research papers

A neural network model of a quasiperiodic elliptically polarizing undulator in universal mode

Ryan Sheppard,[a,b] **Cameron Baribeau,**[a] **Tor Pedersen,**[a] **Mark Boland**[a,c] **and Drew Bertwistle**[a,c]*

[a]Canadian Light Source, 44 Innovation Blvd, Saskatoon, Saskatchewan, Canada, [b]McGill University, 817 Sherbrooke St West, Montreal, Quebec, Canada, and [c]University of Saskatchewan, Department of Physics and Engineering Physics, 116 Science Place, Saskatoon, Saskatchewan, Canada. *Correspondence e-mail: drew.bertwistle@lightsource.ca
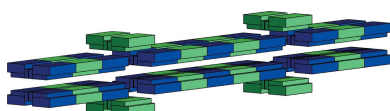
Machine learning has recently been applied and deployed at several light source facilities in the domain of accelerator physics. Here, an approach based on machine learning to produce a fast-executing model is introduced that predicts the polarization and energy of the radiated light produced at an insertion device. This paper demonstrates how a machine learning model can be trained on simulated data and later calibrated to a smaller, limited measured data set, a technique referred to as transfer learning. This result will enable users to efficiently determine the insertion device settings for achieving arbitrary beam characteristics.

1. Introduction

For decades, synchrotron light source facilities have produced highly brilliant and tunable photon beams for experiments across many scientific disciplines, in particular through the use of insertion devices (IDs). At the Canadian Light Source, the Quantum Materials Spectroscopy Center (QMSC) beamline uses an elliptically polarizing undulator (EPU) type ID with a magnetic period of 180 mm to produce soft X-rays with variable polarization in the energy range 15–200 eV.

In materials science, having the ability to probe the orbital structure of electronic states with linear and circular dichroism measurements is critical to understanding the underlying physics in the system under study. Angle-resolved photo-emission is one technique that can extract additional information from a sample by utilizing arbitrary polarization at low photon energies (Day *et al.*, 2019). However, 100% circular polarization is difficult to achieve due to the beamline optics altering the polarization of low-energy photons as they propagate from the ID to the experiment endstation (Wurtz *et al.*, 2014; Marcouille *et al.*, 2007). This introduces the requirement for arbitrary polarization of the light at the EPU, along with the corresponding requirement of knowing the EPU operating parameters that will deliver photons of a certain energy and polarization on demand.

A planar ID has its gap as one degree of freedom. In this case it is straightforward to build a one-dimensional look-up table relating the energy of the radiated photon beam to the device gap, where the look-up table is typically generated from magnetic or beam-based measurements. Operating an EPU in arbitrary polarization requires a multi-dimensional look-up table to relate its parameters to the energy and polarization state of the photon beam. Moreover, the overall system may drift over months or years, for example due to changes in characteristics of the undulator or the beamline optics. Look-

up tables built from measured data are limited to replacing their data one point in the ID's configuration space at a time, and hence the total time necessary to (re-)measure data for multi-dimensional look-up tables from either beam-based or magnetic data becomes prohibitively large. The measurement time can be sidestepped by instead computing the undulator output polarization at any arbitrary point in configuration space from a model, for example using *RADIA* (Elleaume *et al.*, 1997; Chubar *et al.*, 1998). However, such calculations remain time-consuming and the result is then limited by the accuracy of the model. The most attractive outcome is a fast-executing model that can be calibrated from a measured data set that is small compared with the size of a multi-dimensional look-up table. In this article we propose that neural networks can be just such a model, providing rapid accurate predictions of the beam characteristics from a complex undulator.

2. Background

2.1. Elliptically polarizing undulators and polarization

The QMSC undulator is a quasiperiodic APPLE-II type EPU. A section of its modelled magnet arrays is shown in Fig. 1. Certain magnet blocks are offset vertically to incorporate a quasiperiodic magnetic structure, which reduces contamination of the harmonics present in the undulator spectrum (Chavanne *et al.*, 1998). Gap adjustments symmetrically change the vertical distance between the upper and lower magnet arrays. Independent longitudinal motion of the four girders are described with two independent parameters for operating the device, called the elliptical phase φ_E and composite linear phase φ_L (Sigrist *et al.*, 2019). By adjusting these three operating parameters (gap, φ_E, φ_L), the strength and orientation of the undulator's magnetic field can be controlled, which in turn controls the energy and polarization of the radiated photons.

The polarization of the light radiated from the EPU can be described using the Stokes parameters $S1$, $S2$ and $S3$. For this application, the Stokes parameters are normalized and dimensionless, satisfying equation (1), where each parameter ranges from −1 to 1,

$$S1^2 + S2^2 + S3^2 = S0^2 \leq 1. \tag{1}$$

2.2. Machine learning

Machine learning (ML) techniques have been studied for various particle accelerator applications. Recently, ML-based surrogate models have obtained accurate and fast-executing representations of the relevant beam dynamics from a sparse sampling of the physics simulation (Edelen *et al.*, 2020). Neural networks (NN), a sub-type of ML, have been trained to automatically tune and control large complex systems such as particle accelerators and insertion devices (Leemann *et al.*, 2019; Scheinker *et al.*, 2019). Their ability to be trained off-line using simulation data from computationally expensive codes and updated with measurement data has been demonstrated
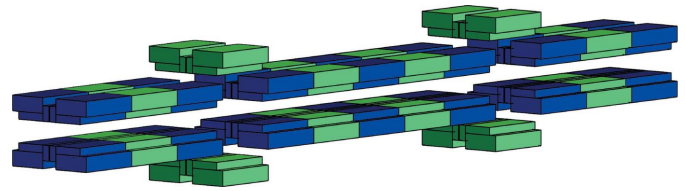

Figure 1
Section of the QMSC insertion device (EPU 180 mm) to illustrate the magnet array.

for multiple applications (Edelen *et al.*, 2010, 2020). This type of ML algorithm is referred to as supervised learning because the model is trained on labelled data sets. In this sense, ground truth outputs exist for each input (Arpaia *et al.*, 2021). In contrast to the simulation software from which ML models are trained, ML models can execute in fractions of a second with comparable accuracy in predicting the resulting beam parameters (Edelen *et al.*, 2020). Additionally, the ability of ML models to be updated with new measurement data ensures that they remain accurate as the characteristics of the modelled device changes (Edelen *et al.*, 2020).

With these advantages in mind, accomplishing the objective of this work entails acquiring a large training data set from simulations. ML models are able to learn complex nonlinear relationships using large amounts of training data; however, producing a large training data set is computationally expensive (Leemann *et al.*, 2019). In practice, the training data size depends on the complexity of the problem and complexity of the ML algorithm. Similar ML scenarios determined the amount of training data required by empirically evaluating the performance of their models with respect to the number of data points (Edelen *et al.*, 2020). This technique was used to determine the size of the required training data set. By varying the resolution of the ID settings in the training data, the size of the data set would change without affecting the equal representation of the operating modes of the ID within the data. The difficulty for ML models to interpolate between training points increases for complex, many-parameter systems (Scheinker *et al.*, 2019); therefore the data size was chosen such that the ID settings have sub-millimetre resolutions.

3. Methods

3.1. Modelling the undulator as a periodic device

As an initial proof of concept for this work, training data were generated from a *RADIA* model of the undulator built as a periodic device. In this simplified case, the photon beam characteristics are derived from the undulator's effective and nominal fields (Sigrist *et al.*, 2019) using equations (2), (5), (6) and (7) (Sigrist, 2018).

The effective field is an approximation of the undulator's peak field, $\hat{B}$, and is obtained via Fourier series decomposition of the modelled field profiles, $B_x(y)$ and $B_z(y)$. An example of one of these magnetic field profiles is shown in Fig. 2. Equation (3) shows the Fourier series decomposition over harmonic i. The effective field is specific to an EPU's

Ryan Sheppard *et al.* · Neural network model of a polarizing undulator

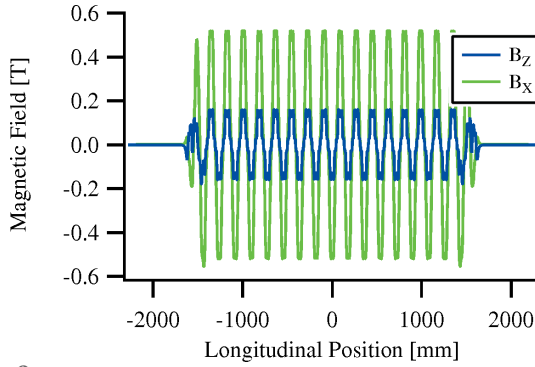| Symbol	Term	Unit
E_γ	Photon energy	eV
E	Electron energy	GeV
λ	Undulator period	mm
k	Wave number $= 2\pi\lambda^{-1}$	mm^{-1}
ϕ_E	Elliptical phase	mm
ϕ_L	Linear phase	mm
K	Deflection parameter	
$B_{x,z\mathrm{eff}}$	Effective horizontal, vertical field	T
$B_{x,z0}$	Nominal horizontal, vertical field	T

Figure 2
Magnetic field profile in EPU 180 mm, modelled as a periodic device, calculated at gap = 15 mm, $\phi_E = 0$ mm and $\phi_L = -60$ mm.

(gap, φ_E, φ_L) settings. Nominal fields are gap-dependent phase-independent terms, as per equation (4). B_{z0} is given by $B_{z\mathrm{eff}}$ at a horizontal polarization and similarly B_{x0} by $B_{x\mathrm{eff}}$ at a vertical polarization, such as $\varphi_E = \pm\lambda/2$ (see Table 1 for a list of the variables),

$$E_\gamma \,[\mathrm{eV}] = 9.50 \, \frac{E^2 \,[\mathrm{GeV}]}{\lambda \,[\mathrm{mm}] \left\{ 1 + (K^2/2) \right\}},$$

$$K = 0.0934 \, \lambda \,[\mathrm{mm}] \, \hat{B} \,[\mathrm{T}], \qquad (2)$$

$$\hat{B}^2 \simeq B_{x\mathrm{eff}}^2 + B_{z\mathrm{eff}}^2,$$

$$B_{x,z\mathrm{eff}} = \left(\sum_{x,zi}^{\infty} b_i^2 / i^2 \right)^{1/2}, \qquad (3)$$

$$B_{x,z0} = M \exp\left(c_1 \frac{\mathrm{gap}^1}{\lambda^1} + c_2 \frac{\mathrm{gap}^2}{\lambda^2} + \ldots \right), \qquad (4)$$

$$S1 = \frac{B_{x\mathrm{eff}}^2 - B_{z\mathrm{eff}}^2}{B_{x\mathrm{eff}}^2 + B_{z\mathrm{eff}}^2}, \qquad (5)$$

$$S2 = \begin{cases} \frac{1}{2} \frac{B_{x0} B_{z0} \sin^2 k\phi_L}{B_{x\mathrm{eff}}^2 + B_{z\mathrm{eff}}^2}, & \text{if } \phi_L \geq 0, \\[2mm] -\frac{1}{2} \frac{B_{x0} B_{z0} \sin^2 k\phi_L}{B_{x\mathrm{eff}}^2 + B_{z\mathrm{eff}}^2}, & \text{if } \phi_L < 0, \end{cases} \qquad (6)$$

$$S3 = \begin{cases} \frac{\left(16 B_{x\mathrm{eff}}^2 B_{z\mathrm{eff}}^2 - B_{x0}^2 B_{z0}^2 \sin^4 k\phi_L \right)^{1/2}}{2 \left(B_{x\mathrm{eff}}^2 + B_{z\mathrm{eff}}^2 \right)}, & \text{if } \phi_E \geq 0, \\[3mm] \frac{\left(16 B_{x\mathrm{eff}}^2 B_{z\mathrm{eff}}^2 - B_{x0}^2 B_{z0}^2 \sin^4 k\phi_L \right)^{1/2}}{2 \left(B_{x\mathrm{eff}}^2 + B_{z\mathrm{eff}}^2 \right)}, & \text{if } \phi_E < 0. \end{cases} \qquad (7)$$
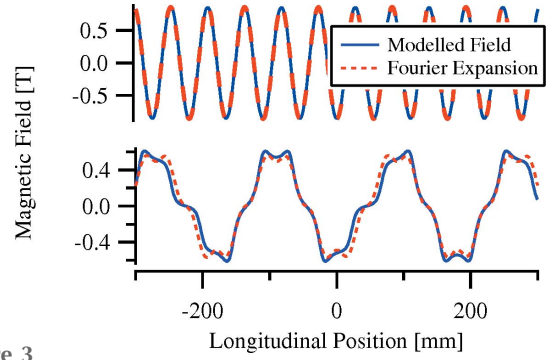

Figure 3
Magnetic field profile (solid) and effective field (dashed) calculated from a Fourier expansion thereof. Top: 55 mm periodic EPU. Bottom: 180 mm quasiperiodic EPU.

However, describing the field profile in terms of effective fields introduces an approximation that holds poorly for quasiperiodic undulators. This point is illustrated in Fig. 3, which shows modelled undulator fields and their Fourier-determined effective equivalents for two cases. The upper plot shows a periodic undulator with a 55 mm period, where the effective field closely matches the undulator field; the lower plot shows the 180 mm quasiperiodic device under consideration, where the effective and undulator fields do not match. Calculating photon energy for the $n = 1$ harmonic from the effective field for this configuration yields 10.7 eV, whereas a more direct calculation (see next section) yields 9.4 eV. These results differ by 12%, which highlights the inapplicability of Fourier decomposition for studying quasiperiodic fields.

3.2. Modelling the undulator as a quasiperiodic device

The photon beam characteristics can be determined without the approximation inherent to the effective field. This is achieved by modelling the undulator in its quasiperiodic configuration and exporting magnetic field data for analysis in the *Synchrotron Radiation Workshop* (*SRW*) code (Chubar & Elleaume, 1998). Undulator radiation spectra are calculated at an observation window 8 mm by 8 mm in size and 18 m downstream of the undulator. The calculation uses a non-filament electron beam defined for a straight section in the Canadian Light Source (CLS) storage ring; see Table 2 for the beam characteristics. The spectra are calculated separately for the total (*S*0), horizontal (0°), vertical (90°), inclined linear (45°, 135°) and left- and right-circular polarizations. Stokes parameters are then obtained by comparing the flux at the $n = 1$ harmonic for the different polarizations.

Scripting was developed in *IGOR Pro* to generate the large data set for training ML models (Wavemetrics, 2018). The script can import and process magnetic field data for any number of EPU configurations. For each configuration, undulator radiation spectra are computed across an energy range near the undulator's $n = 1$ harmonic; the expected energy is calculated using Fourier-determined effective fields. The precise photon energy of the $n = 1$ undulator harmonic is determined by fitting a curve to the total photon flux. The photon beam characteristics and their corresponding ID settings describe a single case for the ML model.

Table 2
Electron beam characteristics defined 2.125 m upstream of the centre of the ID straight section in the CLS storage ring.

Parameter	Value	Unit
Energy	2.9	GeV
Current	220	mA
Horizontal emittance	22.7	nm
Horizontal beta function	9.664	m
Horizontal alpha	0.2315	m
Vertical emittance	0.1017	nm
Vertical beta function	4.331	m
Vertical alpha	0.8285	m
Relative energy spread	0.001	
Horizontal dispersion	0.15	m
Vertical dispersion	0	m
Dispersion derivatives	0	

Two example sets of undulator spectra are shown in Fig. 4. The script's 'information pipeline' and overall procedure for training an ML model is illustrated in Fig. 5 (Edelen *et al.*, 2020).

Lastly, it is important to note that this methodology amounts to training an ML model based on the output of a *RADIA* model. Prior to this work, the *RADIA* model was refined with bench-based magnetic measurements of the actual undulator using a Hall probe and flipping coil setup. The *RADIA* model's tuning process considered 45 EPU
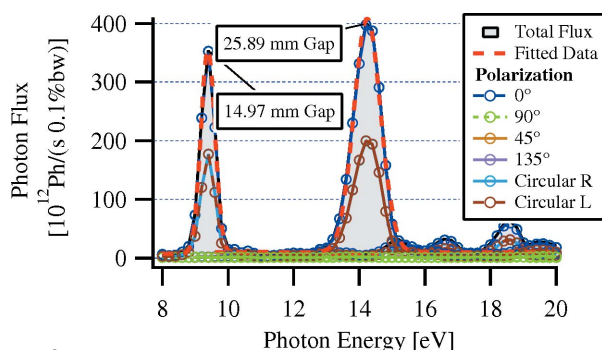

Figure 4
Radiation spectra calculated for various polarization modes in *SRW* for two configurations of the quasiperiodic *RADIA* model. The configurations are at different gaps and are both in planar polarization ($\phi_E = \phi_L = 0$ mm).
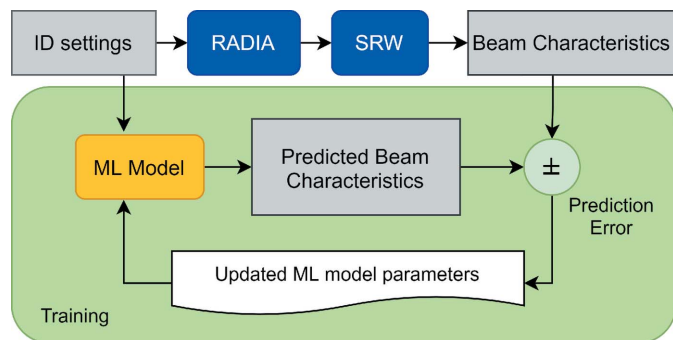

Figure 5
The procedure for training an ML model using the physics simulation as the ground truth (Edelen *et al.*, 2020). The ML model parameters are dynamically updated during training. The hyperparameters of the model are manually adjusted until the model reaches a target performance.

configurations, with priority given to planar and vertical polarization modes across various gap settings. Across the considered configurations, the typical relative difference between modelled magnetic fields compared with bench-based measurements on- and off-axis is 1%.

4. Developing a neural network model

A neural network model was created to predict four outputs, namely the photon beam energy and Stokes parameters $S1$, $S2$, $S3$. A neural network is composed of individual neurons that accept multiple inputs and produce a single output. These neurons are arranged in layers to form a connected network (Smith, 1997). The developed neural network is a feed-forward network in that the data propagate from input to output without looping between intermediate layers. The created model is hereafter referred to as NN4, as its final layer has four output neurons corresponding to the beam parameters. The neural network was implemented using *Keras* with *Tensorflow* 2.0 backend and open-source *scikit-learn* packages (Abadi *et al.*, 2016; Pedregosa *et al.*, 2011).

The architecture of the neural network is a four hidden layer (128–64–32–16), fully connected neural network with a rectified linear unit activation function for each layer. The model was trained using backpropagation with the Adam optimizer (LeCun *et al.*, 1989). The mean squared error (MSE) was used both as a loss function and metric to monitor the performance of the model; the MSE compares the model output, namely scaled photon energy and Stokes parameters, to the training data. The neural network model used scaled inputs in the range $(0, 1)$ and scaled outputs in the range $(-1, +1)$. The data set not used for training is divided equally, resulting in a 60–20–20 split of the training, validation and testing data, respectively.

Although a deep (many hidden layers) and wide (many nodes per layer) NN generally provides better fitting on training data, it is prone to overfitting (Leemann *et al.*, 2019). This issue was minimized by shuffling the data, implementing a learning schedule, adjusting the number of epochs (number of times the model is trained on a subset of data), and adjusting the batch size (the subset data size shown during training).

The simulated training data for the NN model contained 4175 cases that sampled the EPU's operating modes: planar, vertical (ϕ_E and ϕ_L), circular (helicity 1 and 2), elliptical, linear, inclined (helicity 1 and 2), and a selection of universal modes near circular at photon energies of interest to the beamline. To cover the total configuration space of the device, an additional 1000 cases were randomly generated for each quadrant formed by ϕ_E and ϕ_L, for a total of 8175 cases. The 4175 and 8175 case data sets are shown in Figs. 6 and 7, respectively. A single operating mode, E45:L45 ($\phi_E = 45$ mm, $\phi_L = 45$ mm, gap = 15 mm), was not included in the smaller data set for reasons explained in Section 5.

The train/test splitting technique for sorting the simulated data was employed with the ML models. This method entails dividing the data so that one group is used to train the model and a separate group is used to test the model. This compu-
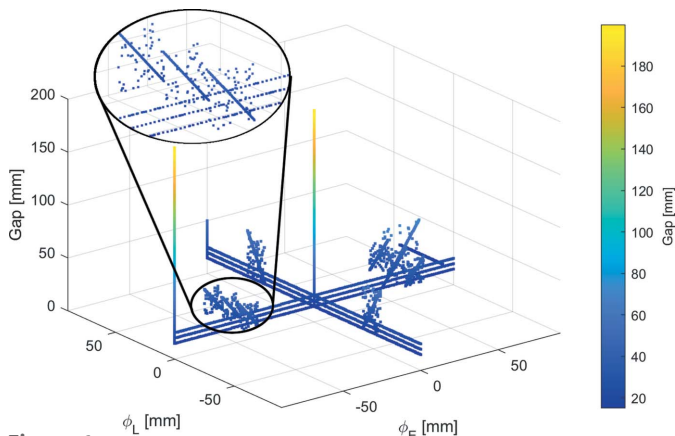
Figure 6
Distribution of 4175 simulated cases covering the operating modes of the EPU. A zoomed-in portion of the figure is provided for clarifying that the apparent 'lines' on the figure are composed of individual cases.
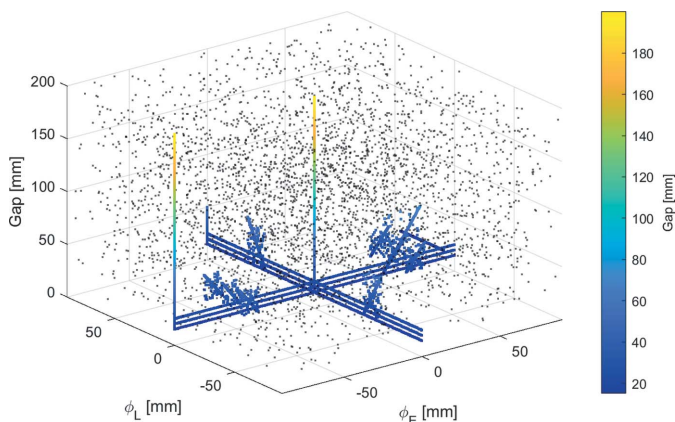

Figure 7
Total distribution of simulated data. The randomly generated cases are represented as black points to distinguish them from the initial 4175 cases.

tationally efficient approach was suitable because the data were shuffled prior to sorting, thereby guaranteeing the configuration space of the device was equally represented in the training and testing data sets. An appropriate distribution of the training and testing data is critical for effectively evaluating model performance. Note that this approach is equivalent to performing a k-fold cross-validation procedure with $k = 2$ (Stone, 1974). The unscaled inputs occupy the following ranges: gap = 15 mm to 200 mm; ϕ_E and ϕ_L = -90 mm to 90 mm. The unscaled outputs occupy the following ranges: E_γ = 6 eV to 400 eV, Stokes parameters = -1 to 1.

5. Predicting modelled EPU beam characteristics

The following results are drawn from the model's performance on the simulated data sets for the EPU. The model used batch sizes of 16, a customized decaying learning rate schedule, and trained for 1500 epochs.

5.1. Configuration space

The first iteration of the NN model was trained on the simulated operating-modes data set, totalling 4175 unique
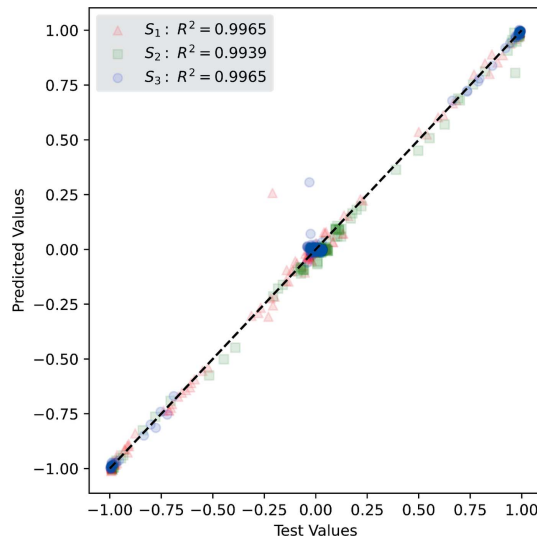

Figure 8
Comparison between predicted and test output Stokes parameters for measured data from the NN. Excellent agreement is shown for the majority of test cases except for the E45:L45 case (predicted $S1$ value of ~ 0.26 versus a test value of ~ -0.21; predicted $S2$ value of ~ 0.8 versus a test value of ~ 0.97; predicted $S3$ value of ~ 0.31 versus a test value of ~ 0).

cases. After testing the model on cases from the measured data set, it was apparent that the model did not generalize well to domains in configuration space not covered in the training data. In particular, a single operating mode (E45:L45) contained in the measured data set was not encompassed by the domain of the simulated operating-modes data set (Fig. 6) used to train the model. This E45:L45 case was intentionally set aside from the simulated operating-mode data set to observe the model's ability to extrapolate for new domains. The predictions made by the NN model on the measured data set are shown in Fig. 8 to demonstrate how the single E45:L45 case stands apart from other predicted cases. Although strict agreement between the predicted and test cases is not expected because the simulated and measured data sets are inherently unique, general agreement is expected.

To test the prediction that the model requires training on each domain for which it will be tested, the model was trained on the operating-modes data set (4175 cases) with one additional quadrant of randomly generated data. The model accurately predicted the E45:L45 case when the extra quadrant data encompassed the E45:L45 case and poorly otherwise.

Since we desire a ML model that may be used to predict the EPU beam characteristics for any operating mode, current or future, the second iteration of training the model was performed on the complete data set shown in Fig. 7, which will be referred to as the simulated data set from now on.

5.2. Results

The ML model was evaluated based on its MSE, mean norm of the Stokes error vector (MSEV) shown in equation (8), the variance of their relative error in predictions, the mean absolute percentage error (MAPE) for the predicted photon

Table 3
The model's performance in predicting EPU beam characteristics.

The metrics in this table are computed on the scaled predicted values which lie in the range (-1 to 1). The metrics of the best performing model from the compilation are included to illustrate the model's performance potential.

Model	MSE	MSEV
NN4	Average: 2.86×10^{-4} Best: 2.26×10^{-4}	Average: 2.01×10^{-2} Best: 1.76×10^{-2}

Table 4
Properties of the model's predicted beam characteristics.

The regression scores are based on the scaled predicted values and the variance scores are computed from the unscaled predicted values.

Target	R^2	Variance σ^2
E_γ	0.9998	4.87 eV^2
$S1$	0.9992	3.49×10^{-4}
$S2$	0.9982	2.68×10^{-4}
$S3$	0.9986	3.88×10^{-4}

energies and whether it satisfied the QMSC beamline's error threshold shown in equation (9),

$$\text{MSEV} = \frac{1}{n} \sum_{i}^{n} || \mathbf{S}_{\text{pred}_i} - \mathbf{S}_{\text{true}_i} ||. \qquad (9)$$

Equation (8) uses the Stokes parameters in vector notation where $\mathbf{S}_{\text{pred}}$ and $\mathbf{S}_{\text{true}}$ are the predicted and target Stokes vectors, respectively. The MSEV then represents the norm of the Stokes error vector averaged over n test cases,

$$\begin{cases} \text{if } |S3| = 1, \\ \text{residual}: \ S1^2 + S2^2 \ \leq \ 0.03. \end{cases} \qquad (9)$$

Equation (9) states that, for operation in circular mode, $|S3| = 1$, the residual components of the Stokes vector must be less than a threshold of 0.03.

To evaluate how well the model generalized to the input configuration space, the predictions made by the model were compared with the target values. This step is performed using the testing data, which is 'unseen' by the model during its training. The NN model was compiled 30 times to establish its average performance. A summary of the model's performance at predicting EPU beam characteristics from simulated data is given in Table 3.

The QMSC beamline's error threshold was satisfied by the model; see Fig. 9. The predicted residuals follow the same trend (magnitude and frequency) as the residuals from the test values. Cases were selected for inclusion in Fig. 9 using the tolerance $|S3| = 1 \pm 0.01$. In the majority of such cases, the
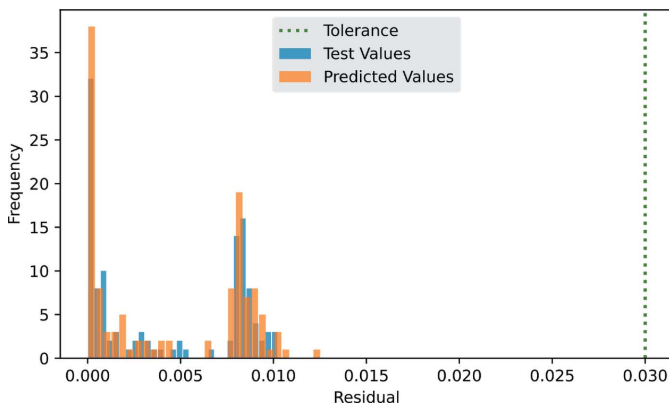

Figure 9
Histogram comparing the residuals for predictions on $|S3| = 1$ cases by the NN4 model with the residuals from the test data set. These cases correspond to cases where the test value of $|S3| = 1$ within a tolerance of ± 0.01.

undulator linear phase is zero and gap and elliptical phase are coordinated; these are the typical usage cases and form the first peak near zero residual. The selected cases also include randomly generated undulator configurations where a small linear phase setting may inflate the residual; these cases form the second peak near 0.08.

The accuracy of the 1633 predictions made by the NN4 model on the test data is shown in Table 4. A regression score, R^2, is calculated for each output beam characteristic to indicate the correlation between predicted and test values. The variances of the relative errors from the predicted test cases of the NN4 model are included to represent the distribution of errors. The averaged MAPE indicates that the NN4 model predicted the photon energy within 2.80%. The near-unity R^2 values for each output indicate that the NN4 model accurately predicts the EPU beam characteristics. The small variances of the relative errors in predictions indicate that the predictions made by the NN4 model are tightly distributed around the mean (zero).

6. Updating the neural network model using transfer learning

The ability of an ML model to be updated with new data, as mentioned in Section 1, was investigated to determine whether the model could predict the Stokes parameters derived from the magnet measurement data. The limited measured data mentioned in Section 3 were used to update the ML model.

Since the beam characteristics are only slightly different between the simulated and measured data sets, and the measured data set is small, the calibrated neural network model (Calibrated NN) used the entire NN4 model as the base model. This methodology involving the bottleneck layer of a trained model in transfer learning applications has been demonstrated by several computer vision works (Wang et al., 2020). Since the bench-based magnetic measurements capture a small subset of the EPU configuration space, the updated model fits the measured data better with fewer trainable parameters. The measured data set contains 169 cases that proportionally represent the operating modes of the device; the 45 cases used to tune the *RADIA* model are a subset of this data set. The Calibrated NN model was trained on 60% of these data (101 cases) and tested on the remaining 40% (68 cases). The model used batch sizes of 4, a customized decaying learning rate schedule, and trained for 300 epochs.
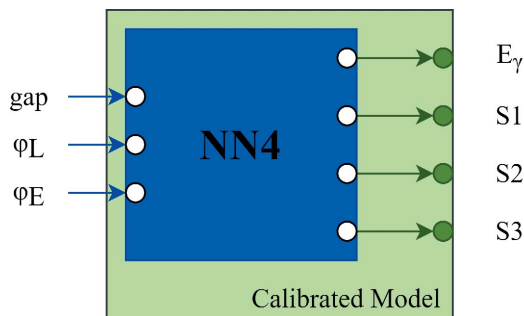
Figure 10
Architecture of the Calibrated NN model. The NN4 model acts as the base model with four outputs and the additional layer is added on top of NN4 (shown in green).

This updated neural network (Calibrated NN) adds one additional layer, identical in structure to the base model output layer (size 4, fully connected, using the linear activation function). A diagram is provided in Fig. 10 to show the architecture of the Calibrated NN model. Similar techniques to those described in Section 4 were employed to optimize the Calibrated NN model.

6.1. Transfer learning results

The Calibrated NN model was evaluated with the same metrics as described in Section 5, although it is now evaluated on the measured data set. The Calibrated NN model was compiled 60 times to establish an average performance. It is important to note that the performance of the Calibrated NN model may only be compared with the NN4 model when they are evaluated on similar data sets. To that end, the performance of the NN4 model was averaged over 60 trials and evaluated on the measured data set, rather than the simulated data on which it was trained. For comparison, a separate neural network model was also created and trained solely on the measured data. This Limited NN model was optimized to fit the measured data set and its performance was averaged over 60 trials. Results for these three models are listed in Table 5.

Like the NN4 model's performance on the simulated data, the Calibrated NN model also satisfied the QMSC beamline's error threshold on the measured data. The largest residual for the six $|S3| = 1$ test cases was 0.005. The Calibrated NN model's prediction accuracy on the measured data set is included in Table 6.

6.2. Discussion

The results in Table 5 indicate how the Calibrated NN model outperforms the Limited NN model by roughly two orders of magnitude. This comparison demonstrates the advantage of applying transfer learning to a base model that was first thoroughly trained on simulated data when the measurement-based data set is small. The improvement of the Calibrated NN model upon the NN4 model is identified by the smaller errors in predictions. However, the MAPE for the Calibrated NN is larger than that for the NN4 model when

Table 5
Summarized performance of the neural networks on the measured data set.

The metrics in this table are computed on the scaled predicted values which lie in the range -1 to 1. The Limited NN model was trained using bench-based magnetic measurement data. The NN4 model was trained using simulated data. The Calibrated NN used NN4 as a base model and was then updated using the measured data via transfer learning.

Model	MSE	MSEV	MAPE
Limited NN	Average: 1.07×10^{-2} Best: 2.34×10^{-3}	Average: 1.16×10^{-1} Best: 7.15×10^{-2}	32.7%
Calibrated NN	Average: 1.31×10^{-4} Best: 9.72×10^{-5}	Average: 1.86×10^{-2} Best: 1.57×10^{-2}	4.57%
NN4	Average: 5.32×10^{-4} Best: 4.20×10^{-4}	Average: 3.57×10^{-2} Best: 3.21×10^{-2}	3.56%

Table 6
Properties of the Calibrated model's predicted beam characteristics from the 68 test cases.

The regression scores are based on the scaled predicted values and the variance scores are computed from the unscaled predicted values.

Target	R^2	Variance σ^2
E_γ	0.9995	2.76 eV^2
$S1$	0.9997	1.56×10^{-4}
$S2$	0.9987	1.96×10^{-4}
$S3$	0.9997	8.02×10^{-5}

evaluated against the measured data. This indicates that the transfer learning somewhat reduced the Calibrated NN model's accuracy in predicting the photon energy of the beam, despite improving the accuracy of the predicted Stokes parameters.

The similar performance characteristics between Tables 4 and 6 indicate that the Calibrated NN model has a comparable prediction accuracy on the measured data to the NN4 model on the simulated data. The near-unity regression scores in Table 6 indicate that the model is accurately predicting the beam characteristics and the small variances imply that the relative errors are small and closely distributed around zero.

7. Conclusion

The results of this work demonstrate the feasibility of generating a ML model to accurately predict the photon beam characteristics of a quasiperiodic EPU. More specifically, this work demonstrates the ability of a neural network to accurately model the complex, multi-parameter functions of an ID. This outcome was achieved by optimizing the neural network model to fit a large simulated data set. The importance of properly sampling the configuration space in the development of a neural network was also demonstrated by the E45:L45 case.

Secondly, the successful application of transfer learning demonstrates how the neural network model was easily adapted to a measured data set. This stage was accomplished by building a separate neural network model, referred to as the Calibrated NN model, based on the NN4 model. This model was then trained on the limited magnetic measured

data set to provide more accurate predictions of the radiated light at the ID. The predictions produced by the Calibrated NN model satisfy the QMSC beamline's error threshold and the relative errors in predictions were shown to be within an acceptable threshold. The photon energy was predicted more accurately by the NN4 than the Calibrated NN model, as indicated by the slightly smaller averaged MAPE. However, this accuracy difference is small. The Calibrated NN model showed promising improvement in predicting the Stokes vector. The MSEV was determined to be 1.86×10^{-2}, indicating the predicted Stokes vectors closely agree with the test Stokes parameters.

Thirdly, the deployment of this updated neural network model provides a synchrotron beamline with a fast-executing model for producing look-up tables and/or predicting single ID cases.

Future work for this project includes the development of an ML model that will predict the beam characteristics at the endstation by following a similar training and calibration approach. Polarization measurements will be acquired using a polarimeter located at the endstation and used to calibrate a neural network from this work. The completion of this work will provide users with an efficient tool for predicting the endstation beam characteristics for arbitrary ID configurations.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. & Zheng, X. (2016). *arXiv*:1605.08695

Arpaia, P., Azzopardi, G., Blanc, F., Bregliozzi, G., Buffat, X., Coyle, L., Fol, E., Giordano, F., Giovannozzi, M., Pieloni, T., Prevete, R., Redaelli, S., Salvachua, B., Salvant, B., Schenk, M., Camillocci, M. S., Tomás, R., Valentino, G., Van der Veken, F. & Wenninger, J. (2021). *Nucl. Instrum. Methods Phys. Res. A*, **985**, 164652.

Chavanne, J., Elleaume, P. & Vaerenbergh, P. V. (1998). *Proceedings of the Sixth European Particle Accelerator Conference (EPAC'98)*, 22–26 June 1998, Stockholm, Sweden, pp. 2213–2215. MOP07F.

Chubar, O. & Elleaume, P. (1998). *Proceedings of the Sixth European Particle Accelerator Conference (EPAC'98)*, 22–26 June 1998, Stockholm, Sweden, pp. 1177–1179. THP01G.

Chubar, O., Elleaume, P. & Chavanne, J. (1998). *J. Synchrotron Rad.* **5**, 481–484.

Day, R. P., Zwartsenberg, B., Elfimov, I. S. & Damascelli, A. (2019). *NPJ Quantum Mater.* **4**, 54.

Edelen, A., Erdstrom, D. R., Piot, P. R., Halavanu, A., Edelen, J. & Biedron, S. (2010). *Proceedings of the Ninth International Particle Accelerator Conference (IPAC'18)*, 29 April–4 May 2018, Vancouver, Canada. Poster SUSPL054.

Edelen, A., Neveu, N., Frey, M., Huber, Y., Mayes, C. & Adelmann, A. (2020). *Phys. Rev. Accel. Beams*, **23**, 044601.

Elleaume, P., Chubar, O. & Chavanne, J. (1997). *Proceedings of the 1997 Particle Accelerator Conference (PAC'97)*, 16 May 1997, Vancouver, BC, Canada. 9P027.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. (1989). *Neural Comput.* **1**, 541–551.

Leemann, S. C., Liu, S., Hexemer, A., Marcus, M. A., Melton, C. N., Nishimura, H. & Sun, C. (2019). *Phys. Rev. Lett.* **123**, 194801.

Marcouille, O., Brunelle, P., Chubar, O., Marteau, F., Massal, M., Nahon, L., Tavakoli, K., Veteran, J. & Filhol, J. (2007). *AIP Conf. Proc.* **879**, 311–314.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.

Scheinker, A., Bohler, D., Tomin, S., Kammering, R., Zagorodnov, I., Schlarb, H., Scholz, M., Beutner, B. & Decking, W. (2019). *Phys. Rev. Accel. Beams*, **22**, 082802.

Sigrist, M. (2018). *Calculating Photon Polarisation for an APPLE-II EPU*. Technical Report. Canadian Light Source, Saskatoon, Canada.

Sigrist, M. J., Baribeau, C. K. & Pedersen, T. M. (2019). *Proceedings of the 10th International Particle Accelerator Conference (IPAC 2019)*, 19–24 May 2019, Melbourne, Australia, pp. 1687–1690. TUPRB005.

Smith, S. W. (1997). *The Scientist and Engineer's Guide to Digital Signal Processing*, ch. 26, pp. 451–480. San Diego: California Technical Publisher.

Stone, M. (1974). *J. R. Stat. Soc. Ser. B*, **36**, 111–147.

Wang, W., Seraj, F., Meratnia, N. & Havinga, P. J. (2020). *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom 2020)*, 23–27 March 2020, Austin, TX, USA, pp. 197–204.

Wavemetrics, (2018). *Igor Pro*, https://www.wavemetrics.com.

Wurtz, W. A., Bertwistle, D., Dallin, L. O. & Sigrist, M. J. (2014). *Proceedings of the 5th International Particle Accelerator Conference (IPAC2014)*, 15–20 June 2014, Dresden, Germany, pp. 1995–1997.