# AN EXPERIMENT-AWARE BAYESIAN OPTIMIZATION WORKFLOW FOR NOISY MIXED-INPUTS SETTINGS

**Yuhao Zhang**
School of Natural Sciences, Physics Department
Technical University of Munich
85748 Garching, Germany
Munich Center for Machine Learning (MCML)
85747 Garching, Germany
Department of Applied Physics
Aalto University
02150 Espoo, Finland
yuhao2.zhang@tum.de

**Ti John**
Department of Computer Science
Aalto University
02150 Espoo, Finland
ti.john@aalto.fi

**Matthias Stosiek**
School of Natural Sciences, Physics Department
Technical University of Munich
85748 Garching, Germany
Munich Center for Machine Learning (MCML)
85747 Garching, Germany
matthias.stosiek@tum.de

**Patrick Rinke**
School of Natural Sciences, Physics Department
Technical University of Munich
85748 Garching, Germany
Atomistic Modeling Center, Munich Data Science Institute
Technical University of Munich
85748 Garching, Germany
Munich Center for Machine Learning (MCML)
85747 Garching, Germany
patrick.rinke@tum.de

## ABSTRACT

Optimizing expensive black-box objectives over mixed and discretized search spaces is a common challenge in experimental materials science. Bayesian optimization (BO) offers sample-efficient decision-making, but in highly discretized and noisy experimental settings, standard GP-based BO workflows frequently stagnate due to repeated sampling of identical or near-identical conditions. Although alternative surrogate models for mixed-variable BO exist, GP-based surrogates remain most effective for BO in the low-data regime of experimental settings. We therefore present an Experiment-Aware Bayesian Optimization (EABO) workflow that directly targets resampling-induced stagnation while retaining the inherent benefits of GP uncertainty modelling. Rather than replacing GP surrogates, our approach focuses on mitigating this dominant practical failure mode encountered in experimental deployments. The workflow introduces explicit stagnation detection and a dynamic exploration fallback that enforces global exploration when redundant acquisitions occur. We demonstrate its effectiveness on noisy, highly discretized benchmark problems representative of mixed search spaces encountered in experimental materials science. Our work thereby provides a ready-to-use tool for mixed-variable experimental materials design and autonomous laboratories.

## 1 INTRODUCTION

Optimization problems within material science often involve black-box objectives with high evaluation costs, such as time-consuming and expensive laboratory experiments. Sun et al. (2021b); Löfgren et al. (2022); Lampe et al. (2023); Pedersen et al. (2021); Zhang et al. (2025); Sanin et al. (2025); Diment et al. (2025); Miranda-Valdez et al. (2025). In these settings, the goal is to minimize the number of experiments required to optimize a target property as a function of processing or synthesis conditions. Such optimization tasks commonly rely on design of experiments (DOE) method-

ologies or sampling strategies based on experience and intuition Montgomery (2017); Greenhill et al. (2020). For multidimensional design spaces, experience- and intuition-based sampling methods are subject to bias and lack of exploration leading to the risk of missing the optimum solution. Meanwhile, conventional DOE approaches sample the design space in a predetermined fashion which prevents information gained from new measurements to improve the sampling strategy.

BO addresses these limitations by providing an adaptive, data-efficient alternative that balances exploration and exploitation through a probabilistic surrogate model and acquisition function (AF). This adaptivity has made BO particularly attractive for experimental and computational settings where evaluations are costly, and has led to its widespread adoption in materials discovery Sun et al. (2021a); Wu et al. (2024); Bisbo & Hammer (2020); Lampe et al. (2023). For fully continuous spaces GPs are commonly used as surrogate models due to their flexibility and principled Bayesian uncertainty quantification, which explicitly models observation noise separately from predictive uncertainty and is well suited to data-scarce and noisy experimental settings. This is particularly the case when the underlying objective function can be reasonably approximated as smooth, leveraging the naturally smooth posterior predictions induced by kernel covariance functions Rasmussen & Williams (2006).

However, experimental optimization problems faced in materials science are not always continuous with smooth objective landscapes. From factors such as experimental limitations, search spaces are typically mixed, involving continuous, integer, discrete, and categorical variables. Throughout this work, we consider mixed variable search spaces containing both continuous and non continuous variables. The latter includes integer, discrete, and categorical variables, where discrete variables take values from finite sets that are not necessarily equidistant, in contrast to integer variables.

While alternative surrogates models such as tree-based methods and Bayesian Neural Networks can naturally extend BO to mixed-variable domains, prior studies have shown that GP surrogates combined with appropriate variable transformation strategies often achieve stronger empirical performance within BO Daulton et al. (2022); Zhang et al. (2022); Cuesta Ramirez et al. (2022); Allec & Ziatdinov (2025). However, studies on mixed-variable BO, including GP-based approaches, are most commonly evaluated on noiseless and synthetic benchmarks that do not reflect experimental realities. As a result, the behavior of BO in noisy, mixed experimental settings remains insufficiently characterized, and failure modes specific to such conditions can go unnoticed.

In this work, we focus on a failure mode that is particularly problematic in experimental settings, namely repeated sampling of identical or near identical conditions caused by the interaction between observation noise, non continuous search spaces, and discontinuous objective landscapes. Unlike continuous domains, discrete spaces lack a natural neighborhood structure that allows BO to exploit locally while making progress, and repeated evaluations can lead to stagnation or wasted experimental budget. Despite its practical importance, this issue is rarely addressed explicitly in existing mixed variable BO workflows.

To address this challenge, we present the EABO workflow shown in Figure 1, designed to handle mixed and highly discontinuous objective landscapes encountered in materials science. Our workflow builds on a state of the art GP implementation for mixed variable search spaces proposed by Daulton et al. (2022), utilizes kernel constructions that have demonstrated strong performance in materials optimization problems, and introduces explicit mechanisms to detect stagnation and dynamically enforce exploration. By targeting workflow level failure modes induced by noisy and discretized experimental settings, the proposed approach provides a practical and ready-to-use solution for deploying BO in real-world experiments and autonomous laboratory settings.

## 2 Experiment-Aware Bayesian Optimization Workflow

### 2.1 Mixed Space Surrogate Modeling for Experiments

To apply GP-based BO in mixed-variable settings, a common strategy is to transform the original mixed search space into a continuous latent space. This transformation enables the exploitation of smoothness correlations encoded by the GP covariance function and facilitates gradient-based optimization of the AF, see A.1 Appendix for more details. Such reparameterization-based approaches are widely adopted in mixed-variable materials optimization due to their compatibility with struc-

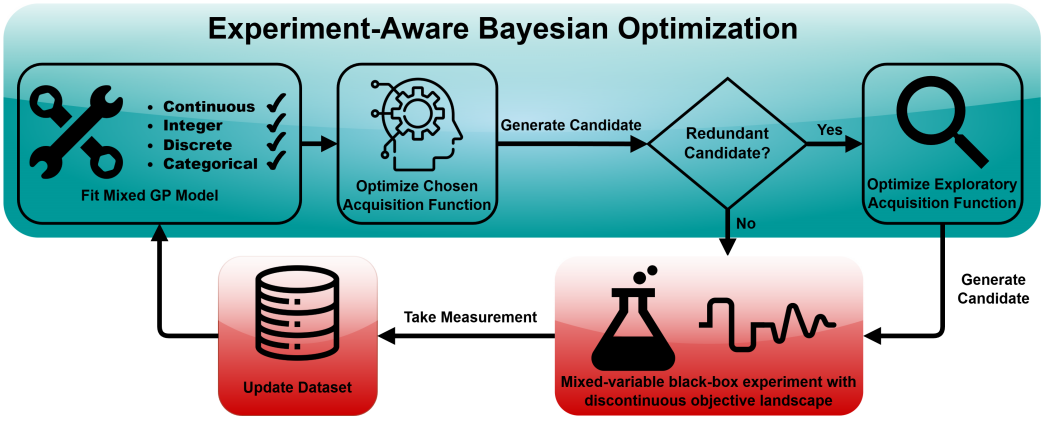


Figure 1: Schematic of the proposed EABO workflow for noisy and discontinuous mixed-variable optimization problems. The mixed GP surrogate is fitted to noisy experimental observations from a non-continuous search space. A redundancy check at the candidate generation stage enables temporary switching to a pure exploration strategy when repeated or near-duplicate candidates are proposed.

tured kernel designs and existing BO frameworks Cuesta Ramirez et al. (2022); Zhang et al. (2022; 2020). For this purpose, we implement the Probabilistic Reparameterization (PR) method proposed by Daulton et al. (2022), a state-of-the-art approach that has been shown to outperform alternative GP-based variable transformation strategies for GP-based BO in mixed-variable settings. Within our framework, we extend this PR method to explicitly handle discrete variables, enabling support for continuous, integer, discrete, and categorical variables, all commonly encountered in experimental materials optimization tasks. Throughout this work, we refer to our generalized PR-based surrogate model as mixed GP.

Building on our mixed GP foundation, we adopt the default product kernel formulation used by the BOSS package Todorović et al. (2019), which constructs a composite kernel by assigning a kernel to each input dimension and combining them multiplicatively. This kernel design was specifically designed to be flexible and tailored for objective landscapes commonly encountered in materials optimization and has been successfully deployed across a wide range of experimental and computational studies, including polymer and perovskite systems Todorović et al. (2019); Zhang et al. (2025); Li et al. (2024). Details of the kernel construction are provided in section A.3 Appendix.

The choice of our surrogate model and its kernel formulation within the EABO workflow is guided by two considerations central to experimental optimization tasks. First, our mixed GP surrogate is builds upon the state-of-the-art PR framework for mixed-variable modeling, and is generalized to additionally handle discrete variables. This GP-based formulation encodes smoothness assumptions that remain informative across many materials science problems, particularly those involving multiple continuous processing variables, even when the search space is partially non-continuous. Second, the BOSS kernel design has been empirically validated in a wide range of experimental materials optimization tasks Zhang et al. (2025); Iannacchero et al. (2025); Löfgren et al. (2022), providing a robust and well-tested foundation for the proposed workflow.

### 2.2 Experiment-Aware Bayesian Optimization Loop

Under non-negligible observation noise, resampling in non-continuous optimization spaces has been observed in the PR approach of Daulton et al. (2022), as well as in alternative mixed and discrete GP-based methods reported in experimental studies Garrido-Merchán & Hernández-Lobato (2020); Zhang et al. (2025). Importantly using standard AFs, this behavior is implementation-agnostic and inherent to GP-based surrogates under noisy conditions in purely discrete domains. In such settings, non-negligible observation noise results in a residual predictive variance at previously evaluated points, which may cause AFs to repeatedly select the same configurations and lead to stagnation. Without explicit mechanisms to handle such failure cases, the performance gains offered by state of the art mixed-variable GP surrogates may not translate into effective optimization in practice,

as repeated sampling can lead to stagnation. A more detailed explanation with example of the resampling behaviour is provided in section A.4 Appendix.

To address this issue, we modify the standard BO loop by introducing a minimum Euclidean distance threshold between newly proposed candidate points and previously evaluated configurations. When the AF proposes a candidate within this threshold, it is deemed redundant, as illustrated in Figure 1, and replaced by a new candidate generated using a purely exploratory AF. This exploratory step selects the feasible point with the highest predictive uncertainty, ensuring global exploration and preventing repeated sampling of identical or near identical configurations.

## 3 EXPERIMENTS

To evaluate the proposed EABO workflow, we focus on isolating the impact of workflow-level modifications rather than benchmarking surrogate model performance. Accordingly, we evaluate EABO on two highly discontinuous synthetic test functions, Discontinuous Unsmoothed Step-like Test 1 and Test 2 (*DUST1* and *DUST2*) which are visualized in Figures 4 and 5 in the Appendix. These benchmarks are designed to emulate real-world objective landscapes in mixed spaces that exhibit large flat regions and step-like features, which commonly arise in experimental materials optimization due to effects such as phase transitions or discretization imposed by experimental constraints. *DUST2* is designed to be more complex than *DUST1*, featuring a larger design space with a greater number of discrete levels.

Although no observation noise was injected into the DUST benchmark evaluations, we initialized the mixed GP surrogate with a fixed noise variance of 2. This represents a non-negligible noise level relative to the objective ranges and is chosen to reflect noise magnitudes commonly encountered in experimental settings when fitting surrogate model hyperparameters to real, noisy data. This choice is substantially larger than values (e.g., $10^{-6}$-$10^{-3}$) typically used in other BO benchmarks, and therefore provides a more realistic stress test of BO behavior under experimental conditions.

To evaluate the proposed EABO workflow we included vanilla GP-based BO, Sobol sampling, and RF-surrogate BO into the benchmarks. All methods use identical Sobol initializations and are evaluated over 10 runs with shared random seeds for generating the initial dataset. The commonly used Expected Improvement (EI) and Lower Confidence Bound (LCB) AFs were considered for all surrogate models, and performance is measured by convergence and final regret. The RF surrogate is used with default scikit-learn settings (200 trees) to reflect realistic experimental deployment under limited data.

## 4 RESULTS AND DISCUSSION

Figure 2 shows the convergence behavior of the evaluated methods on the *DUST1* and *DUST2* benchmarks. Across both benchmarks, the vanilla BO workflow using the mixed GP surrogate (MGP+V_BO) exhibits a consistent failure mode where the optimizer becomes trapped in local minima and repeatedly samples identical or near-identical configurations. Resampling behavior is observed for both the Expected Improvement (EI) and Lower Confidence Bound (LCB) AFs. Consequently, despite ten independently initialized runs, the optimization fails to progress beyond the initial iterations and is outperformed by Sobol sampling at later iterations as seen in Figure 2 (a-d). The resulting convergence behavior is comparable to that of BO using a random forest surrogate, which likewise underperforms relative to Sobol sampling. Additionaly the relatively small design space of *DUST1*, characterized by a limited number of discrete levels, allows Sobol sampling to achieve competitive performance. In such settings, the search space can be covered efficiently even by quasi-random sampling. However, for the more complex *DUST2*, the advantage of BO becomes more pronounced as evidenced by steeper initial decrease in the regret curves of the BO models due to their ability to guide the search more effectively.

In contrast, the proposed EABO workflow (MGP+EA_BO) consistently avoids this resampling behavior, as exemplified in Figure 6. By detecting near-duplicate candidate proposals and temporarily switching to a purely exploratory acquisition function, EABO enforces global exploration when standard acquisition optimization fails. This mechanism enables the optimization to escape local minima and continue making progress across the design space, resulting in consistently superior

convergence compared to the vanilla BO workflow (MGP+V_BO) on both *DUST1* and the more challenging *DUST2* benchmarks. Notably, EABO also outperforms the RF surrogate-based BO as well as Sobol sampling. Importantly, EABO does not degrade performance on well-behaved optimization problems. In regimes where the objective landscape is sufficiently smooth or with little discretization, redundant sampling rarely occurs and the exploratory safeguard is seldom activated. In these cases, EABO recovers the performance of standard BO. Thus, the proposed method acts as an additional failsafe rather than a replacement for standard BO workflows, preserving exploitation efficiency while preventing continuous resampling.

As seen in Figure 2, we observe a characteristic low-iteration supremacy effect for MGP+V_BO workflows, where they initially converge faster than their EABO counterparts under both EI and LCB AFs. This behavior is more pronounced on *DUST2* and reflects the ability of BO to efficiently exploit local structure early in the optimization. However, this advantage is short-lived as once the optimizer enters a local minimum, the absence of redundancy handling leads to repeated sampling. In contrast, EABO trades some early exploitation for sustained progress, yielding superior long-term convergence. Across both benchmarks, with the exception of the RF surrogate, EI- and LCB-based BO workflows exhibit comparable convergence behavior, suggesting that in highly discontinuous and discretized settings, performance differences between acquisition function choices are negligible when using GP-based surrogates.

Overall, the DUST benchmarks represent a challenging stress test that mimics settings that may be encountered in mixed-space materials optimization problems. Despite the highly discretized and discontinuous nature of these landscapes, the EABO workflow consistently demonstrates robust convergence behavior. Crucially, EABO enables direct deployment of mixed-GP-based BO without prior knowledge or manual assessment of the optimization landscape. These results highlight the importance of workflow-level safeguards for deploying BO in noisy, non-continuous experimental settings, where even state-of-the-art surrogate quality alone is insufficient to guarantee effective optimization.

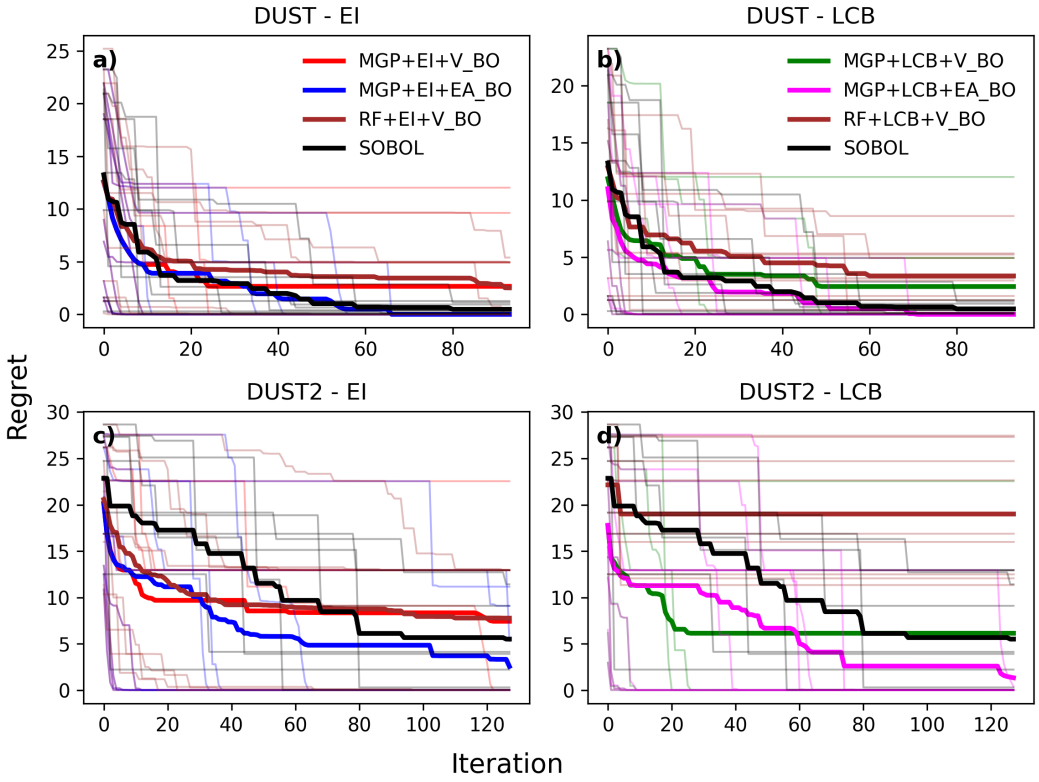


Figure 2: Convergence plots on the *DUST1* and *DUST2* benchmark functions. Performance of the proposed EABO workflow using a mixed GP surrogate (MGP+EA_BO) compared to a standard vanilla BO loop with the same surrogate (MGP+V_BO). Results are shown for both Expected Improvement (EI) and Lower Confidence Bound (LCB) AFs, indicated within the workflow names. Sobol sampling (black) and a random forest surrogate baseline (maroon) are included for reference. Bold curves denote the mean convergence across 10 runs initialized from different Sobol point sets, with thin curves indicating individual runs.

## 5 Conclusion

In this work, we identified a practical yet underexplored failure mode of GP-based BO in noisy, mixed-variable, experimental settings with discontinuous landscapes in the form of repeated sampling of identical or near-identical configurations. While recent advances in mixed-variable GP surrogates have improved modeling capability, our results show that surrogate quality alone is insufficient to guarantee reliable optimization performance in such regimes.

To address this issue, we propose EABO, which augments standard BO loops with explicit redundancy detection and a dynamic exploration fallback. When redundant candidates are proposed, the workflow temporarily switches to a purely exploratory acquisition strategy, preventing stagnation while maintaining informative sampling for GP model updates. Benchmark results on the DUST functions demonstrate that EABO consistently outperforms standard BO workflows and exhibits robust convergence behavior under conditions representative of those encountered in materials science involving mixed search spaces and discontinuous objective landscapes.

The proposed workflow is intentionally simple, surrogate-agnostic, and compatible with existing BO extensions, including batch acquisition, multiobjective optimization, and multi-fidelity settings. As such, it provides a practical tool for deploying BO in autonomous laboratories and other real-world optimization scenarios where noise, mixed spaces, and limited data are unavoidable. Future work should explore adaptive redundancy thresholds, more flexible exploration–exploitation schedules, and integration with human-in-the-loop decision strategies.

## REFERENCES


Sarah I. Allec and Maxim Ziatdinov. Active and transfer learning with partially bayesian neural networks for materials and chemicals. *Digital Discovery*, 4:1284–1297, 2025. doi: 10.1039/D5DD00027K. URL http://dx.doi.org/10.1039/D5DD00027K.

Malthe K. Bisbo and Bjørk Hammer. Efficient global structure optimization with a machine-learned surrogate model. *Phys. Rev. Lett.*, 124:086102, 02 2020. doi: 10.1103/PhysRevLett.124.086102. URL https://link.aps.org/doi/10.1103/PhysRevLett.124.086102.

John Cook. Determining distribution parameters from quantiles. *UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series*, 01 2010.

Jhouben Cuesta Ramirez, Rodolphe Le Riche, Olivier Roustant, Guillaume Perrin, Cédric Durantin, and Alain Glière. A comparison of mixed-variables bayesian optimization approaches. *Advanced Modeling and Simulation in Engineering Sciences*, 9(1):6, 2022. doi: 10.1186/s40323-022-00218-8. URL https://doi.org/10.1186/s40323-022-00218-8.

Samuel Daulton, Xingchen Wan, David Eriksson, Maximilian Balandat, Michael A. Osborne, and Eytan Bakshy. Bayesian optimization over discrete and mixed spaces via probabilistic reparameterization. *arXiv preprint arXiv:2210.10199*, 2022. URL https://arxiv.org/abs/2210.10199.

Daryna Diment, Joakim Löfgren, Marie Alopaeus, Matthias Stosiek, MiJung Cho, Chunlin Xu, Michael Hummel, Davide Rigo, Patrick Rinke, and Mikhail Balakshin. Enhancing lignin-carbohydrate complexes production and properties with machine learning. *ChemSusChem*, 18(8):e202401711, 2025. doi: https://doi.org/10.1002/cssc.202401711. URL https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cssc.202401711.

Eduardo C. Garrido-Merchán and Daniel Hernández-Lobato. Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing*, 380:20–35, 2020. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2019.11.004. URL https://www.sciencedirect.com/science/article/pii/S0925231219315619.

Stewart Greenhill, Santu Rana, Sunil Gupta, Pratibha Vellanki, and Svetha Venkatesh. Bayesian optimization for adaptive experimental design: A review. *IEEE Access*, PP:1–1, 01 2020. doi: 10.1109/ACCESS.2020.2966228.

Matteo Iannacchero, Joakim Löfgren, Mithila Mohan, Patrick Rinke, and Jaana Vapaavuori. Machine learning-assisted development of polypyrrole-grafted yarns for e-textiles. *Materials & Design*, 249:113528, 2025. ISSN 0264-1275. doi: https://doi.org/10.1016/j.matdes.2024.113528. URL https://www.sciencedirect.com/science/article/pii/S0264127524009031.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106.

Carola Lampe, Ioannis Kouroudis, Milan Harth, Stefan Martin, Alessio Gagliardi, and Alexander S. Urban. Rapid data-efficient optimization of perovskite nanocrystal syntheses through machine learning algorithm fusion. *Advanced Materials*, 35(16):2208772, 2023. doi: https://doi.org/10.1002/adma.202208772. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202208772.

Jingrui Li, Fang Pan, Guo-Xu Zhang, Zenghui Liu, Hua Dong, Dawei Wang, Zhuangde Jiang, Wei Ren, Zuo-Guang Ye, Milica Todorović, and Patrick Rinke. Structural disorder by octahedral tilting in inorganic halide perovskites: New insight with bayesian optimization. *Small Structures*, 5(11):2400268, 2024. doi: https://doi.org/10.1002/sstr.202400268. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sstr.202400268.

Joakim Löfgren, Dmitry Tarasov, Taru Koitto, Patrick Rinke, Mikhail Balakshin, and Milica Todorović. Machine learning optimization of lignin properties in green biorefineries. *ACS Sustainable Chemistry & Engineering*, 10(29):9469–9479, 2022. doi: 10.1021/acssuschemeng.2c01895. URL https://doi.org/10.1021/acssuschemeng.2c01895.

Isaac Y. Miranda-Valdez, Tero Mäkinen, Juha Koivisto, and Mikko J. Alava. Bayesian optimization to infer parameters in viscoelasticity. *Journal of Rheology*, 69(6):1059–1066, 10 2025. ISSN 0148-6055. doi: 10.1122/8.0001068. URL https://doi.org/10.1122/8.0001068.

D.C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, Incorporated, 2017. ISBN 9781119113478. URL https://books.google.fi/books?id=Py7bDgAAQBAJ.

Jack K. Pedersen, Christian M. Clausen, Olga A. Krysiak, Bin Xiao, Thomas A. A. Batchelor, Tobias Löffler, Vladislav A. Mints, Lars Banko, Matthias Arenz, Alan Savan, Wolfgang Schuhmann, Alfred Ludwig, and Jan Rossmeisl. Bayesian optimization of high-entropy alloy compositions for electrocatalytic oxygen reduction. *Angewandte Chemie International Edition*, 60(45):24144–24152, 2021. doi: https://doi.org/10.1002/anie.202108116. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202108116.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. ISBN 026218253X. URL http://www.GaussianProcess.org/gpml.

Binxin Ru, Ahsan Alvi, Vu Nguyen, Michael A. Osborne, and Stephen Roberts. Bayesian optimisation over multiple continuous and categorical inputs. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8276–8285. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/ru20a.html.

Alexey Sanin, Jackson K. Flowers, Tobias H. Piotrowiak, Frederic Felsen, Leon Merker, Alfred Ludwig, Dominic Bresser, and Helge Sören Stein. Integrating automated electrochemistry and high-throughput characterization with machine learning to explore si-ge-sn thin-film lithium battery anodes. *Advanced Energy Materials*, 15(11):2404961, 2025. doi: https://doi.org/10.1002/aenm.202404961. URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aenm.202404961.

Shijing Sun, Armi Tiihonen, Felipe Oviedo, Zhe Liu, Janak Thapa, Yicheng Zhao, Noor Titan P. Hartono, Anuj Goyal, Thomas Heumueller, Clio Batali, Alex Encinas, Jason J. Yoo, Ruipeng Li, Zekun Ren, I. Marius Peters, Christoph J. Brabec, Moungi G. Bawendi, Vladan Stevanovic, John Fisher, and Tonio Buonassisi. A data fusion approach to optimize compositional stability of halide perovskites. *Matter*, 4(4):1305–1322, 2021a. ISSN 2590-2385. doi: https://doi.org/10.1016/j.matt.2021.01.008. URL https://www.sciencedirect.com/science/article/pii/S2590238521000084.

Shijing Sun, Armi Tiihonen, Felipe Oviedo, Zhe Liu, Janak Thapa, Yicheng Zhao, Noor Titan P. Hartono, Anuj Goyal, Thomas Heumueller, Clio Batali, Alex Encinas, Jason J. Yoo, Ruipeng Li, Zekun Ren, I. Marius Peters, Christoph J. Brabec, Moungi G. Bawendi, Vladan Stevanovic, John Fisher, and Tonio Buonassisi. A data fusion approach to optimize compositional stability of halide perovskites. *Matter*, 4(4):1305–1322, 2021b. ISSN 2590-2385. doi: https://doi.org/10.1016/j.matt.2021.01.008. URL https://www.sciencedirect.com/science/article/pii/S2590238521000084.

Milica Todorović, Michael U. Gutmann, Jukka Corander, and Patrick Rinke. Bayesian inference of atomistic structure in functional materials. *npj Computational Materials*, 5(1):35, 3 2019. ISSN 2057-3960. doi: 10.1038/s41524-019-0175-2. URL https://doi.org/10.1038/s41524-019-0175-2.

Yifan Wu, Aron Walsh, and Alex M. Ganose. Race to the bottom: Bayesian optimisation for chemical problems. *Digital Discovery*, 3(6):1086–1100, 2024. doi: 10.1039/D3DD00234A. URL http://dx.doi.org/10.1039/D3DD00234A.

Hengrui Zhang, Wei (Wayne) Chen, Akshay Iyer, Daniel W. Apley, and Wei Chen. Uncertainty-aware mixed-variable machine learning for materials design. *Scientific Reports*, 12(1):19760, Nov 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-23431-2. URL https://doi.org/10.1038/s41598-022-23431-2.

Yichi Zhang, Daniel W. Apley, and Wei Chen. Bayesian optimization for materials design with mixed quantitative and qualitative variables. *Scientific Reports*, 10(1):4924, 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-60652-9. URL https://doi.org/10.1038/s41598-020-60652-9.

Yuhao Zhang, Maija Vaara, Azin Alesafar, Duc Bach Nguyen, Pedro Silva, Laura Koskelo, Jussi Ristolainen, Matthias Stosiek, Joakim Löfgren, Jaana Vapaavuori, and Patrick Rinke. Data-efficient optimization of thermally-activated polymer actuators through machine learning. *Materials & Design*, 253:113908, 2025. ISSN 0264-1275. doi: https://doi.org/10.1016/j.matdes.2025.113908. URL https://www.sciencedirect.com/science/article/pii/S0264127525003284.

# A Appendix

## A.1 Gaussian Process for Mixed Spaces

In this section, we provide the GP posterior formulation used used for mixed variable BO tasks. We consider the problem of optimizing a black-box objective function $f : \mathcal{X} \times \mathcal{Q} \to \mathbb{R}$ over a bounded mixed search space $\mathcal{X} \times \mathcal{Q}$, where $\mathcal{X} = \mathcal{X}^{(1)} \times \cdots \times \mathcal{X}^{(c)}$, is the domain of $c \geq 0$ continuous parameters and $\mathcal{Q} = \mathcal{Q}^{(1)} \times \cdots \times \mathcal{Q}^{(d)}$ is the domain of $d \geq 0$ non-continuous variables (integer, discrete and categorical). As a physical example, consider optimizing a multilayer material for a desired property such as strength, conductivity, or transparency. Integer variables may represent the number of layers, categorical variables may correspond to material type, stacking order, or substrate, and discrete variables may take values from a predefined set, such as layer thicknesses or process parameters constrained by fabrication limits. The AF $\alpha : \mathcal{X} \times \mathcal{Q} \to \mathbb{R}$ assigns a real value to each candidate pair $(\boldsymbol{x}, \boldsymbol{q})$ where $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{q} \in \mathcal{Q}$. Maximizing $\alpha$ in this mixed space is therefore necessary but presents a non-trivial challenge. Generally mixed variable BO with GPs perform the following: let $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^m$ be continuous variables, where $m \geq c + d$. We then define a mapping:

$$g : \Theta \to \mathcal{X} \times \mathcal{Q}, \quad \boldsymbol{\theta} \mapsto (\boldsymbol{x}, \boldsymbol{q}), \tag{1}$$

where $\Theta \subseteq \mathbb{R}^m$ is a continuous search space, and $(\boldsymbol{x}, \boldsymbol{q}) \in \mathcal{X} \times \mathcal{Q}$ denotes the corresponding mixed-variable configuration. Using this mapping, we define a reparameterized AF:

$$\tilde{\alpha}(\boldsymbol{\theta}) = \alpha(g(\boldsymbol{\theta})) = \alpha(\boldsymbol{x}, \boldsymbol{q}), \tag{2}$$

which allows us to optimize the AF entirely in the continuous domain $\Theta$ using a GP with kernel $\kappa$. Let $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N] \in \Theta^N$ denote the matrix of training inputs, and $\boldsymbol{\Theta}_* = [\boldsymbol{\theta}_{*1}, \ldots, \boldsymbol{\theta}_{*N_*}] \in \Theta^{N_*}$ the test inputs. In this domain the predictive posterior mean and covariance on the test inputs is given by:

$$\mu_* = \mathbf{k}_*^T \left(\mathbf{K} + \sigma_0^2 \mathbf{I}\right)^{-1} \mathbf{y}. \tag{3}$$

$$\Sigma_* = k_{**} - \mathbf{k}_*^T \left(\mathbf{K} + \sigma_0^2 \mathbf{I}\right) \mathbf{k}_*, \tag{4}$$

where $\mathbf{k}_* = \kappa(\boldsymbol{\Theta}, \boldsymbol{\Theta}_*) \in \mathbb{R}^{N \times N_*}$ is the covariance matrix between training and test inputs, $k_{**} = \kappa(\boldsymbol{\Theta}_*, \boldsymbol{\Theta}_*) \in \mathbb{R}^{N_* \times N_*}$ is the covariance matrix between test inputs, $\mathbf{K} = \kappa(\boldsymbol{\Theta}, \boldsymbol{\Theta}) \in \mathbb{R}^{N \times N}$ is the covariance matrix between training inputs and $\sigma_0^2$ is the fitted noise (variance) hyperparameter that models the underlying noise (aleatoric uncertainty) within the data.

AFs are optimized in the continuous latent space $\Theta$, with candidate points mapped back to the mixed-variable domain via the transformation $g$.

### A.2 GENERALIZED PROBABILISTIC REPARAMETERIZATION

In relation to the mapping $g$ presented in Section A.1, the probabilistic reparameterization method induces a continuous latent space transformation between $\mathcal{Q}$ and $\Theta$ via a discrete probability distribution $p(\boldsymbol{Q} \mid \boldsymbol{\theta})$ over integer or categorical random variables $\boldsymbol{Q}$. This distribution is parameterized by a vector of continuous parameters $\boldsymbol{\theta}$. Given this reparameterizaion, the probabilistic objective for is defined as:

$$\mathbb{E}_{\boldsymbol{Q} \sim p(\boldsymbol{Q} \mid \boldsymbol{\theta})} \left[ \alpha \left( \boldsymbol{x}, \boldsymbol{Q} \right) \right] = \sum_{\boldsymbol{z} \in \mathcal{Z}} p \left( \boldsymbol{q} \mid \boldsymbol{\theta} \right) \alpha \left( \boldsymbol{x}, \boldsymbol{Q} \right), \tag{5}$$

where $\alpha$ is the AF. Since $p \left( \boldsymbol{Q} \mid \boldsymbol{\theta} \right)$ is a discrete probability distribution, its expectation can be expressed as a linear combination where each discrete option is weighted by its probability mass. The authors have formally proven that maximizing this PO for $(\boldsymbol{x}, \boldsymbol{\theta})$ is equivalent to maximizing $\alpha$ over the original mixed-variable space $\mathcal{X} \times \mathcal{Q}$. Also, by choosing $p \left( \boldsymbol{Q} \mid \boldsymbol{\theta} \right)$ to be a discrete distribution over $\mathcal{Z}$ means that all sampled instances $\boldsymbol{q}$ are feasible non-continuous values given any $\boldsymbol{\theta}$.

The original PR implementation already handles binary, integer (termed ordinal in the original study), and categorical variables. Our generalized PR method directly adopts their formulation without changes and extends it to additionally handle discrete variables. Table 1 summarizes the full reparameterizations for all four variable types used by the PO, as defined in equation 5.

Table 1: Summary of probabilistic reparameterizations for different variable types. We denote the $(C-1)$-simplex as $\Delta^{C-1}$.

| **Parameter type** | **Random variable** | **Continuous parameter** |
|---|---|---|
| Binary | $Q \sim B\big(b(\theta)\big)$ | $\theta \in [0, 1]$ |
| Integer | $Q = \lfloor \theta \rfloor + B\big(\mathcal{I}(\theta)\big)$ | $\theta \in [0, I-1]$ |
| Discrete | $Q = d_i + (d_{i+1} - d_i)\, B\big(\mathcal{D}(\theta)\big)$ | $\theta \in \{d_1, d_2, \ldots, d_D\}$ |
| Categorical | $Q \sim \mathrm{CAT}\big(C(\boldsymbol{\theta})\big),\ \boldsymbol{\theta} = (\theta^{(1)}, \ldots, \theta^{(C)})$ | $\theta \in \Delta^{C-1}$ |

Within Table 1 $B$ denotes a Bernoulli random variable, $I$ and $D$ represents the number of allowed integer and discrete levels for the respective parameter type. Before sampling the mixed variable parameters, the optimized continuous parameter $\theta$ (or $\boldsymbol{\theta}$ for categorical variables) is passed through the following functions: $b\left(\theta\right) = \sigma\left(\left(\theta - \frac{1}{2}\right)/\tau\right)$, $\mathcal{I}\left(\theta\right) = \sigma\left(\left(\theta - \lfloor\theta\rfloor - 0.5\right)/\tau\right)$, $\mathcal{D}\left(\theta\right) = \sigma\left(\left(\theta - d_i - \frac{d_{i+1}-d_i}{2}\right)/\tau\right)$ and $C\left(\boldsymbol{\theta}\right) = \mathrm{softmax}\left(\left(\boldsymbol{\theta} - 0.5\right)\tau\right)$ corresponding to the binary, ordinal, discrete, and categorical parameter types, respectively where $\sigma$ is the sigmoid function. $\tau$ is the temperature parameter set to 0.1 for all variable types following the original PR implementation. For the discrete variables, $d_i$ and $d_{i+1}$ are the allowed discrete levels such that $d_i \leq \theta \leq d_{i+1}$, the discrete formulation is therefore analogous to the integer case but takes into account non-equal distances.

We also note that in the original PR implementation, the reparameterizations described above apply only when sampling $Q$ from already optimized $\theta$ values. During backpropagation, a differentiable $\tanh$ function is used in place of the non-differentiable Bernoulli distribution. For consistency, we follow the same procedure. The Adam optimizer was used to perform gradient descent on $\boldsymbol{\theta}$ Kingma & Ba (2014).

### A.3 BOSS KERNEL

Table 2: Matérn-5/2 kernel specification with product formulation and Gamma priors.

| Kernel property | Specification |
|---|---|
| Kernel name | BOSS_on_gam_Mat52 |
| Base kernel | Matérn-$5/2$ |
| Formulation | $k_{\text{scale}}\left(k_{\text{cont}}^{\text{Prod}} \times k_{\text{int}}^{\text{Prod}} \times k_{\text{discr}}^{\text{Prod}} \times k_{\text{cat}}^{\text{ARD}}\right)$ |
| Scale prior | $\text{Gam}\left(2, \left(\frac{1}{2(\max(\boldsymbol{Y})-\min(\boldsymbol{Y}))}\right)^2\right)$ |
| Lengthscale prior | $\text{Gam}(\alpha, \lambda)$ |
| Comments | – |

$k_{\text{scale}}$ denotes a scaling kernel applied to the underlying kernel. The subscripts $cat$, $cont$, $int$, and $discr$ refer to kernels defined over categorical, continuous, integer, and discrete variables, respectively. All kernels take the form of the base kernel except for the categorical kernels. The categorical kernel ($k_{\text{cat}}$) is based on the Hamming distance, following the formulation by Ru et al. Ru et al. (2020), as employed by Daulton et al. (2022). The superscript $ARD$ indicates that the kernel employs automatic relevance determination for the corresponding variable types, assigning a separate lengthscale to each dimension. The superscript $Prod$ denotes a product kernel, where the overall kernel is constructed as the product of individual one-dimensional kernels for each dimension of the specified variable type. The shape and rate parameters, $\alpha$ and $\lambda$, are determined by specifying lower and upper quantiles using the method proposed by Cook et al. Cook (2010), with $p_1 = 0.05$ and $p_2 = 0.5$ set relative to the input bounds for the given kernel.

### A.4 RESAMPLING

We illustrate the resampling phenomenon in Figure 3 c), where during the third BO iteration of a noisy one-dimensional integer optimization problem, the AF selects a previously evaluated point. Although the GP surrogate has converged to the global optimum in this illustrative example, in practical settings the optimizer may instead become trapped at a local optimum. Once this occurs, the resampling behavior can persist indefinitely across subsequent optimization iterations, as observed in this case, unless the GP hyperparameters change sufficiently to induce a new acquisition-function maximum elsewhere in the domain.

We note that this behavior was not reported in Daulton et al. (2022), likely because their benchmarks included at least one continuous parameter and employed a very small noise hyperparameter ($10^{-6}$). In contrast, under non-negligible observation noise and fully discrete search spaces, resampling becomes significantly more pronounced.

Crucially, the resampling observed in the PR implementation arises from a different mechanism than that associated with other continuous relaxation or latent variable methods as reported by Daulton et al. (2022). It is not a limitation of the PR framework itself, but rather a consequence of the GP covariance structure in the presence of observation noise. As shown in Equation 4, the GP covariance includes a fitted noise term $\sigma_0^2$ on the diagonal, which ensures that previously sampled points retain a predictive variance that is lower bounded by the noise level. When this noise is sufficiently large, the resulting residual uncertainty contributes to the exploration term of common acquisition functions, thereby increasing the likelihood of revisiting already evaluated points.

In continuous optimization problems, such revisits typically manifest as evaluations in the immediate neighborhood of existing samples, enabling local exploitation and subsequent hyperparameter updates. However, in fully discrete domains no such neighborhood structure exists. As a result, the optimizer may repeatedly select exactly the same discrete configuration, with no natural mechanism to escape. This makes resampling particularly detrimental in discrete spaces, where it can lead to complete stagnation of the optimization process.

This failure mode is not unique to the PR framework. Similar resampling behavior is observed in the actuator optimization study of Zhang et al. (2025) when using a kernel rounding method of Garrido-

Merchán & Hernández-Lobato (2020), reinforcing the conclusion that this phenomenon is a general limitation of GP-based BO under noisy, fully discrete-variable optimization tasks.

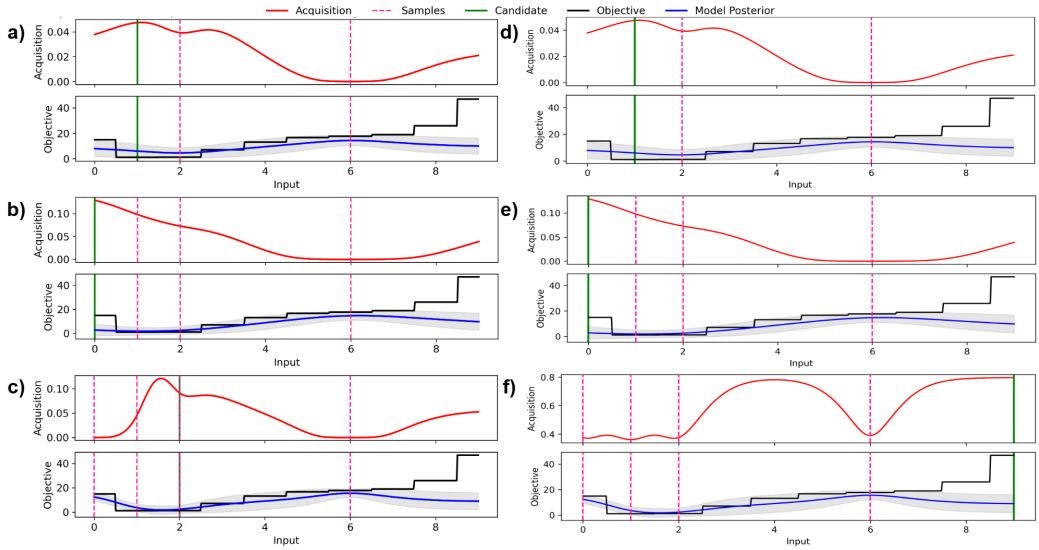


Figure 3: Mixed GP surrogate (PR implementation, noise = 0.2) BO on a dummy 1D problem. Top panels show the acquisition function and bottom panels show the posterior. Pink dashed lines indicate existing acquisitions, and the solid green line marks the proposed candidate. The standard BO workflow (left, a–c) exhibits repeated sampling at the third iteration (c), whereas the PR-based EABO workflow (right, d–f) avoids redundant evaluations by using an exploratory AF during the third iteration (f).

### A.5 DUST BENCHMARKS

*DUST1* consists of one continuous parameter with bounds $[5, 25]$, one integer parameter that is effectively binary with values $\{0, 1\}$, and one discrete parameter with allowed values $\{2, 4, 7, 8\}$. *DUST2* represents a more complex setting, with one continuous parameter bounded by $[5, 50]$, the same binary parameter, and one discrete parameter taking values $\{2, 3, 5, 6, 9, 10, 11, 12, 16, 19\}$.

For the EABO workflow, the minimum Euclidean distance threshold set for *DUST1* and *DUST2* were set to 0.1 and 0.05 respectively.

For *DUST1* and *DUST2* all models were fitted on 6 and 12 datapoints to initiate BO respectively.

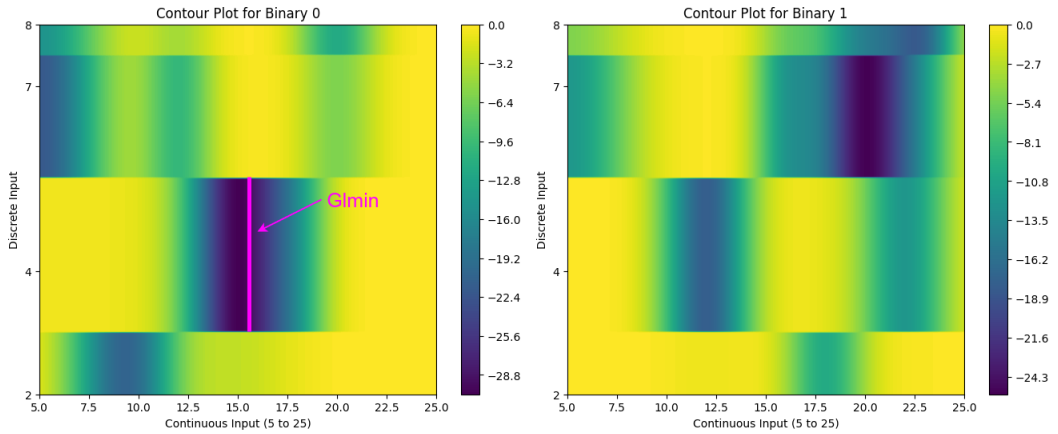


Figure 4: Objective Landscape of the *DUST1* function with the global minima (Glmin) indicated in Pink.

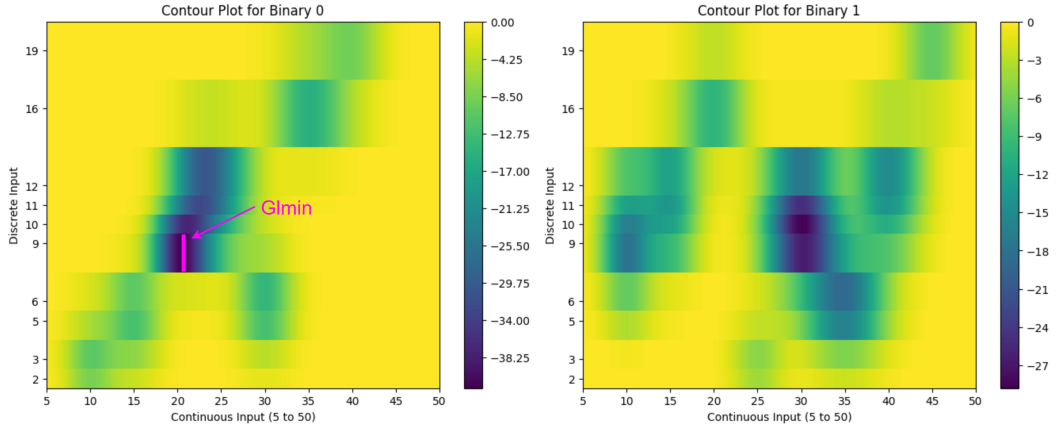


Figure 5: Objective Landscape of the *DUST2* function with the global minima (Glmin) indicated in Pink.

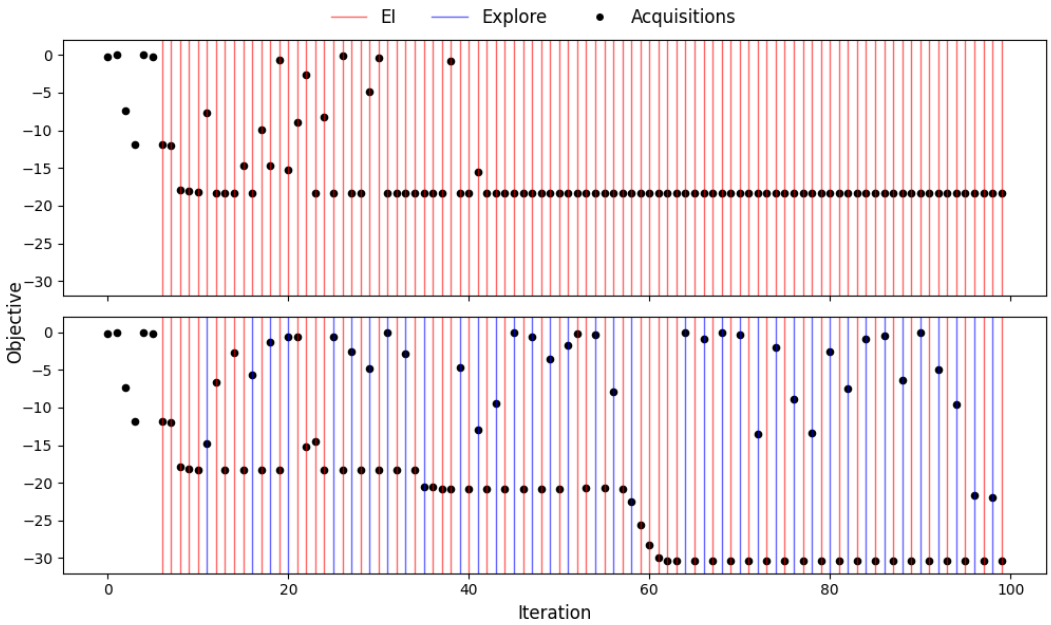


Figure 6: All acquisition points in terms of objective value from run 9 of the MGP+EI+V_BO approach (up) versus the MGP+EI+EA_BO approach (down). The AF used at each iteration is shown using red and blue vertical lines corresponding to the EI and pure exploration AF respectively. The global minima objective value is -30