MACHINE-LEARNING–ASSISTED BAYESIAN UNCERTAINTY QUANTIFICATION FOR ACCELERATOR DIGITAL TWIN MODELING AND CONTROL*

W. Lin†[1], K. Brown[1], J. Edelen[2], E. Hamwi[1], G. Hoffstaetter[1,3], C. Kelly[1], N. Urban[1]
[1]Brookhaven National Laboratory, Upton, NY, United States
[2]RadiaSoft, Boulder, CO, United States
[3]Cornell University (CLASSE), Ithaca, NY, United States

Abstract

Digital twins of particle accelerators are increasingly used for experiment planning, machine studies, and model-based control. Achieving high-fidelity predictions requires knowledge of machine properties that are difficult to measure directly, such as magnet alignments, transfer function variations, nonlinearities, and stray fields. In this work, we introduce parameterizations to capture these effects and employ Bayesian inference to estimate their values and uncertainties by calibrating a digital twin to orbit response measurements from the AGS Booster at Brookhaven National Laboratory. A machine-learning emulator trained on a perturbed ensemble of Bmad simulations enables computationally efficient sampling of the high-dimensional posterior. The resulting joint parameter distribution incorporates BPM uncertainties and provides data-constrained variations that, when inserted back into the digital twin, significantly improve agreement with measured beam orbits while yielding uncertainty estimates on both parameters and predictions.

INTRODUCTION

Digital twins provide a powerful framework for predictive control, online optimization, and machine studies planning in particle accelerators. Their accuracy, however, is limited by uncertainties in component properties such as magnet alignments, saturation effects, current-to-field transfer functions, and stray fields. In the AGS Booster, comparisons between measured and simulated orbit response matrices (ORMs) show systematic discrepancies larger than the beam position monitor (BPM) noise level, pointing toward deficiencies in the magnetic model and motivating a formal uncertainty quantification (UQ) approach. Bayesian UQ is particularly well suited to accelerator modeling because it provides a principled way to quantify correlated parameter uncertainties and propagate them through a physics-based model, yielding interpretable corrections and uncertainty-aware predictions.

In this work, we introduce a Bayesian UQ framework that infers magnet-specific variations to quadrupole strengths using ORM data. To make the inference tractable, we employ a machine-learning surrogate model of the Bmad [1] digital twin capable of fast and differentiable predictions, enabling efficient Hamiltonian Monte Carlo (HMC) sampling in a 48-dimensional parameter space. The resulting posterior distributions provide both parameter estimates and uncertainty bounds, while the inferred variations yield substantial improvements when integrated into the digital twin.

* Work supported by Brookhaven Science Associates, LLC under Contract No. DE-SC0012704, No. DE-SC0024287, and No. DE-SC0025351 with the U.S. Department of Energy.
† wlin1@bnl.gov

METHODS

Orbit Response Data

ORM data were collected by stepping each of the Booster's corrector magnets sequentially through 22 A, 0 A, and −22 A. Each setting included multiple (2–5) repeated BPM measurements at the 92 ms flat-top of the magnet cycle. After automated outlier removal, the dataset consisted of 23 horizontal and 14 vertical perturbed ORM measurements using 15 and 18 valid BPMs, respectively. Orbit differences (perturbed minus baseline) were used to suppress static orbit offsets and reduce sensitivity to unrelated systematic errors, allowing the inference to isolate the effect of each corrector more cleanly.

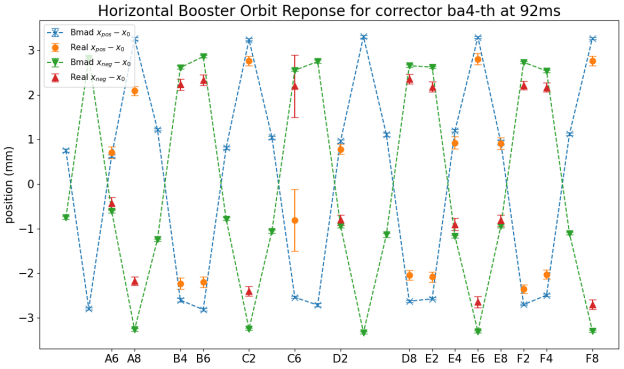

Figure 1: Comparison between measured and simulated horizontal orbit responses with corrector A4 set to ±22 A, data taken at 92 ms in the Booster cycle. The larger error on BPM C6 is due to a known hardware problem.

A comparison of the orbit differences between positive, zero, and negative corrector settings for corrector A4 is shown in Fig. 1. The error bars are estimated based on the fluctuations over the repeated measurements. We observe that the difference between measured and simulated differential orbits are mostly within 1 mm, but still much larger than the error bars of the measurements. This discrepancy implies there are sources of error in the real machine that

are not included in the simulation model. The BPM C6 has much larger measurement fluctuations due to known hardware issues. Prior to performing the inference we trimmed these and other outlier points from the data set.

Quadrupole Uncertainty Parameterization

The Booster contains 48 quadrupoles (24 in each plane) powered in series and modeled with a fifth-order polynomial relating power supply current to magnetic field gradient [2]. To represent unmodeled effects such as magnet specific transfer function variations or calibration errors, we introduce multiplicative correction parameters ("vars") for each quadrupole in each transverse plane. The resulting k_1 of the quadrupoles in the model are therefore defined as:

$$(k'_{1,H})_i = var_i^{(x)} k_{1,H}, \quad (k'_{1,V})_i = var_i^{(y)} k_{1,V},$$

where $i \in \{1 \ldots 24\}$ indexes the quadrupole magnet. These 48 parameters form the target of the Bayesian inference. This parameterization provides a simple but flexible way to capture magnet-specific modeling errors that affect the ORM.

Bayesian Formulation

Bayesian UQ takes expert knowledge of the model parameters θ as priors $p(\theta)$ as well as a probability model of the data-generating process $p(x|\theta)$ called the likelihood function. By constraining the parameters with data samples $\mathcal{D}_n$, Bayesian UQ updates the model and calculates the posterior probability distribution $p(\theta|\mathcal{D}_n)$ of parameters θ given the data. According to Bayes' Theorem, the posterior distribution is given by [3]:

$$P(\theta|x_1, ..., x_n) = \frac{p(\mathcal{D}_n|\theta)p(\theta)}{p(\mathcal{D}_n)} = \frac{\mathcal{L}_n(\theta)p(\theta)}{c_n}.$$

The likelihood function evaluates the expected orbit for given values of parameters θ and magnet currents I:

$$x = m(\theta; I) + \epsilon,$$

where m is the surrogate model prediction of the Bmad simulation, and ϵ is an additive noise term capturing BPM measurement uncertainty and surrogate model error. We assume ϵ follows an independent Gaussian distribution, with its variance determined from BPM noise measurements and increased in quadrature to account for surrogate error in each plane.

Because the deviations between the measurements and simulation are small, we expect the effective strength deviations of individual quadrupoles to be small and centered around the nominal model. Therefore, we assign each parameter a lognormal prior:

$$\text{var}_i \sim \text{Lognormal}(1 - \sigma^2/2, \sigma^2),$$

for which

$$\overline{\text{var}_i} = 1, \quad \text{Var}(\text{var}_i) = \sigma^2 + O(\sigma^4),$$

where we choose a uniform prior width σ for all quadrupoles. The lognormal form ensures the parameters remain positive and centered near unity. The priors are additionally truncated to the interval 0.94–1.06 to prevent the inference from exploring unphysical configurations that would violate tune constraints or surrogate-model validity. Posterior sampling is performed using the `Turing.jl` probabilistic programming environment [4].

Surrogate Model

We use the No-U-Turn Sampler (NUTS) [5], an HMC variant, to explore the posterior distribution. NUTS is well suited for this high-dimensional problem because it adapts step sizes and trajectory lengths and makes efficient use of gradients. However, Bmad simulations do not provide gradients directly. To overcome this, we trained a high-capacity ResNet-style [6] neural network ($\sim 360k$ parameters) to emulate the mapping from magnet currents, corrector perturbations, and quadrupole parameters to predicted BPM readings. The surrogate was trained on ~2 million Sobol-sampled simulations per plane, with samples filtered to maintain realistic tune values and avoid resonance-induced instabilities. This filtering step was essential to prevent the surrogate from being trained on configurations that lead to unstable or resonance-driven orbits, which would otherwise degrade model accuracy. Final validation errors (0.057 mm horizontal, 0.017 mm vertical) are well below BPM noise and are incorporated into the likelihood. The use of the surrogate reduces likelihood evaluation time by orders of magnitude compared to direct Bmad simulations, making full HMC sampling feasible.

Dataset Construction

The available measurements allow many possible orbit response combinations to be used in the inference, so we evaluated several choices for inclusion in the Bayesian analysis. Different constructions emphasize different regions of parameter space, so comparing them helped identify the most informative subset for the inference. The responses constructed from stepping each corrector between −22 A and 22 A consistently yielded the tightest posteriors, likely due to the larger current swing amplifying any mismatch between simulation and measurement. To assess the robustness of the inference, four statistically independent datasets of this type were constructed using different choices among the repeated measurements. Inference is performed separately on each dataset, and the resulting posterior samples are pooled to obtain final parameter estimates and uncertainties.

RESULTS

Inferred Quadrupole Variations

Figure 2 presents the inferred quadrupole-strength variations together with their 68 % and 95 % Bayesian credible intervals, which reflect the central ranges of posterior probability for each parameter. While most inferred variations remain within about ±2.5 % of their nominal values, the

associated uncertainty bands are narrow enough to show that these deviations are statistically well resolved. Taken together with the consistency across the four independent datasets, this indicates that the Bayesian analysis is resolving genuine, systematic machine-model mismatches rather than fitting noise fluctuations. In fact, several quadrupoles exhibit credible intervals that exclude unity with high probability, providing clear evidence of physically meaningful, magnet-specific departures from the nominal digital twin model.

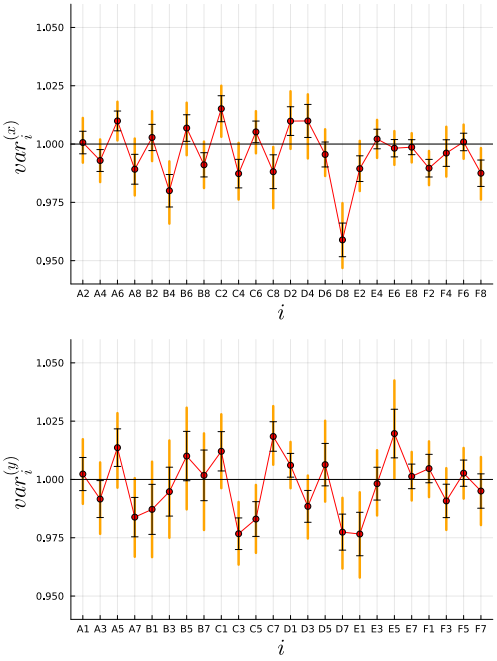

Figure 2: The final posterior results for the inferred $x-$ (upper) and $y-$plane (lower) vars for each of the 24 associated quadrupoles. The darker and lighter error bars indicate the 68 % and 95 % confidence bands, respectively.

Digital Twin Improvement

To assess the practical impact of the inferred quadrupole strength variations, we incorporated the posterior mean values of the vars directly into the Bmad digital twin and regenerated the orbit responses for all four -22 A $\rightarrow$ 22 A datasets. Figure 3 summarizes the resulting distribution of absolute orbit prediction errors across all BPMs and all measurements, comparing the updated model against the nominal one.

We observe a substantial overall reduction in simulation error, particularly in the horizontal plane where the discrepancy between model and measurement is reduced by approximately a factor of two. Improvements are also visible in the vertical plane, though somewhat smaller in magnitude due to the more limited amount of constraining data available for that plane. Importantly, the reduction in error is not localized to any single corrector or BPM. It appears systematically across the full set of orbit responses, indicating that the inferred parameters capture broad, machine-relevant sources of model discrepancy. These improvements demonstrate

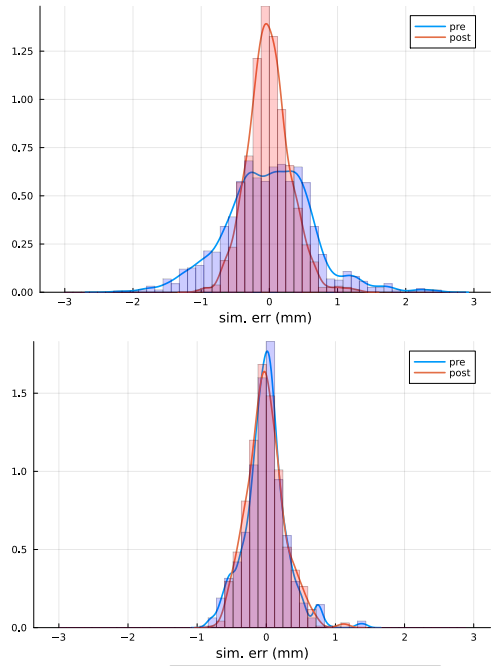

Figure 3: The distribution of the absolute digital twin error in the x- (upper) and y-plane (lower) over all measured orbit responses within the four -22 A $\rightarrow$ 22 A datasets, obtained with (red) and without (blue) incorporating the point estimates of the vars.

that even limited-scope parameterizations, restricted here to quadrupole strengths, can yield meaningful corrections, suggesting that broader uncertainty models (e.g., including alignments or sextupole variations) could further enhance digital twin fidelity.

CONCLUSION

In this work, we demonstrated a machine-learning–assisted Bayesian UQ framework capable of identifying magnet-specific modeling errors in the AGS Booster digital twin. The method produces physically interpretable posterior distributions, reveals significant deviations for a small subset of quadrupoles, and yields tangible reductions in simulation-to-measurement error when the inferred variations are incorporated into the digital twin. Future work will extend this framework to additional magnet families and alignment parameters, and integrate uncertainty propagation into digital-twin–based optimization workflows, moving toward uncertainty-aware model-based control.

REFERENCES

[1] D. Sagan, "Bmad: A relativistic charged particle simulation library", *Nucl. Instrum. Methods Phys. Res. A*, vol. 558, no. 1, pp. 356–359, 2006. doi:10.1016/j.nima.2005.11.001

[2] K. A. Brown, R. Fliller, W. Meng, and W. van Asselt, "A high precision model of AGS Booster tune control", in *Proc. 8th Eur. Part. Accel. Conf. (EPAC'02)*, Paris, France, paper THPLE112, pp. 548–550, Jun. 2002.

[3] L. Wasserman, "Bayesian inference", in *All of Statistics: A Concise Course in Statistical Inference*. New York, NY: Springer New York, 2004, pp. 175–192. doi:10.1007/978-0-387-21736-9_11

[4] H. Ge, K. Xu, and Z. Ghahramani, "Turing: A language for flexible probabilistic inference", in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, vol. 84, pp. 1682–1690, Apr. 2018.

[5] M. D. Hoffman and A. Gelman, "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo", 2011. doi:10.48550/arXiv.1111.4246

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", 2015. doi:10.48550/arXiv.1512.03385