Best Practices for Multi-Fidelity Bayesian Optimization in Materials and Molecular Research

Víctor Sabanza-Gil
EPFL
Switzerland

Daniel Pacheco Gutiérrez
Atinary Technologies Inc.
Switzerland

Jeremy S. Luterbacher
EPFL
Switzerland

Riccardo Barbano
Atinary Technologies Inc.
Switzerland

José M. Hernández-Lobato
University of Cambridge
United Kingdom

Philippe Schwaller
EPFL
Switzerland
philippe.schwaller@epfl.ch

Loïc Roch
Atinary Technologies Inc.
Switzerland
loic.roch@atinary.com

Abstract

Multi-fidelity Bayesian Optimization (MFBO) is a promising framework to speed up materials and molecular discovery as sources of information of different accuracies are at hand at increasing cost. Despite its potential use in chemical tasks, there is a lack of systematic evaluation of the many parameters playing a role in MFBO. In this work, we provide guidelines and recommendations to decide when to use MFBO in experimental settings. We investigate MFBO methods applied to molecules and materials problems. First, we test two different families of acquisition functions in two synthetic problems and study the effect of the informativeness and cost of the approximate function. We use our implementation and guidelines to benchmark three real discovery problems and compare them against their single-fidelity counterparts. Our results may help guide future efforts to implement MFBO as a routine tool in the chemical sciences.

Among Machine Learning (ML) techniques, Bayesian optimization (BO) has emerged as the go-to choice for optimizing the design of experiments in the chemical domain[1]. Bayesian optimization, grounded in a probabilistic framework[2], consists of two main components: a probabilistic model that serves as a proxy for the experimental process being optimized, and a policy that governs the acquisition of new experimental data. For instance, a researcher seeking to maximize the yield of a given reaction would query the model to identify which experimental conditions to be tested to achieve the desired outcome. This methodology has shown success in diverse optimization tasks such as chemical reactions[1,3,4] and functional molecules[5]. Iterating through this learning cycle has recently enabled the rapid identification of optimal conditions within extensive search spaces[6,7].

While canonical BO has recently been popularized among experimentalists, the experiment design may benefit if the practitioner can collect data at different degrees of reliability while paying a lower price. Additional experimental evidence that may be readily available can be integrated within the model representing the process to be optimized (e.g., low-precision experiments conducted with bench-top nuclear magnetic resonance can be integrated into more expensive, high-precision experiments[8]). In classical experimental design, the inclusion of such information sources representing different reliabilities is referred to as multi-fidelity Bayesian Optimization (MFBO). Within this setting, a

AI for New Drug Modalities at NeurIPS 2024.

specific cost is assigned to each information source — hereafter defined as fidelity. The multi-fidelity probabilistic model learns the process of interest by extracting knowledge from data available at different fidelities and understanding their interplay. The policy for querying new experimental data at a given fidelity also takes into account the overall cost. By combining low-fidelity (LF) and high-fidelity (HF) points, the overall optimization cost can be reduced compared to the single-fidelity BO (SFBO). Figure 1 exemplifies how the iterative cycle of MFBO may reduce the overall cost, compared to the standard SFBO.

Although MFBO has garnered interest in the past decade within the ML community, leading to a myriad of model definitions and acquisition policies[9–11], researchers have only lately started to integrate them within their design methodologies. In the chemical domain, there has been recent interest in incorporating cost awareness into the BO loop[12]. In the specific multi-fidelity approach, several studies have successfully applied MFBO to the materials discovery domain[13–16]. However, there is a fundamental lack in the assessment of MFBO performance, and due to this, each work represents MFBO performance in different ways[9,11,14]. Although some metrics have been proposed[13], the lack of clarity and unified criteria on how to assess the benefits of MFBO in the chemical domain has hindered its widespread adoption among the experimental community[17]. Several factors come into play when assessing its impact, yet the application of MFBO can also be detrimental to the overall optimization process[18]. It has been also shown how in the long run the advantage of MFBO over SFBO can be lost[17].

In this work, we propose a series of guidelines on when to use MFBO within the experimental design pipeline. We conduct an exhaustive experimental investigation on both standard MFBO problems — ubiquitous in BO literature — as well as chemistry-based ones. Initially, we optimize two synthetic problems to assess the behavior of a multi-fidelity model in simulated black-box scenarios, detecting unfavorable situations where the application of MFBO does not offer an advantage over SFBO. We exhaustively scan MFBO experimental parameters (namely, cost ratio and informativeness of the LF source) to identify trends that indicate promising experimental settings for when to apply MFBO. Building on these guidelines, we then progress to chemistry-based problems, tackling three real challenges in molecular and materials optimization, where MFBO successfully outperforms SFBO. This work offers a comprehensive view and a reference guide for applying MFBO in the molecules and materials discovery domain.

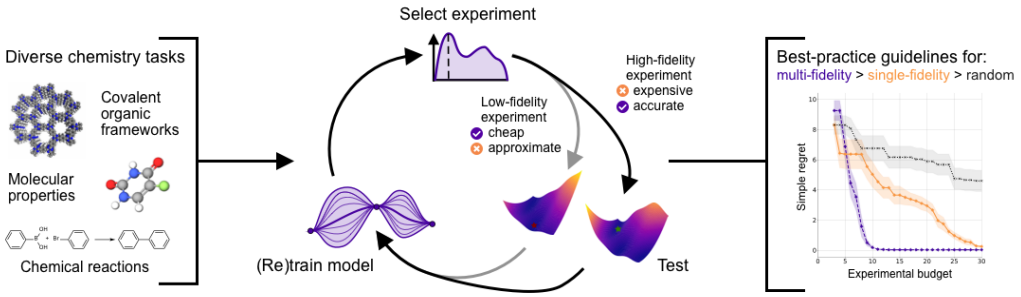

Figure 1: Multi-fidelity Bayesian Optimization (MFBO) combines expensive but informative and cheap but approximate sources of information with a ML model to optimize black-box problems at an overall reduced budget. The high-fidelity source (HF, e.g., real lab experiment) provides accurate information but at a high cost. The low-fidelity source (LF, e.g., computer simulation of real experiment) represents a cheaper approximation of the HF. In the illustration, we show the surfaces corresponding to the Branin function used in this study at the HF and one LF level, where the optima are located in different places. The MFBO loop iteratively selects HF or LF to maximize the information gain while reducing the overall optimization cost.

Methods

Multi-fidelity Bayesian Optimization

BO uses a probabilistic surrogate model and a selection policy (acquisition function) to optimize black-box problems. These problems are normally expensive to query, and BO accelerates the finding

of optimal points by minimizing the number of calls to the target function. The optimization problem can be formulated as

$$\underset{x \in \mathcal{X}}{\operatorname{argmax}}\, f(x)$$

where f is the target black-box function and $\mathcal{X}$ is the space of all possible input candidates. Considering $\mathcal{D} = [(x_1, y_1), (x_2, y_2), ...(x_n, y_n)]$ a dataset containing pairs of inputs-outputs $(x_n, y_n) \in \mathbb{R}^d \times \mathbb{R}$ of our problem, we can use a surrogate model to learn the relationships between the data in the form $y_i = f(x_i) + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The surrogate is commonly a Gaussian Process (GP), a non-parametric model that provides uncertainty quantification[19]. GPs are usually defined as $f(x) \sim \mathcal{GP}(0, k(x, x'))$, placing a zero-mean prior and a covariance given by a kernel function $cov[f(x), f(x')] = k(x, x')$ that measures the similarity between inputs. Given a dataset $\mathcal{D}$, the predictive mean and variance are given by the following expressions

$$\mu(x) = k_x(K + \sigma_\epsilon^2 I)^{-1} y$$
$$\sigma^2(x) = k(x, x) - k_x(K + \sigma_\epsilon^2 I)^{-1} k_x^\top$$

where $k_x = [k(x, x_1), ..., k(x, x_n)]^\top \in \mathbb{R}^n$, $K \in \mathbb{R}^{n \times n} = (k(x_i, x_j))_{1 \leq i,j \leq n}$ and $y = [y_1, ..., y_n]$. In order to optimize the target problem, after training the surrogate with the available data, a decision is made to select the next points by maximizing the acquisition function. Acquisition functions are heuristics that decide the next point to query by balancing exploration and exploitation. The next inputs are given by

$$\underset{x \in \mathcal{X}}{\operatorname{argmax}}\, \texttt{acqf}(x)$$

where `acqf` is the acquisition function.

Multi-fidelity BO extends the previous setup by considering an extra input parameter for the surrogate model, the fidelity l. This parameter indicates which source of information the other inputs correspond to (e.g., high fidelity, HF, or low fidelity, LF). In our case, we use a discrete fidelity setting, where $l \in [l_0, l_1, ...l_m]$ and l_m corresponds to the highest fidelity among the m levels. In all cases, we limit the analysis to two levels of fidelity (l_0 and l_1 and $m = 1$) , but the methodology can be extended to more levels. Additionally, these models assume that the cost only depends on the fidelity level l and not on the search space x. We employed a modified GP proposed in[10] that extends the previously described surrogate modelling to the multi-fidelity space of $\mathbb{R}^{n+1}$ dimensions, chosen due to its popularity amongst MFBO applications[10,14,18]. The modified kernel is defined as $k((x, l), (x', l'))$, where fidelity is introduced by defining a separate kernel to model the input space, and a kernel to model the correlation and interaction between the different fidelity levels. The mathematical expression is:

$$k\left((x, l), (x', l')\right) = k_{\mathrm{I}}(x, x') \times k_{\mathrm{IS}}(l, l')$$
$$k_{\mathrm{I}}(x, x') = \exp\left(-\frac{1}{2}\sum_{i=1}^{d} \lambda_i^{-1}(x_i - x_i')^2\right)$$
$$k_{\mathrm{IS}}(l, l') = c + (1 - l)^{1+\delta}(1 - l')^{1+\delta}$$

Given a dataset of inputs with their associated fidelities and outputs, the model can be trained using standard maximum log-likelihood optimization to obtain the optimal hyperparameters λ_i, c and δ. In the experiments, we set the l values of the HF and LF levels to 1 and 0 respectively following previous studies that used this kernel in a binary setting[14,18].

In the multi-fidelity case, acquisition functions must compute the information gained over the cost of a query at a given fidelity level. The previous expression for the acquisition function maximization can be therefore generalized as $\texttt{acqf}(x, l) = \texttt{acqf}(x) \cdot \texttt{cost}(l)^{-1}$. The next pair of input and fidelity query is given by

$$\underset{(x,l) \in \mathcal{X} \times [m]}{\operatorname{argmax}}\ \texttt{acqf}(x, l)$$

In the experiments we use two families of acquisition functions, with their single and multi-fidelity versions: Maximum Entropy Search (MES)[11] and Expected Improvement (EI)[20]. We also initially tested Knowledge Gradient (KG)[10], but we discarded it due to its computational expense.

In all cases, the experiments are run for 20 independent seeds, and the number of initial samples corresponds to 10% of the total optimization budget, sampled using a Latin hypercube strategy. In the MFBO case, the budget is split to sample 50% of HF and 50% of LF initial points. In the synthetic functions, the total budget corresponds to 50 HF queries. In the chemistry and materials benchmarks, the total budget corresponds to 30 HF queries. Data collection is sequential. All the experiments are run using BOTorch[21] as a standard Bayesian Optimization framework.

Metrics

High-fidelity regret calculation in MFBO Setting

Due to the lack of standardized metrics in the MFBO field[17], we define a standardized simple regret and a discount metric to quantify the performance of MFBO over SFBO. Let f be the HF target problem. At a given step t of the SFBO run, the simple regret is defined as $r_t = f^* - f_t^*$, where f^* is the known global optimum and f_t^* is the best value of f found so far. The simple regret is computed only at the highest fidelity level in the multi-fidelity runs. To compare multi-fidelity and single-fidelity runs it is necessary to find the corresponding high-fidelity simple regret in the multi-fidelity setting, which can query both the low-fidelity or high-fidelity at a given step. We define a standard algorithm to identify the corresponding HF simple regret at a given budget of an MFBO campaign. The algorithm aligns the r values from the MFBO to a common cost scale given by the SFBO cost steps. This makes it possible to compare the performance of the MFBO method at each step of the associated SFBO run. Algorithm 1 explains in detail the procedure to compute simple regret both for single-fidelity and multi-fidelity runs.

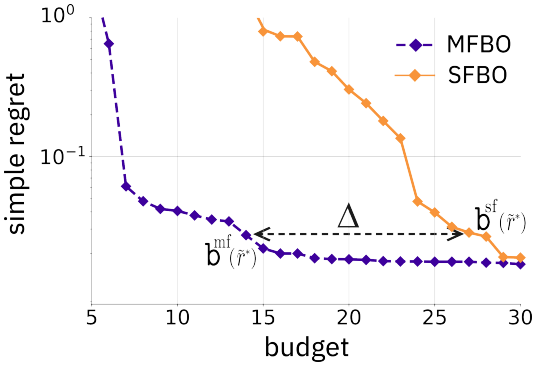

Figure 2: Standardized metrics to compare MFBO and SFBO. Simple regret traces are computed using algorithm 1 and provide a direct comparison between the optimization progress in each method by computing the equivalent number of steps from SFBO in the MFBO trace. Discount (Δ) reflects the savings in optimization cost by computing the normalized difference in budget spent by each algorithm ($\mathtt{b}^{\mathrm{sf}}(\tilde{r}^*)$ and $\mathtt{b}^{\mathrm{mf}}(\tilde{r}^*)$) to find $\tilde{r}^*$ (horizontal arrow, see equation 1).

In the plots, we compute the average r and standard deviation of an optimization run using the proposed algorithm from all the different repeats. Figure 2 shows a graphical explanation of the associated regret traces depicted in orange and purple which have an equal number of datapoints.

Discount as a utility metric for MFBO

We also propose a discount Δ metric to estimate the savings provided by the MFBO method compared to the SFBO (see figure 2 for a graphical explanation of this metric). The discount variable reflects the difference of budgets required to reach a reference regret value between the multi-fidelity and single-fidelity methods. Specifically, it is computed over two individual SF and MF runs as follows:

$$\Delta = \frac{\mathtt{b}^{\mathrm{sf}}(\tilde{r}^*) - \mathtt{b}^{\mathrm{mf}}(\tilde{r}^*)}{\mathtt{b}^{\mathrm{sf}}(\tilde{r}^*)}, \tag{1}$$

where $\tilde{r}^*$ is the *corrected best multi-fidelity regret*, defined as $\tilde{r}^* = r_t^{\mathrm{mf}}\tau$. The τ correction accounts for situations where the MFBO method gets to low regrets faster than the SFBO at the beginning of the optimization, but its performance plateaus in the long term as noted in[17]. In our experiments, we set $\tau = 2$ as a reasonable adjustment, given that r_t^{mf} tends to be small. A positive Δ indicates cost savings with MFBO, meaning it achieves the minimum regret with a lower budget than SFBO. Conversely, a negative Δ implies that MFBO was more expensive than SFBO. If the SFBO method

is unable to reach $\tilde{r}^*$, Δ is capped at 1. Note that equation 1 compares individual runs started at the same seed as displayed in Figure 2, and we report the average Δ over all the seeds onwards.

LF level cost and informativeness estimation

We use cost ratio and informativeness metrics to compute the characteristics of the low-fidelity approximation with respect to the high-fidelity. Cost is characterized by the cost ratio ρ, which is obtained by dividing the LF cost and the HF cost. Informativeness is characterized by the R^2 of the LF approximation. To compute this metric, we uniformly sample 100 points for the given problem and extract their associated values at the HF and LF levels. Then, the R^2 between a linear fitting of the LF prediction and the true HF value is computed. Supplementary Information A.3 shows the plots with the sampled points and the computed R^2 of each of problems used in this study.

Synthetic functions

Synthetic functions are simulated black-box problems where a mathematical expression is used to generate points from a surface to be optimized. In the multi-fidelity case, there are lower accuracy approximations of the target function at a lower cost. These can either have a fixed expression or be generated by biasing the original function using a given parameter α. In our case, we use the Branin[18] function previously used in MFBO works. We also modify a Park function used in MFBO[22] by incorporating a parameter α to modulate the bias with a similar behaviour to the Branin case. The expressions of the functions can be found in the Supplementary Information A.1. The cost of querying the high-fidelity level is set to a value of 1, and the cost of the low-fidelity approximations is set to a fraction of this value.

Chemistry and materials design benchmarks

We use or adapt previously reported benchmarks for real experimental optimization problems in the chemistry domain. The Covalent Organic Frameworks (COFs) dataset was used in a previous work on MFBO[14], and consists of 608 candidate COFs encoded in a 14-dimensional vector accounting for their composition and crystal structure. The high-fidelity simulation uses a Markov chain Monte Carlo simulation to compute the adsorption of Kr and Xe in the material, and the associated Kr/Xe selectivity. This simulation is expensive, with an average running time of 230 minutes. In this case, a low-fidelity approximation of the selectivity can be obtained using Henry's law, reducing the computing time to 15 minutes, giving a ρ of 0.065 for this problem.

The second benchmark is derived from the Alexandria library[23], and it was used in previous work on MFBO for materials discovery[13]. It comprises 1134 molecules encoded using RDKit 2D-descriptors[24] and reduced to a 10-D vector using PCA. The high-fidelity is the experimental polarizability, whereas the low-fidelity is the computed polarizability at the Hartree-Fock 6-31G+ level of theory. ρ is set up to 0.167 in this case following the previous study.

The last benchmark is extracted from the FreeSolv library[25]. It comprises 641 molecules, encoded using RDKit 2D-descriptors and reduced to a 10-D vector using PCA. The high-fidelity is the experimental free solvation energy, whereas the low-fidelity is the computed solvation energy using molecular dynamics. ρ is set up to 0.1 in this case.

Results

Shedding light into MFBO failure modes

We test the performance of MFBO methods over their SFBO counterparts on two different synthetic functions, Branin and Park (with 2 and 4 dimensions respectively), on two different scenarios (favorable and unfavorable). Synthetic functions are commonly employed as black-box problems to test the performance of BO algorithms (see Synthetic functions). In the case of MFBO, the output of the functions can be biased using a parameter α to provide lower accuracy information sources. This parameter is related to the informativeness of the LF source (in our case, we quantify the informativeness of the LF by computing the R^2 with respect to the HF, see Metrics). For the favorable scenario, we tune the α parameter to provide a highly informative LF source (LF $R^2 > 0.9$). We

set the cost of querying a HF point to 1 and the cost of a LF point to 0.1 ($\rho = 0.1$). Figure 3 shows the results for the favorable scenario. In this case, both MFBO methods outperform their SFBO counterparts, with maximum discounts of 0.5 and 0.58 for Branin and Park respectively. The MFBO runs get to lower regrets quicker, spending less resources by exploiting the use of LF and HF sources of information. The resulting optimization runs are therefore desirable as they effectively leverage the access to the LF to guide the optimization of the HF level.

Although under the previous settings MFBO provides a higher performance than SFBO, a change in the LF conditions dramatically affects the result of the optimization. In the unfavorable scenario, we set the cost of the LF source to 0.5 (only half of the HF cost, $\rho = 0.5$) and decrease the informativeness of the functions ($R^2 < 0.75$). In this case, the performance of both methods is similar (minimum Δ of -0.07 and -0.1 for Branin and Park respectively), and the MFBO does not provide a quicker convergence to low regret values. The previous results illustrate how MFBO performance is affected by the cost and the informativeness of the LF source, losing its advantage over SFBO if the LF source is not informative and cheap enough.

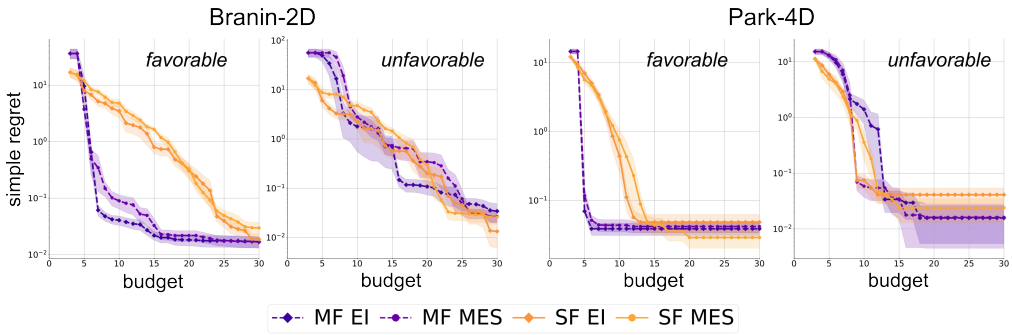

Figure 3: MFBO results are dependent on problem conditions. In the favorable scenario (highly informative and cheap LF source, $\rho = 0.1$, $R^2 > 0.9$), MFBO offers better performance than SFBO on the synthetic functions (maximum Δ of 0.5 and 0.58 for Branin and Park, respectively). The optimization traces show how lower regrets are found at lower budgets under this scenario. When the conditions are changed and a low informative and more expensive LF is used ($\rho = 0.5$, $R^2 < 0.75$), MFBO performance is degraded. Under this scenario, the optimization traces are similar, and MFBO does not offer an advantage over SFBO (minimum Δ of -0.07 and -0.1 for Branin and Park, respectively).

Finding suitable scenarios for MFBO

We investigate in detail the effect of the LF cost and informativeness on MFBO performance (measured by Δ, see Metrics) for the two previous synthetic functions. We run the MFBO method with different combinations of LF cost and informativeness and compare it to SFBO. Figure 4 shows the computed Δ for each run as a result of the previous parameters. The x axis of the heatmap represent the α value that was used to bias the synthetic function, and the upper plot shows the computed R^2 for each value. In both cases, a gradient can be observed where the progression towards cheaper and more informative LF sources provides better MFBO performances (higher Δ). In the case of Branin, the lower values of R^2 in the synthetic functions are associated with a less informative LF approximation, which translates to lower values of Δ. In the same way, in Park Δ is also decreased when the R^2 is lower, giving negative results when its value is close to 0. This result is consistent as a more inaccurate source of information is likely to reduce the performance of the MFBO method. In terms of cost, there is also an inverse correlation in both cases, where lower ρ (cheaper LF sources) generate higher discounts. This is an expected trend, as the model can access the approximate sources at a lower cost, allowing a wider exploration of the problem space for the same price.

The results follow the same trend in both cases independently of the acquisition function. Although there are differences in the absolute Δ values between the two problems (due to the specific landscape of each synthetic function), the regions of high performance are localized in the same areas of the heatmaps. These results provide a qualitative understanding of the effect of cost and informativeness on the performance of MFBO. The general high-performance region is located around cheap and highly informative LF sources, and a good experimental setup for MFBO application should aim to

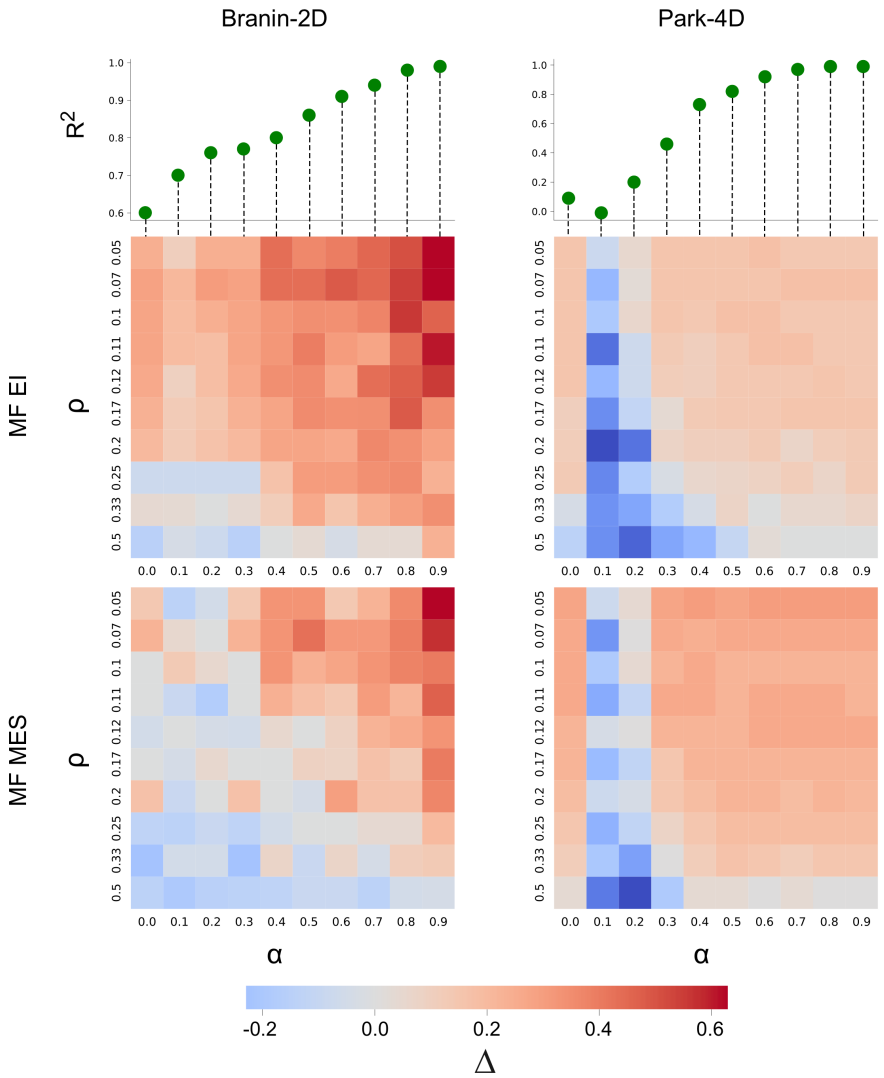

Figure 4: Change of Δ with the cost and informativeness of the LF source. Each heatmap shows the computed Δ for each synthetic function and acquisition function family. The heatmap's y axis represents the cost of the LF source (ρ), and the x axis its informativeness. The informativeness of each function is tuned using a parameter α in the definition of the function. Because α cannot be known in a real experiment, the upper plots show the estimated R^2 value for each α value (see Methods for an explanation of R^2 estimation). In all the cases, cheaper and informative LF sources provide the highest discounts (a gradient in Δ increasing when ρ decreases and R^2 increases can be observed).

find suitable LF sources. In general, a user should first verify that the cost of the auxiliary experiment is cheap enough to run MFBO. In the same way, they should check and estimate the degree of informativeness of the auxiliary experiment to decide if it provides a reasonable approximation of the real problem (in this case, we propose R^2 as a guiding metric, but other informativeness measurements could be employed). Figure 5 displays a guiding flowchart following the previous considerations. Using these conditions does not guarantee superior performance over SFBO, but provides a guiding principle for the successful application of MFBO in real scenarios.

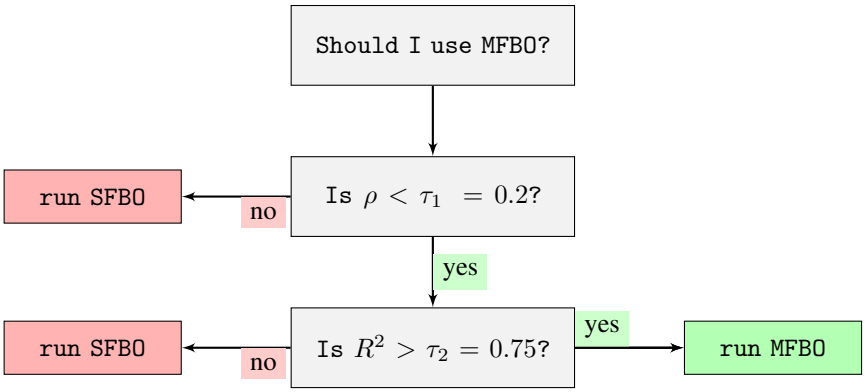

Figure 5: Proposed guidelines to run MFBO. Evaluation of the informativeness and cost of the LF approximation provides a reasonable estimation of the potential success of the MFBO approach over standard SFBO. The flowchart illustrates a potential analysis of the scenario to decide if running MFBO or not by using a threshold of ρ (τ_1) and R^2 (τ_2). If the LF source is cheap enough ($\rho < \tau_1$) and informative enough ($R^2 > \tau_2$), we recommend running MFBO. We propose $\tau_1 = 0.2$ and $\tau_2 = 0.75$ as a soft threshold for this study based on our empirical investigation.

Application of MFBO guidelines to chemistry problems

We translate the previous recommendations to three benchmarks in chemistry and materials science to showcase their utility in real experimental situations. The tasks correspond to scenarios where a cheap and informative LF source is available under the proposed guidelines (see Methods for a detailed explanation of each benchmark). Figure 6a illustrates the results of MFBO and SFBO in these scenarios. In all cases, the MFBO method can get to lower regrets at a lower cost than the SFBO by leveraging the cheaper source of information at hand. In the COFs benchmark (problem settings: ρ = 0.065, R^2 = 0.98), the MFBO gets to the COF with the highest computed Kr/Xe selectivity in the allocated budget, whereas the SFBO does not find it. In the polarizability benchmark (problem settings: ρ = 0.167, R^2 = 0.99), the MFBO provides a maximum discount of 0.6 over the SFBO. In this case, the use of a cheap Hartree-Fock computation is successfully combined with the experimental measurements to find the molecule with the highest polarizability. In the solvation energy benchmark (problem settings: ρ = 0.1, R^2 = 0.88), the MFBO finds the molecule with the highest solvation energy but the SFBO not, as in the COFs case. This example also integrates computed values with experimental measurements to reduce the spent budget. Figure 6b shows the proportion of queries to each fidelity level in the different tasks. By distributing the available budget between the HF and LF levels, the MFBO methods can integrate the cheaper information of the LF approximations to improve the optimization results. The number of calls to each query depends on the specific acquisition function and task. However, in all cases the proportion of calls to the HF level is lower than 0.4, suggesting that this proportion can provide a threshold for a successful MFBO optimization. In these real scenarios, the MFBO methods can exploit the low accuracy but cheap LF levels by querying them more often than the HF level.

We also include a negative example where the conditions of one of these benchmarks are changed to an unfavorable scenario and the associated MFBO performance is overridden. We change the LF conditions of the molecule polarizability benchmark (ρ = 0.5 and R^2 = 0.49) to illustrate how Δ is reduced with respect to the previous case. Figure 6c shows the result for the optimization, where Δ is reduced from a maximum value of 0.67 to 0.03, providing almost no advantage over SFBO if compared with the previous favorable scenario (similar to the results described for the synthetic functions, see Results). The previous benchmarks show how the application of MFBO in suitable scenarios provides an advantage over standard SFBO cases and effectively decreases the optimization cost. By locating experiments where the cost and informativeness of the LF source are adequate, we can successfully translate the advantages of MFBO into a realistic experimental task.

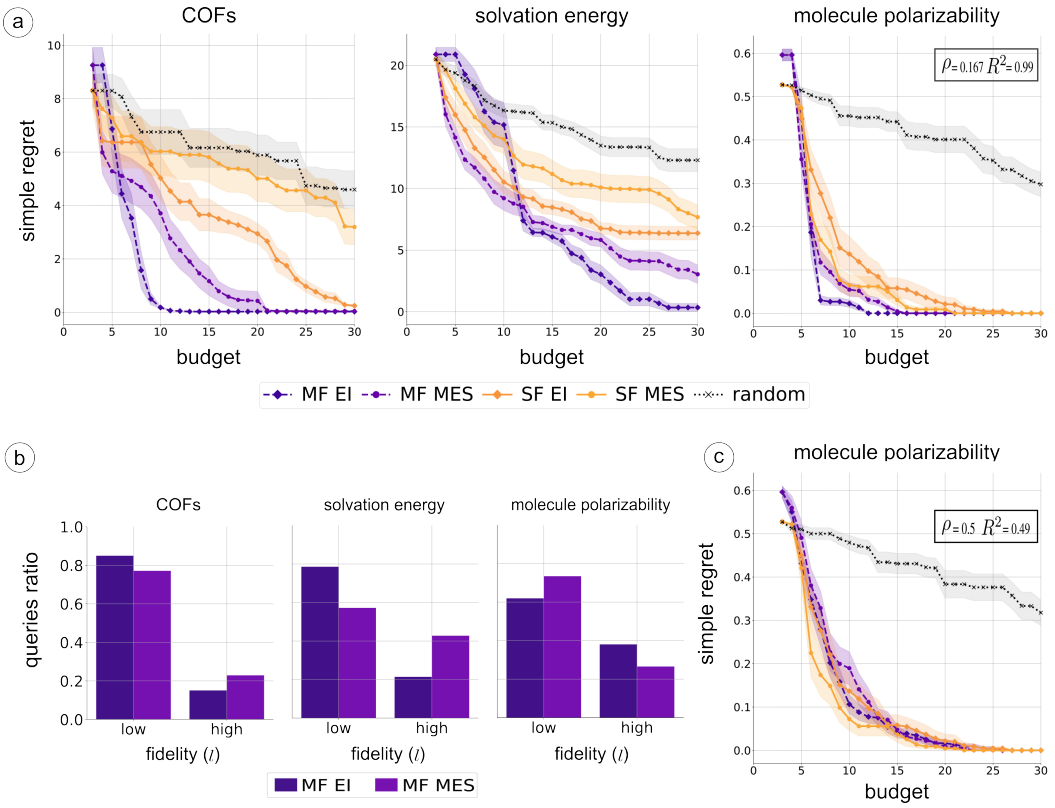

Figure 6: Molecules and materials benchmark results. a) Optimization performance for the MFBO and the SFBO methods in each benchmark (see Methods for a description of each task). The maximum Δ in each case is 1, 1 and 0.67 for the COFs, solvation energy and polarizabiliry tasks, respectively. In all the cases, MFBO is able to achieve lower regrets than SFBO by spending less budget. b) Proportion of queries to each fidelity source in the MFBO setting. The MFBO method balances information gain and budget spent by querying more often the LF values (between 0.6 and 0.8 in the ratio of queries) while also using information from HF values. c) Experimental negative example. Molecule polarizability benchmark conditions were changed so the LF source was expensive and less informative. The resulting Δ is 0 and 0.03 for the MES and EI families respectively, showing how unfavorable conditions can override the MFBO advantage over SFBO in a real experimental setting.

Discussion

Multi-fidelity Bayesian optimization (MFBO) offers a promising approach to reducing costs in optimization by leveraging inexpensive, approximate sources of information. In the chemical sciences, MFBO can accelerate optimization while maintaining cost efficiency. However, its utility is not universal; in some cases, it may prove less effective or even counterproductive[18]. To address these concerns, we conducted a comprehensive study to determine under which conditions MFBO outperforms standard single-fidelity Bayesian optimization (SFBO).

We introduce two key metrics for comparison: a standardized regret, r, and a discount metric, Δ (see Metrics). Our analysis of synthetic benchmarks demonstrates that MFBO performance is highly dependent on the characteristics of the low-fidelity (LF) source. Specifically, not all LF conditions are conducive to the effective use of MFBO. To systematically identify favorable conditions, we performed a grid-based exploration of LF cost and informativeness against the discount metric, Δ. The results indicate that both ρ and R^2 significantly influence the discount, with the most favorable outcomes achieved when the LF source is both highly informative and low-cost. Based on these findings, we developed a flowchart to guide the application of MFBO (see Figure 5). These guidelines were validated using three real-world experimental benchmarks in molecular and materials discovery,

demonstrating that MFBO consistently outperforms SFBO in reducing optimization costs, provided the LF source is sufficiently inexpensive and informative.

Our work provides a structured decision-making framework for determining the applicability of MFBO. While our analysis focused on a specific surrogate model and two acquisition functions, the approach is extensible to other surrogate models and acquisition function families. Future research will expand this framework to include models such as Bayesian neural networks (BNNs)[26], and explore additional experimental applications, evaluating the impact of varied feature spaces or additional fidelity levels. This study offers valuable insights for practitioners seeking to integrate MFBO into their experimental workflows, paving the way for more routine application of MFBO in chemical and materials optimization.

References

[1] Florian Häse, Loïc M. Roch, Christoph Kreisbeck, and Alán Aspuru-Guzik. Phoenics: A bayesian optimizer for chemistry. *ACS Central Science*, 4(9):1134–1145, 2018.

[2] Roman Garnett. *Bayesian optimization*. Cambridge University Press, Cambridge, England, February 2023.

[3] Florian Häse, Loïc M Roch, and Alán Aspuru-Guzik. Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chemical science*, 9(39):7642–7655, 2018.

[4] Jeff Guo, Bojana Ranković, and Philippe Schwaller. Bayesian optimization for chemical reactions. *Chimia*, 77(1-2):31–38, February 2023.

[5] Ryan-Rhys Griffiths, Jake L Greenfield, Aditya R Thawani, Arian R Jamasb, Henry B Moss, Anthony Bourached, Penelope Jones, William McCorkindale, Alexander A Aldrick, Matthew J Fuchter, and Alpha A Lee. Data-driven discovery of molecular photoswitches with multioutput gaussian processes. *Chem. Sci.*, 13(45):13541–13551, November 2022.

[6] Manu Suvarna, Tangsheng Zou, Sok Ho Chong, Yuzhen Ge, Antonio J Martín, and Javier Pérez-Ramírez. Active learning streamlines development of high performance catalysts for higher alcohol synthesis. *Nat. Commun.*, 15(1):5844, July 2024.

[7] Xiaobo Li, Yu Che, Linjiang Chen, Tao Liu, Kewei Wang, Lunjie Liu, Haofan Yang, Edward O Pyzer-Knapp, and Andrew I Cooper. Sequential closed-loop bayesian optimization as a guide for organic molecular metallophotocatalyst formulation discovery. *Nat. Chem.*, June 2024.

[8] Bernhard Blümich. Low-field and benchtop NMR. *J. Magn. Reson.*, 306:27–35, September 2019.

[9] Kirthevasan Kandasamy, Gautam Dasarathy, Junier B Oliva, Jeff Schneider, and Barnabas Poczos. Gaussian process bandit optimisation with multi-fidelity evaluations. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[10] Jian Wu, Saul Toscano-Palmerin, Peter I. Frazier, and Andrew Gordon Wilson. Practical multi-fidelity bayesian optimization for hyperparameter tuning. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 788–798. PMLR, 22–25 Jul 2020.

[11] Shion Takeno, Hitoshi Fukuoka, Yuhki Tsukada, Toshiyuki Koyama, Motoki Shiga, Ichiro Takeuchi, and Masayuki Karasuyama. Multi-fidelity Bayesian optimization with max-value entropy search and its parallelization. 119:9334–9345, 13–18 Jul 2020.

[12] Alexandre Schoepfer, Jan Weinreich, Ruben Laplaza, Jerome Waser, and Clemence Corminboeuf. Cost-informed bayesian reaction optimization. April 2024.

[13] Clyde Fare, Peter Fenner, Matthew Benatan, Alessandro Varsi, and Edward O Pyzer-Knapp. A multi-fidelity machine learning approach to high throughput materials screening. *Npj Comput. Mater.*, 8(1), December 2022.

[14] Nickolas Gantzler, Aryan Deshwal, Janardhan Rao Doppa, and Cory M Simon. Multi-fidelity bayesian optimization of covalent organic frameworks for xenon/krypton separations. *Digit. Discov.*, 2(6):1937–1956, 2023.

[15] Jungtaek Kim, Mingxuan Li, Yirong Li, Andrés Gómez, Oliver Hinder, and Paul W Leu. Multi-BOWS: multi-fidelity multi-objective bayesian optimization with warm starts for nanophotonic structure design. *Digit. Discov.*, 3(2):381–391, 2024.

[16] Jose Pablo Folch, Robert M Lee, Behrang Shafei, David Walz, Calvin Tsay, Mark van der Wilk, and Ruth Misener. Combining multi-fidelity modelling and asynchronous batch bayesian optimization. *Comput. Chem. Eng.*, (108194):108194, February 2023.

[17] Gbetondji J-S Dovonon and Jakob Zeitler. Long-run behaviour of multi-fidelity bayesian optimisation, 2023.

[18] Petrus Mikkola, Julien Martinelli, Louis Filstroff, and Samuel Kaski. Multi-fidelity bayesian optimization with unreliable information sources. 206:7425–7454, 25–27 Apr 2023.

[19] Carl Edward Rasmussen and Christopher K I Williams. *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning Series. MIT Press, London, England, 2019.

[20] D Huang, T T Allen, W I Notz, and R A Miller. Sequential kriging optimization using multiple-fidelity evaluations. *Struct. Multidiscipl. Optim.*, 32(5):369–382, September 2006.

[21] Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems 33*, 2020.

[22] Shifeng Xiong, Peter Z G Qian, and C F Jeff Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, February 2013.

[23] Mohammad M Ghahremanpour, Paul J van Maaren, and David van der Spoel. The alexandria library, a quantum-chemical database of molecular properties for force field development. *Sci. Data*, 5(1):180062, April 2018.

[24] RDKit: Open-source cheminformatics. `http://www.rdkit.org`.

[25] David L Mobley and J Peter Guthrie. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput. Aided Mol. Des.*, 28(7):711–720, July 2014.

[26] Yucen Lily Li, Tim G. J. Rudner, and Andrew Gordon Wilson. A study of bayesian neural network surrogates for bayesian optimization, 2024.

Appendix A Supplementary information

A.1 Synthetic functions expressions

Branin-2D

$$f(\mathbf{x}, \alpha) = (x_2 - (\frac{5.1}{4\pi^2} - 0.1(1-\alpha))x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})cos(x_1) + 10 \tag{2}$$

defined over $[-5, 10] \times [0, 15]$, and $\alpha \in [0, 1]$ is the bias term.
Park-4D

$$\begin{aligned} f(\mathbf{x}, \alpha) = \frac{x_1}{2}(\sqrt{1 + \frac{(x_2 + x_3^2)x_4}{x_1^2}} - 1) \\ + (x_1 + (3 - 1.5(1-\alpha))x_4)\exp(1 + \sin(x_3)) \end{aligned} \tag{3}$$

defined over $[0, 1]^4$, and $\alpha \in [0, 1]$ is the bias term.

A.2 Simple regret algorithm

Algorithm for the standardized simple regret computation. Single-fidelity and multi-fidelity cumulative cost values (c^{sf} and c^{mf}) represent the accumulated cost at each step of the optimization run.

Algorithm 1: Algorithm for computing multi-fidelity simple regret

Input: $t \in \mathbb{N}_+$ `// Number of single-fidelity optimization steps`
$n \in \mathbb{N}_+$ `// Total number of multi-fidelity steps`
$y^{\text{mf}} \in \mathbb{R}^n$ `// Multi-fidelity output values`
$l \in \{0, 1\}^n$ `// Fidelity values (1 : high-fidelity, 0 : low-fidelity)`
$c^{\text{sf}} \in \mathbb{R}^t$ `// Single-fidelity cumulative cost values`
$c^{\text{mf}} \in \mathbb{R}^n$ `// Multi-fidelity cumulative cost values`
$f^* \in \mathbb{R}$ `// High-fidelity global optimum`
Output: $r \in \mathbb{R}^t$ `// Multi-fidelity regret over` t `single-fidelity steps`

Function `ComputeSimpleRegret(`y^{mf}, l, c^{sf}, c^{mf}, f^*, t`):`
  `// Select high-fidelity outputs`
  $y^{\text{hf}} \leftarrow [y_i^{\text{mf}}]_{\forall i | l_i = 1}$
  `// Compute high-fidelity simple regret`
  $r^{\text{hf}} \leftarrow f^* - y^{\text{hf}}$
  `// Initialize simple regret list`
  $r \leftarrow [\,]$
  for $i \leftarrow 1$ **to** t **do**
    `// Find minimum regret among indices where` $c_k^{\text{mf}} \leq c_i^{\text{sf}}$
    $r_{\min} \leftarrow \min\left([r_k^{\text{hf}}]_{\forall k | c_k^{\text{mf}} \leq c_i^{\text{sf}}}\right)$
    `// Append to simple regret list`
    Append $r_{\min}$ to r
  end
return r

A.3 Informativeness measurement

The informativeness of the different problems is computed using the R^2 as explained in .

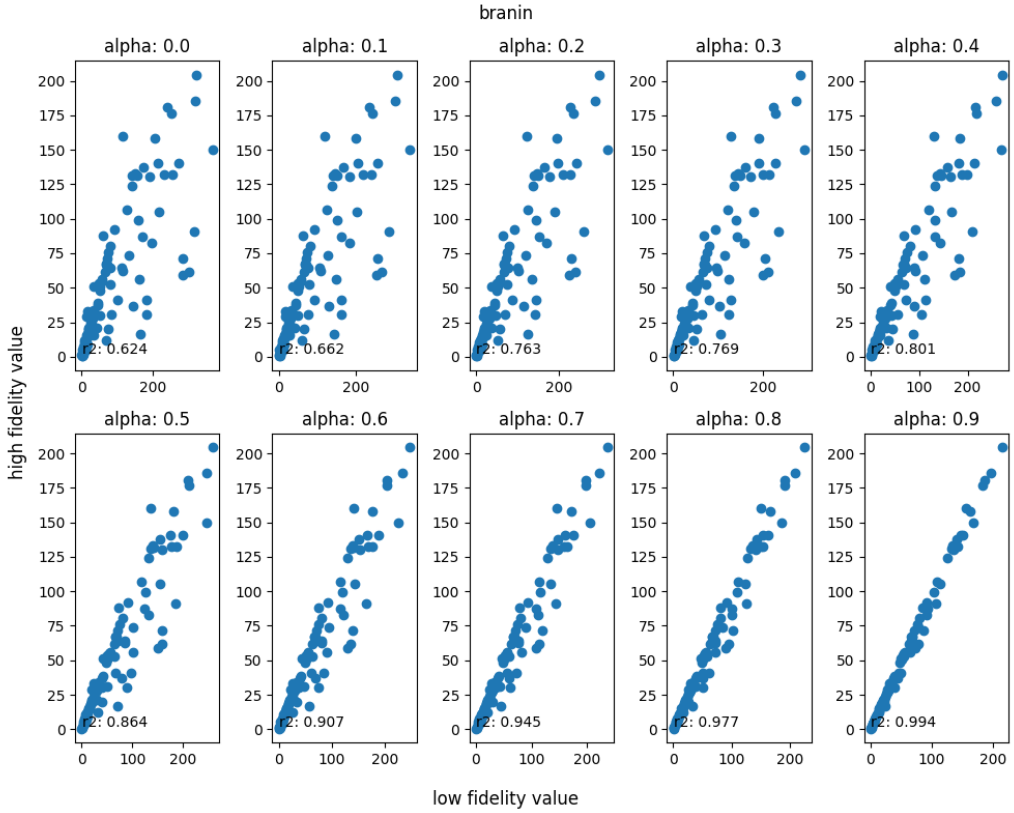

Figure 7: Informativeness of Branin-2D LF approximation varying with the α parameter

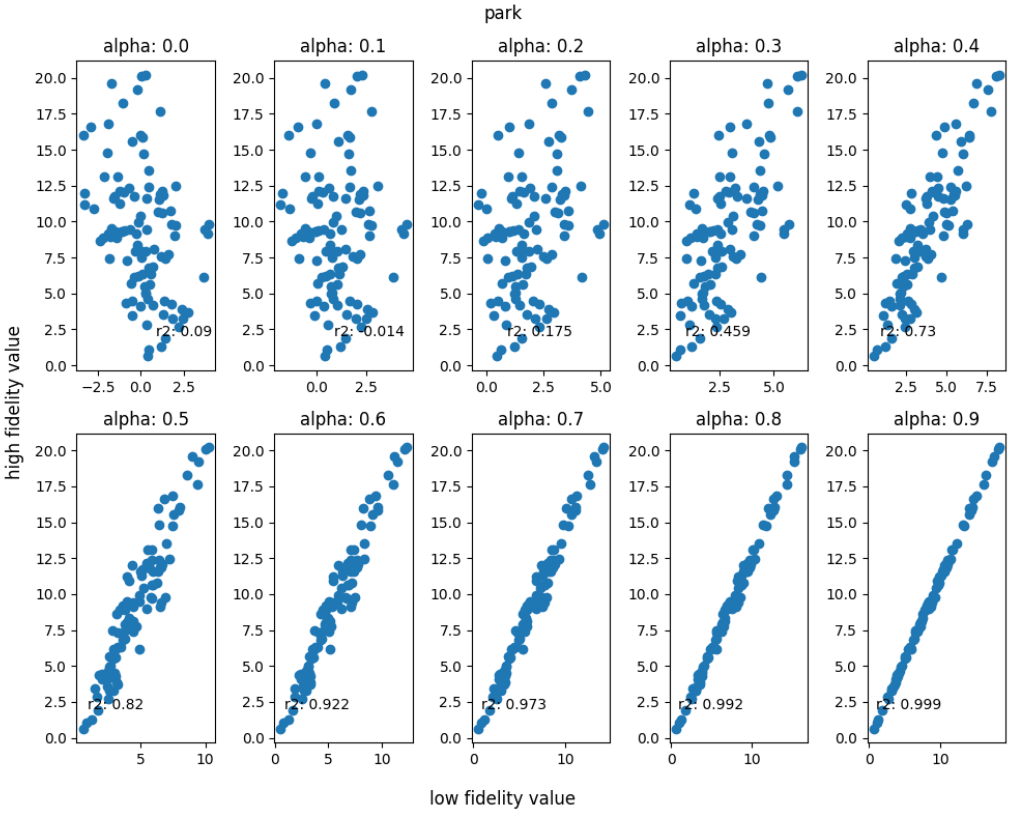

Figure 8: Informativeness of Park-2D LF approximation varying with the α parameter

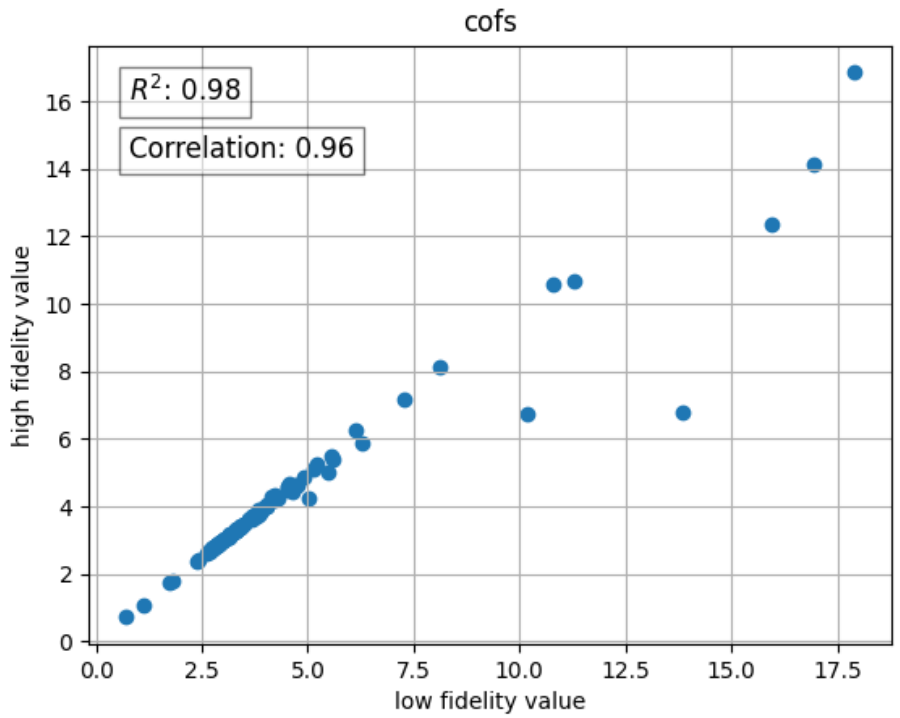

Figure 9: Informativeness of COFs LF approximation

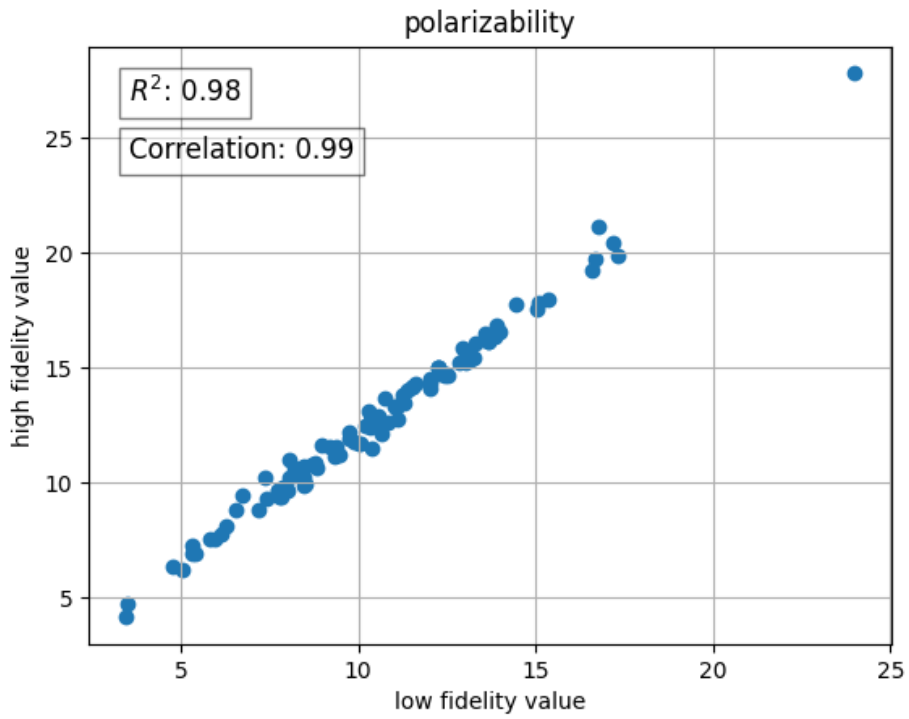

Figure 10: Informativeness of polarizability LF approximation

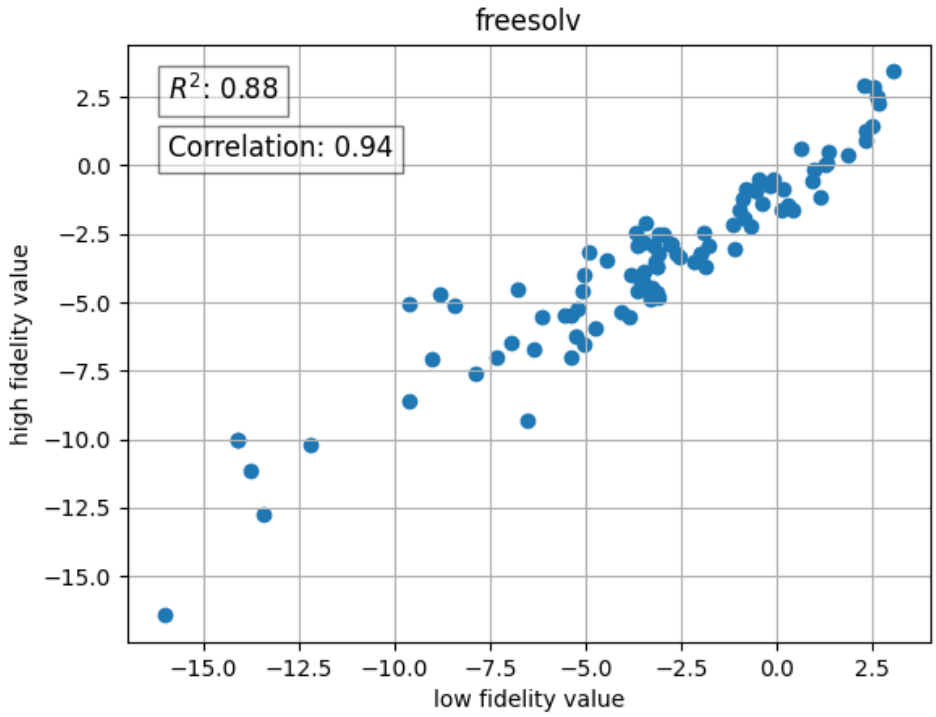

Figure 11: Informativeness of solvation energy LF approximation