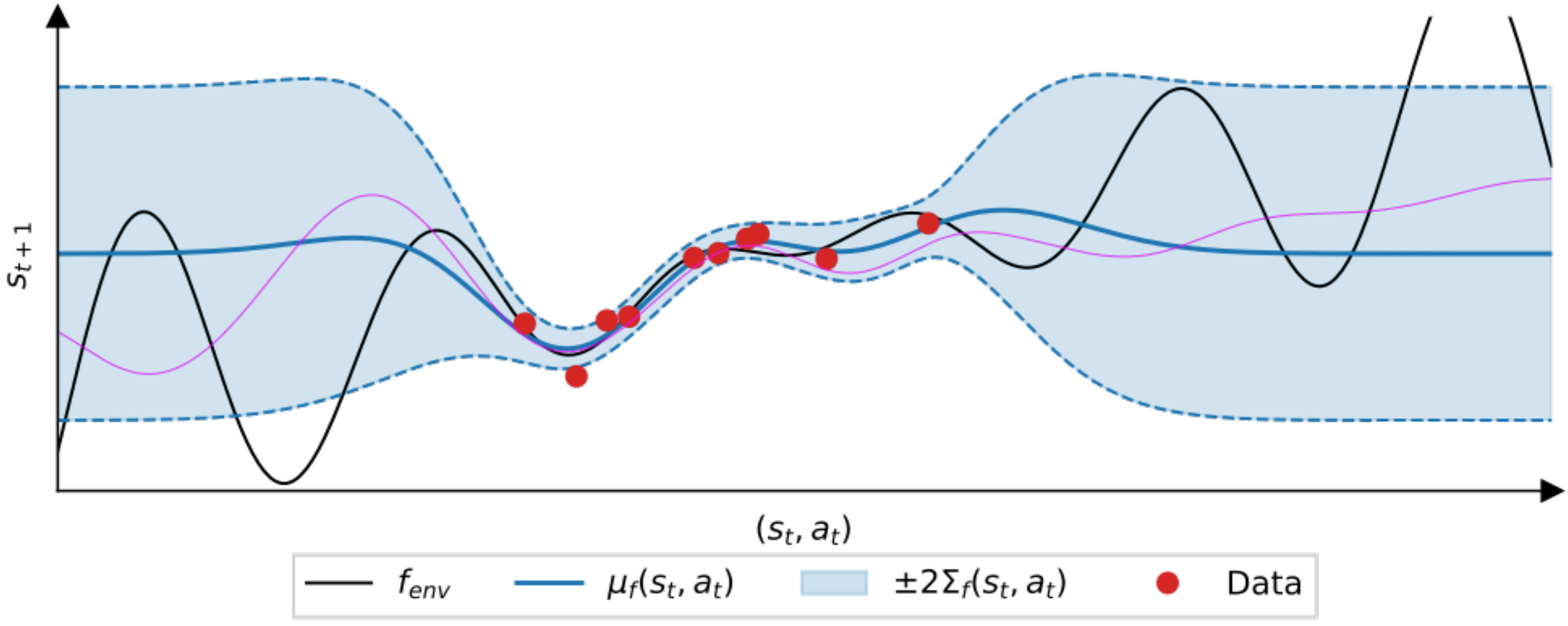
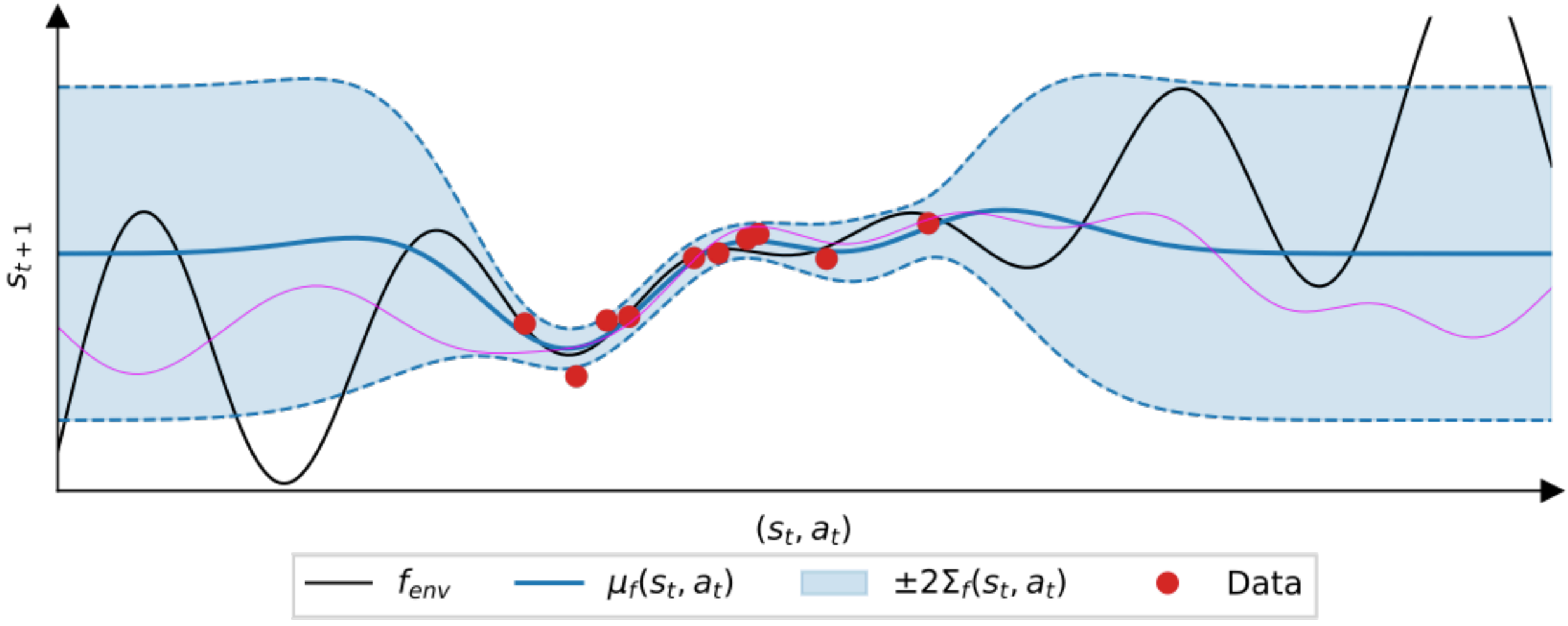
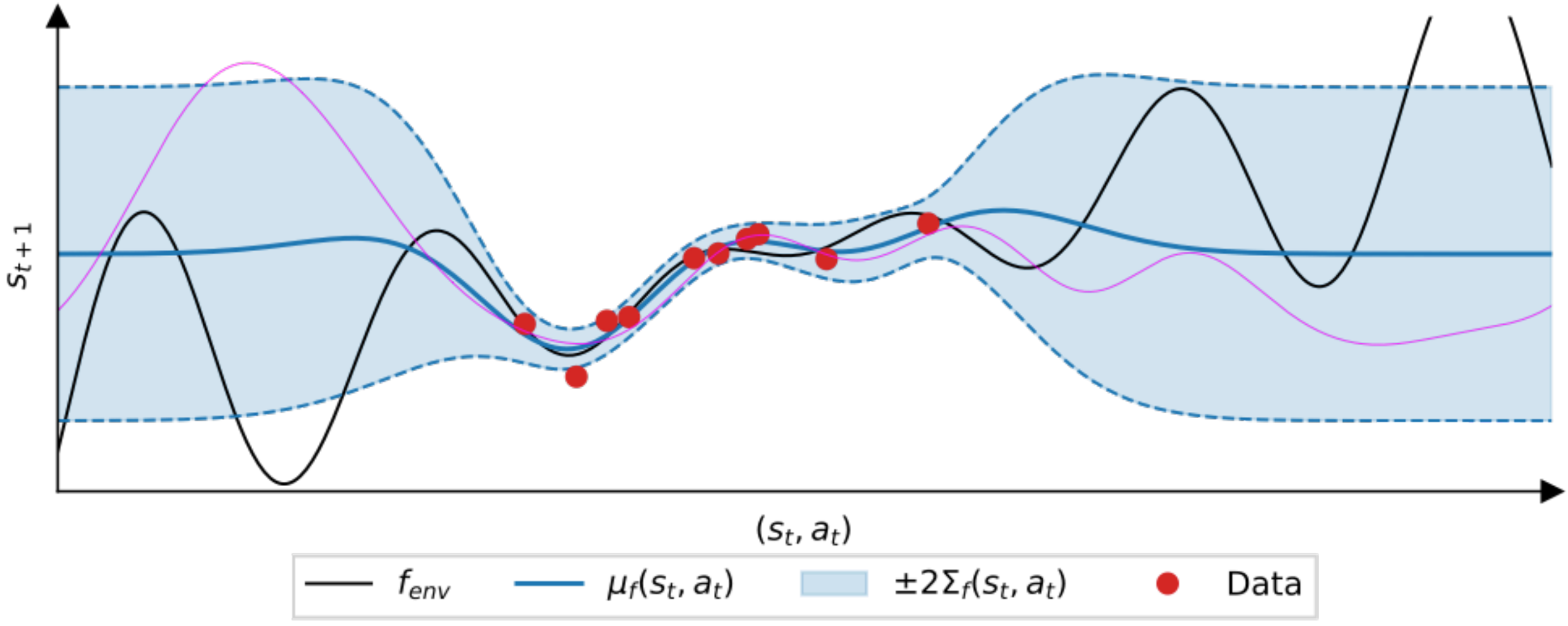

s_{t+1}
(s_t, a_t)
f_{env}
$\mu_f(s_t, a_t)$
$\pm 2\Sigma_f(s_t, a_t)$
Data


s_{t+1}
(s_t, a_t)
f_{env}
$\mu_f(s_t, a_t)$
$\pm 2\Sigma_f(s_t, a_t)$
Data

Exploration via Posterior Sampling

41


s_{t+1}
(s_t, a_t)
f_{env}
$\mu_f(s_t, a_t)$
$\pm 2\Sigma_f(s_t, a_t)$
Data

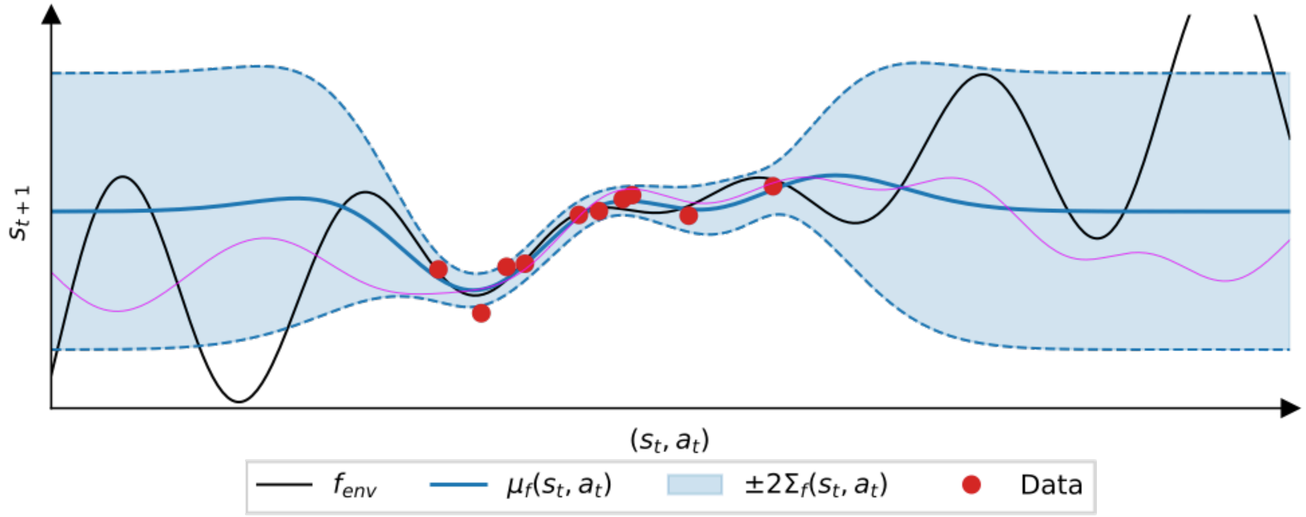
$$\pi^{PS} = \arg\max_{\pi} J(\pi; \tilde{f}), \quad \tilde{f} \sim p(f \mid \mathscr{D})$$

Osband et al. (2013). (More) Efficient Reinforcement Learning via Posterior Sampling. NeurIPS.

No extra hyperparamters

Exploration via Posterior Sampling

$$\pi^{PS} = \arg\max_{\pi} J(\pi; \tilde{f}), \quad \tilde{f} \sim p(f \mid \mathcal{D})$$



No extra hyperparamters

Osband et al. (2013). (More) Efficient Reinforcement Learning via Posterior Sampling. NeurIPS.

Exploration via Upper Confidence Bound

$$\pi^{UCB} = \arg\max_{\pi} \max_{f \in \mathcal{M}} J(\pi; f)$$