

Multi-Fidelity Bayesian Optimization with Unreliable Information Sources

Petrus Mikkola, **Julien Martinelli**, Louis Filstroff, Samuel Kaski

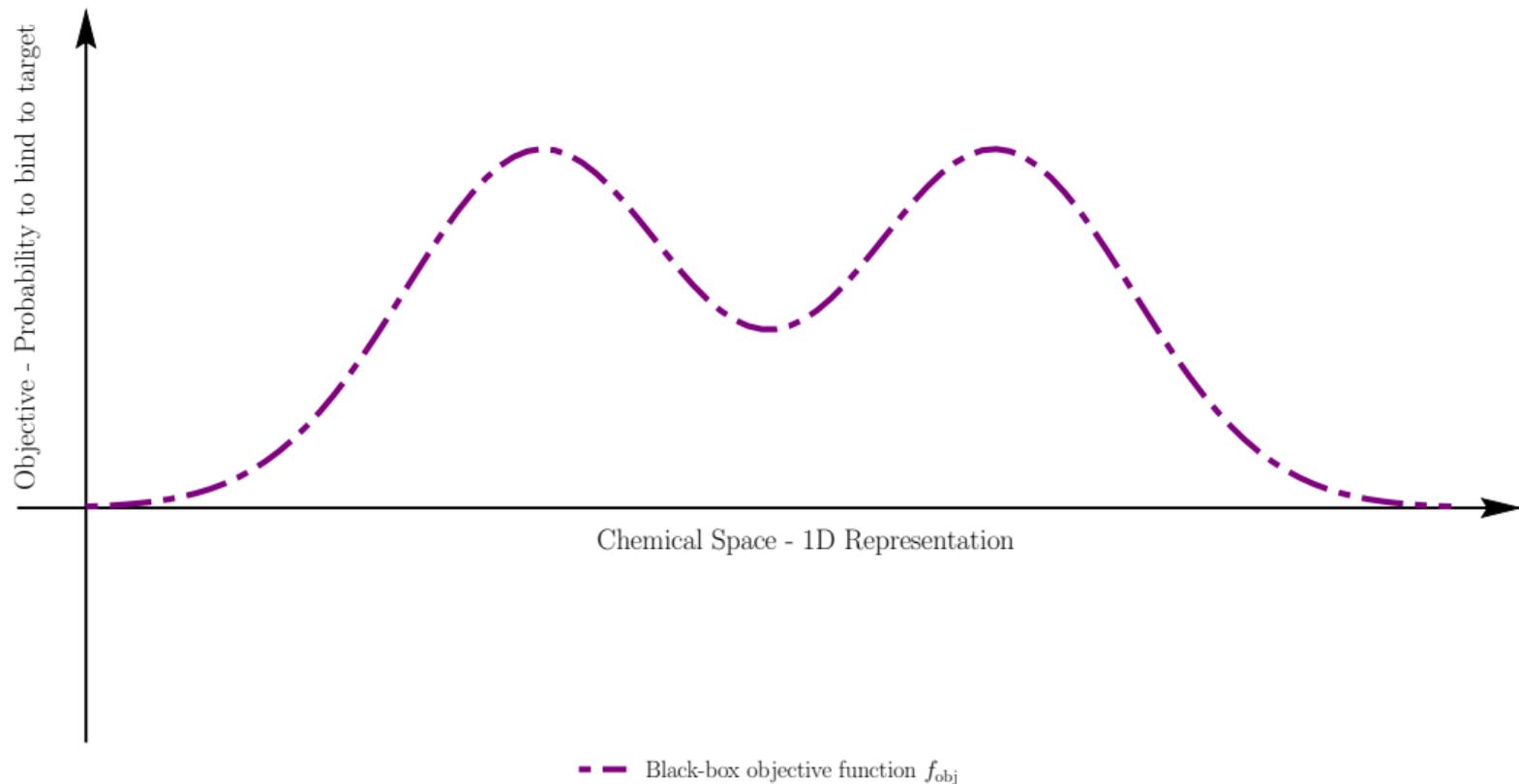
AISTATS 2023



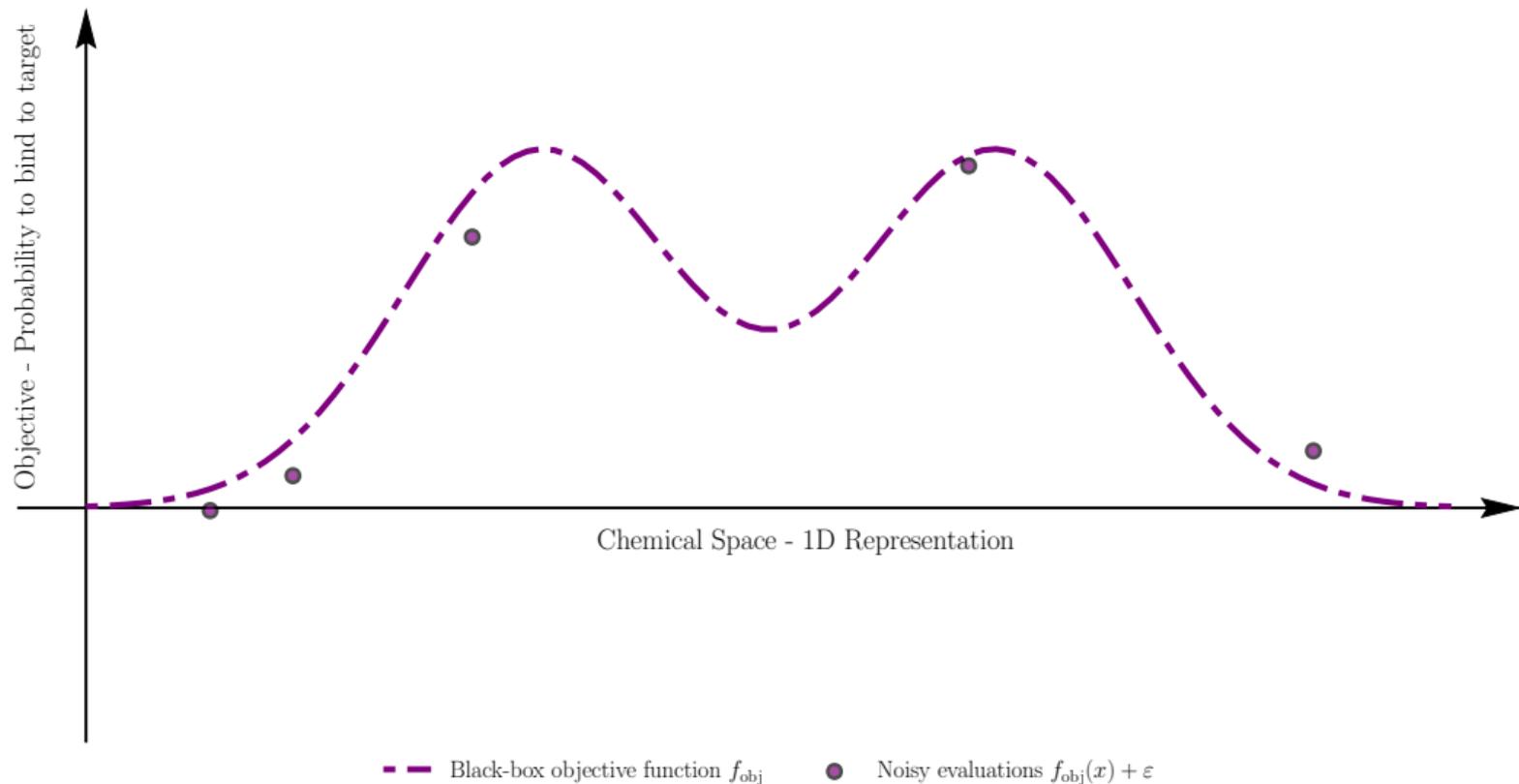
March 25th, 2025

Bayesian Optimization 101

Bayesian Optimization 101

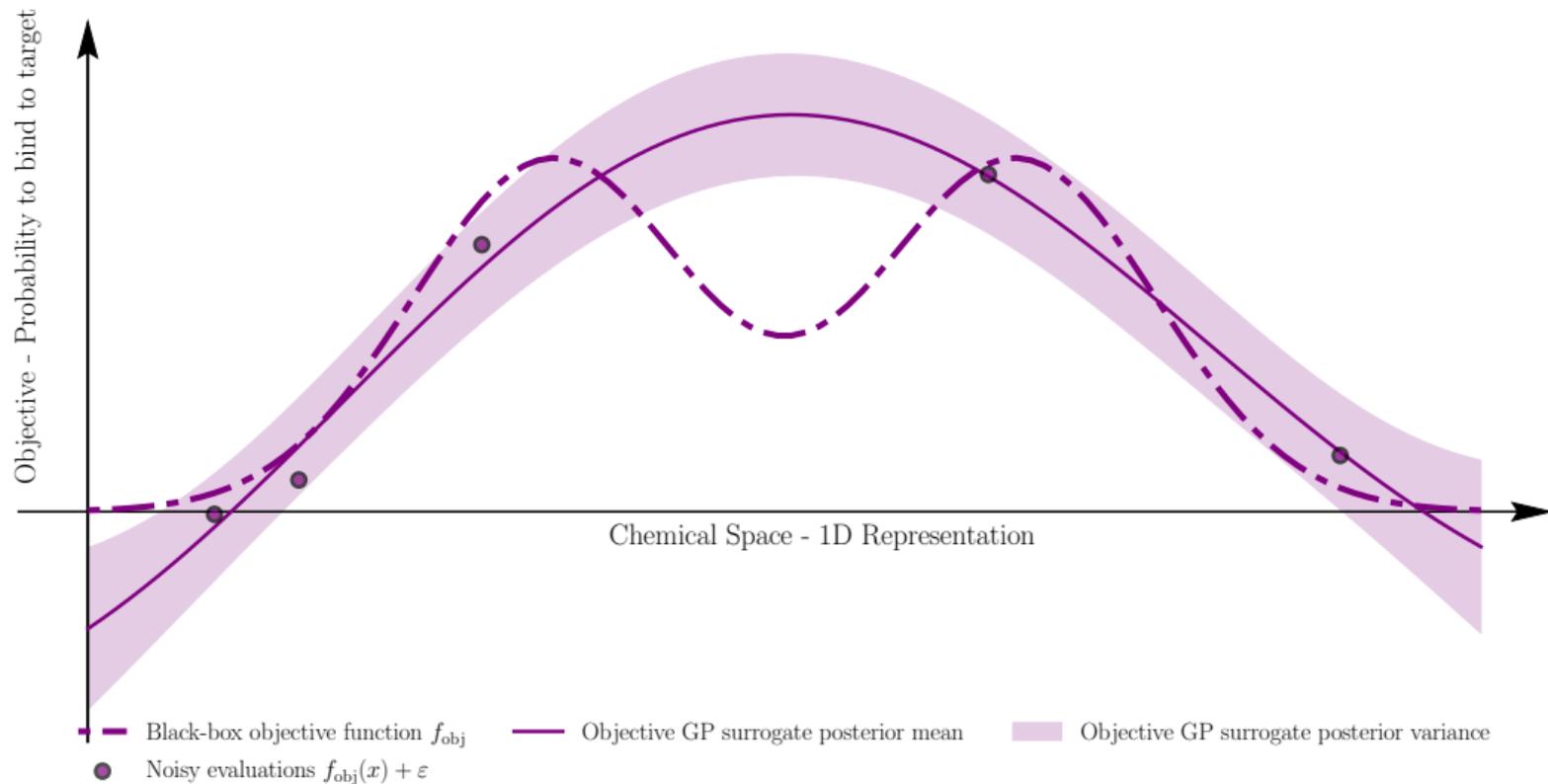


Bayesian Optimization 101



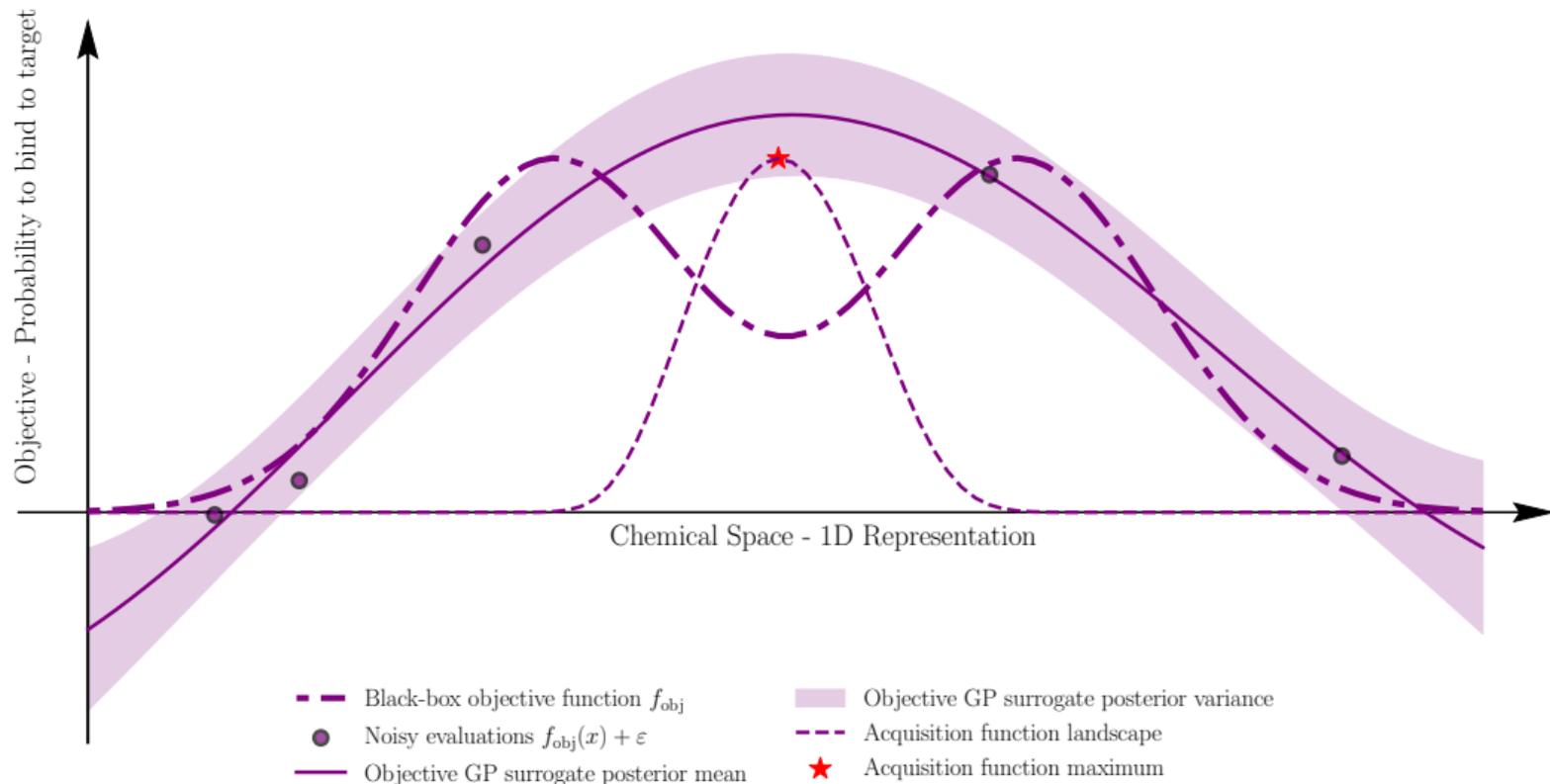
Bayesian Optimization 101

Budget = 20



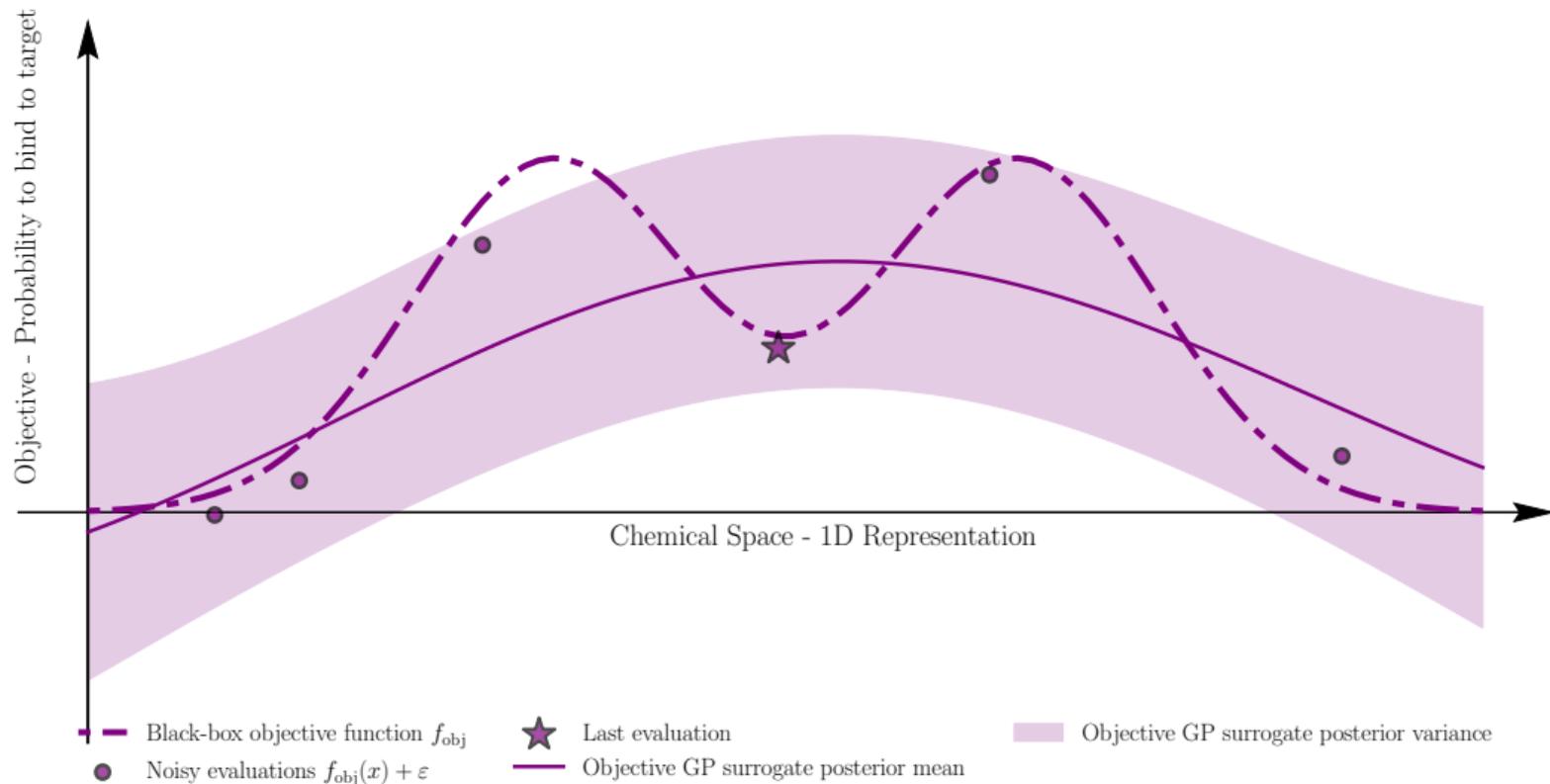
Bayesian Optimization 101

Budget = 20



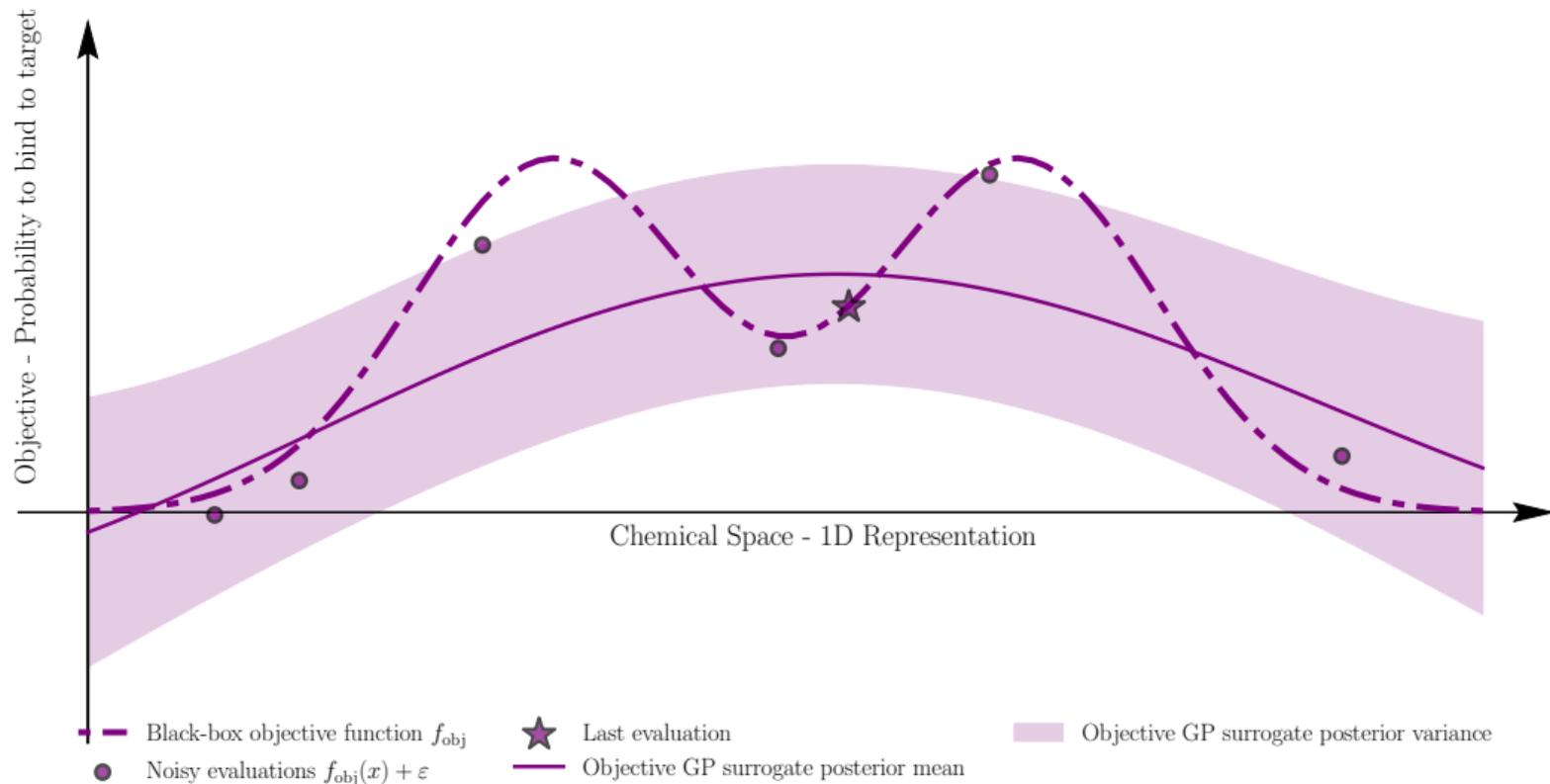
Bayesian Optimization 101

Budget = 19



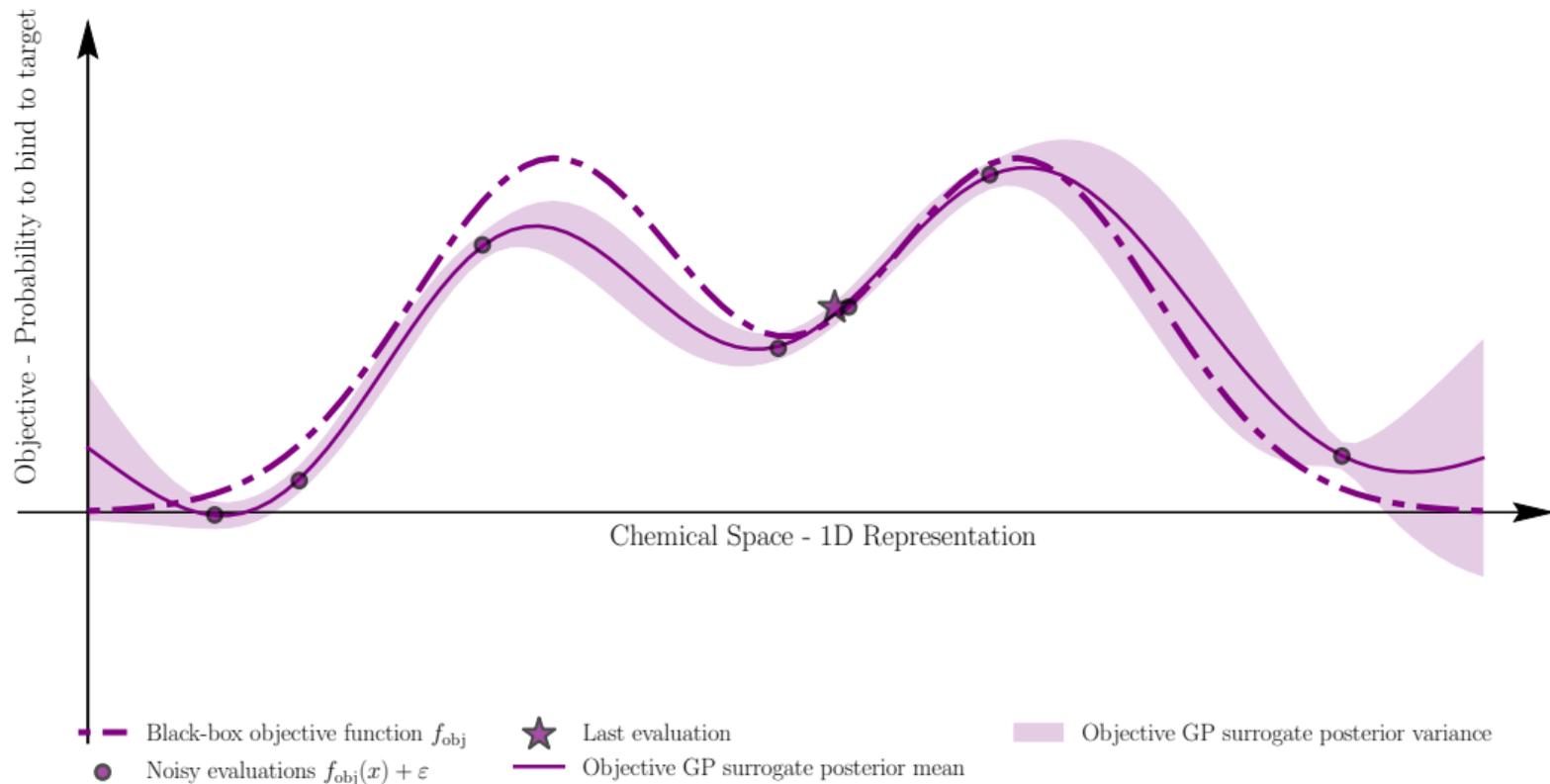
Bayesian Optimization 101

Budget = 18



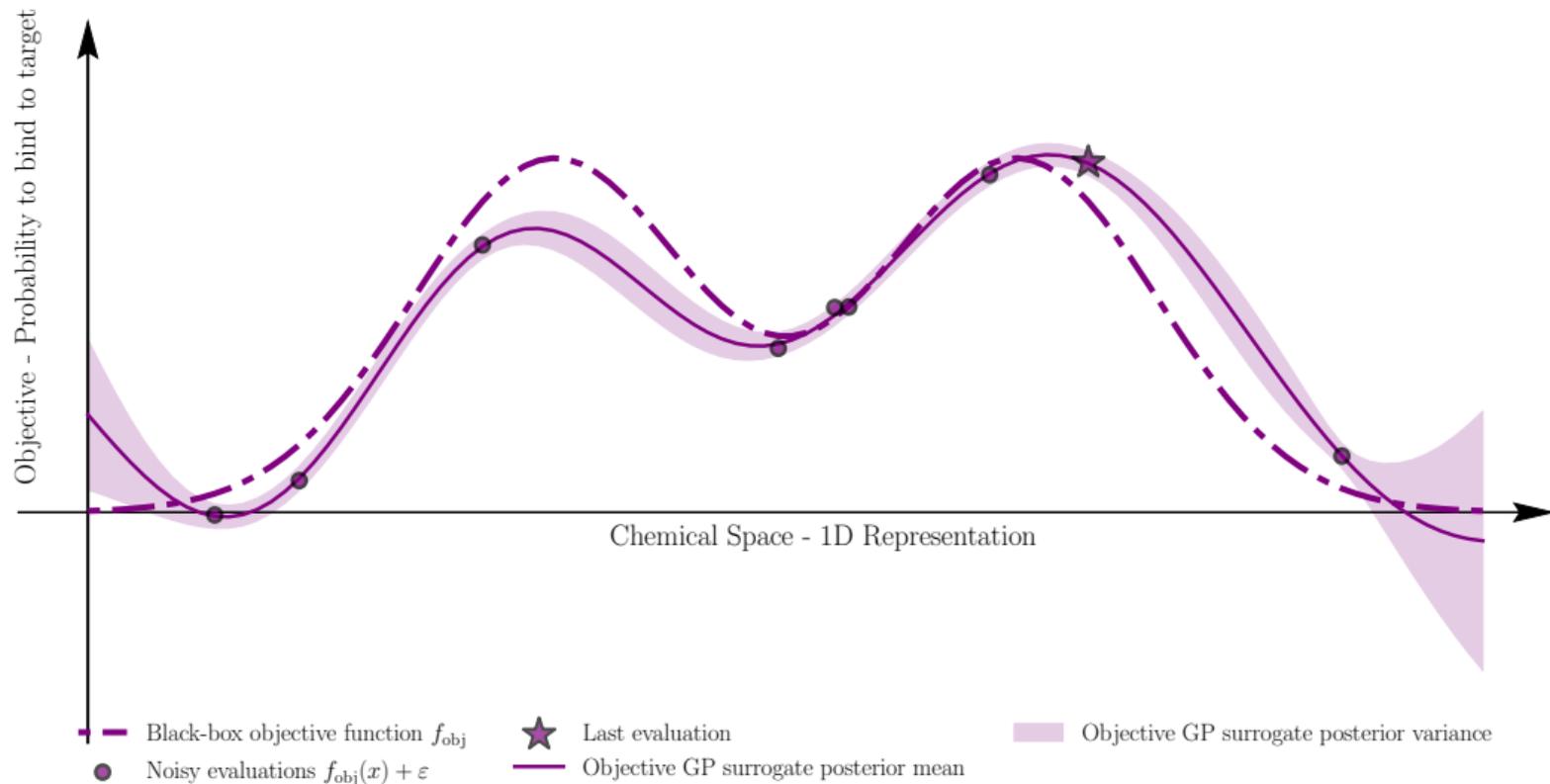
Bayesian Optimization 101

Budget = 17



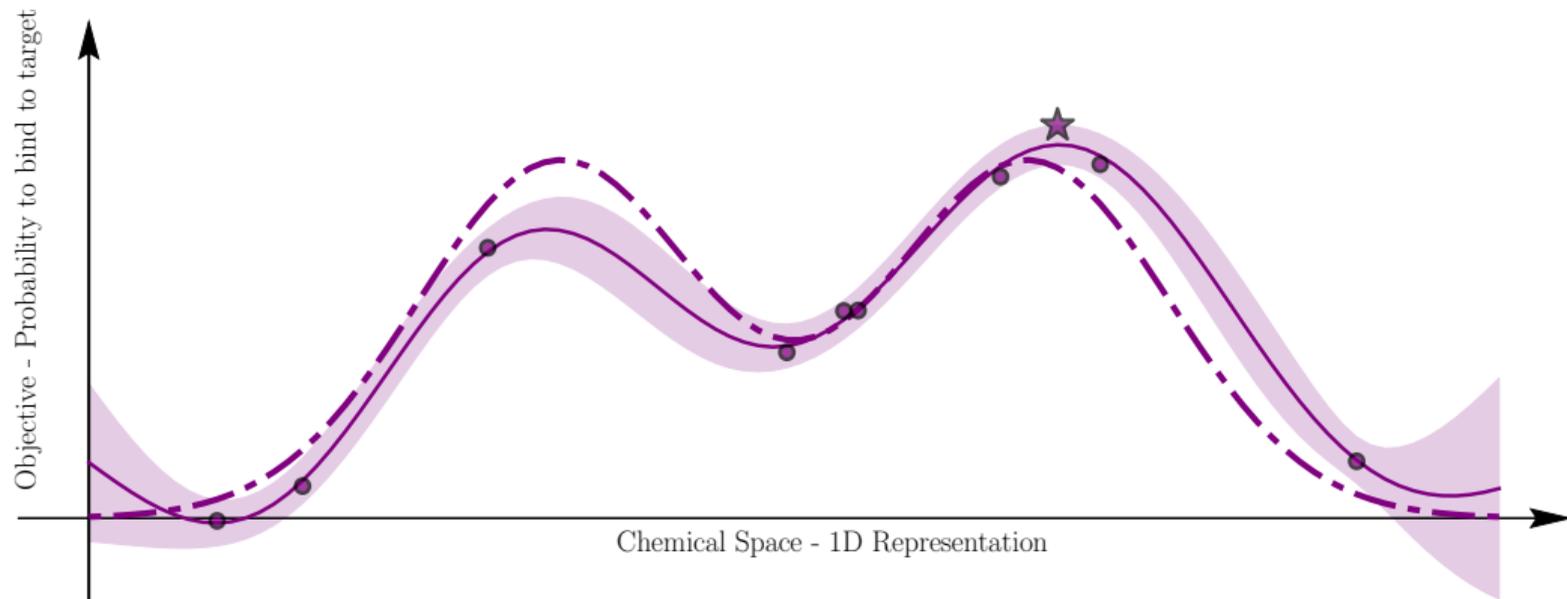
Bayesian Optimization 101

Budget = 16



Bayesian Optimization 101

Budget = 15



— Black-box objective function f_{obj}

● Noisy evaluations $f_{\text{obj}}(x) + \epsilon$

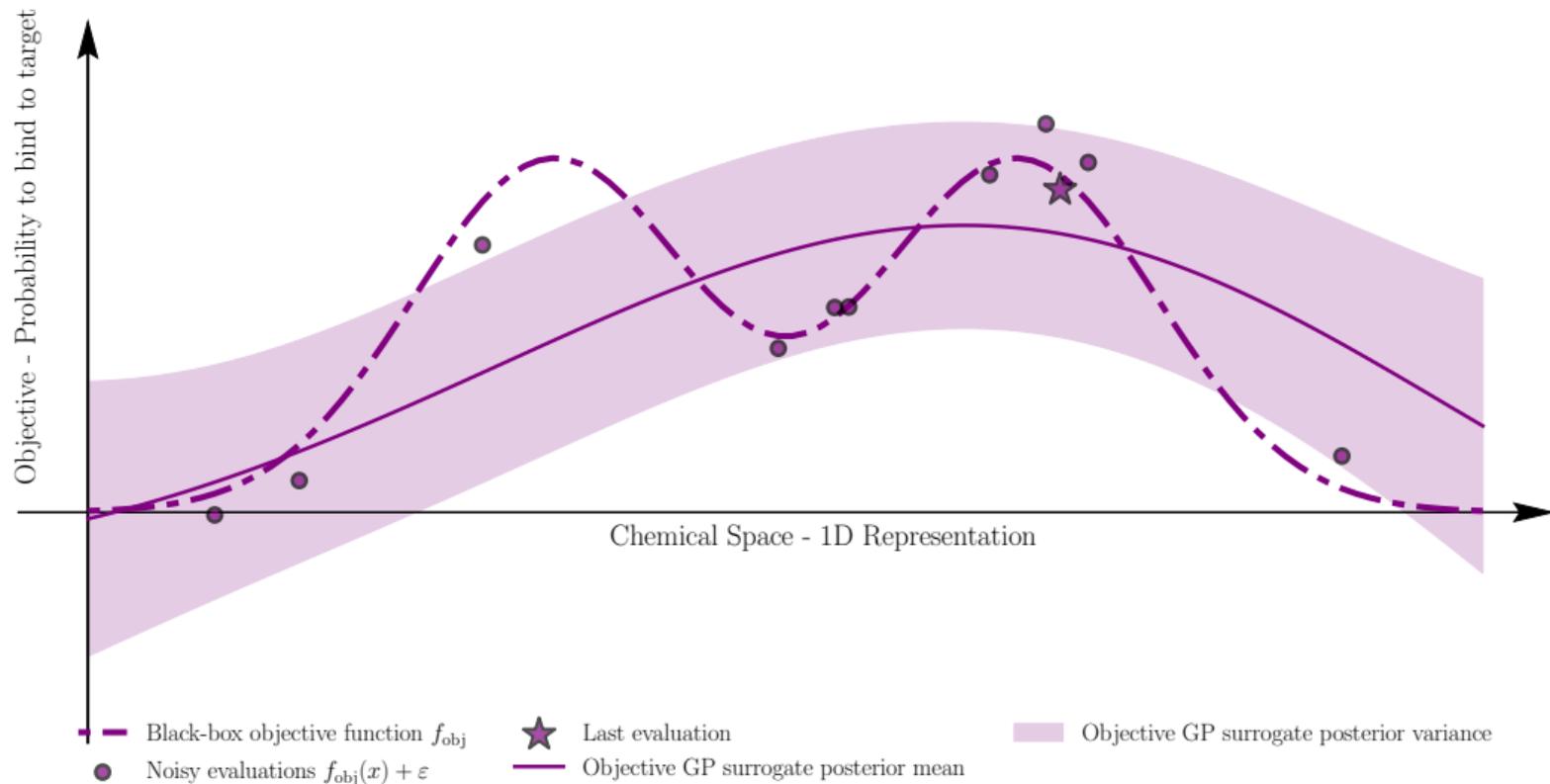
★ Last evaluation

— Objective GP surrogate posterior mean

■ Objective GP surrogate posterior variance

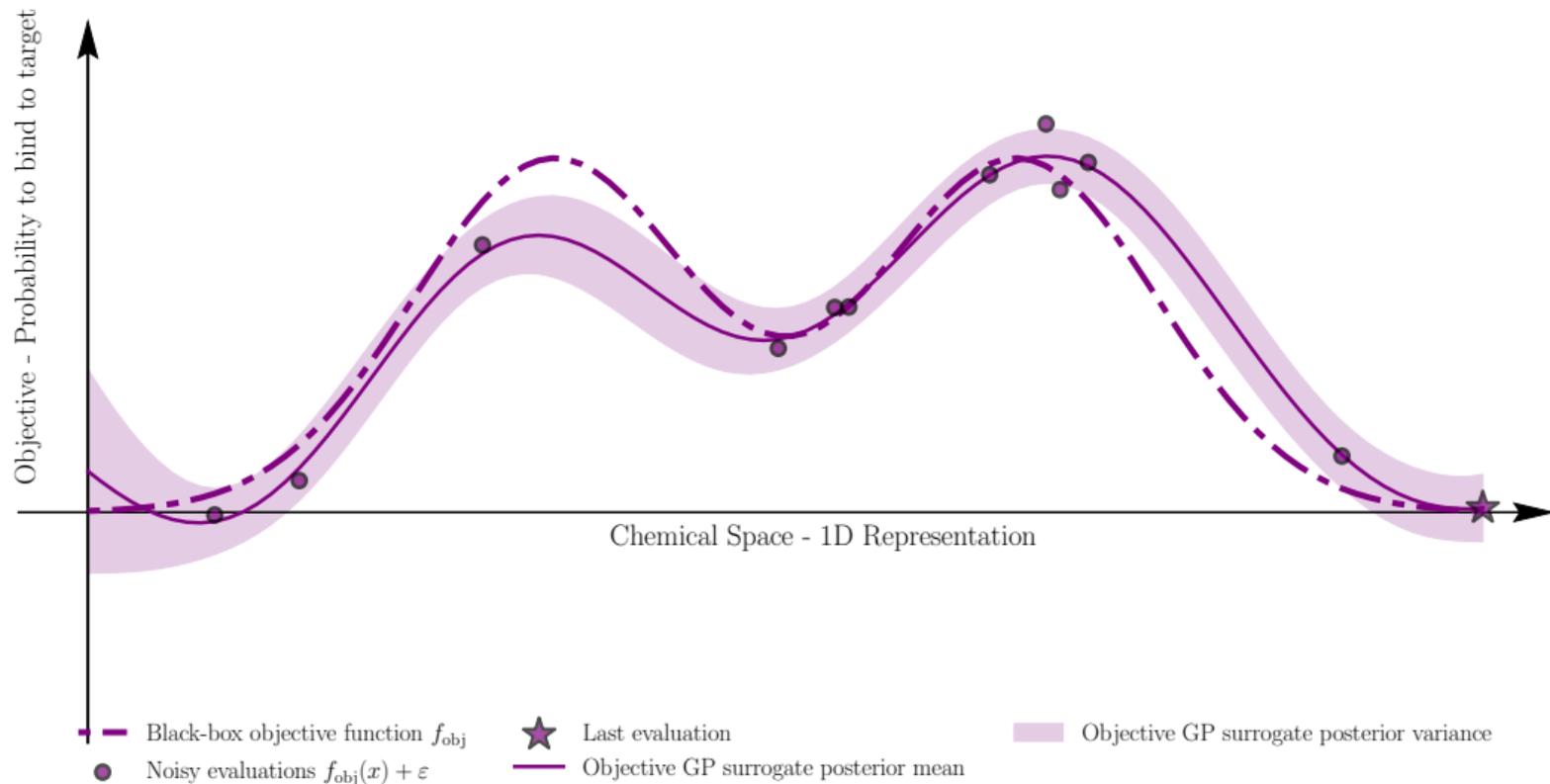
Bayesian Optimization 101

Budget = 14



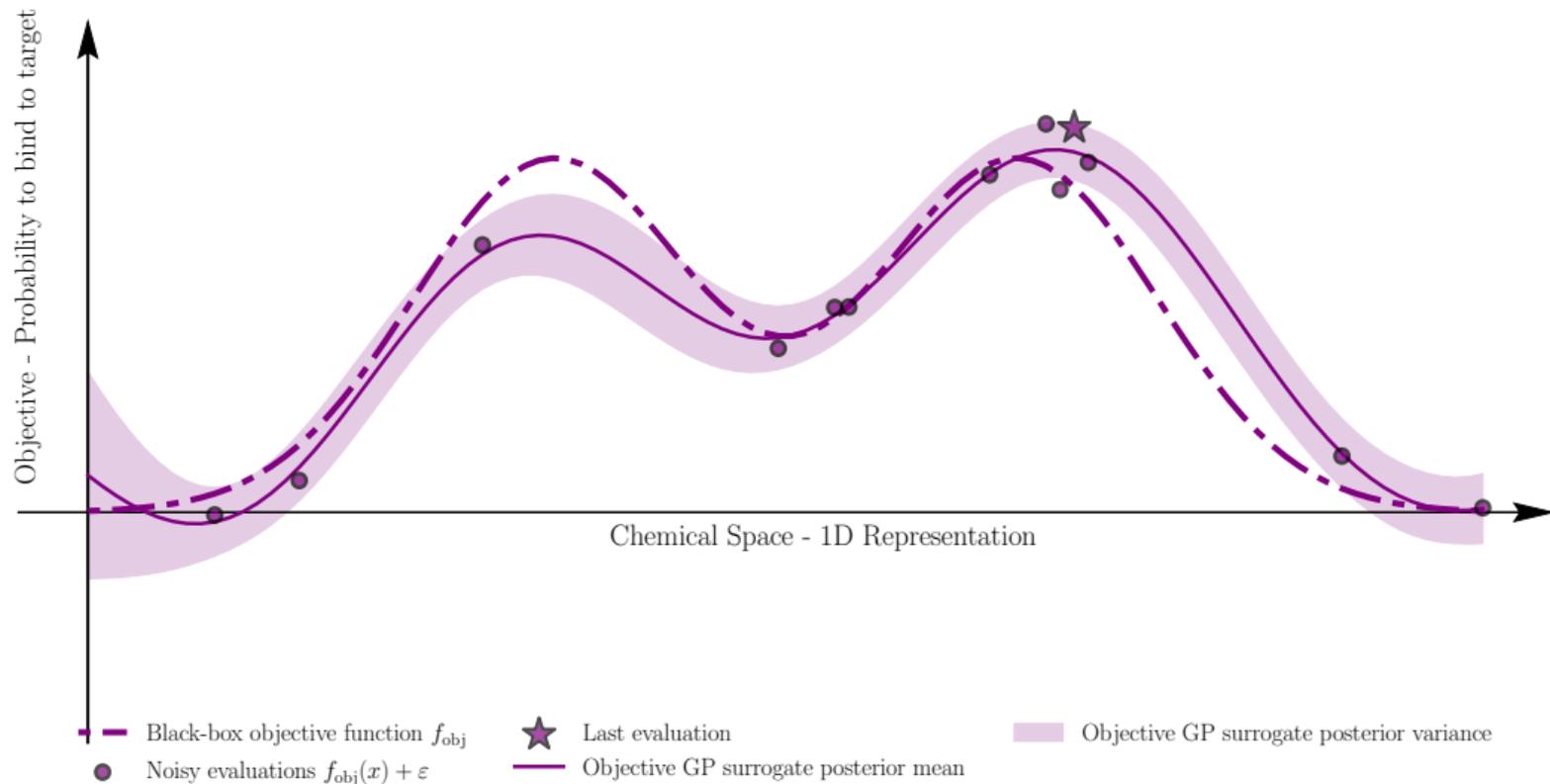
Bayesian Optimization 101

Budget = 13

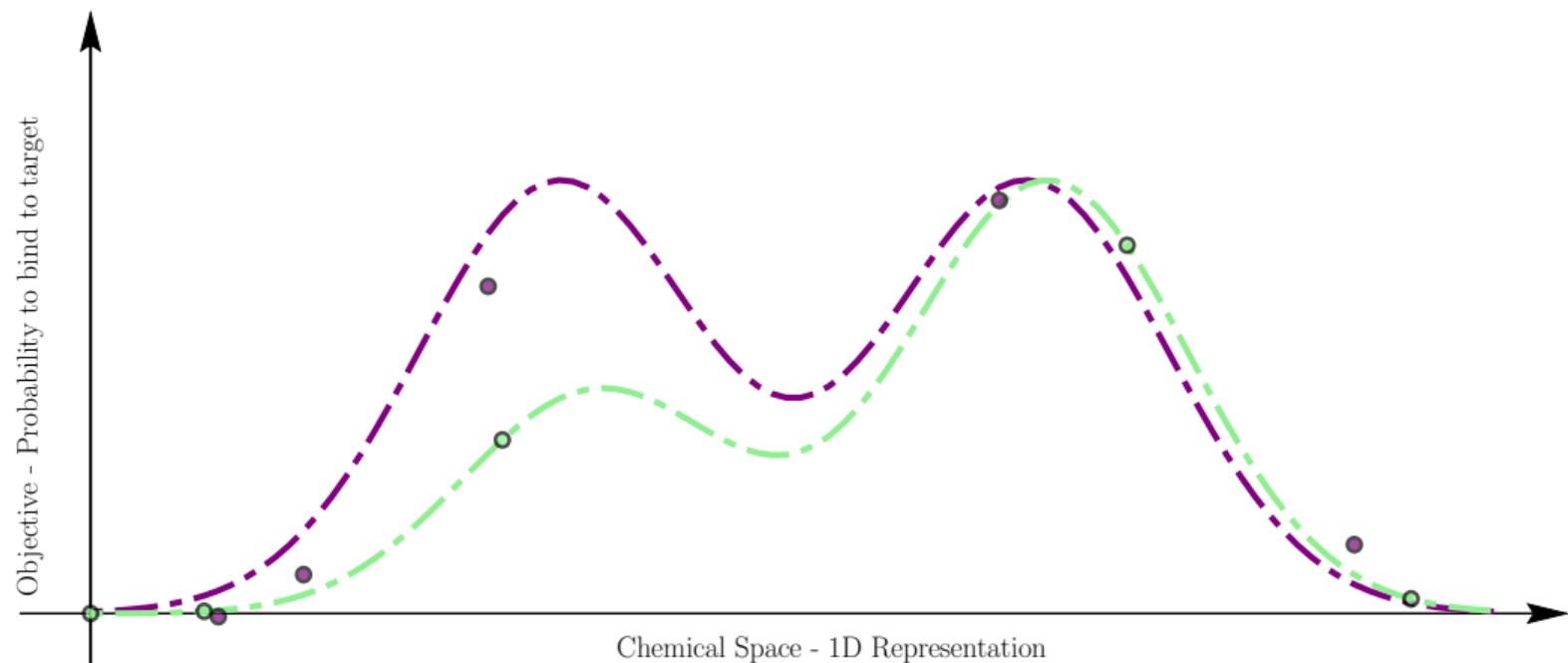


Bayesian Optimization 101

Budget = 12



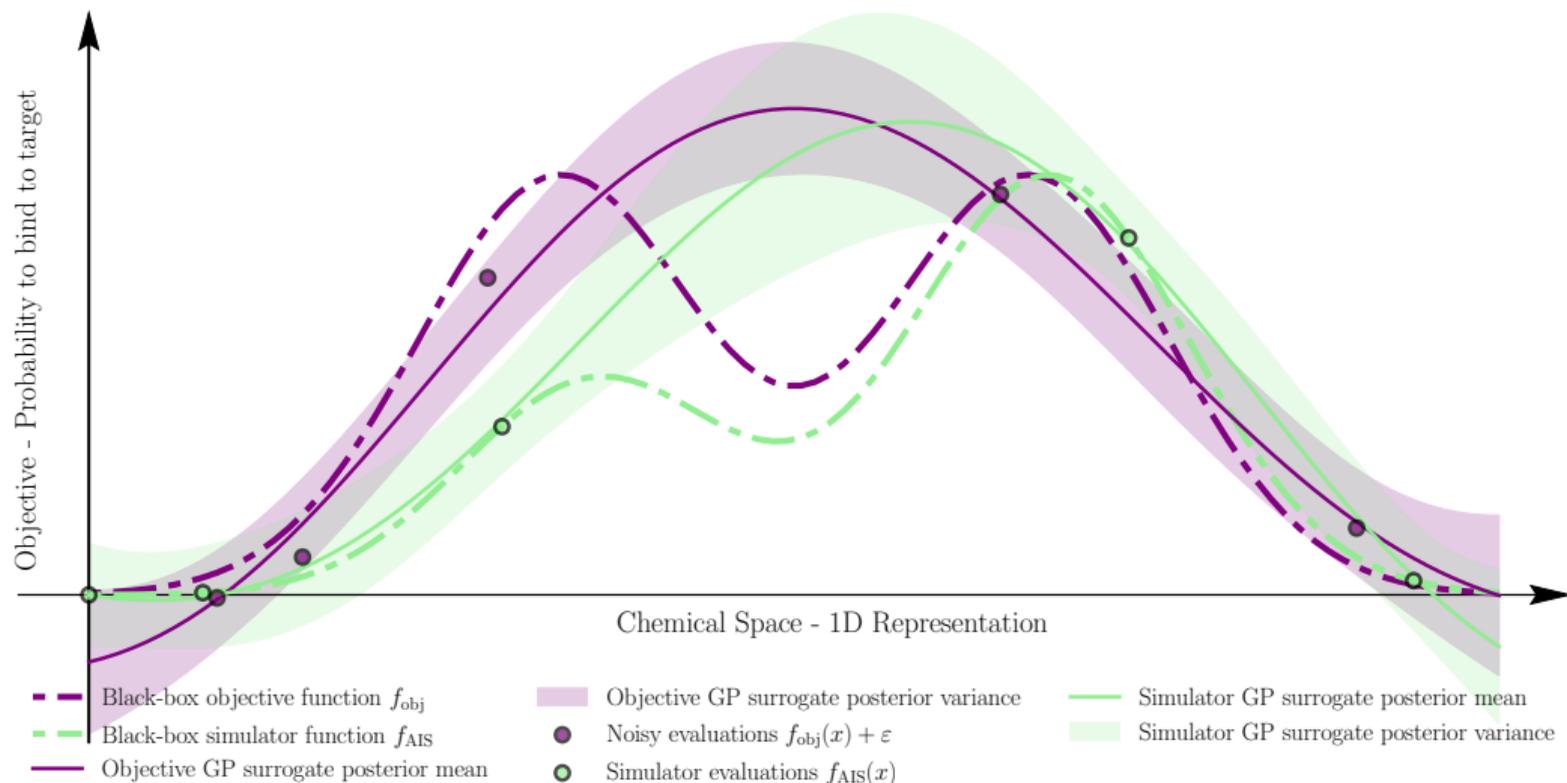
Multi Fidelity Bayesian Optimization 101



- Black-box objective function f_{obj}
- Black-box simulator function f_{AIS}
- Noisy evaluations $f_{obj}(x) + \varepsilon$
- Simulator evaluations $f_{AIS}(x)$

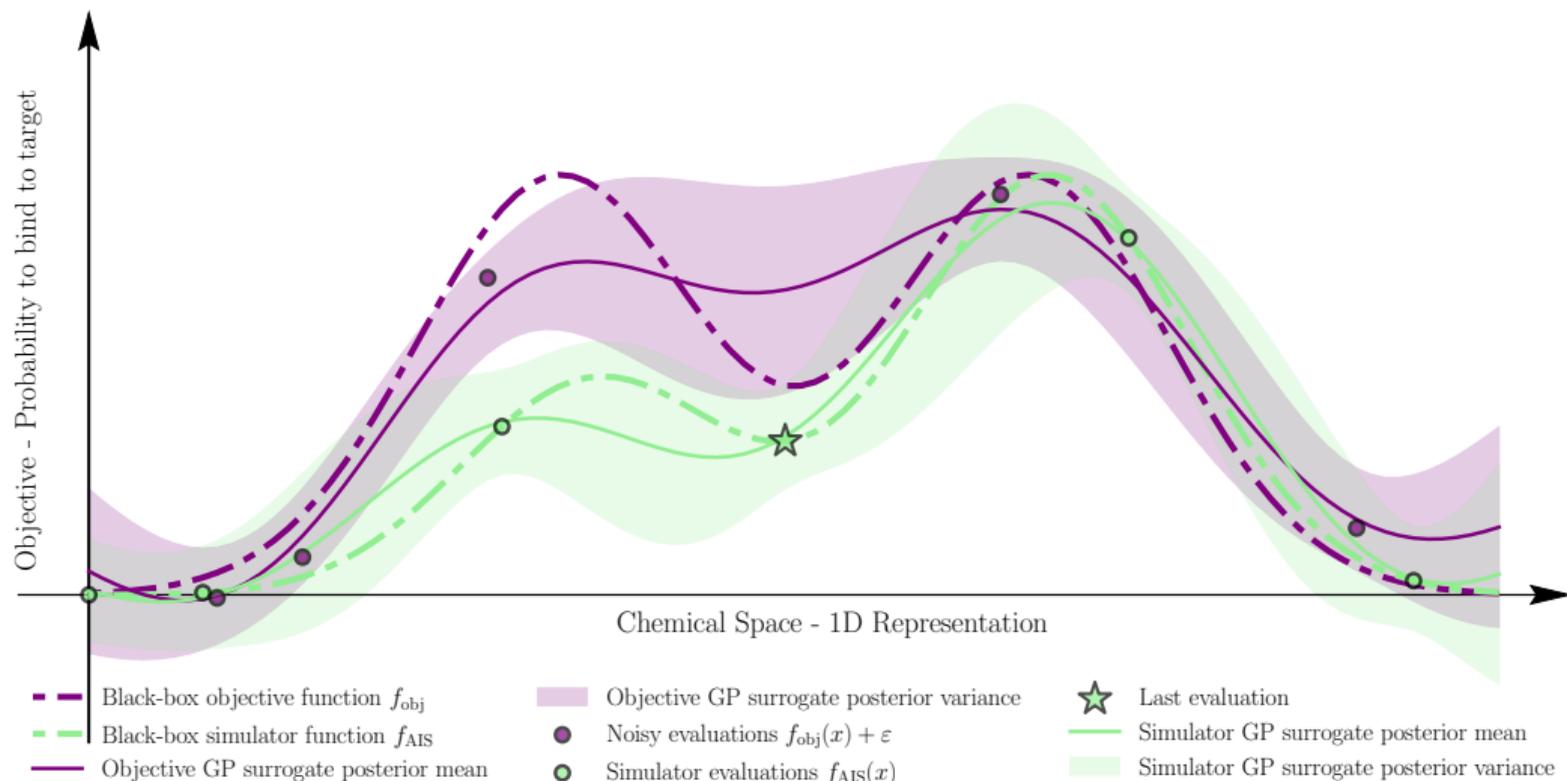
Multi Fidelity Bayesian Optimization 101

Budget = 20



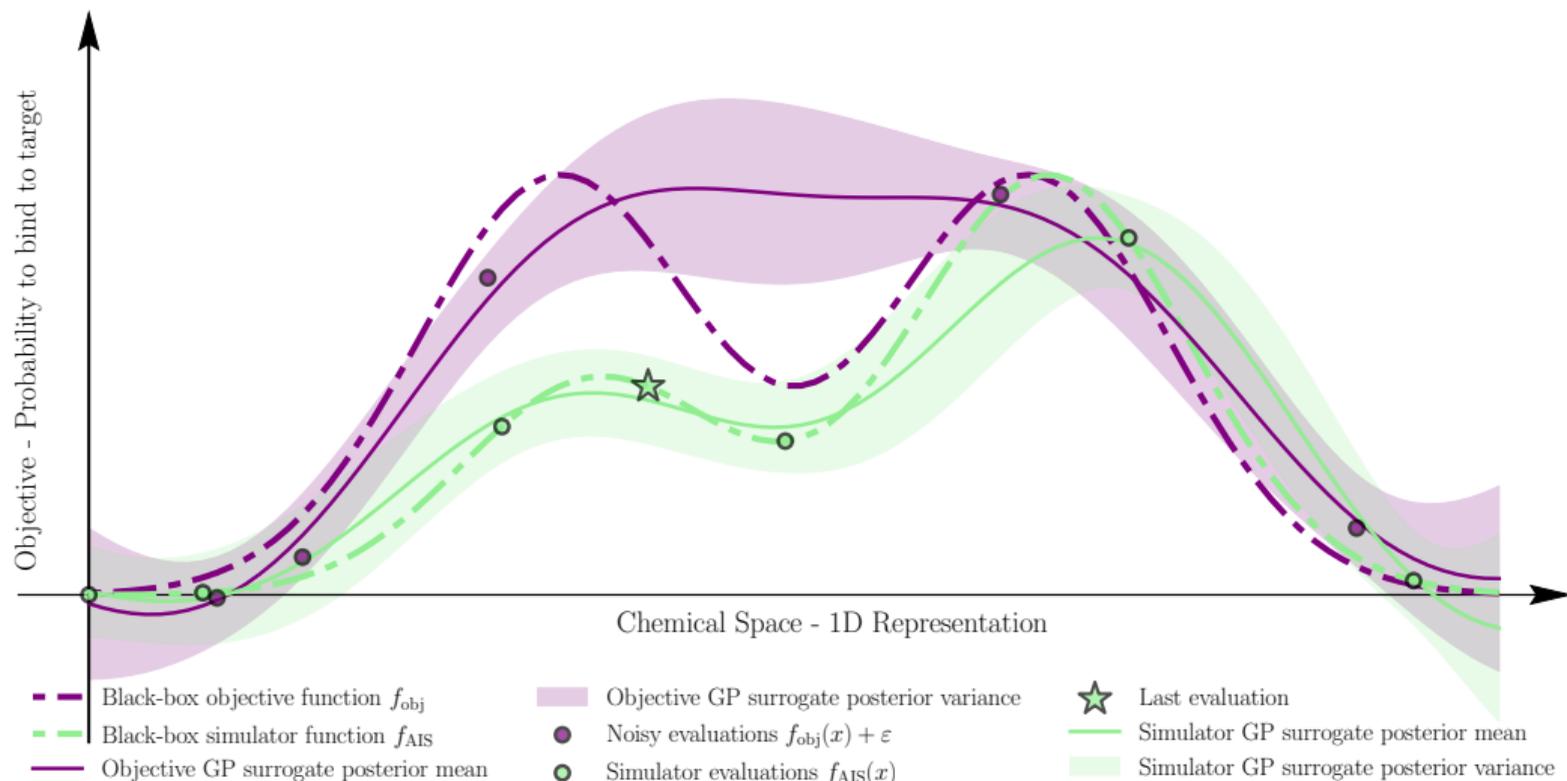
Multi Fidelity Bayesian Optimization 101

Budget = 19.8



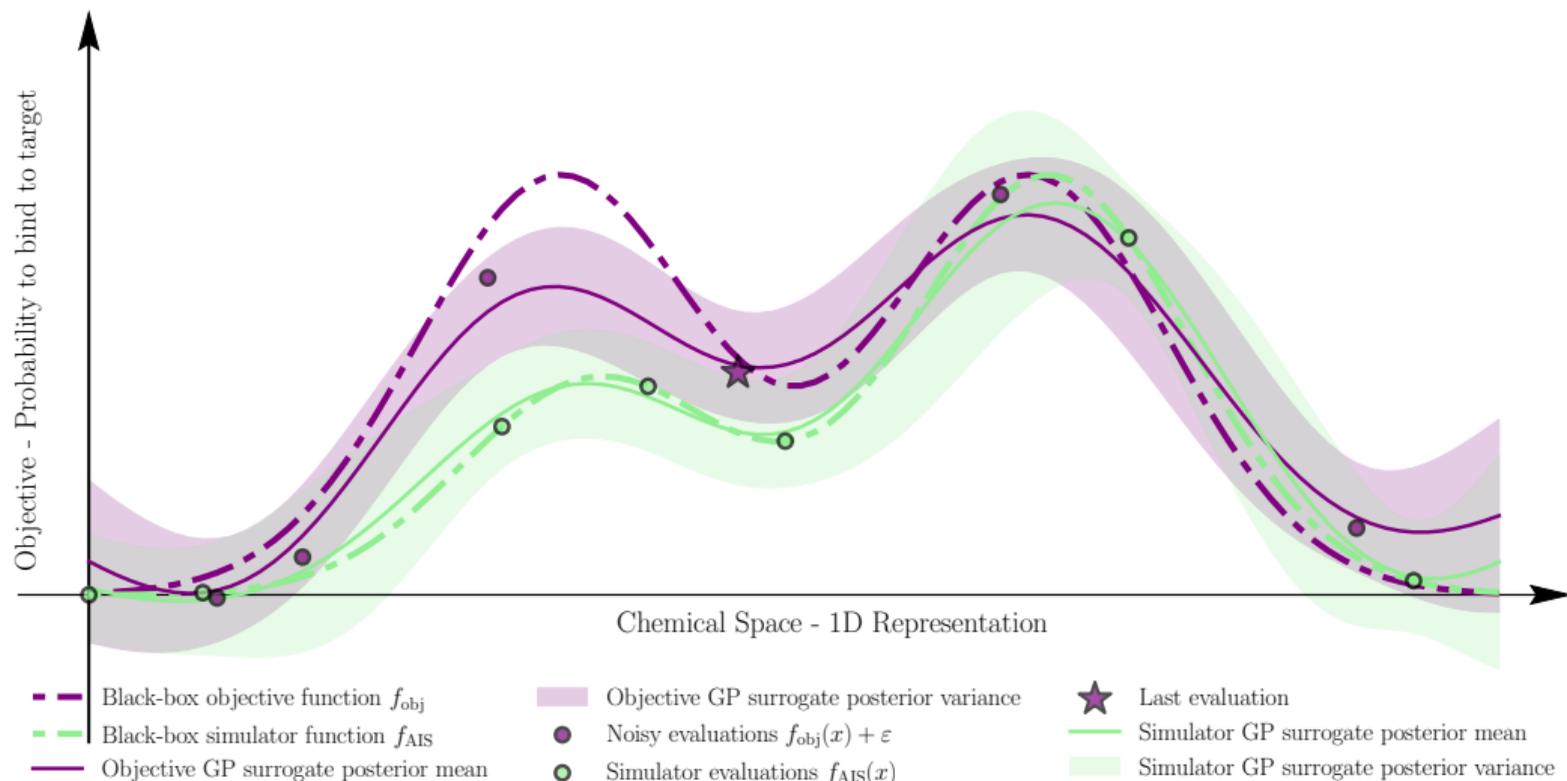
Multi Fidelity Bayesian Optimization 101

Budget = 19.6



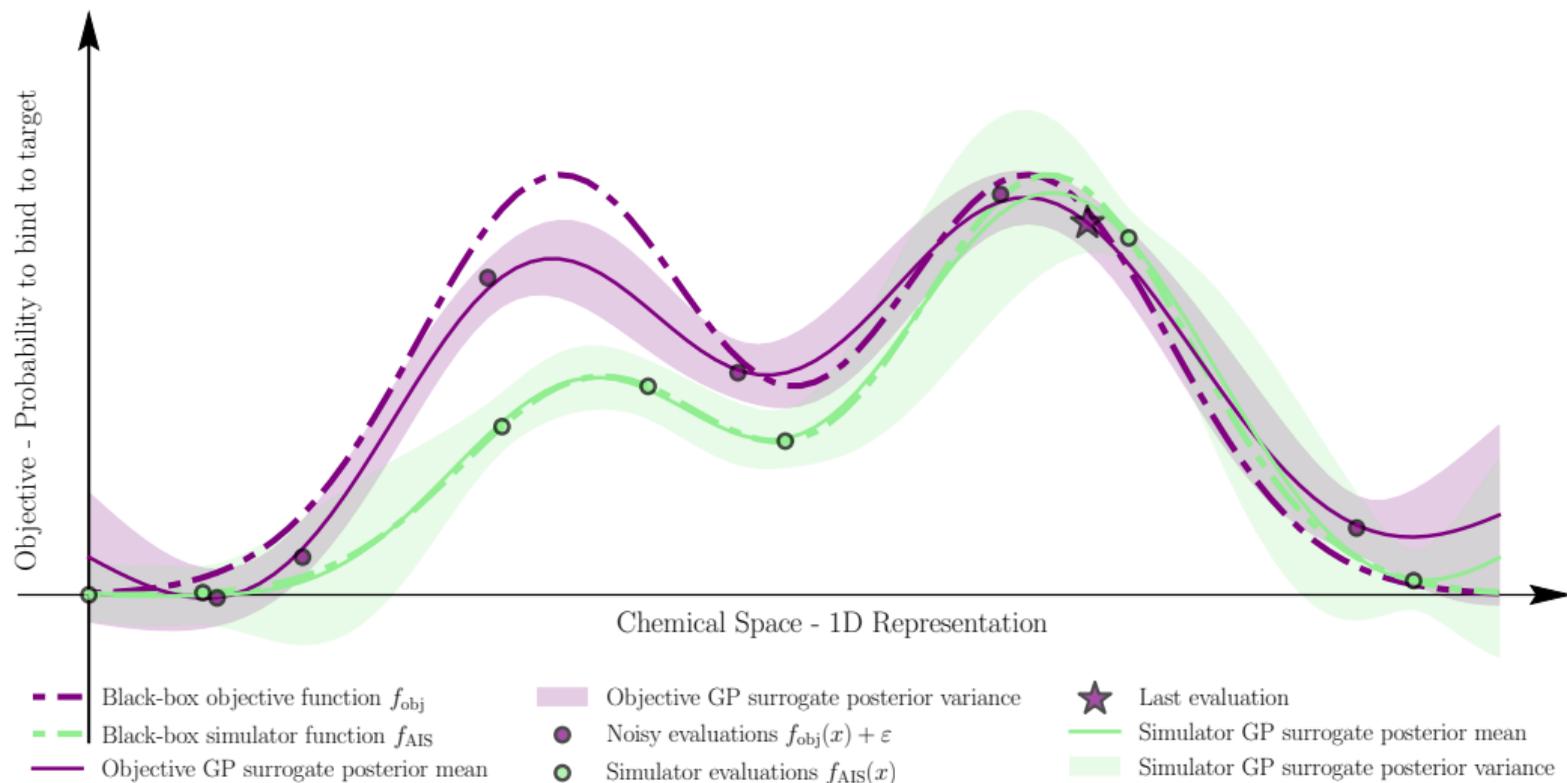
Multi Fidelity Bayesian Optimization 101

Budget = 18.6



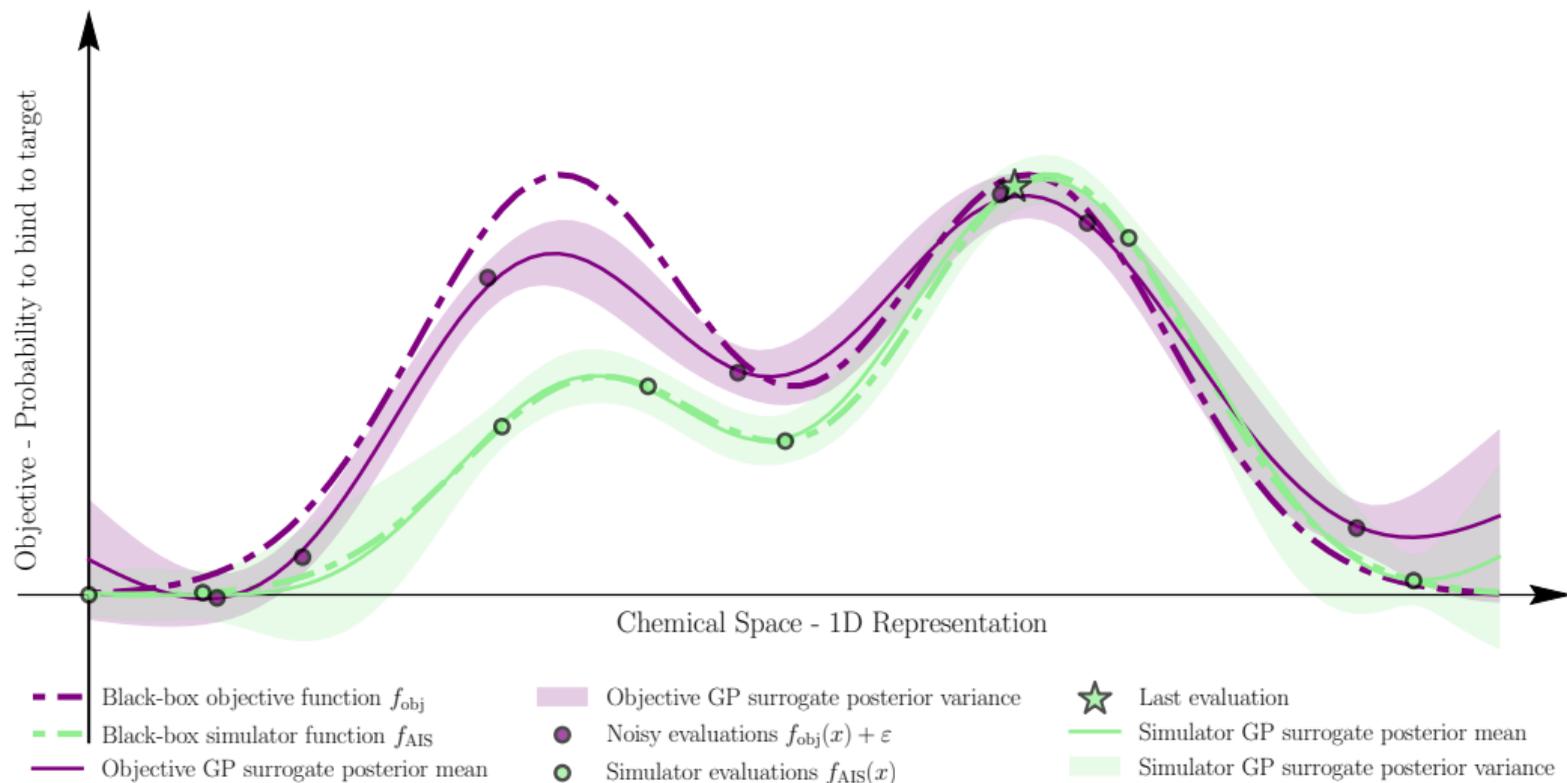
Multi Fidelity Bayesian Optimization 101

Budget = 17.6



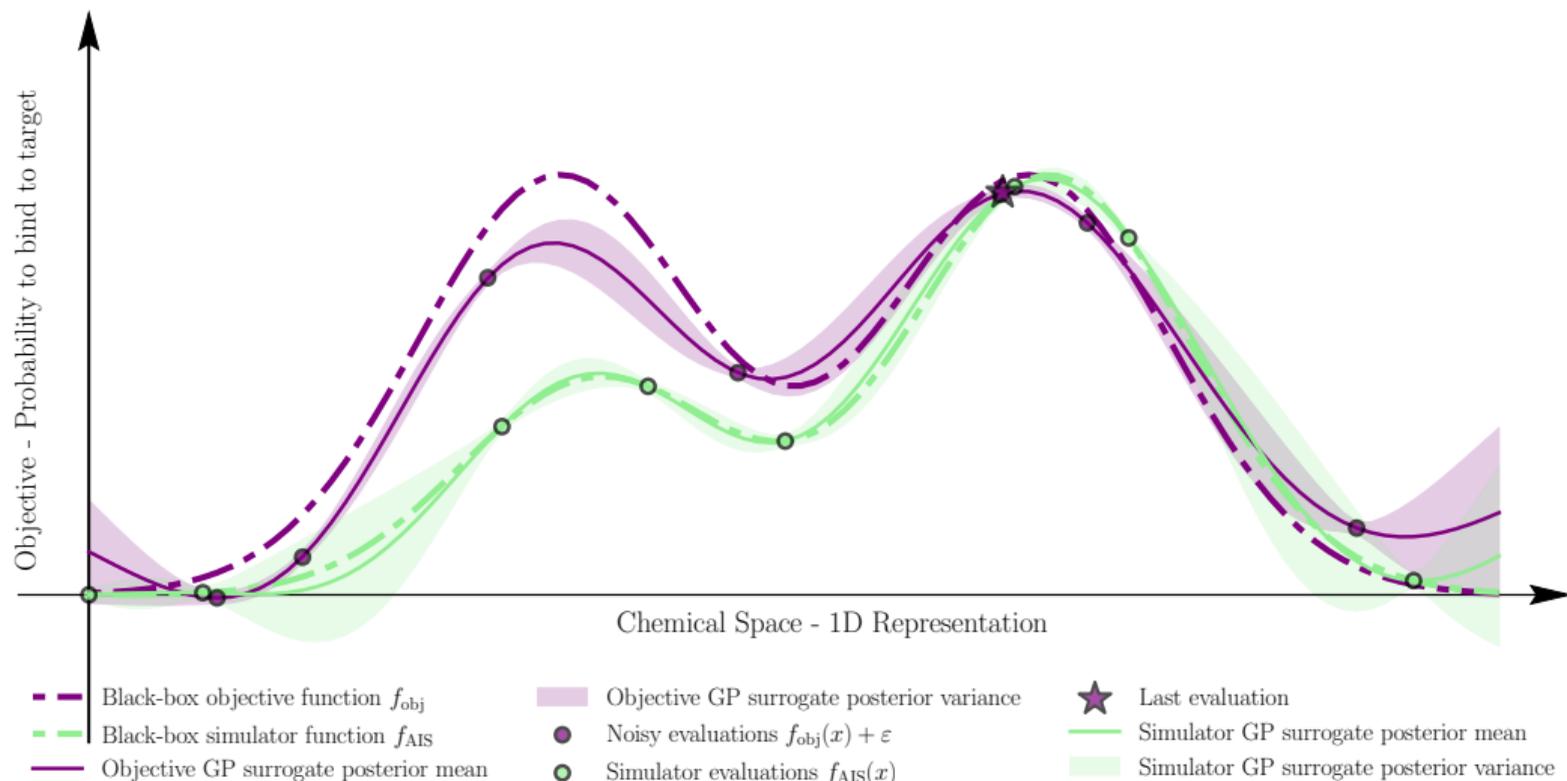
Multi Fidelity Bayesian Optimization 101

Budget = 17.4



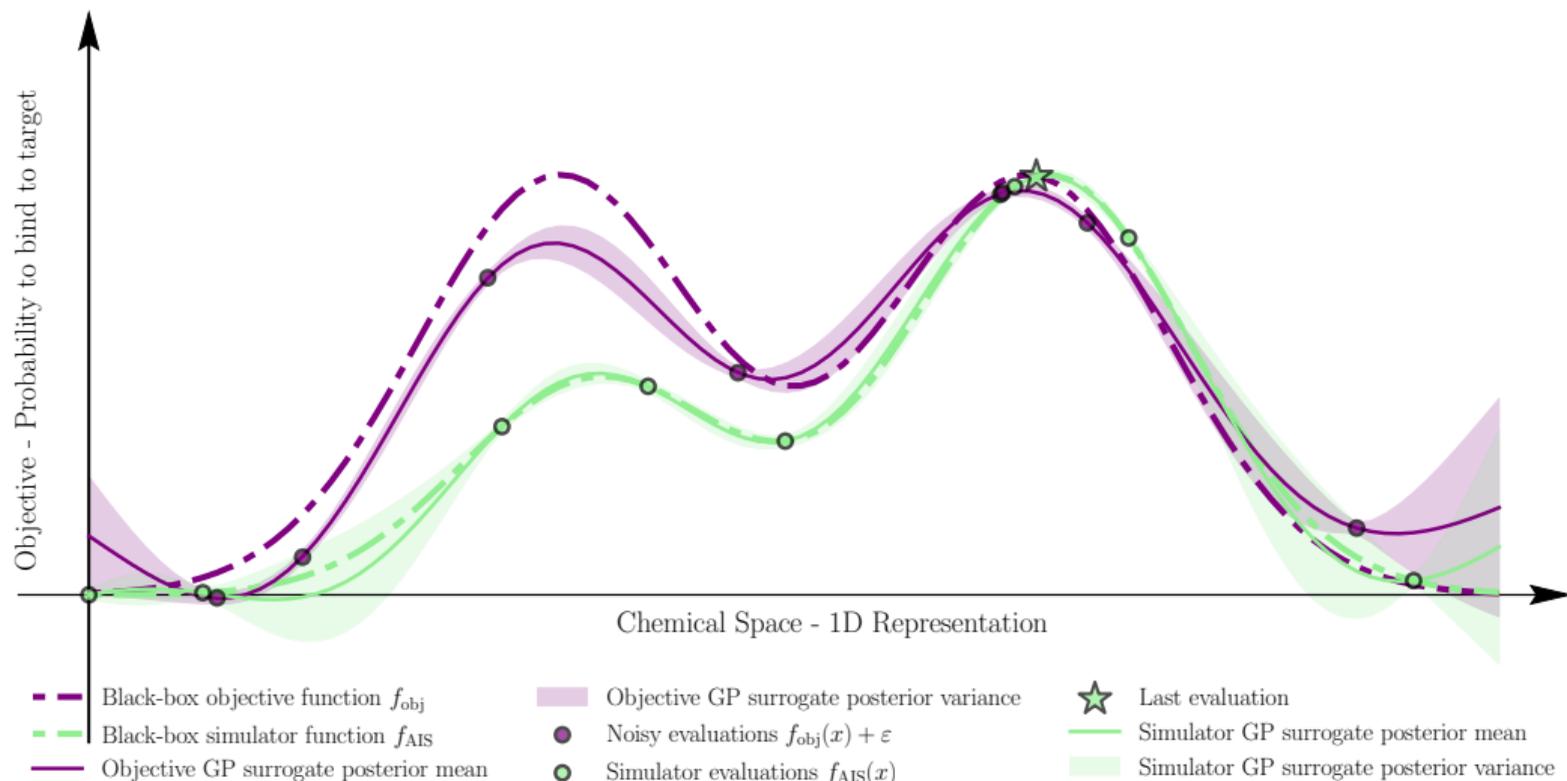
Multi Fidelity Bayesian Optimization 101

Budget = 16.4



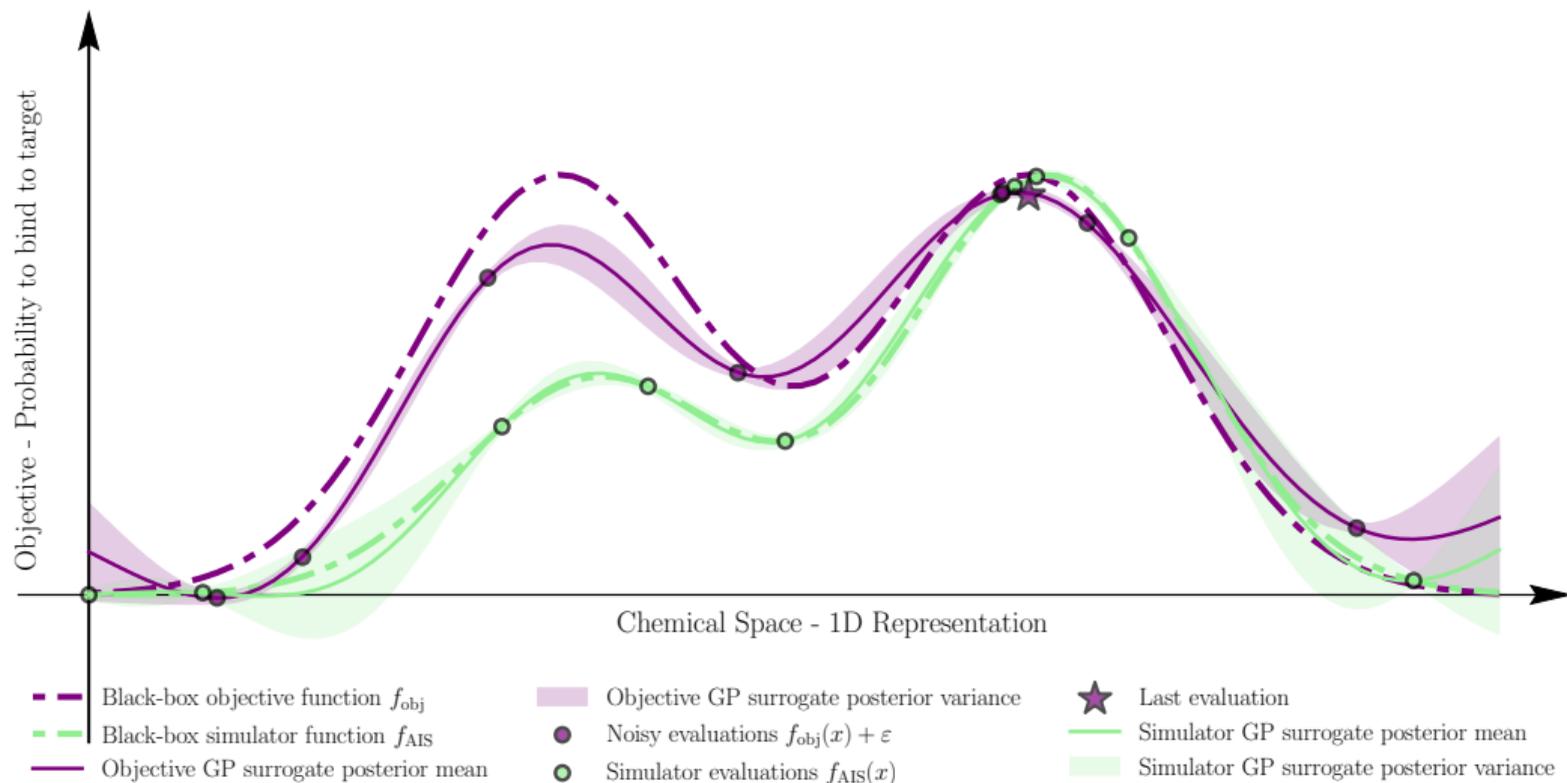
Multi Fidelity Bayesian Optimization 101

Budget = 16.2

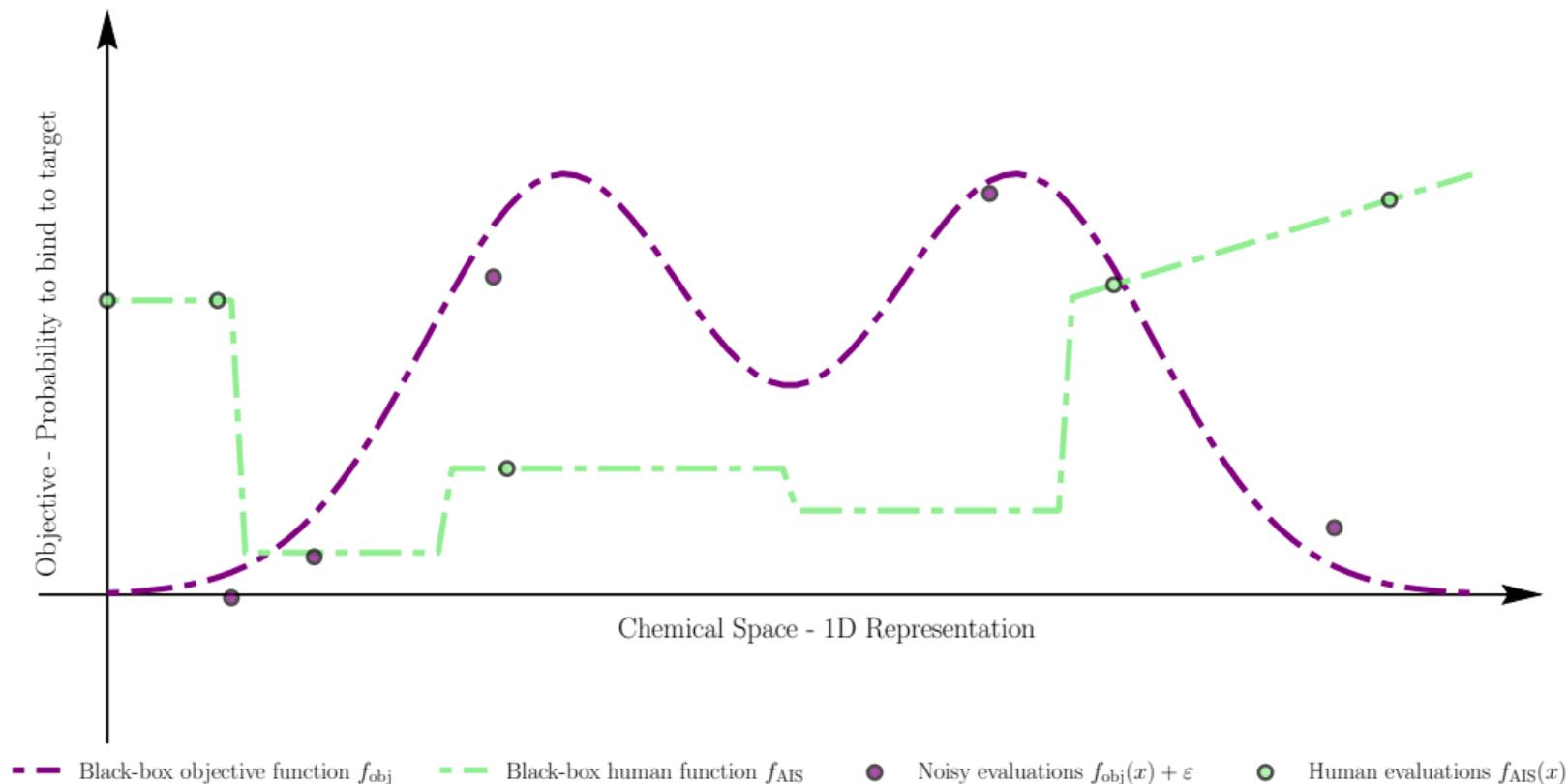


Multi Fidelity Bayesian Optimization 101

Budget = 15.2

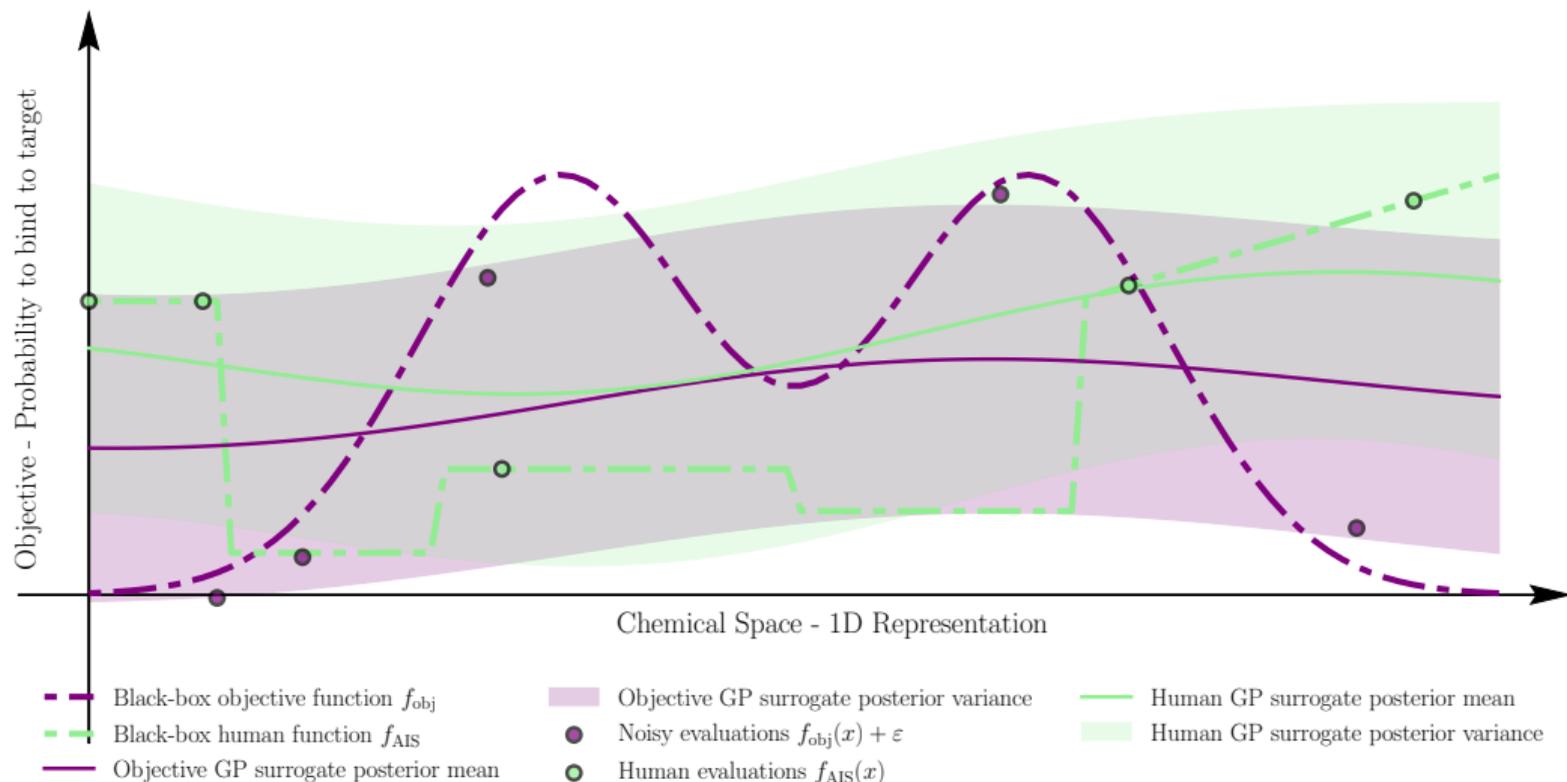


Multi Fidelity Bayesian Optimization with Unreliable Sources



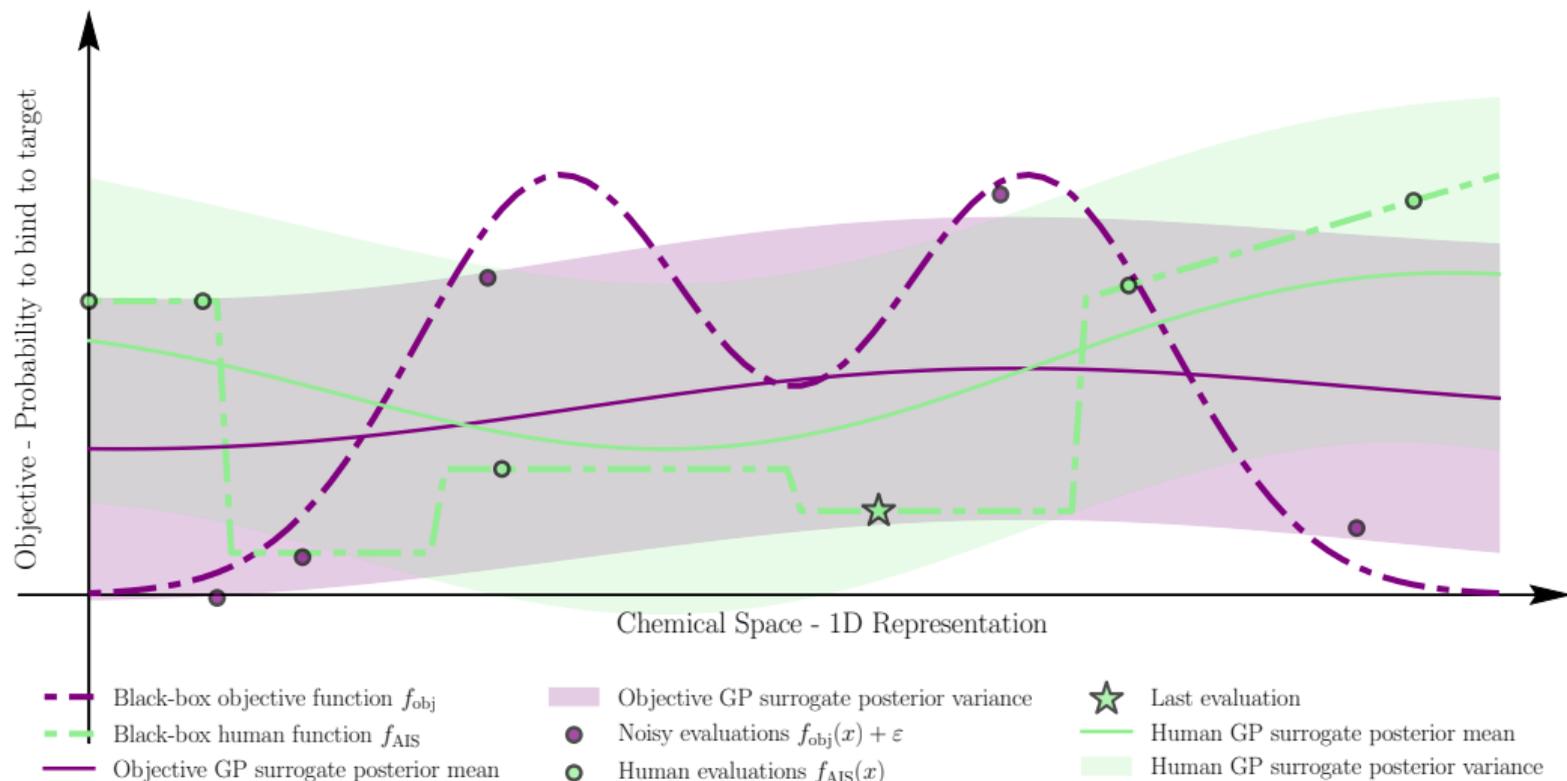
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 20



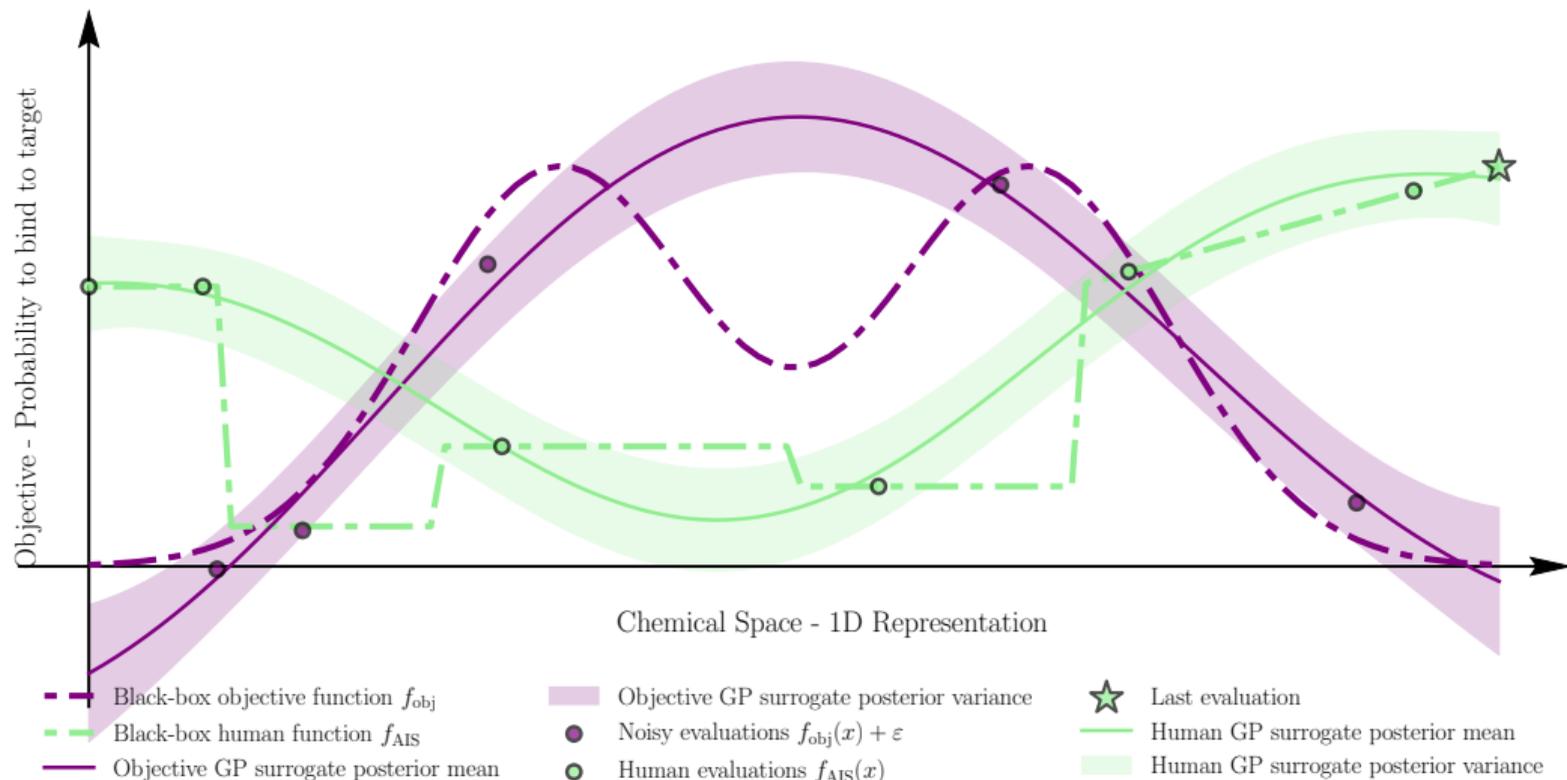
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 19.9



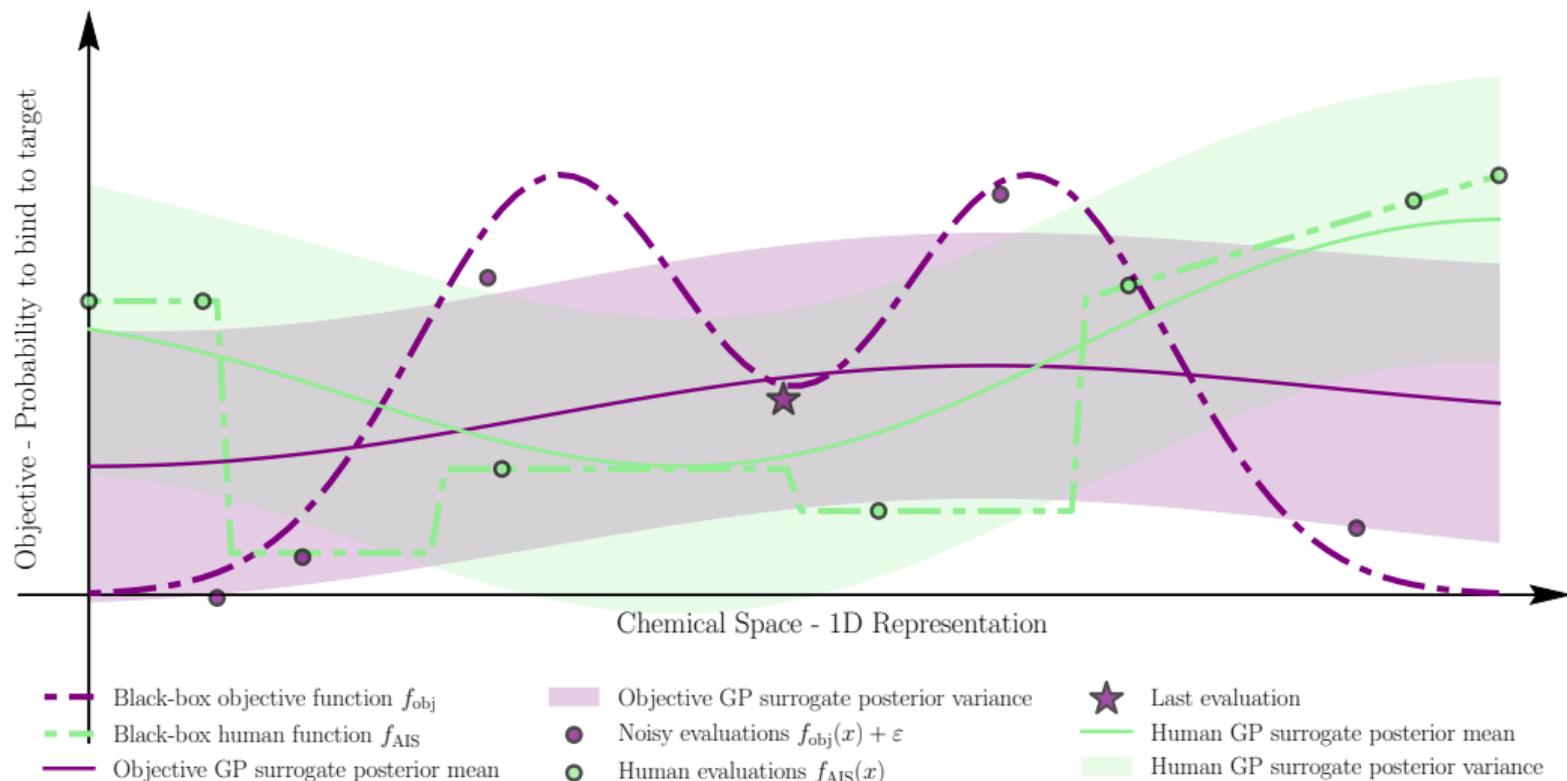
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 19.8



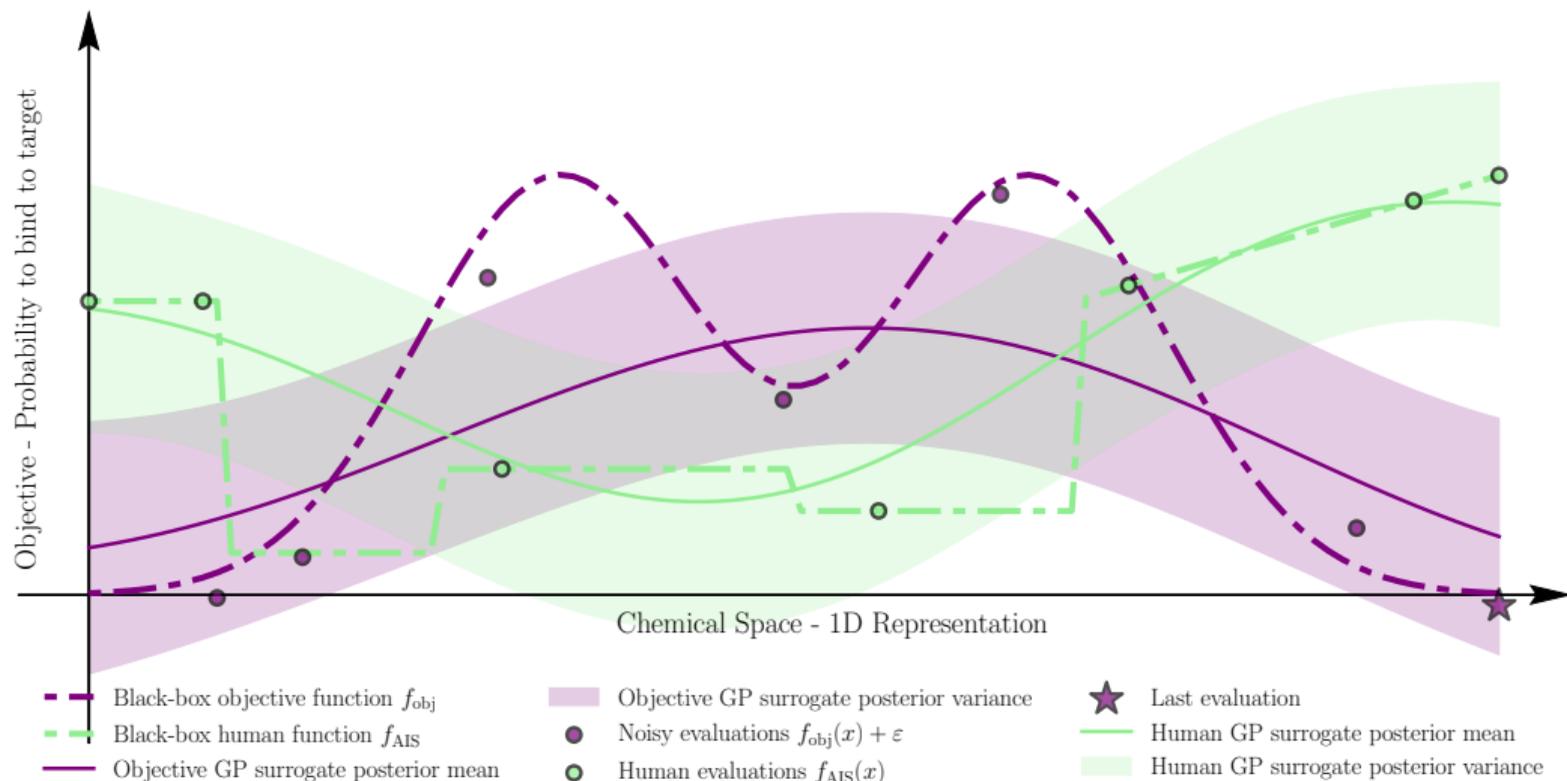
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 18.8



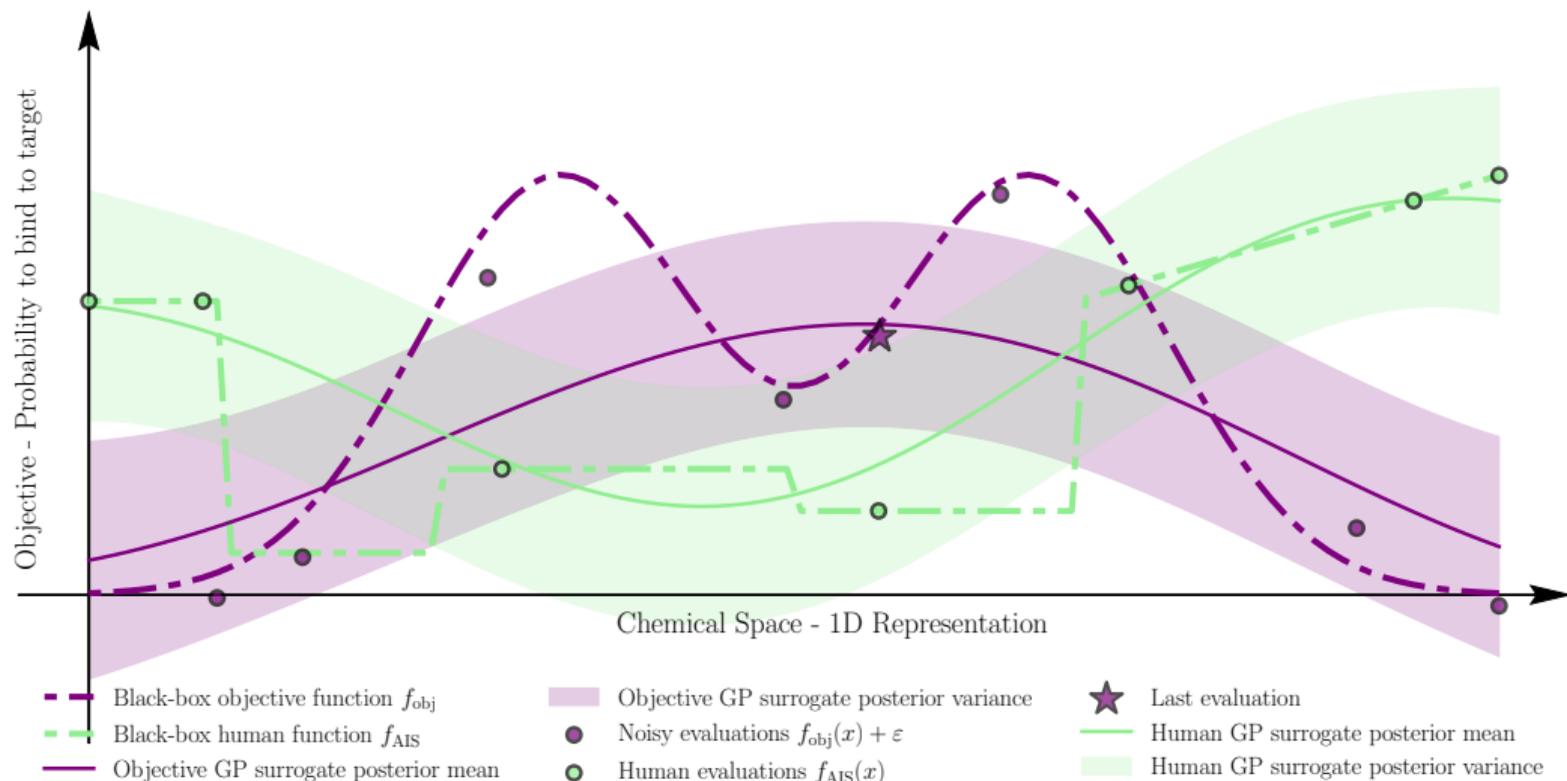
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 17.8



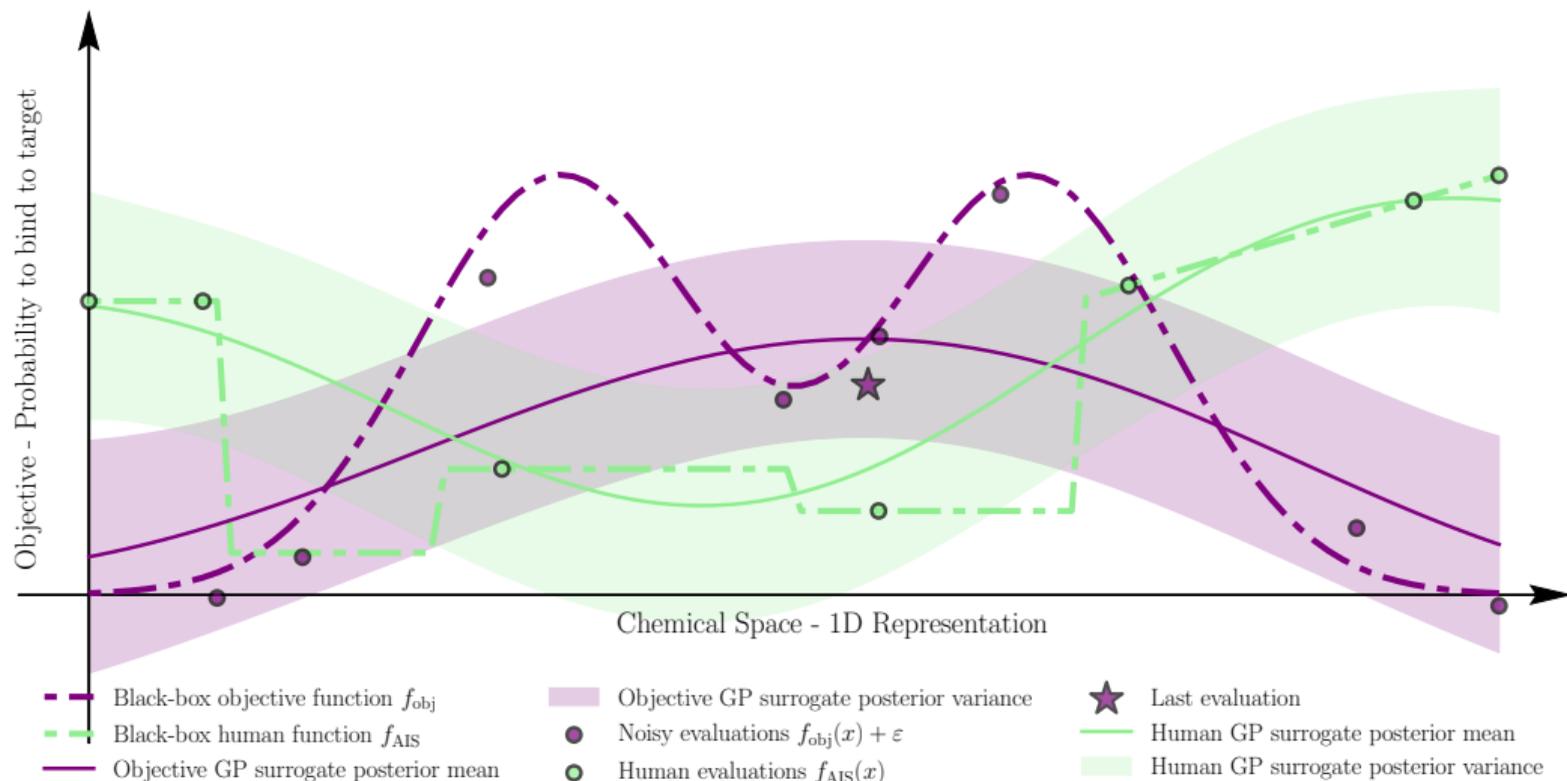
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 16.8



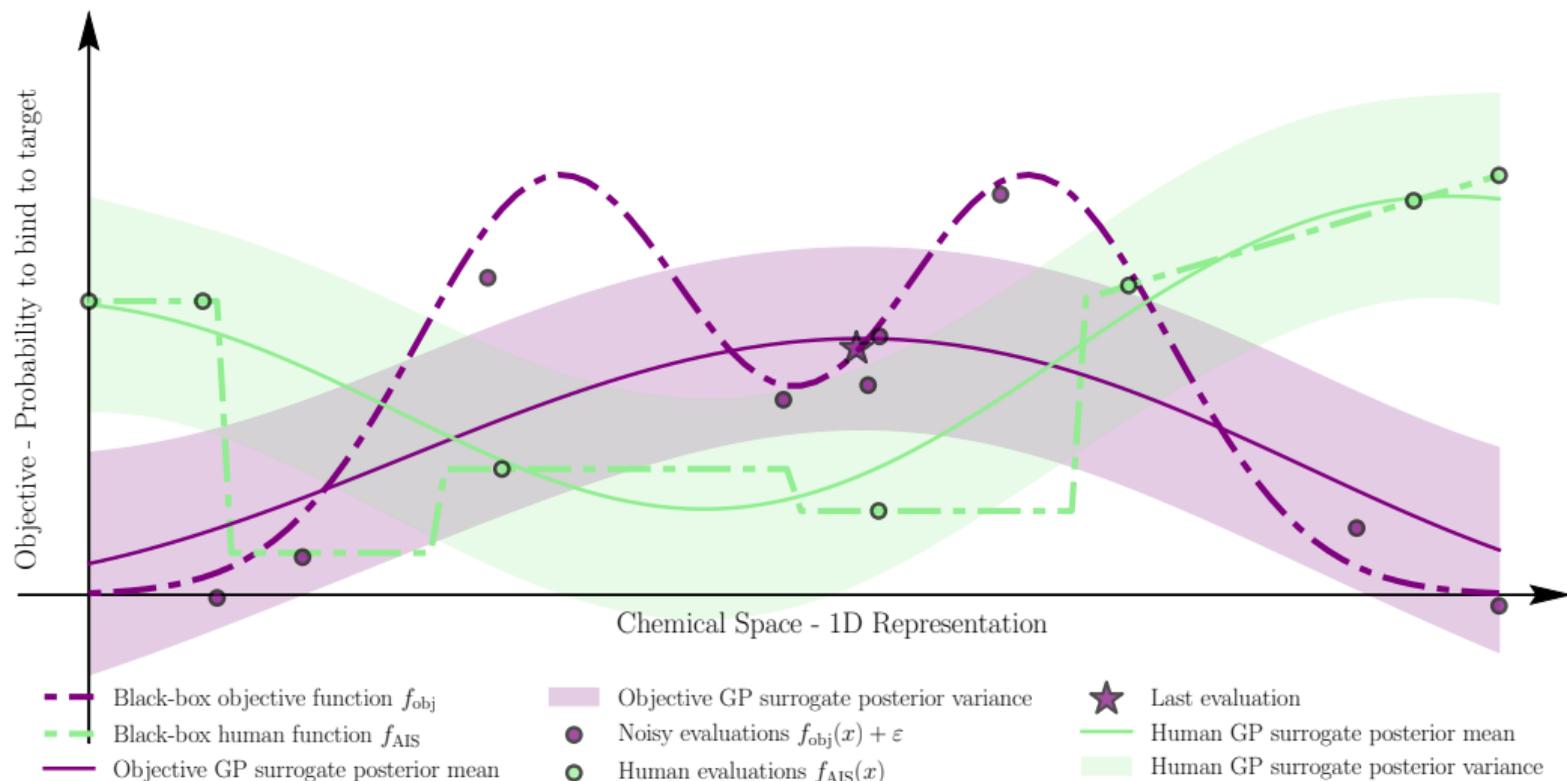
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 15.8



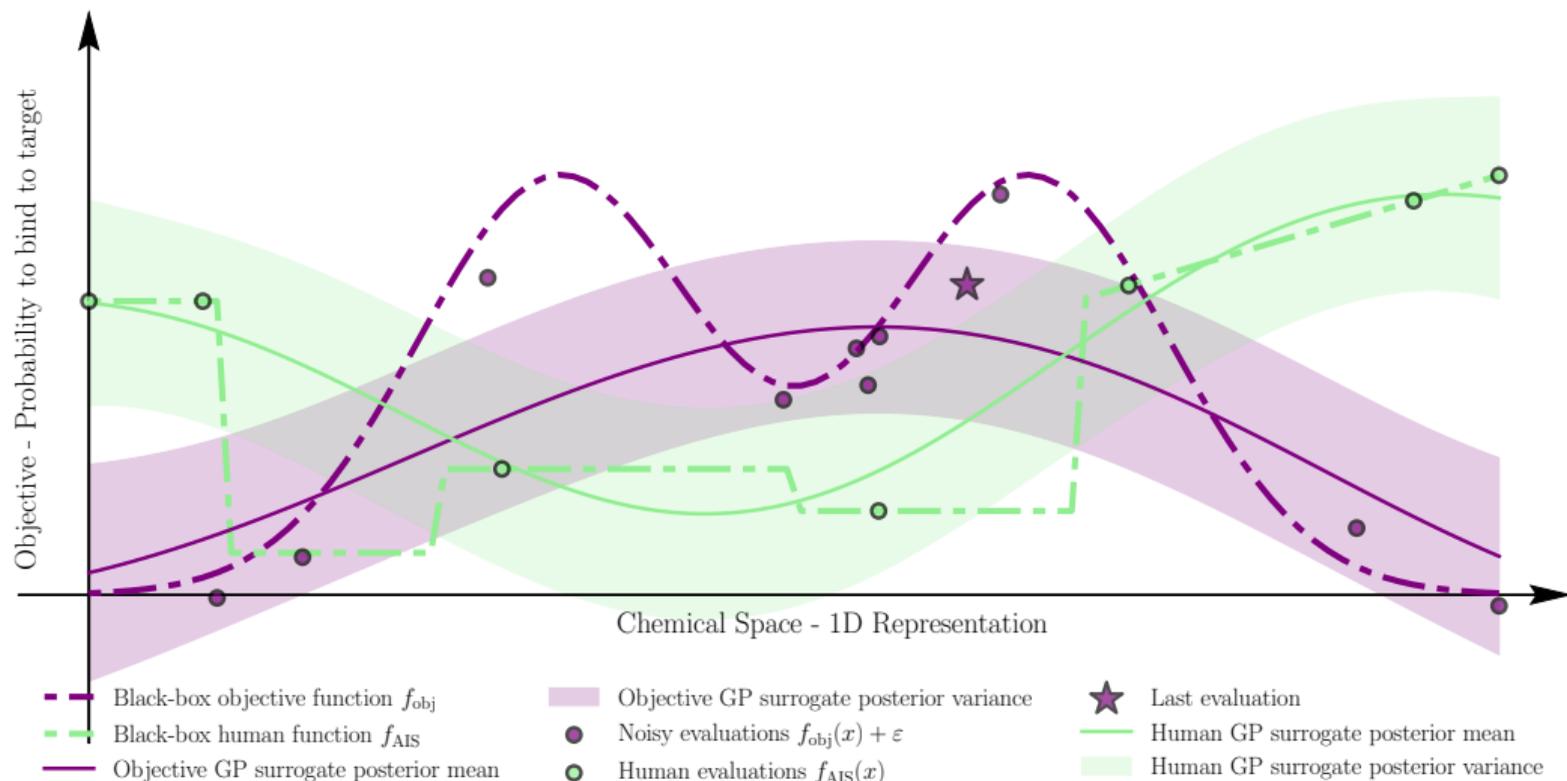
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 14.8



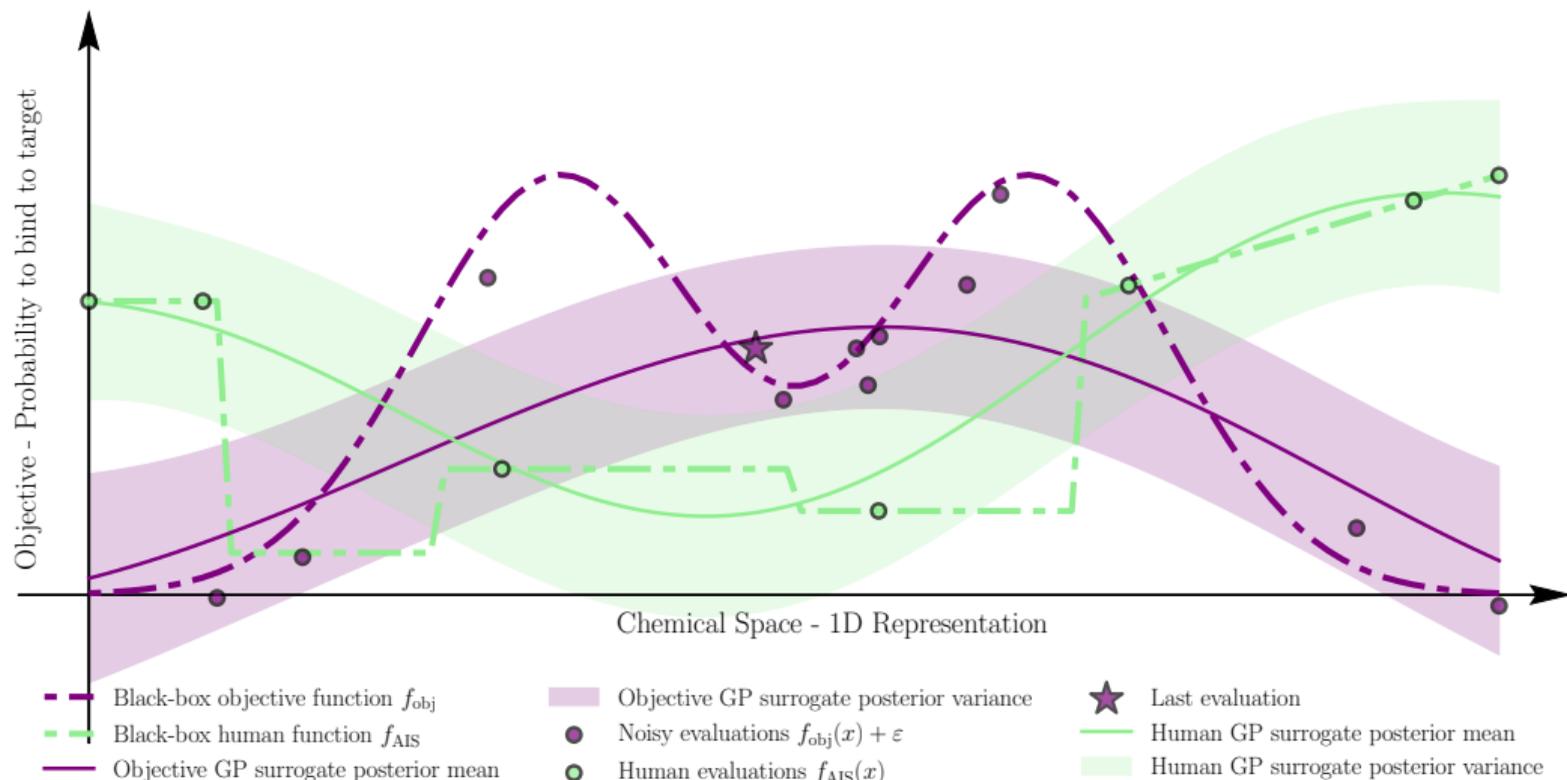
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 13.8



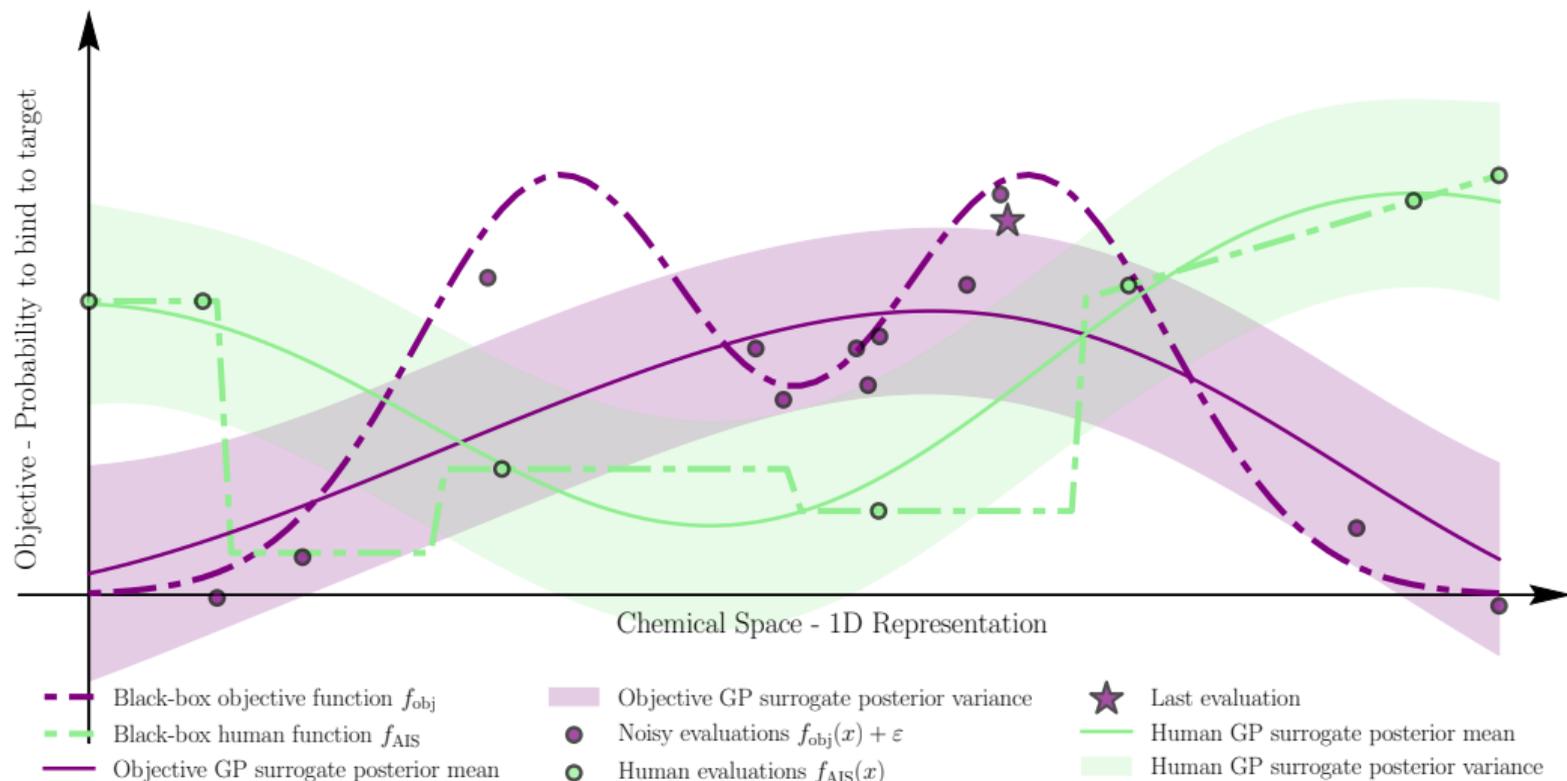
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 12.8



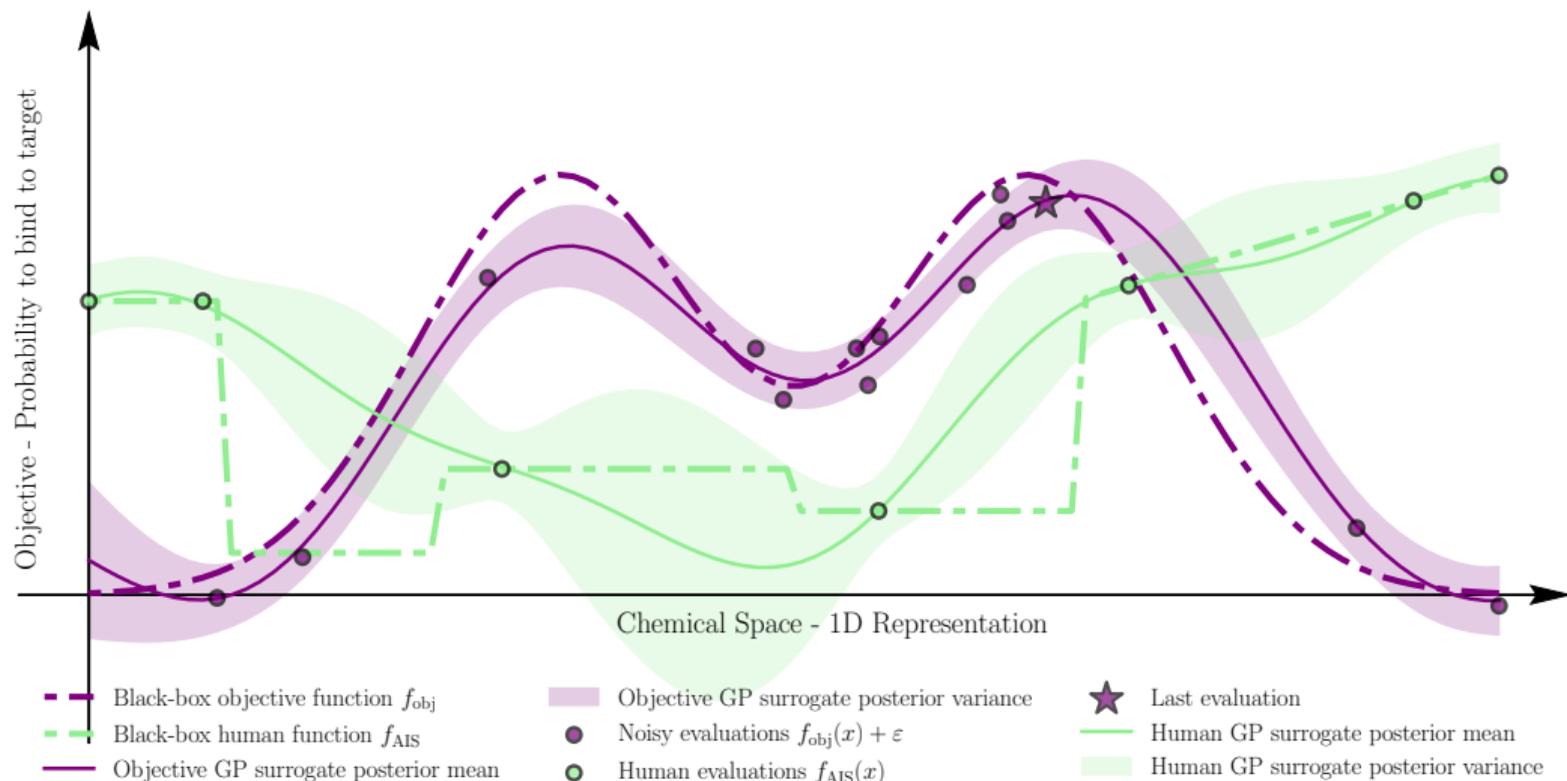
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 11.8



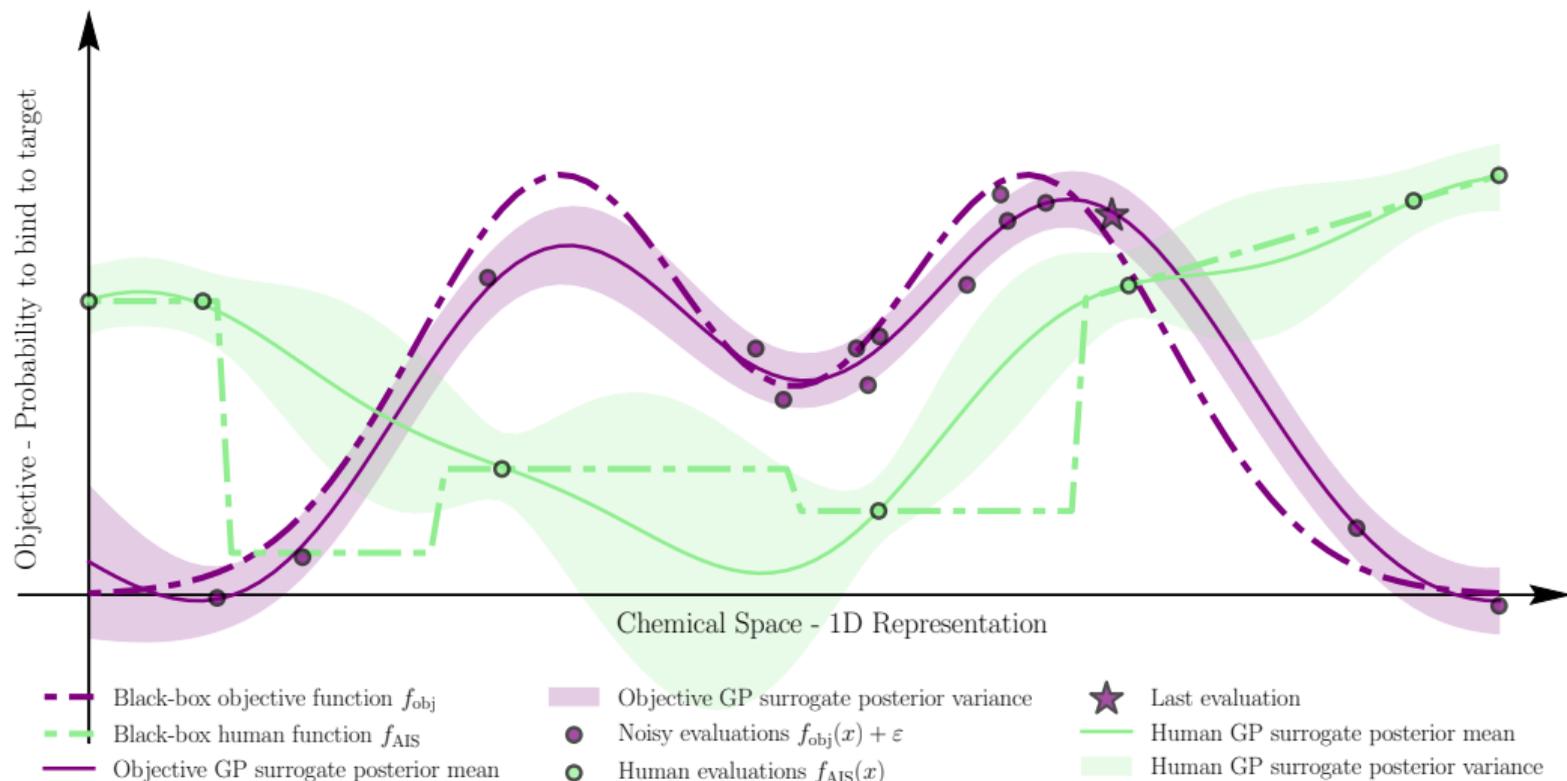
Multi Fidelity Bayesian Optimization with Unreliable Sources

Budget = 10.8

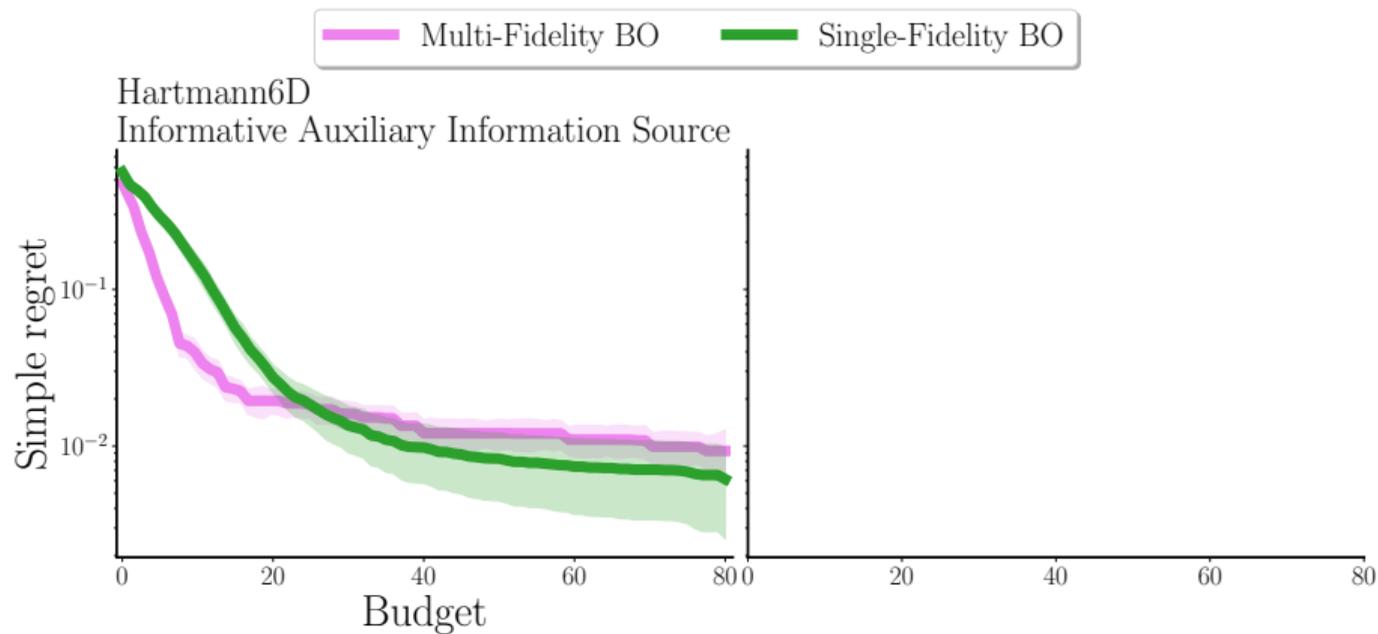


Multi Fidelity Bayesian Optimization with Unreliable Sources

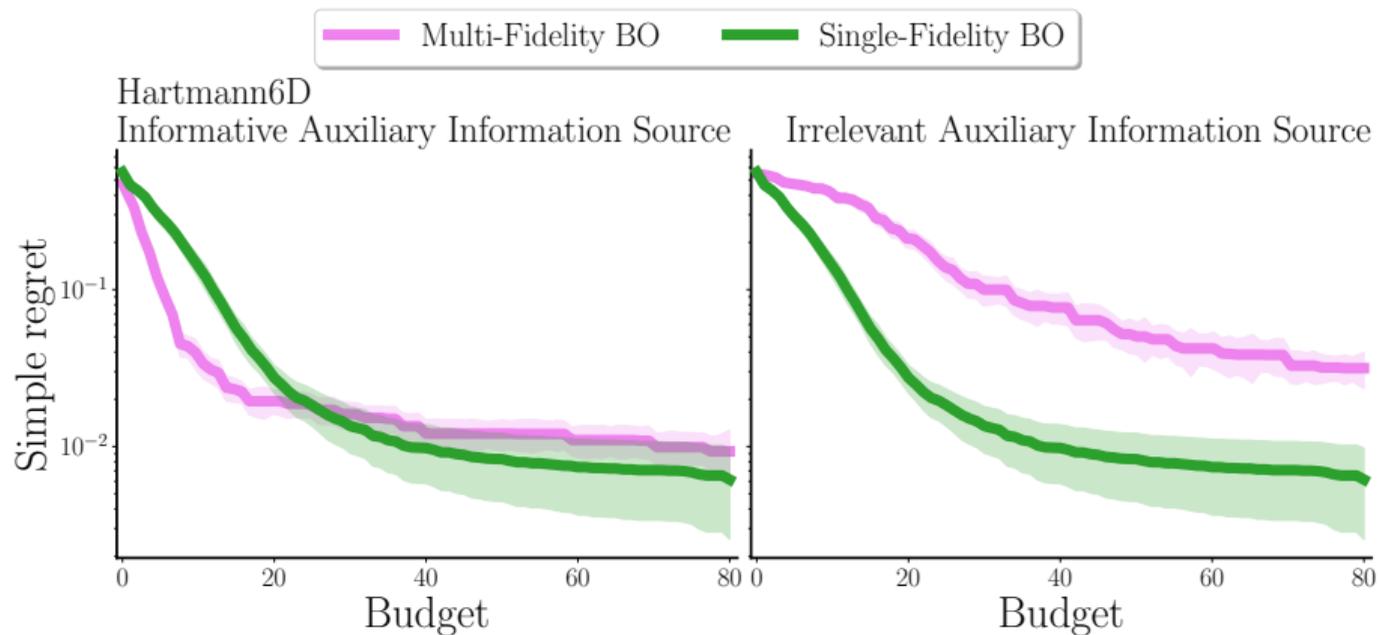
Budget = 9.8



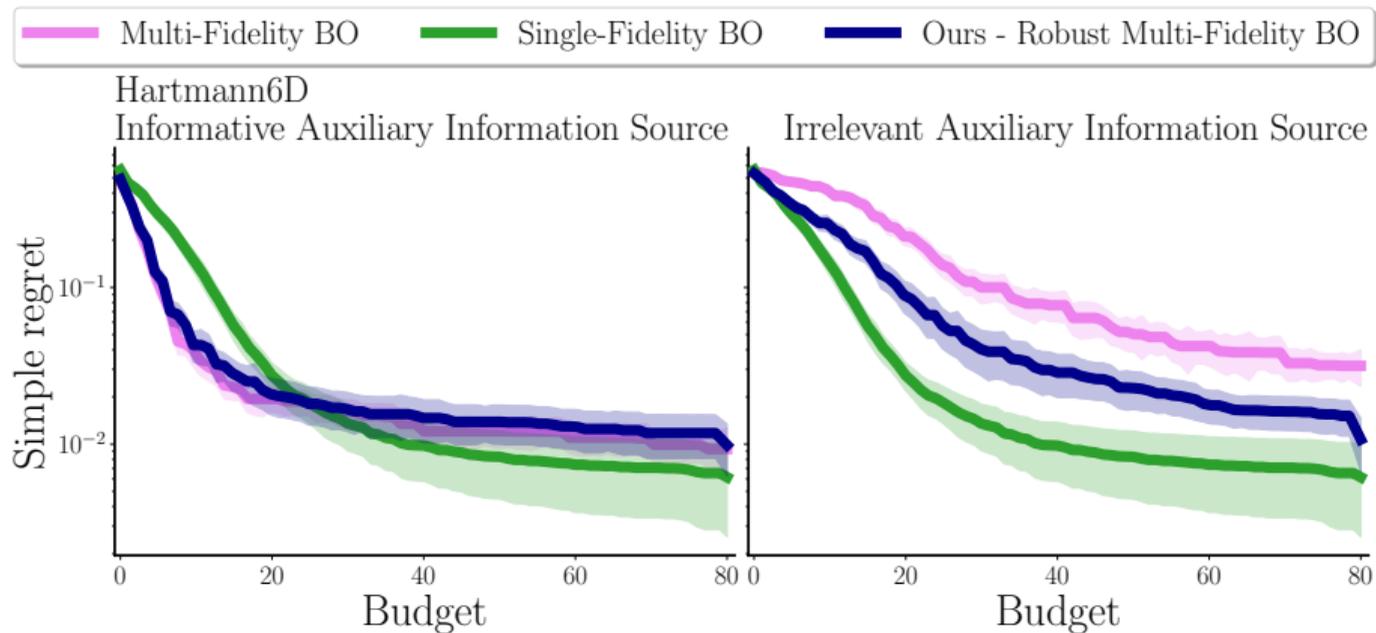
Multi-Fidelity BO is not robust to unreliable Information Sources



Multi-Fidelity BO is not robust to unreliable Information Sources



Multi-Fidelity BO is not robust to unreliable Information Sources



Not far from a well-known problem in transfer learning: **negative transfer**

- Main aim of our contribution: **robustness** to irrelevant AIS...
- ...While still accelerating convergence for relevant AIS

Introducing robust MFBO (rMFBO)

- In a nutshell, carry MFBO ... **BUT** ...
- ...Keep track of what would have looked like the acquisition trajectory without AIS using a single-output GP \implies **HOW?**

At iteration t , choose between two queries:

$$(x_t^{\text{MF}}, \ell_t) = \operatorname{argmax}_{x \in \mathcal{X}, \ell \in \{\text{obj}, \text{AIS}\}} \alpha(x, \ell | \mu_{\text{MF}}, \sigma_{\text{MF}}, \mathcal{D}^{\text{MF}}) / \lambda_{t_t}$$

$$(x_t^{\text{pSF}}, \text{obj}) = \operatorname{argmax}_{x \in \mathcal{X}} \alpha(x | \mu_{\text{SF}}, \sigma_{\text{SF}}, \mathcal{D}^{\text{pSF}})$$

If $\sigma_{\text{MF}}(x_t^{\text{pSF}}, \text{obj}) \leq c_1 \implies$ joint model reliable at x_t^{pSF} , i.e. $\mu_{\text{MF}}^{\text{obj}}(x_t^{\text{pSF}}) \approx f^{\text{obj}}(x_t^{\text{pSF}})$

Therefore $\mathcal{D}^{\text{pSF}} \leftarrow (x_t^{\text{pSF}}, \mu_{\text{MF}}^{\text{obj}}(x_t^{\text{pSF}}))$: creating a *pseudo* single fidelity track

Summary

$$\sigma_{\text{MF}}(x_t^{\text{pSF}}, \text{obj}) \leq c_1 \implies \begin{cases} \text{pick } (x_t^{\text{MF}}, \ell_t) \\ \mathcal{D}^{\text{MF}} \leftarrow (x_t^{\text{MF}}, f^{(\ell_t)}(x_t^{\text{MF}})) \\ \mathcal{D}^{\text{pSF}} \leftarrow (x_t^{\text{pSF}}, \mu_{\text{MF}}^{\text{obj}}(x_t^{\text{pSF}})) \end{cases}$$

Summary

$$\sigma_{\text{MF}}(x_t^{\text{pSF}}, \text{obj}) \leq c_1 \implies \begin{cases} \text{pick } (x_t^{\text{MF}}, \ell_t) \\ \mathcal{D}^{\text{MF}} \leftarrow (x_t^{\text{MF}}, f^{(\ell_t)}(x_t^{\text{MF}})) \\ \mathcal{D}^{\text{pSF}} \leftarrow (x_t^{\text{pSF}}, \mu_{\text{MF}}^{\text{obj}}(x_t^{\text{pSF}})) \end{cases}$$

because x_t^{pSF} only brings little information

Summary

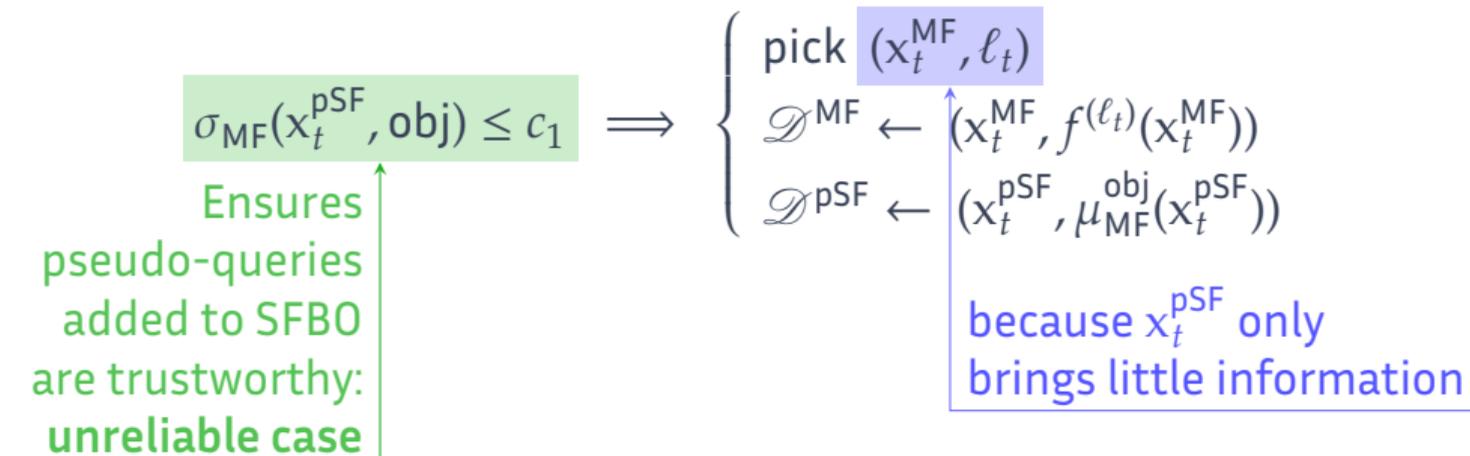
$$\sigma_{\text{MF}}(x_t^{\text{pSF}}, \text{obj}) \leq c_1$$

Ensures
pseudo-queries
added to SFBO
are trustworthy:
unreliable case

$$\implies \begin{cases} \text{pick } (x_t^{\text{MF}}, \ell_t) \\ \mathcal{D}^{\text{MF}} \leftarrow (x_t^{\text{MF}}, f^{(\ell_t)}(x_t^{\text{MF}})) \\ \mathcal{D}^{\text{pSF}} \leftarrow (x_t^{\text{pSF}}, \mu_{\text{MF}}^{\text{obj}}(x_t^{\text{pSF}})) \end{cases}$$

because x_t^{pSF} only
brings little information

Summary



If not satisfied:

- 1 Pick (x_t^{pSF}, obj)
- 2 $\mathcal{D}^{MF} \leftarrow (x_t^{pSF}, f^{obj}(x_t^{pSF}))$
- 3 $\mathcal{D}^{pSF} \leftarrow (x_t^{pSF}, f^{obj}(x_t^{pSF}))$

\mathcal{D}^{pSF} and \mathcal{D}^{SF} only differ at the points where we inputted $\mu_{MF}^{obj}(x_t^{pSF})$!

rMFBO regret can be tied to that of SFBO

Assumptions:

- f^{obj} is drawn from a GP with zero-mean and covariance function $\kappa(x, x')$
- κ is known and twice differentiable
- $\mathbb{P}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial f^{\text{obj}}}{\partial x_j} \right| > L\right) \leq ae^{-(L/b_j)^2} \quad \forall j \in \{1, \dots, d\}, \text{ for } a, b_j > 0$

rMFBO regret can be tied to that of SFBO

Assumptions:

- f^{obj} is drawn from a GP with zero-mean and covariance function $\kappa(x, x')$
- κ is known and twice differentiable
- $\mathbb{P}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial f^{\text{obj}}}{\partial x_j} \right| > L\right) \leq ae^{-(L/b_j)^2} \quad \forall j \in \{1, \dots, d\}, \text{ for } a, b_j > 0$
- The mapping $(x, \mathcal{D}) \mapsto \alpha(x|p(f|\mathcal{D}))$ is twice differentiable

rMFBO regret can be tied to that of SFBO

Assumptions:

- f^{obj} is drawn from a GP with zero-mean and covariance function $\kappa(x, x')$
- κ is known and twice differentiable
- $\mathbb{P}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial f^{\text{obj}}}{\partial x_j} \right| > L\right) \leq ae^{-(L/b_j)^2} \quad \forall j \in \{1, \dots, d\}, \text{ for } a, b_j > 0$
- The mapping $(x, \mathcal{D}) \mapsto \alpha(x|p(f|\mathcal{D}))$ is twice differentiable

Theorem:

for any AIS, the difference in regrets achieved by SFBO and rMFBO can be bounded.

rMFBO regret can be tied to that of SFBO

Assumptions:

- f^{obj} is drawn from a GP with zero-mean and covariance function $\kappa(x, x')$
- κ is known and twice differentiable
- $\mathbb{P}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial f^{\text{obj}}}{\partial x_j} \right| > L\right) \leq ae^{-(L/b_j)^2} \quad \forall j \in \{1, \dots, d\}, \text{ for } a, b_j > 0$
- The mapping $(x, \mathcal{D}) \mapsto \alpha(x|p(f|\mathcal{D}))$ is twice differentiable

Theorem:

for any AIS, the difference in regrets achieved by SFBO and rMFBO can be bounded.

No assumption on the amount of information that f^{AIS} can provide about f^{obj} !

rMFBO regret can be tied to that of SFBO

Assumptions:

- f^{obj} is drawn from a GP with zero-mean and covariance function $\kappa(x, x')$
- κ is known and twice differentiable
- $\mathbb{P}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial f^{\text{obj}}}{\partial x_j} \right| > L\right) \leq ae^{-(L/b_j)^2} \quad \forall j \in \{1, \dots, d\}, \text{ for } a, b_j > 0$
- The mapping $(x, \mathcal{D}) \mapsto \alpha(x|p(f|\mathcal{D}))$ is twice differentiable

Theorem:

for any AIS, the difference in regrets achieved by SFBO and rMFBO can be bounded.

No assumption on the amount of information that f^{AIS} can provide about f^{obj} !

$$R(\Lambda, x_T^{\text{rMF}}) \leq R(\Lambda, x_T^{\text{SF}}) + \varepsilon \max\{T\hat{M}_T d^{T+1}, 2\} \text{ with probability } \geq q \left(1 - da \exp\left(-\frac{1}{b^2}\right)\right)$$

$$c_1(\varepsilon, q) = \frac{\varepsilon}{\sqrt{-2 \log(1-q)}}.$$

Interested? Have a look at the paper 

Petrus Mikkola, **Julien Martinelli**, Louis Filstroff, Samuel Kaski
Multi-Fidelity Bayesian Optimization with Unreliable Sources. AISTATS 2023.

Some stuff (continued) $\backslash_(\u2197)_/$

$$k_{\text{MISO}}((x, \ell), (x', \ell')) = \begin{cases} k_{\text{input}}(x, x') + k_{\ell}(x, x') & \ell = \ell' \neq 1 \\ k_{\text{input}}(x, x') & \text{otherwise} \end{cases}$$

$$k_{\text{LT}}((x, \ell), (x', \ell')) = \begin{cases} k_{\text{input}}(x, x') + (1 - \ell)(1 - \ell')k_{\text{IS}}(x, x') & \ell \neq 1, \ell' \neq 1 \\ k_{\text{input}}(x, x') & \text{otherwise} \end{cases}$$

$$k_{\text{DS}}((x, \ell), (x', \ell')) = \begin{cases} ck_{\text{input}}(x, x') + (1 - \ell)(1 - \ell')k_{\text{input}}(x, x') & \ell \neq 1, \ell' \neq 1 \\ ck_{\text{input}}(x, x') & \text{otherwise} \end{cases}$$