorithm for the Online Joint Bid/Budget Optimization of Pay-per-Click Advertising Campaigns. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[3] Åkerblom et al. Online learning of energy consumption for navigation of electric vehicles. In *Artificial Intelligence*, 2023.

4. Experimental results

- Real-world road network data combined with energy model.
- Edge features: length, incline and speed limit.
- Bayesian inference (BI) with UCB, BayesUCB and TS as baseline [3].
- **Contextual features reduce regret.**
- **GP-BayesUCB is more flexible than GP-UCB.**
- **GP-TS has lowest regret.**



FFI WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM