Fair Model-Based Reinforcement Learning Comparisons with Explicit and Consistent Update Frequency

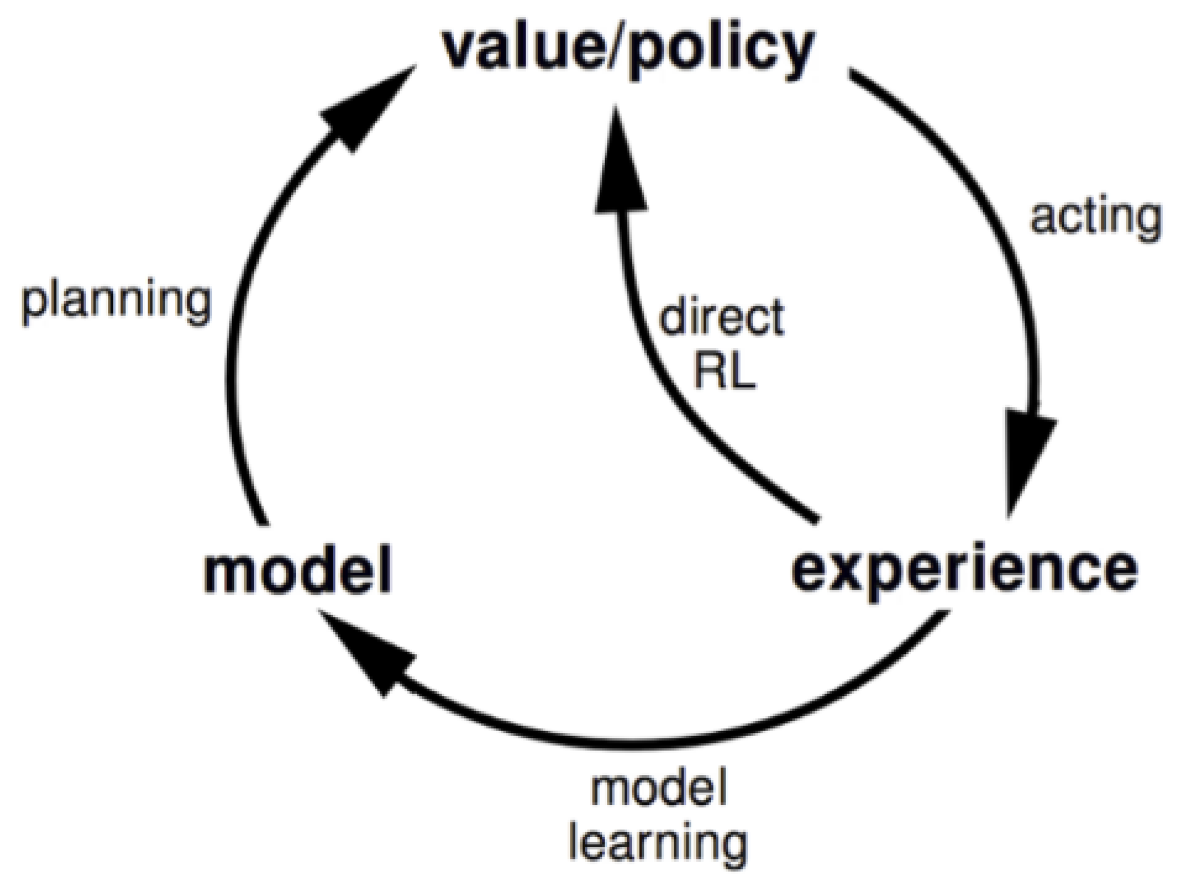
Albert Thomas*
Abdelhakim Benechehab*°
Giuseppe Paolo*
Balázs Kégl*

* Huawei, Noah's Ark Lab
° EURECOM, Department of Data Science

Model-based Reinforcement Learning

A loop that alternates between:

- Experience Collection
- Model learning
- Policy learning



The update frequency

The number of steps before updating the policy parameters.

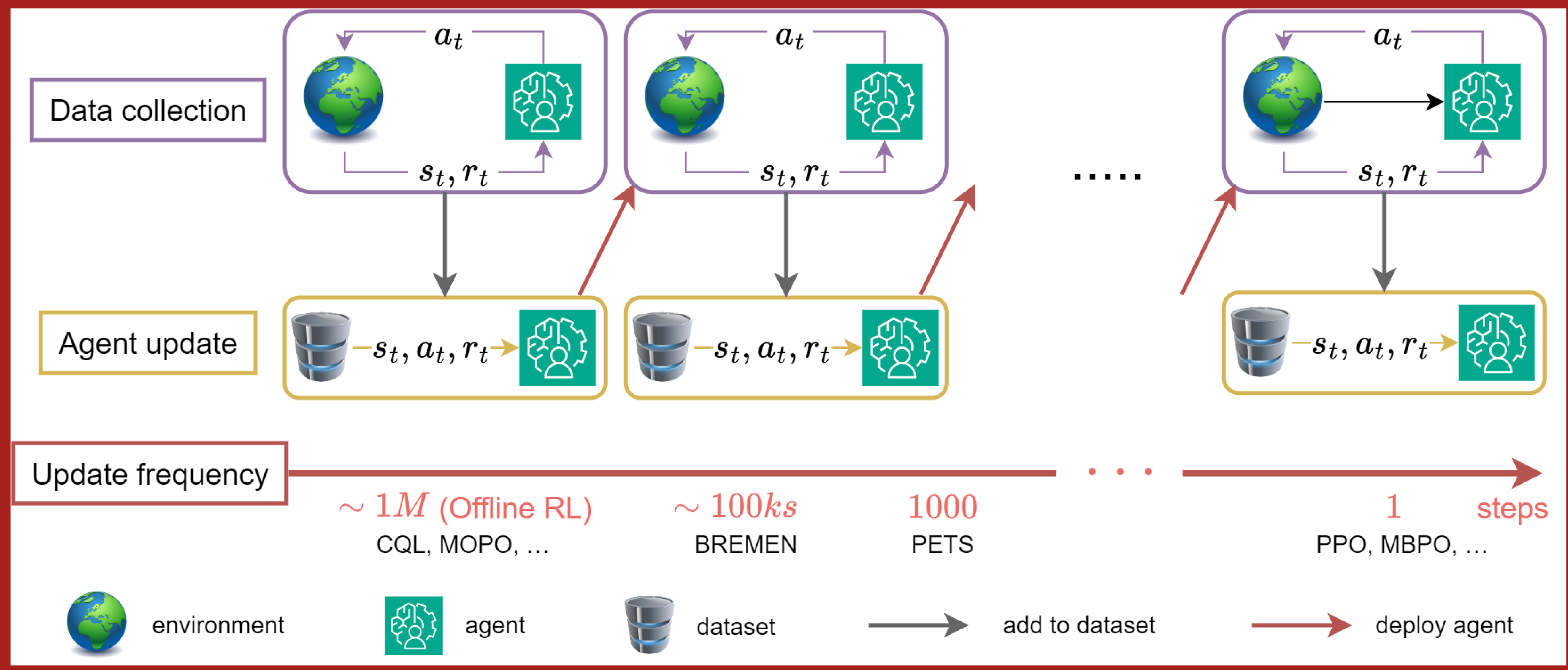
Examples:

- Systems constraining the policy to be updated every once in a while.
- Real-time learning for Robotics applications.

IN MBRL, THE UPDATE FREQUENCY IS:

- IMPLICIT
- CONFOUNDING
- UNDERSTUDIED

THIS IS A PROBLEM.



Specifically:

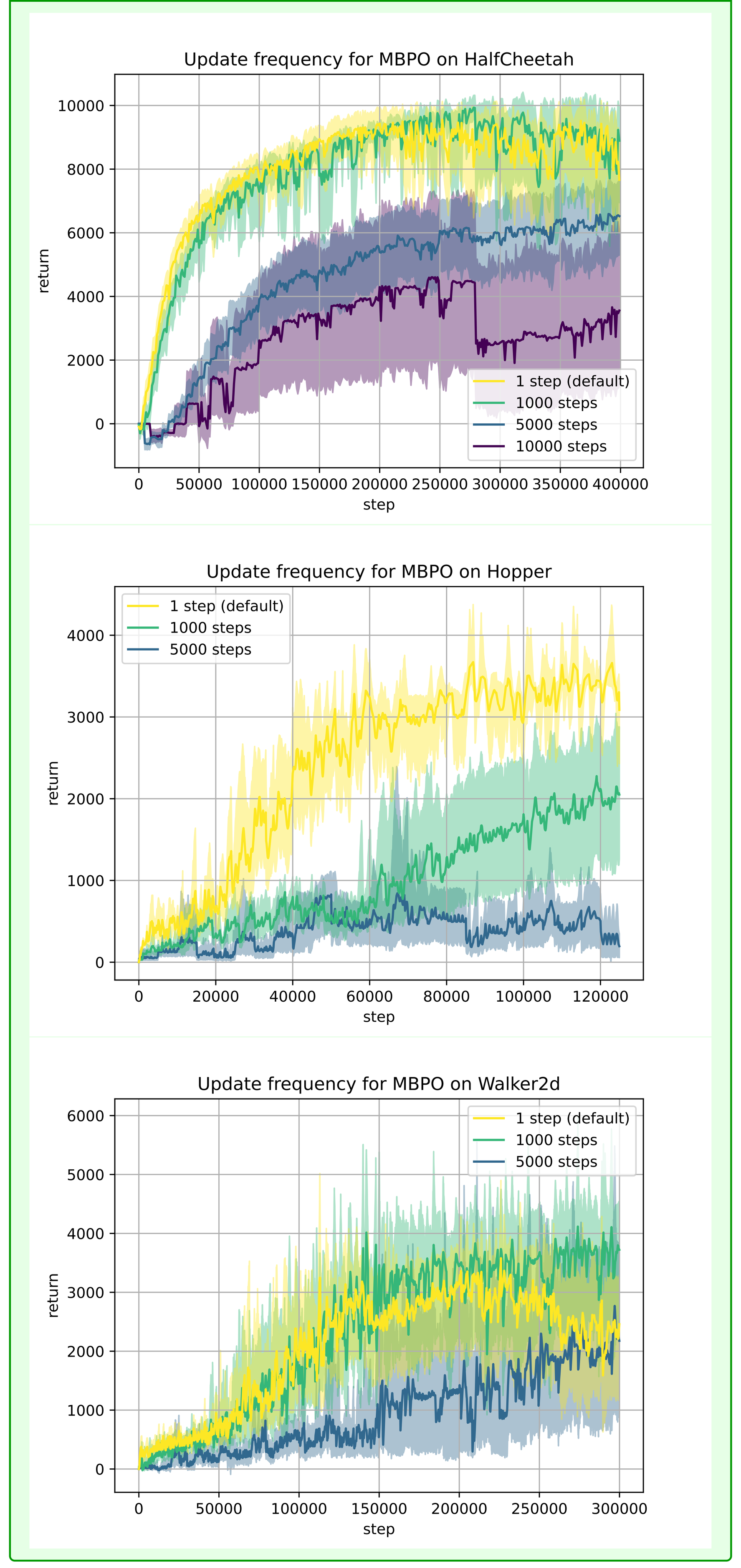
- The update frequency choice is rarely made **explicit** in the literature
- Comparisons between algorithms often do not **fix** the update frequency
- MBRL papers often lack **ablation** studies on the update frequency

Read the full blogpost

Guidelines for the update frequency

- The update frequency should also be regarded as a **system constraint** rather than a **design choice**
- Make it **explicit**
- **Fix** it when **comparing** with other algorithms
- If possible, provide an **ablation** study showing its **impact** on the performance of your algorithm

Experiments with MBPO



References

- **[MBPO]** (Janner et al. 2019) When to trust your model: model-based policy optimization.
- **[BREMEN]** (Matsushima et al. 2020) Deployment-Efficient Reinforcement Learning via Model-Based Offline Optimization.