HIGH-DIMENSIONAL BAYESIAN OPTIMIZATION FOR SPARSE OBJECTIVES: AN APPLICATION FOR AUTOMATED BEAM COMMISSIONING IN THE LOW ENERGY ION RING AT CERN

B. Rodriguez Mateos*, T. Argyropoulos, V. Kain,
M. Schenk, M. Slupecki†, European Organization for Nuclear Research, Geneva, Switzerland

Abstract

Recent advances in high-dimensional Bayesian Optimization have opened the door to new tools for beam commissioning. At the CERN Low Energy Ion Ring (LEIR), several tuning challenges arise from the complex parameter space governing beam transfer and accumulation dynamics. In this paper we benchmark several state-of-the-art High Dimensional Bayesian Optimization methods to optimize the transfer from Linac3 to LEIR and maximize the accumulated beam inside the ring. We evaluate algorithms based on different strategies: trust region approaches (TuRBO), sparse axis-aligned subspace priors (SAASBO), nested embeddings for mixed spaces (Bounce), and length-scale-adapted priors in regular Bayesian Optimization. Our results demonstrate the relative strengths of each method in the context of particle accelerator optimization, where sample efficiency is critical, the objective function exhibits sparsity in relevant dimensions, and the parameter space contains both local and global structures. The benchmarking provides practical insights for selecting appropriate algorithms for beam commissioning tasks, considering factors such as convergence speed, computational overhead, and robustness to noisy observations.

INTRODUCTION

Modern particle accelerators are complex systems with a large number of tunable parameters that must be carefully adjusted to achieve optimal beam performance. Beam commissioning often requires the simultaneous tuning of tens of parameters spanning transfer lines, injection elements, radio-frequency (RF) systems, and correction magnets, a task traditionally performed manually by expert operators through iterative adjustments. As facilities become increasingly complex and demand ever-higher performance, there is a growing need for automated tuning procedures that can reliably navigate these high-dimensional parameter spaces [1].

Among data-driven approaches, Bayesian Optimization (BO) has demonstrated particular success in a wide range of accelerator physics applications [2], including injection efficiency tuning [3, 4], Free Electron Laser (FEL) performance optimization [5], and trajectory correction [6]. BO treats the objective as a black-box, handles noisy measurements gracefully, and is sample-efficient, a critical feature when each evaluation corresponds to a machine cycle [7]. However, the performance of standard BO with Gaussian Process (GP) surrogates is known to degrade rapidly as the dimensionality of the search space grows: distances between candidate points become increasingly similar, reducing the ability of the GP kernel to distinguish nearby from distant points [8, 9].

Several high-dimensional Bayesian Optimization (HDBO) algorithms have been developed to address this limitation, each with a different inductive bias. Trust-region approaches such as TuRBO [10] restrict the optimization to local hyper-rectangular regions. Sparse-axis-aligned methods such as SAASBO [11] impose a hierarchical sparsity prior on the kernel lengthscales, assuming only a small subset of dimensions is relevant. Nested embedding methods such as Bounce [12] progressively build low-dimensional embeddings, originally designed for combinatorial and mixed spaces but adaptable to continuous problems. Recent work [13] has also shown that standard BO can be made competitive in high-dimensional spaces by rescaling the lengthscale prior with the dimensionality. For comparison, we include CMA-ES [14], a derivative-free evolutionary optimizer widely used as a baseline to probabilistic approaches [15].

The CERN Low Energy Ion Ring (LEIR) provides an ideal testbed for benchmarking these algorithms. LEIR accumulates up to eight multi-turn injections of Pb^{54+} ions from Linac3 through phase-space painting combined with electron cooling, before RF capture and extraction to the Proton Synchrotron [16]. Its high repetition rate, low beam energy, and the regular need for re-tuning due to Linac3 stripper foil aging [17] make it a practical platform for developing and validating data-driven tuning strategies. Previous work at LEIR demonstrated that standard BO with Upper Confidence Bound (UCB) and Log Noisy Expected Improvement (LogNEI) acquisition functions can outperform operational tuning for both injection efficiency and RF capture losses when the number of dimensions is moderate [4]. Extending this to full beam commissioning, however, requires simultaneously tuning on the order of 40 parameters without operational priors, pushing the problem firmly into the high-dimensional regime.

In this paper, we benchmark a representative set of state of the art HDBO algorithms on the task of commissioning the Linac3-to-LEIR transfer and injection starting from an unoptimized initial configuration and without prior knowledge. We consider two operationally relevant beam types, the NOMINAL cycle with 43 tunable parameters and the EARLY cycle with 38, both optimized with respect to the accumulated beam intensity before RF capture, measured with

* borja.rodriguez.mateos@cern.ch
† also at Department of Communications and Computer Engineering, University of Malta

the ring beam current transformer (BCT). All algorithms are evaluated under matched initialization budgets and identical objective functions to provide a fair comparison, offering practical guidance for selecting HDBO methods in accelerator commissioning.

METHODOLOGY

Optimization objective

LEIR beam intensity is maximized by adjusting the beam position with respect to the injection septa, matching the injected energy distribution with the electron cooler parameters in the longitudinal plane, and tuning the working point and chromaticity to mitigate space-charge driven losses. We define the objective as the injection efficiency $o(i_{\text{injected}}, i_{\text{L3}}) = \frac{i_{\text{injected}}}{i_{\text{L3}}} = \frac{i_{\text{injected}}}{\sum_{j=1}^{8} i_j}$ where i_{injected} is the LEIR ring intensity at the peak of the injection plateau (1700 ms after cycle start for NOMINAL, 600 ms for EARLY) and i_j is the intensity of the j-th Linac3 pulse measured at the end of the Linac3-to-LEIR transfer line. Normalizing by the total delivered charge isolates the contribution of the controlled parameters from upstream fluctuations originating in the ECRIS source [18] and along the linac. This is operationally critical: ion source afterglow intensity, transport jitter, and pulse-to-pulse energy spread can vary by up to five percent on the timescale of a single optimization run. The same normalization was used in our previous BO studies [4]. The benchmark covers the two operational cycles: 43 parameters for NOMINAL (eight stacked injections) and 38 for EARLY (single injection). The action space spans the low-level RF parameters at the linac exit (ramping cavity phase, debunching cavity start and end phases) [19], transfer line correctors, the four orbit bump settings around the electron cooler, the injection bump settings, the electron gun voltage program defining the electron energy, and the LEIR tune and chromaticities at flat bottom.

Bayesian Optimization framework

All BO-based methods considered in this work share the same skeleton [2, 7]. A GP prior is placed on the objective $f : \mathcal{X} \subset \mathbb{R}^D \to \mathbb{R}$, $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, with a constant mean function $m(\mathbf{x})$ and a Matérn-5/2 kernel function $k(\mathbf{x}, \mathbf{x}')$ with automatic relevance determination (ARD) lengthscales $\{l_i\}_{i=1}^{D}$. Given n observations $\mathcal{D}_n = \{(\mathbf{x}_i, y_i)\}$ with $y_i = f(\mathbf{x}_i) + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ denotes the zero-mean Gaussian observation noise with variance σ_n^2, the posterior $p(f \mid \mathcal{D}_n)$ is closed-form Gaussian. The next query point is selected by maximizing an acquisition function. Throughout this work we use LogNEI [20] unless stated otherwise, for its numerically stable gradients in noisy regimes. Initialization is common to all GP-based methods: half of the budget is drawn from a Sobol low discrepancy sequence [21] and the remaining half from a pure exploration acquisition $\alpha_{\text{PE}}(\mathbf{x}) = \sigma(\mathbf{x})$, following [4].

High-Dimensional Bayesian Algorithms

Standard BO performance degrades in high-dimensional parameter spaces because pairwise distances concentrate and ARD lengthscales become poorly identified from limited data [22]. The four HDBO algorithms benchmarked here address this with distinct inductive biases. We also include CMA-ES as a non-BO baseline.

Lengthscale-rescaled standard BO [13] attributes the failure of standard BO in high dimensions to a misspecified lengthscale prior. Placing a dimensionality-aware log-normal prior on each ARD lengthscale,

$$l_i \sim \mathcal{LN}\left(\mu_0 + \frac{\log D}{2},\ \sigma_0\right), \tag{1}$$

where $\mathcal{LN}$ denotes the log-normal distribution. This makes the prior-expected lengthscale grow with $\sqrt{D}$, counteracting the shrinkage of fitted lengthscales as dimensionality increases. We use $\mu_0 = \sqrt{2}$ and $\sigma_0 = \sqrt{3}$ as recommended. The method retains a single global GP and standard acquisition optimization.

TuRBO (Trust Region Bayesian Optimization) [10] performs optimization within hyper-rectangular trust regions centered on the current best observation. A local GP is fit within each region and Thompson sampling selects candidates. The side length L is contracted after τ_{fail} consecutive failures to improve and expanded after τ_{succ} successes, with restarts triggered when $L < L_{\text{min}}$. We use the default hyperparameters of the reference implementation.

SAASBO (Sparse Axis-Aligned Subspace BO) [11] keeps a single global GP but imposes a hierarchical sparsity prior on the per-dimension inverse squared lengthscales $\rho_i = l_i^{-2}$, with $\tau \sim \mathcal{HC}^+(\alpha)$, $\rho_i \mid \tau \sim \mathcal{HC}^+(\tau)$, where $\mathcal{HC}^+$ is the half-Cauchy distribution and τ a global shrinkage parameter. This horseshoe-style prior concentrates posterior mass on configurations in which only a few dimensions are active. Posterior inference uses the No-U-Turn Sampler (NUTS), making SAASBO substantially slower per cycle than the other methods.

Bounce [12] optimizes in a low-dimensional latent space $\mathbb{R}^{d_e}$ with $d_e \ll D$, mapped into the original space through a structured random projection. The embedding dimensionality is increased in stages: when optimization stalls at a given d_e, the embedding is enlarged and the GP refit. Originally designed for combinatorial and mixed spaces, we apply it here to the purely continuous case with initial $d_e = 25$ based on preliminary scans.

EARLY cycle optimization (38 parameters)

Figure 1 shows the traces for the 38-dimensional performance optimization of the EARLY cycle. CMA-ES converges most rapidly in the first ~25 cycles (in part by skipping the initial exploration budget), reaching an injection efficiency of approximately 0.28, but stagnates thereafter and is overtaken by all GP-based methods after cycle 100. This behavior is consistent with the expected strengths of the algorithm: a derivative-free evolutionary search makes fast

initial progress on smooth landscapes but lacks the global probabilistic model needed to escape local optima. TuRBO and Bounce reach the highest final efficiencies (≈0.42), with TuRBO crossing the operational target of 2×10^{10} charges in under 100 cycles. Lengthscale-rescaled standard BO is competitive but converges to a slightly lower plateau (≈0.40), suggesting that for this dimensionality the additional inductive bias of trust regions or learned embeddings provides a measurable advantage over a global GP, even with a properly scaled prior. SAASBO is the slowest to make progress, only beginning to improve after ~150 cycles and not reaching the performance of the other BO variants within the 250 cycle budget. The sparse-axis-aligned prior assumes the objective depends on a small subset of dimensions, an assumption that appears to be poorly matched to the EARLY cycle where transverse, longitudinal, and cooler-bump parameters all contribute meaningfully to the objective.

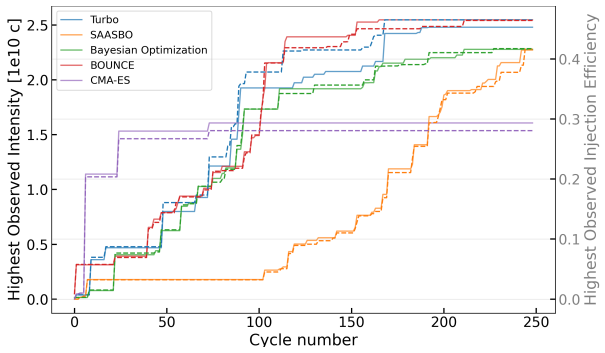

Figure 1: Highest observed injection efficiency (dashed, right axis) and ring intensity (solid, left axis) vs. cycle number for the five algorithms on the EARLY cycle (38 parameters). All runs were recorded under similar experimental conditions.

NOMINAL cycle optimization (43 parameters)

The 43-dimensional optimization problem on the NOMINAL cycle is the more challenging benchmark, both because of the higher dimensionality and because injection efficiency depends on the stacking of eight successive pulses rather than one. The added dimensions correspond mainly to per-pulse trims of the ETL.BHN10 transfer line corrector magnet, applied independently to each of the eight injections, which extends the number of "sparse" directions in the sense described above: each per-pulse trim can independently steer that pulse out of the LEIR acceptance, producing a flat zero contribution to the objective over a substantial fraction of its range. Figure 2 shows the corresponding traces. TuRBO clearly outperforms all other algorithms, reaching the operational target of 12×10^{10} charges (or 0.27 injection efficiency) in under 200 cycles and saturating near ≈0.30 (or 14×10^{10} charges) thereafter. Lengthscale-rescaled standard BO is the second best method, plateauing around 0.25 after ~250 cycles. Bounce, which performed competitively on the EARLY cycle, struggles in this setting and saturates near 0.15. We attribute this to the additional sparse directions introduced by the per-pulse ETL.BHN10 trims: Bounce optimizes within a low-dimensional embedding constructed from random axis combinations, and a single mismatched mixing of a sparse direction with non-sparse directions can effectively block exploration of the relevant subspace. As the number of sparse directions grows, the probability that the embedding aligns favorably with all of them drops sharply, making Bounce increasingly fragile in higher dimensions. CMA-ES again converges quickly in the early cycles but stagnates at a sub-optimal point (≈0.14), confirming that without a probabilistic surrogate the method cannot effectively navigate the partitioned landscape of the NOMINAL cycle. SAASBO eventually reaches ≈0.22 by cycle 400, but does so substantially more slowly than TuRBO or standard BO.

A practical caveat applies to SAASBO. Because posterior inference uses NUTS at every acquisition step, SAASBO incurs a substantially higher computational cost than the other methods, on the order of minutes per step on our hardware. Machine cycles continue to elapse during inference, so SAASBO effectively operates at a lower duty cycle in wall-clock time. The cycle counts in Fig. 2 therefore underestimate its real-time disadvantage; the wall-clock gap to TuRBO is even larger. This matters operationally, since beam commissioning has a fixed time budget and per-cycle optimizer cost is a first-class concern.

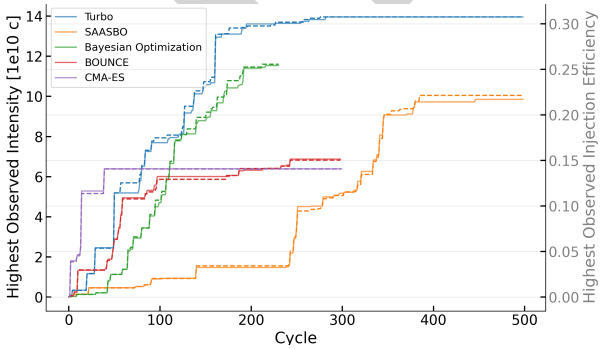

Figure 2: Highest observed injection efficiency (dashed, right axis) and ring intensity (solid, left axis) vs. cycle number for the five algorithms on the NOMINAL cycle (43 parameters). All runs were recorded under similar experimental conditions.

CONCLUSION

Across both benchmarks, TuRBO offers the best combination of asymptotic performance, sample efficiency, and computational overhead, and is retained for LEIR beam commissioning. Its trust region structure suits the injection landscape: fitting local GPs only within a contracting hyper-rectangle avoids the global identifiability problems caused by sparse directions while still exploiting the objective's smoothness in the high-reward basin. Combined with Sobol and pure exploration initialization, TuRBO consistently brings the machine from cold start to operational injection efficiency in under 200 cycles for both beam types.

REFERENCES

[1] A. Edelen *et al.*, "Opportunities in Machine Learning for Particle Accelerators", Nov. 2018. doi:10.48550/arXiv.1811.03172

[2] R. Roussel *et al.*, "Bayesian optimization algorithms for accelerator physics", *Phys. Rev. Accel. Beams*, vol. 27, no. 8, pp. 084801, Aug. 2024. doi:10.1103/PhysRevAccelBeams.27.084801

[3] S. Kato and G. Mitsuka, "Implementation and Deployment of an Injection Tuning Tool Using Bayesian Optimization at the SuperKEKB Accelerator", Oct. 2025. doi:10.48550/arxiv.2510.09176

[4] B. Rodriguez Mateos *et al.*, "Operational results of data-driven automated intensity optimization at CERN's LEIR", in *Proc. IPAC'25*, Taipei, Taiwan, Jun. 2025, pp. 2913–2916. doi:10.18429/JACoW-IPAC2025-THPM109

[5] J. Duris *et al.*, "Bayesian Optimization of a Free-Electron Laser", *Phys. Rev. Lett.*, vol. 124, no. 12, pp. 124801, Mar. 2020. doi:10.1103/PhysRevLett.124.124801

[6] S. Hirlaender and N. Bruchon, "Model-free and Bayesian Ensembling Model-based Deep Reinforcement Learning for Particle Accelerator Control Demonstrated on the FERMI FEL", Jan. 2020. doi:10.48550/arXiv.2012.09737

[7] B. Shahriari *et al.*, "Taking the Human Out of the Loop: A Review of Bayesian Optimization", *Proc. of the IEEE*, vol. 104, pp. 148-175, 2016. doi:10.1109/JPROC.2015.2494218

[8] M. González-Duque *et al.*, "A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences", Nov. 2024. doi:10.48550/arXiv.2406.04739

[9] R. Bellman, *Dynamic Programming*. Dover Publications, 1957.

[10] D. Eriksson *et al.*, "Scalable Global Optimization via Local Bayesian Optimization", Feb. 2019. doi:10.48550/arXiv.1910.01739

[11] D. Eriksson and M. Jankowiak, "High-Dimensional Bayesian Optimization with Sparse Axis-Aligned Subspaces", Jun. 2021. doi:10.48550/arXiv.2103.00349

[12] L. Papenmeier, L. Nardi, and M. Poloczek, "Bounce: Reliable High-Dimensional Bayesian Optimization for Combinatorial and Mixed Spaces", in *Advances in Neural Information Processing Systems*, vol. 36, Curran Associates, Inc., 2023, pp. 1764-1793.

[13] C. Hvarfner, E. O. Hellsten, and L. Nardi, "Vanilla Bayesian Optimization Performs Great in High Dimensions", in *Proc. 41st International Conference on Machine Learning*, vol. 235, PMLR, Jul. 2024, pp. 20793-20817. https://proceedings.mlr.press/v235/hvarfner24a.html

[14] N. Hansen, "The CMA Evolution Strategy: A Tutorial", Mar. 2016. doi:10.48550/arXiv.1604.00772

[15] J. Kirschner *et al.*, "Tuning particle accelerators with safety constraints using Bayesian optimization", *Phys. Rev. Accel. Beams* , vol. 25, no. 6, Jun. 2022. doi:10.1103/PhysRevAccelBeams.25.062802

[16] M. Benedikt *et al.*, "LHC Design Report. 3. The LHC injector chain", CERN, Geneva, Switzerland, Rep. CERN-2004-003-V-3, CERN-2004-003, Dec. 2004. doi:10.5170/CERN-2004-003-V-3

[17] S. Hirlaender *et al.*, "LEIR Injection Efficiency Studies as a Function of the Beam Energy Distribution from Linac3", in *Proc. IPAC'18*, Vancouver, Canada, Apr.-May 2018, pp. 758–760. doi:10.18429/JACoW-IPAC2018-TUPAF034

[18] V. Coco *et al.*, "Acceleration of Several Charge States of Lead Ion in CERN LINAC3", 2004. https://cds.cern.ch/record/925515

[19] B. Rodriguez Mateos *et al.*, "Online Reinforcement Learning for Stripper Foil Aging Compensation at the CERN Low Energy Ion Ring", presented at the IPAC'26, Deauville, France, May 2026, paper WEP6021, this conference.

[20] S. Ament *et al.*, "Unexpected Improvements to Expected Improvement for Bayesian Optimization", Jan. 2025. doi:10.48550/arXiv.2310.20708

[21] I. M. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals", *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 4, pp. 86-112, 1967. doi:10.1016/0041-5553(67)90144-9

[22] L. Papenmeier, M. Poloczek, and L. Nardi, "Understanding High-Dimensional Bayesian Optimization", Jun. 2025. doi:10.48550/arXiv.2502.09198