ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-MOPS65

ENHANCING CERN-SPS SLOW EXTRACTION EFFICIENCY: META BAYESIAN OPTIMISATION IN CRYSTAL SHADOWING

F. M. Velotti*, L. S. Esposito, M. Fraser, V. Kain, E. Matheson,
S. Solis Paiva, CERN, Geneva, Switzerland

Abstract

The Super Proton Synchrotron at CERN serves the fixed-target experiments of the North Area, providing protons and ions via slow extraction, and employs the crystal shadowing technique to significantly minimise losses. Over the past three operational years, the use of a crystal, positioned upstream of the electrostatic septum to shadow its blade, has allowed to achieve a 25% reduction in losses. Additionally, a novel non-local shadowing technique, utilising a different crystal location, has successfully halved these losses. While using a single crystal in this location resulted in a temporary 50% reduction in slow extraction losses at nominal intensity, this effect was not sustainable beyond a few hours. This limitation is primarily attributed to the magnetic non-reproducibility and hysteresis inherent to the SPS main dipoles and quadrupoles. In this paper, we introduce the application of the Rank-Weighted Gaussian Process Ensemble to the setup of shadowing. We demonstrate its superior efficiency and effectiveness in comparison to traditional Bayesian optimisation and other numerical methods, particularly in managing the complex dynamics of local and non-local shadowing.

INTRODUCTION

The Super Proton Synchrotron (SPS) at CERN is a versatile accelerator that provides beams to a variety of fixed-target experiments in the North Area [1]. The SPS is the last accelerator in the LHC injector chain, and it accelerates protons and heavy ions to high energies. The slow extraction of the SPS is performed by means of third-integer resonant slow extraction [2, 3]. The slow extraction process is a critical operation for the SPS, as it is the main source of beam losses. The losses are mainly due to the beam particles that are not extracted by the septum and are instead lost in the machine.

The slow extraction losses are the main limitation to the maximum intensity that can be extracted from the SPS. In this context, a significant R&D effort has been dedicated to the development of techniques to improve the slow extraction efficiency. One of the most successful techniques is crystal shadowing[4]. The crystal shadowing technique consists in placing a crystal [5] upstream of the electrostatic septum to shadow the septum wires. The crystal is oriented in such a way that the particles interacting with it are deflected away from the septum wires. This technique has been successfully used in the last three years of operation of the SPS, resulting in about 25% reduction in losses [6].

Furthermore, a novel non-local shadowing technique has been developed, which consists in placing a crystal (TECA) in a different location, upstream of the extraction channel. This technique has been successfully demonstrated in the SPS to reduce the losses by 50% [7]. However, the effect of the non-local shadowing is very sensitive to the SPS closed orbit variations, which are mainly due to the magnetic non-reproducibility and hysteresis of the SPS main dipoles and quadrupoles.

In this paper, we introduce the application of Bayesian Optimisation (BO) aided with Rank-weighted Gaussian Process Ensemble (RGPE) to the setup of non-local shadowing. We demonstrate the potential gain in efficiency and we discuss possible operational challenges. Finally, we report on the latest results obtained with non-local shadowing in the SPS and the challenges that should be addressed to make this technique operational.

METHOD

The RGPE [8] is a method designed to enhance BO by utilising information from previous optimisation tasks. It involves combining multiple base models and a target model to predict the function of interest, using ensemble methods weighted by the performance of each model on the target task. In simpler words, the RGPE method will attempt to exploit the available previous knowledge (via the base models), to speed up the search of the global optimum. This is done by ranking the prediction of each individual base models to decide for the next point to explore and then by building a surrogate of the target function via weighting of the predictions by their quality.

More in details, RGPE constructs an ensemble model from multiple base Gaussian Process (GP) models and one target GP model. The target GP model is specific to the new task, while the base models are pre-trained on related tasks. The prediction for the new input points is a weighted sum of predictions from all these models, forming an ensemble.

The ensemble prediction $\bar{f}$ is calculated as follows:

$$\bar{f}(\mathbf{x}|\mathcal{D}) = \sum_{i=1}^{t} w_i f^i(\mathbf{x}|\mathcal{D}_i) \tag{1}$$

where $f^i(\mathbf{x}|\mathcal{D}_i)$ is the prediction of the i-th base model on the new input $\mathbf{x}$, $\mathcal{D}_i$ is the dataset used to train the i-th base model, and w_i is the weight assigned to the i-th base model. The weights are calculated based on the performance of each base model on the target task.

Weights w_i for each model are determined based on the ”ranking loss”, which assesses how well the model's predictions align with the ranking of the observed outcomes from

* francesco.maria.velotti@cern.ch

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-MOPS65

the target task:

$$\mathcal{L}(f^i, \mathcal{D}_t) = \sum_{j=1}^{n_t}\sum_{k=1}^{n_t} \mathbf{1}\left[\left(f^i(\mathbf{x}_j^t) < f^i(\mathbf{x}_k^t)\right) \oplus \left(y_j^t < y_k^t\right)\right] \tag{2}$$

where $\mathcal{L}$ denotes the ranking loss function, $(f^i, \mathcal{D}_t)$ specifies that the loss is calculated for the i-th model on the dataset $\mathcal{D}_t$, the sum is done over all pairs of observations j and k in the target dataset $\mathcal{D}_t$, $\oplus$ denotes the exclusive OR operation and checks if exactly one of the two conditions is true.

The ensemble model itself is a Gaussian Process:

$$\bar{f}(\mathbf{x}|\mathcal{D}) \sim \mathcal{N}\left(\sum_{i=1}^{t} w_i \mu_i(\mathbf{x}), \sum_{i=1}^{t} w_i^2 \sigma_i^2(\mathbf{x})\right) \tag{3}$$

where $\mu_i(\mathbf{x})$ and $\sigma_i^2(\mathbf{x})$ are the mean and variance of the predictions of each model.

This method leverages the strengths of individual models and mitigates their weaknesses, leading to a more robust prediction, especially in cases where data on the new task is limited but related historical data is plentiful. The implementation of this method is openly available on the official BoTorch website [9].

APPLICATION TO NON-LOCAL SHADOWING

In our case, we have a very accurate model of the SPS slow extraction with a crystal shadowing setup, which is used to generate the data to train a deep neural network (DNN) model (Fig. 1). The DNN model is a fully connected neural network with two hidden layers, each with 512 neurons and hyperbolic tangent activation functions. The input to the DNN model is the crystal angle and position, and the output is the slow extraction losses.

The DNN model is then used to generate the data to train the base models, by shifting the response in the two input dimensions, i.e. the crystal angle and the crystal position. The optimisation task is the unchanged DNN model, which is sampled at every step of the optimisation routine.

The RGPE method is then used to optimise the crystal angle and position to minimise losses. All the results that we present in this paper are obtained using exclusively the DNN model and future work will include the test carried out at the CERN-SPS.

RESULTS

The RGPE method was applied to the non-local shadowing setup, together with classic BO and random search, as baseline. For both RGPE and BO, the acquisition function used was the Expected Improvement (EI). The upper confidence bound (UCB) was also tested but the results were not as good as with EI, in all the cases.

In Fig. 2 the results of three optimisation runs are shown. For all methods, the optimisation was performed 5 times, with 50 iterations each. The results show that RGPE converges faster than the other methods, and both RGPE and BO

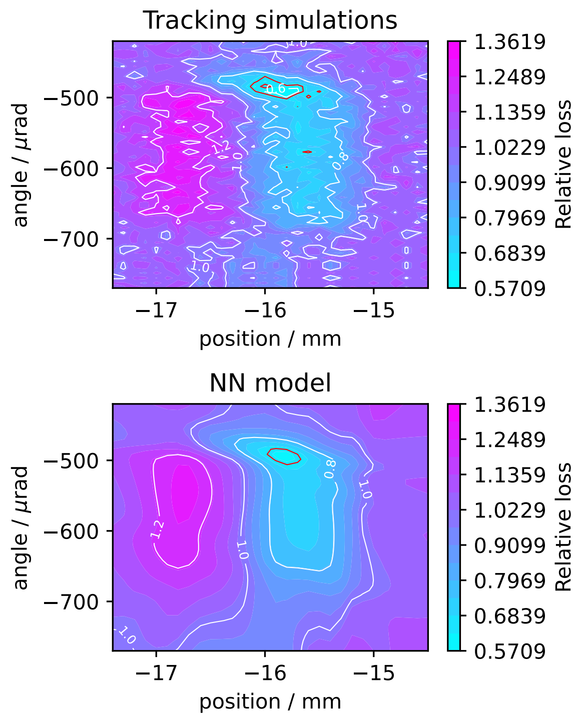

Figure 1: (Top) Relative losses in the SPS slow extraction as a function of the crystal angle and position obtained from numerical tracking simulations. (Bottom) DNN model trained on the numerical data.

achieve similar final results, which is in line with the previous experience on local shadowing. RGPE converges to the global optimum (blue dashed line) after about 30 machine iterations, instead the vanilla BO at around 50 iterations. The difference recorded at 30 iterations between the two methods is indeed significant, as only RGPE actually reaches the deep of full loss reduction highlighted by a red circle in Fig 1. Random search is the slowest to converge and achieves the worst final result.

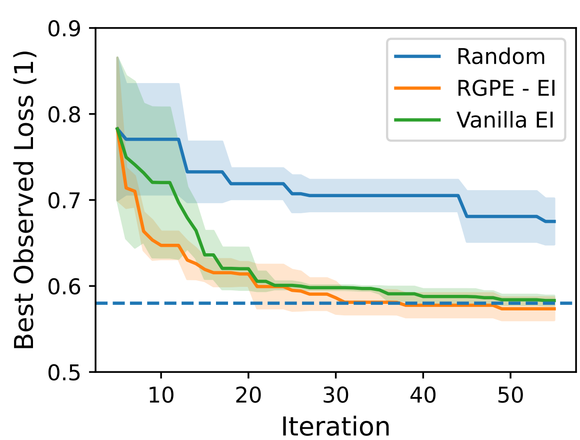

Figure 2: Comparison of RGPE, vanilla BO and random search on the non-local shadowing setting up task. The best observed value at each iteration is plotted.

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-MOPS65

From these results, we would expect a significant improvement in the speed of convergence of the RGPE method compared to BO, when applied to the real machine. The speed of convergence for RGPE significantly depends on the quality of the base models. In our case, the base models are generated only by shifting the input space. Prior knowledge of the correlation between the base models and the real response could drastically improve the performance of the RGPE method.

If this method shows the same performance observed on simulations on the real SPS, it can be a significant improvement in the context of the SPS operation, where the time to perform the optimisation is a critical factor to maximise the physics throughput.

EXPERIENCE WITH NON-LOCAL SHADOWING IN OPERATION

The non-local shadowing technique has been tested in the SPS [7] and has shown promising results. Loss reduction up to a factor 2 was demonstrated during dedicated beam time [7]. Furthermore, this technique was tested in realistic physic conditions, with up to 2×10^{13} p extracted to the North Area. The main challenge that was encountered was the stability of shadowing in time.

In order to study this effect, a dedicated test was performed in the SPS, where the TECA was aligned in shadowing via classic BO. Once the optimal crystal angle and position were found, the losses were monitored for 1 hour. The results are summarised in Fig. 3. The first observation is that the TECA cannot be placed exactly at the optimal configuration found via BO due to the limits in the crystal motor movements - now set to a minimum of 8 μrad and 0.1 mm for the angle and position, respectively. Once the crystal is set up, the losses drift in time, and after about 30 minutes the losses are significantly increased, and the shadowing effect is almost completely lost. As a possible explanation, the last plot in Fig. 3 shows the change in beam position measured in the proximity of the crystal and the change of the beam momentum offset[1]. These two quantities are very well correlated, and their change in time is also correlated with the change in losses.

FUTURE WORK

A full solution to the observed behaviour is two-fold. First, RGPE can be used to reduce the time needed to find the optimal configuration of the crystal, and second, a real-time controller, that would continuously monitor the losses and beam position while adjusting the crystal angle and position accordingly, can be implemented.

Further studies are needed to deploy RGPE in the SPS and this will be the main focus of the upcoming machine development period.

Finally, a possible method for the real-time control of the crystal could be the implementation of a Model Predictive Control (MPC) algorithm based on the available DNN surrogate model and the available numerical model of the SPS.

[1] The momentum offset is calculated via the requested change in machine reference momentum, exploiting the COSE extraction methodology [10].

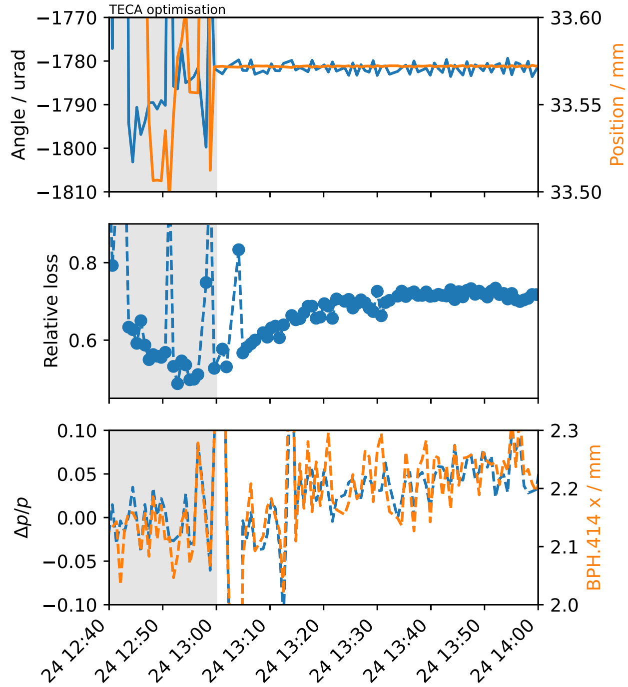

Figure 3: (Top) Time evolution of the crystal angle and position during the test. (Middle) Relative losses at the extraction septum as a function of time. (Bottom) Time evolution of the relative momentum offset and horizontal beam position (BPM.41408).

CONCLUSION

In this paper, we have introduced the application of the Rank-Weighted Gaussian Process Ensemble to the setup of shadowing in the CERN-SPS. We have demonstrated its superior efficiency and effectiveness in comparison to traditional BO, particularly in managing the complex dynamics of non-local shadowing. We have shown that the RGPE method can provide a significant improvement in the speed of convergence of the optimisation routine, which is a critical factor in the context of the SPS operation. Future work will focus on machine development tests of the RGPE method and the implementation of a real-time controller to adjust the crystal angle and position based on the current beam conditions.

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-MOPS65

REFERENCES

[1] D. Banerjee *et al.*, "The North Experimental Area at the Cern Super Proton Synchrotron," *CERN internal note*, 2021. doi:10.17181/CERN.GP3K.0S1Y

[2] F. Velotti *et al.*, "Characterisation of the SPS Slow-extraction Parameters," *7th International Particle Accelerator Conference*, THPOR055, 2016. doi:10.18429/JACoW-IPAC2016-THPOR055

[3] M. A. Fraser *et al.*, "Slow extraction at the SPS: extraction efficiency and loss reduction studies," in *Injector MD Days 2017*, 2017, pp. 87–94. doi:10.23727/CERN-Proceedings-2017-002.87

[4] F. M. Velotti *et al.*, "Septum shadowing by means of a bent crystal to reduce slow extraction beam loss," *Phys. Rev. Accel. Beams*, vol. 22, p. 093 502, 9 2019. doi:10.1103/PhysRevAccelBeams.22.093502

[5] K. Elsener *et al.*, "Proton extraction from the CERN SPS using a bent crystal," CERN, Tech. Rep., 1995. https://cds.cern.ch/record/292372

[6] F. Velotti *et al.*, "Exploitation of Crystal Shadowing via Multi-Crystal Array, Optimisers and Reinforcement Learning," in *Proc. IPAC'22*, Bangkok, Thailand, 2022, paper WEPOST013, pp. 1707–1710. doi:10.18429/JACoW-IPAC2022-WEPOST013

[7] F. M. Velotti *et al.*, "Exploitation of Crystal Shadowing via Multi-Crystal Array, Optimisers and Reinforcement Learning," in *Proc. IPAC'22*, Bangkok, Thailand, 2022, pp. 1707–1710. doi:10.18429/JACoW-IPAC2022-WEPOST013

[8] M. Feurer, B. Letham, F. Hutter, and E. Bakshy, "Practical transfer learning for bayesian optimization," 2022. doi:10.48550/arXiv.1802.02219

[9] M. Balandat *et al.*, "Botorch: A framework for efficient monte-carlo bayesian optimization," 2020. doi:10.48550/arXiv.1910.06403

[10] V. Kain *et al.*, "Resonant slow extraction with constant optics for improved separatrix control at the extraction septum," *Phys. Rev. Accel. Beams*, vol. 22, p. 101 001, 10 2019. doi:10.1103/PhysRevAccelBeams.22.101001