ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST048

DEVELOPMENT OF A VIRTUAL DIAGNOSTIC FOR ESTIMATING KEY BEAM DESCRIPTORS

K. R. L. Baker*, I. D. Finch, S. Lawrie, A. Saoulis,
ISIS Neutron and Muon Source, Rutherford Appleton Laboratory, U.K.
S. Basak, J. Cha, J. Thiyagalingam
Scientific Machine Learning, Rutherford Appleton Laboratory, U.K.

Abstract

Real-time beam descriptive data such as emittance, envelope and loss, are central to accelerator operations, including online diagnostics, maintenance and beam quality control. However, these cannot always be obtained without disrupting user runs. Physics-based simulations, such as particle tracking codes, can be leveraged to provide estimates of these beam descriptors. However, such simulation-based methods are computationally intensive requiring access to high performance computing facilities, and hence, they are often non-realistic for real-time purposes. The proposed work explores the feasibility of using machine learning to replace these simulations with fast-executing inference models based on surrogate modelling. The approach is intended to provide the operators with estimates of key beam properties in real time. Bayesian optimisation is used to generate a synthetic dataset to ensure the input space is efficiently sampled and representative of operating conditions. This is used to train a surrogate model to predict beam envelope, emittance and loss. The methodology is applied to the ISIS MEBT as a case study to evaluate the performance of the surrogate model.

INTRODUCTION

An ultimate goal at accelerator facilities is to produce a high quality, efficient and reliable beam. However, due to the quantity and complexity of the systems involved, significant operator intervention is required to maximise these objectives. If key beam descriptors such as the envelope, emittance, bunch length and beam loss levels are known, it is easier for operators to identify and rectify the cause of a problem to restore the machine back to optimum performance levels. In reality, accessing these descriptors can be difficult for a variety of reasons. The measurements may be physically impossible to take, require disruptive measurements to the beam which would result in user down-time or there may be limited space available for the necessary diagnostic equipment.

Physics simulations can be used to estimate these beam parameters given a set of control parameters. Their dependence on fast Fourier transforms (FFTs) makes them impractical for use in real-time applications as a single simulation can take on the order of days to complete.

A possible solution is to make use of data-driven techniques such as machine learning to replicate these expensive calculations at much shorter timescales. These models can then be linked into control systems to produce real-time estimates for key beam descriptors to allow operators to diagnose and respond to problems more accurately.

For the purpose of this study, we take the new Medium Energy Beam Transport (MEBT), to be installed in the ISIS linear accelerator [1] as part of an ongoing upgrade [2] ,as a case study. As ISIS has evolved organically over time [3], the physical space available for the MEBT is limited; even more so for additional diagnostic equipment. It is therefore an ideal candidate to test the practicality of a surrogate model.

DATA GENERATION

For a virtual diagnostic to be useful, the input space used to train the machine learning model must be comparable to that seen during operation. We therefore define the controllable parameters in the MEBT (cavity and quadrupole strengths) as Gaussian variables from which we can sample to generate the input space. The mean of each Gaussian is the optimum setting found during the MEBT's design phase, while the standard deviation is calculated using a minimum and maximum operating range that is assumed to represent 95% of the data.

Parameters over which the machine operators have no control, such as the incoming beam position, emittance and beam current are defined as uniform random variables that can take any value between a given maximum and minimum. Combining both sets of inputs results in 17 different input features for the machine learning model.

Randomly combining accelerator machine settings will inevitably lead to unfavourable combinations which lead to losses well above the permitted operational levels. To circumvent this issue we intelligently sample the input space using a Bayesian optimisation to select combinations of inputs that would result in low losses that reflect expected operating conditions more accurately. As running each new set of inputs selected by the Bayesian optimisation loop is embarrassingly parallel, we were able to generate a data set of 2136 simulations using the lume-astra [4] Python interface to Astra [5] by utilising multiple CPUs

DATA PROCESSING

Each simulation used a different combination of 17 scalar input parameters and produces spatial 200 length arrays of the emittance (ϵ_{xyz}), transverse envelope (σ_{xy}), bunch length (σ_z), longitudinal energy spread (ΔE) and loss along the length of the MEBT.

* k.baker@stfc.ac.uk

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST048

For simplicity, the arrays were converted to scalar values by reassigning the distance along the MEBT (the z-value) as one of the inputs, a similar approach to that used by Pestourie et al [6] to reduce the complexity and dimensionality of the problem. Applying this transformation allows us to apply traditional data processing techniques for regression problems to our data, such as logging, clipping and z-scaling, as well as expand the number of machine learning architectures that could be used. This generated a data set of 18 input values and 6 scalar output values for each of the 200 z-values along the length of the MEBT, multiplying the size of the 2136 simulation data set by 200.

After the first models were developed using this data set, it was noticed that increasing the last quadrupole strength towards the exit of the MEBT would result in the beam distribution changing near the entrance, a non-physical result. We diagnosed this as correlations the model was learning between inputs, potentially introduced by the active learning process used to generate the data set. To mitigate this effect we masked the downstream inputs for any given z-value by setting their value to the mean of the data set.

SURROGATE MODEL

The virtual diagnostic used a simple feed forward neural network architecture, consisting of a model with 4 branches, one for each beam 'descriptor'. Each branch contained 5 layers of 128 nodes followed by 1 layer with 16 nodes and a final output layer of 1 or 2 nodes depending on the number of outputs in a given group of descriptors. Each of the layers was separated by a Dropout layer with a rate of 0.1 for regularisation [7], as well as a Batch Normalisation layer. The output of each of these branches is concatenated to form the final output layer of the surrogate.

TensorFlow [8] and Keras [9] implementations were used. Training minimised the mean squared error over a maximum of 300 epochs using early stopping and a scheduled decreasing learning rate with the Adam [10] optimiser.

A baseline model that predicted the mean value for each output at each z-position achieved a mean squared error of 0.991 on the validation set which consisted of 25% of the total data set. Our surrogate was able to achieve an error of 0.0275. Predictions for a single simulation (200 data points) could be generated in 39.6 ± 0.9 ms on a i7-10750H CPU. In our experience it took native Astra 1 day to complete the same simulation meaning we are able to achieve a time improvement of more than 6 orders of magnitude.

Table 1: R^2 values computed from the predicted and true values for each of the model's seven outputs

Loss	σ_x	σ_y	ϵ_x	ϵ_y	ΔE	σ_z
0.965	0.995	0.988	0.953	0.920	0.995	0.997

Table 1 shows the results of the model across each of the outputs. It is evident that the model is able to predict σ_{xyz} and ΔE to high level of accuracy, but struggles with ϵ_{xy}

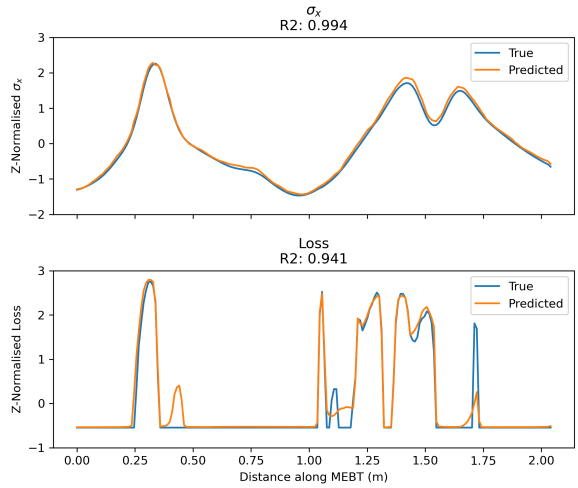

Figure 1: Example output of the virtual diagnostic, predicting the beam loss and σ_x. True output is indicated by the blue line, while model predictions are given by the orange line.

and the loss. Figure 1 illustrates the difference in prediction between the loss and the envelope.

Clearly, surrogate modelling provides a feasible alternative to physics simulations for predicting beam descriptors in real time. However, beam behaviour in the live machine can vary significantly from simulations. In order to evaluate the true efficacy of the virtual diagnostic, it needs to be integrated into the control system and validated against live data. This will not be possible until the MEBT is installed on the ISIS Linac, but will form the basis of future work.

UNCERTAINTY QUANTIFICATION (UQ)

Both accuracy and model uncertainty are of key importance in delivering a useful virtual diagnostic for operators, particularly when using it to inform a decision making process. Although Bayesian Neural Networks and Gaussian Processes are considered the gold standard in uncertainty quantification, the size of our data set (approximately 320,000 training points) made it impractical to apply these methods using the computing resources we had available. Instead, alternatives were evaluated, which consisted of an ensemble model [11, 12], a dropout model [13], a quantile regression model [14] and a variance model [15]. In each case, the same architecture as that of the basic surrogate model was used. Each technique was evaluated against standard UQ metrics such as sharpness, dispersion and calibration and negative log likelihood (NLL), as outlined in Tran et al. [16] to determine if any of these methods could sufficiently represent the trustworthiness of our virtual diagnostic.

On first inspection, the results in Table 2 look promising across all of the models except for the variance model, which was the only model to predict the mean and variance in one shot. This suggests that the basic surrogate's architecture was not sufficient for this model to be able to learn a generalised

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST048

Table 2: Key metrics for evaluating a model's ability to quantify its own uncertainty

Model	R^2	RMSE	Sharpness	Dispersion	NLL	Time (s) (1 CPU)
Ensemble	0.976	0.159	0.051	0.051	9.70e5	29.868
Variance	0.923	0.286	0.100	0.211	2.17e6	5.522
Dropout	0.971	0.174	0.053	0.044	1.61e6	29.817
Quantile	0.964	0.197	0.192	0.264	6.27e5	19.949

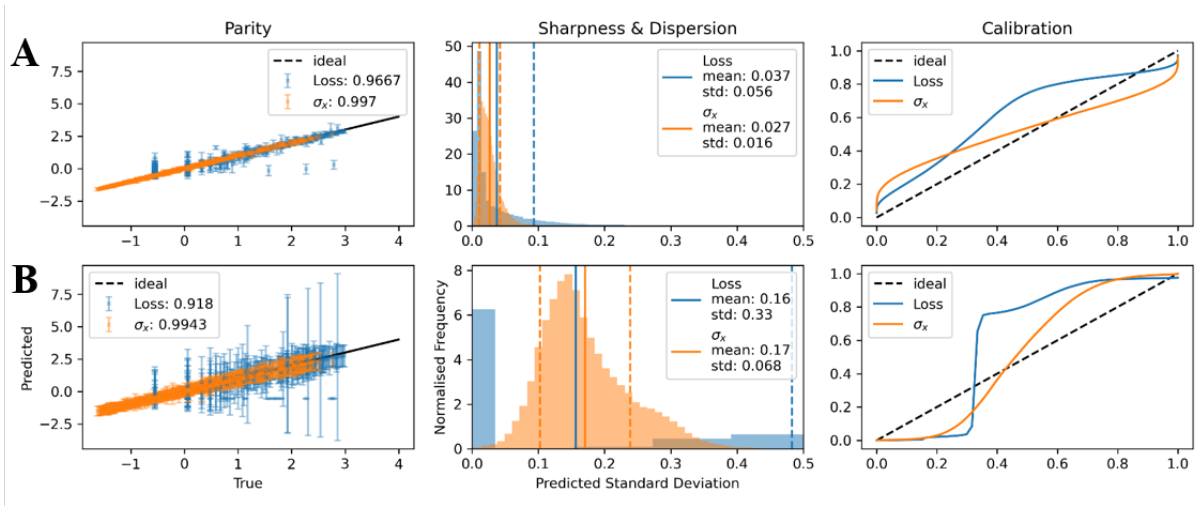

Figure 2: Comparison of UQ metrics for the A) ensemble and B) quantile regression models. Blue lines indicate the beam loss while orange lines represent σ_x. Error bars in the parity plots indicate the 95% confidence interval predicted by the model. The solid line in the central plots represents the sharpness while the dotted line represents the dispersion of the data.

representation of the data as well as the other models, which is reflected in its poor accuracy r esults. However, it showed promise particularly in its execution times compared to other models (which rely on repeat executions), so future work will consist of optimising this architecture to reduce the overfitting seen during training.

Meanwhile, the ensemble, dropout and quantile regression models are all able to predict the outputs to a similar level of accuracy as the model without UQ. As all models struggle to represent the beam loss and emittances accurately, comparison of the uncertainty quantification is redundant as the prediction itself is inaccurate. This is illustrated by the parity plots of the loss in Figs. 2A and 2B.

Although the low sharpness values across all of the models look promising initially, closer examination of Fig. 2A and B shows this is a result of poor calibration. The ensemble and dropout models are highly overconfident, indicated by a line falling below the ideal on the right hand side of the calibration plot in Fig. 2A. In practice, operators should not trust the model's assessment of its own confidence, as the uncertainty in the model is in fact higher than is portrayed. In contrast, the quantile model suffers from under confidence, as shown by its calibration curve falling below the ideal line on the left hand side. Practically, this would be illustrated by a broad 95% confidence interval around the predictions which, if seen regularly, would cause operators to disregard the model's output.

These results tell us that while the models are capable of being used as a virtual diagnostic to predict the beam descriptors, we are not yet confident enough to deploy any of the uncertainty quantification methods in practice. Investigation into recalibration [17] may provide a solution to our calibration issues in future.

SUMMARY AND OUTLOOK

The present work has illustrated that surrogate modelling is a viable alternative to physics simulations for providing estimates of key beam descriptors in real time to a high degree of accuracy. However, none of the uncertainty quantification techniques explored within the work are suitable for use by the operators. Subsequent work will involve validating the efficacy of the virtual diagnostic with live machine data once the MEBT is installed and investigating alternative UQ techniques.

ACKNOWLEDGEMENTS

Computing resources were provided by STFC Scientific Computing Department (SCD)'s SCARF [18] and PEARL [19] clusters. The authors would also like to thank the Scientific Machine Learning group in SCD and the Low Energy Beams group at ISIS, for their support and guidance.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST048

REFERENCES

[1] *A Practical Guide to the ISIS Neutron and Muon Source*. Science and Technology Facilities Council, 2021, https://www.isis.stfc.ac.uk/Pages/A%20Practical%20Guide%20to%20the%20ISIS%20Neutron%20and%20Muon%20Source.pdf

[2] S. R. Lawrie *et al.*, "A pre-injector upgrade for ISIS, including a medium energy beam transport line and an RF-driven H ion source," *Review of Scientific Instruments*, vol. 90, no. 10, p. 103 310, 2019, doi:10.1063/1.5127263

[3] J. Thomason, "The ISIS spallation neutron and muon source—the first thirty-three years," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 917, pp. 61–67, 2019, doi:https://doi.org/10.1016/j.nima.2018.11.129

[4] C. E. Mayes et al., "Lightsource Unified Modeling Environment (LUME), a Start-to-End Simulation Ecosystem", in *Proc. IPAC'21*, Campinas, Brazil, May 2021, pp. 4212-4215. doi:10.18429/JACoW-IPAC2021-THPAB217

[5] K. Floettmann, *A space charge tracking algorithm*, https://www.desy.de/~mpyflo/

[6] R. Pestourie, Y. Mroueh, T. V. Nguyen, P. Das, and S. G. Johnson, "Active learning of deep surrogates for pdes: Application to metasurface design," *npj Computational Materials*, vol. 6, no. 1, p. 164, 2020, doi:10.1038/s41524-020-00431-2

[7] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[8] Martín Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015, https://www.tensorflow.org/

[9] F. Chollet *et al.*, *keras*, https://keras.io, 2015.

[10] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014, doi:10.48550/ARXIV.1412.6980

[11] K. Weinberger. "CS4780 lecture 18: Bagging." (), https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote18.html

[12] J. Schupbach, J. W. Sheppard, and T. Forrester, "Quantifying uncertainty in neural network ensembles using u-statistics," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8, doi:10.1109/IJCNN48605.2020.9206810

[13] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of The 33rd International Conference on Machine Learning*, vol. 48, 2016, pp. 1050–1059, https://proceedings.mlr.press/v48/gal16.html

[14] R. Koenker and K. F. Hallock, "Quantile regression," *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 143–156, 2001, doi:10.1257/jep.15.4.143

[15] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," 2016.

[16] K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi, "Methods for comparing uncertainty quantifications for material property predictions," *Machine Learning: Science and Technology*, vol. 1, no. 2, p. 025 006, 2020, doi:10.1088/2632-2153/ab7e1a

[17] V. Kuleshov, N. Fenner, and S. Ermon, "Accurate uncertainties for deep learning using calibrated regression," in *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, 2018, pp. 2796–2804, https://proceedings.mlr.press/v80/kuleshov18a.html

[18] STFC's Scientific Computing Research Infrastructures Group. "SCARF Scientific Computing Application Resource for Facilities." (2022), https://www.scarf.rl.ac.uk/

[19] "PEARL." (2022), https://www.turing.ac.uk/research/asg/pearl