Predicting the transverse emittance of space charge dominated beams using the phase advance scan technique and a fully connected neural network

F. Mayet,* M. Hachmann, K. Floettmann, F. Burkart, H. Dinter, W. Kuropka, T. Vinatier, and R. Assmann
Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany
(Dated: May 20, 2022)

The transverse emittance of a charged particle beam is an important figure of merit for many accelerator applications, such as ultra-fast electron diffraction, free electron lasers and the operation of new compact accelerator concepts in general. One of the easiest to implement methods to determine the transverse emittance is the phase advance scan method using a focusing element and a screen. This method has been shown to work well in the thermal regime. In the space charge dominated laminar flow regime, however, the scheme becomes difficult to apply, because of the lack of a closed description of the beam envelope including space charge effects. Furthermore, certain mathematical, as well as beamline design criteria must be met in order to ensure accurate results. In this work we show that it is possible to analyze phase advance scan data using a fully connected neural network (FCNN), even in setups, which do not meet these criteria. In a simulation study, we evaluate the perfomance of the FCNN by comparing it to a traditional fit routine, based on the beam envelope equation. Subsequently, we use a pre-trained FCNN to evaluate measured phase advance scan data, which ultimately yields much better agreement with numerical simulations. To tackle the confirmation bias problem, we employ additional mask-based emittance measurement techniques.

I. INTRODUCTION

Many modern particle accelerators are tuned to achieve as small transverse beam emittance as possible. This is due to the fact that most users demand for the highest beam brightness possible. Beam brightness is important for many accelerator applications, such as ultra-fast electron diffraction [1], free electron lasers [2] and the operation of new compact accelerator concepts in general (e.g. [3–6]). A common definition of brightness is [7]

$$B = \frac{\eta I}{\pi^2 \varepsilon_x \varepsilon_y}, \tag{1}$$

where η is a form factor close to unity, I is the beam peak current and $\varepsilon_{x,y}$ the horizontal and vertical transverse emittance respectively. Hence, in order to maximize B, transverse emittance has to be minimal.

There are multiple methods to characterize the transverse emittance. One of the most common techniques is the phase advance scan technique, where the transverse beam size is recorded on a screen vs. the focusing strength of an upstream quadrupole or solenoid magnet [8–10]. The data can then be fitted based on the beam envelope equation. Space charge effects can be included to some extent [11, 12]. Instead of scanning the focusing strength of a magnet, also multiple screens can be used to record the beam size vs. the phase advance. Other – potentially single-shot – methods involve insertion of masks into the beamline, which then, subsequently, can be imaged on a downstream screen [13]. Coupled with advanced reconstruction algorithms these methods are capable of delivering reconstruction of the core 4D phase space [14]. In this work we concentrate on the phase advance scan technique, as this is the easiest one to implement, only requiring standard beamline components.

One of the limitations of the phase advance scan technique is that there is no closed description of the beam envelope for space charge dominated beams [11, 12]. It is therefore difficult to apply the method in this regime. Space charge dominated beams especially occur, for example, in the injector part of high-brightness electron sources, where the beam is still non-relativistic. In order to quantify whether a beam is space charge dominated, the so-called *laminarity parameter* ρ can be calculated [15]. This parameter represents the ratio between the space charge term and the emittance term of the beam envelope equation. It is given by

$$\rho = \frac{I\sigma^2}{2I_\mathrm{A}\gamma\varepsilon_n^2}, \tag{2}$$

where I is the peak current of the beam, $I_\mathrm{A} \approx 17\,\mathrm{kA}$ is the Alfvén current and $\varepsilon_n = \beta\gamma\varepsilon$ is the normalized emittance with the Lorentz factor γ and $\beta = v/c$. In case $\rho \gg 1$, the beam can be considered as space charge dominated (laminar flow regime). Otherwise the evolution of the beam envelope is dominated by the emittance pressure (thermal regime).

In this work we show in simulation that it is possible to successfully analyse phase advance scan data for $\rho \gg 1$ beams using a pre-trained fully connected neural network (FCNN). Subsequently, we apply the method to measured data. Machine learning and neural networks in particular have recently been used in the context of accelerators for various purposes. These include, among others, fault detection of machine components [16], machine stability optimization and analysis [17–19], virtual diagnostics [20–22], beam quality optimization in plasma accelerators [23, 24] and orders of magnitude speed-up

* frank.mayet@desy.de

in multiobjective optimization of accelerator parameters [25]. Here, we focus on the analysis of otherwise difficult to interpret measurement data.

II. MEASUREMENT TECHNIQUE

The state of a single particle with respect to a given design, or reference trajectory, is usually defined by the 6D phase space vector

$$\mathbf{X} = (x, x', y, y', z, \delta)^{\mathrm{T}}, \tag{3}$$

where $x' = p_x/p_z$ and $y' = p_y/p_z$ are the horizontal and vertical divergence respectively, x, y, z the distances of the particle from the reference trajectory and $\delta = \Delta p/p_0$ is the relative deviation of the particle's individual momentum from the reference momentum. p_x, p_y, p_z are the three momentum components and $()^{\mathrm{T}}$ denotes the transpose of a matrix. We are interested in the evolution of the 4D transverse phase space vector

$$\mathbf{X}^{(\mathrm{tr})} = (x, x', y, y')^{\mathrm{T}}. \tag{4}$$

Assuming negligible correlation between the evolution of the phase space coordinates in the x and y planes, it is possible to treat them separately, yielding the two subspace vectors

$$\mathbf{x} = (x, x')^{\mathrm{T}}, \qquad \mathbf{y} = (y, y')^{\mathrm{T}}. \tag{5}$$

A common framework to describe the evolution of a charged particle is linear beam optics, where only linear transformations of $\mathbf{x}$ and $\mathbf{y}$ are taken into account. Each beamline element is represented by a so-called transfer matrix defined by the relation

$$\mathbf{x} = \mathbf{M} \cdot \mathbf{x_0}, \tag{6}$$

where $\mathbf{x_0} = (x_0, x_0')^{\mathrm{T}}$ is the initial phase space coordinate and

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}. \tag{7}$$

The 2×2 matrix for a simple drift is given by

$$\mathbf{M}_{\mathrm{D}}(s) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \tag{8}$$

where s is the drift distance. Applying this matrix to $\mathbf{x_0}$ would result in $x = x_0 + x's, x' = x_0'$, as expected.

In an experiment only the rms beam size $\sigma_x = \sqrt{\langle x^2 \rangle}$ is accessible, where $\langle\rangle$ denotes the second central moment. Applying this to the general equation

$$x = M_{11} x_0 + M_{12} x_0' \tag{9}$$

yields

$$\begin{aligned} \sigma_x^2 =& M_{11}^2 \sigma_{x,0}^2 \\ &+ 2 M_{11} M_{12} \sigma_{x,0} (\sigma_{x,0})' \\ &+ M_{12}^2 \left(\frac{\varepsilon_x^2}{\sigma_{x,0}^2} + (\sigma_{x,0})'^2 \right) \end{aligned} \tag{10}$$

with $\varepsilon_x = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2}$. Equation 10 is the so-called rms envelope equation, which can be used to determine the transverse rms emittance ε_x using a suitable (i.e. tunable) beam transformation $\mathbf{M}$. Note that Eq. 10 does not take any space charge effects into account and is hence only valid in the $\rho \approx 1$ regime.

Figure 1 shows a sketch of a potential measurement scenario. The elements and distances are chosen according to what is installed at the ARES electron linac at DESY, Hamburg [26]. The transfer matrix of the double solenoid magnet can be written as

$$\begin{aligned} \mathbf{M}_{\mathrm{DS}} &= \mathbf{M}_{\mathrm{TL}} \cdot \mathbf{M}_{\mathrm{D}}(l_{\mathrm{D}}) \cdot \mathbf{M}_{\mathrm{TL}} \\ &= \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & l_{\mathrm{D}} \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - l_{\mathrm{D}}/f & l_{\mathrm{D}} \\ (l_{\mathrm{D}} - 2f)/f^2 & 1 - l_{\mathrm{D}}/f \end{pmatrix}, \end{aligned} \tag{11}$$

where l_{D} is the drift distance between the two single solenoids and f the focal length of each solenoid. Here the approximation that f is larger than the length of the solenoid was used, i.e. the thin lens approximation. The focal length of a solenoid is given by [27]

$$f(B_{z,\max}) = \left[\left(\frac{q}{2\overline{p}_z} \right)^2 F_2 \right]^{-1}, \tag{12}$$

where $B_{z,\max}$ is the peak magnetic field, q the particle charge and $\overline{p}_z$ the average longitudinal beam momentum. $F_2 = \int B_z^2 \mathrm{d}z \propto B_{z,\max}^2$ is the second field integral of the on-axis magnetic field. By inserting the expression $\mathbf{M}_{\mathrm{D}}(l_{\mathrm{S}}) \cdot \mathbf{M}_{\mathrm{DS}}$, where l_{S} is the drift between the solenoid and the screen, in Eq. 10, it can be seen that now the M_{ij} elements can be conveniently adjusted in the experiment as $B_{z,\max}$ is varied. The emittance at the position of the solenoid can thus be determined by fitting the recorded $\sigma_{x,i}$ vs. $B_{z,\max,i}$ at the screen with Eq. 10.

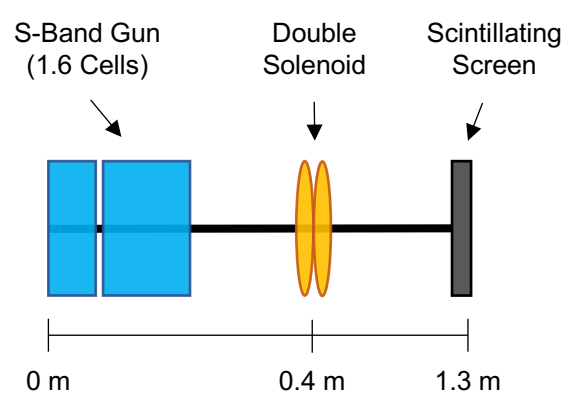

FIG. 1. Schematic of a suitable beamline layout for transverse emittance measurements using the phase advance scan technique. The distances are based on what is installed at the ARES linac at DESY, Hamburg.

It is possible to include transverse space charge forces into the model to some extent. This is done by including

a defocusing term in the drift between the focusing element and the screen. Considering a uniformly charged cylindrical bunch with radius R and length L, the envelope equation then reads in differential form [12]

$$\sigma_x'' - \frac{P}{4\sigma_x} G(\xi, A) - \frac{\varepsilon_x^2}{\sigma_x^3} = 0, \tag{13}$$

where $G(\xi, A) \in [0, 1]$ is a form factor, which depends on the centered longitudinal intra-bunch coordinate ξ and the rest frame aspect ratio $A = R/(\gamma L)$. P is the so-called generalized perveance given by

$$P = \frac{eQ}{2\pi\epsilon_0 m_e c^2 L \beta^2 \gamma^3}, \tag{14}$$

where Q is the total charge of the bunch, ϵ_0 the vacuum permittivity and m_e the electron rest mass. In principle Eq. 13 can now be used to construct a similar fit model to Eq. 10. There are a number of caveats to take into account, however:

- The model is only fully valid for a perfectly cylindrical bunch.
- The form factor G depends on the aspect ratio, which depends on the transverse beam size and bunch length, both not being constant in the experiment (especially in the over-focused part of the scan).
- Since the concept of the envelope equation relies on the emittance being a constant of motion, nonlinear space charge forces are intrinsically neglected in this approach.
- The perveance term depends on the bunch length, which might not be accessible to sufficient precision in the experiment.
- Equation 13 cannot be solved analytically and has to be approximated by a polynomial series (see [11] for a detailed description).

All of these caveats lead to the conclusion, that the envelope equation based data analysis method is not ideal in the $\rho \gg 1$ regime.

Independent of the value of ρ, two more criteria need to be met in order to ensure an accurate fit result. Considering the third term of Eq. 10, the purely mathematical criterion

$$\frac{\varepsilon_x^2}{\sigma_{x,0}^2 \cdot (\sigma_{x,0})'^2} \geq 0.01 \tag{15}$$

can be derived [11]. This criterion ensures the numerical significance of ε_x. The second criterion is based on the fact that the scan needs to include a minimum. The initial beam optics needs to be setup, such that a potential focus of the beam lies behind the screen used to measure the beamsize. At the same time, the focusing element needs to be strong enough to focus the beam onto the screen, which implies a constraint on the distance between focusing element and screen. Using Eq. 10, these considerations can be summarized by the criterion

$$-\frac{\sigma_{x,0} \cdot (\sigma_{\text{eff}})'}{f_{\max}^2 \cdot \left((\sigma_{\text{eff}})'^2 + \frac{\varepsilon_x^2}{\sigma_{x,0}^2} \right)} \leq l_{\text{S}} \leq \frac{\sigma_{x,0}^2}{2\varepsilon_x}, \tag{16}$$

where $(\sigma_{\text{eff}})' = -(\sigma_{x,0}/f_{\max} - (\sigma_{x,0})')$. In case both of the two criteria are fulfilled, the emittance can be retrieved.

Based on the aforementioned considerations, we propose using an alternative way to analyze phase advance scan data. Specifically, we propose using a pre-trained FCNN to overcome the problem of the incomplete fit model in $\rho \gg 1$ cases, as well as the criterion described by Eq. 15. To this end, we have performed a simulation study, which is presented in detail in the following sections. The resulting FCNN was then subsequently applied to real world data, as shown below.

III. SIMULATION STUDY - METHODOLOGY

It has been shown already in 1989 that neural networks with only one unbounded hidden layer can approximate any Borel measureable function from finite dimensional space to another to arbitrary precision [28, 29]. More recent research focuses on the expressiveness (approximation accuracy) of both depth (i.e. the number of hidden layers) and width (i.e. the number of artificial neurons in a layer) bounded networks. In [30], for example, the authors show that any Lebesgue integrable function $f : \mathbb{R}^n \to \mathbb{R}$ on n-dimensional space can be approximated to arbitrary accuracy by a fully connected width-$(d_{\text{in}} + 4)$ ReLU network with respect to the ℓ_1 norm as a measure of approximation quality. In other words, the network represented by the transfer function F satisfies $\int_{\mathbb{R}^n} |f(x) - F(x)| \mathrm{d}x < \epsilon, \forall \epsilon > 0$ (see [30], Theorem 1). ReLU here refers to the the so-called Rectified Linear Unit neuron activation function, definded by ReLU(x) = max(0,x) and d_{in} is the input dimensionality. This width boundary w has since been refined and generalized for example in [31, 32] to be $d_{\text{in}} + 1 \leq w_{\min}(d_{\text{in}}, d_{\text{out}}) \leq d_{\text{in}} + d_{\text{out}}$, for functions of the form $f : [0, 1]^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$, where d_{in} and d_{out} are the input and output dimensionality respectively. Limits on the depth can be estimated in specific cases in terms of the so-called modulus of continuity of f, given by $\omega_f(\varepsilon) = \sup\{|f(x) - f(y)| | |x - y| \leq \varepsilon\}$, where ε is an arbitrarily small change in the argument of f. For continuous functions $f : [0, 1]^{d_{\text{in}}} \to \mathbb{R}_+$ the depth of a $d_{\text{in}} + 2$ wide network $\mathcal{N}$ can, for example, be expressed as $\text{depth}(\mathcal{N}_\varepsilon) = 2 \cdot d_{\text{in}}! / \omega_f(\varepsilon)^{d_{\text{in}}}$, cf. [31]. We note that in practice the specific layout of a neural network is often determined experimentally, as the aforementioned boundaries are merely based on proofs of existence.

In this study, we aim to map the phase advance scan data to the normalized transverse emittance at the fo-

cusing element. Mathematically, this means we assume a connection of the scan data to the physical quantity of the form $f : \mathbb{R}_+^{d_{\text{in}}} \to \mathbb{R}_+$, where the dimensionality d_{in} is given by the number of scan data points. Note that, based on the knowledge of the problem, we can always map (normalize) the input data from $\mathbb{R}_+^{d_{\text{in}}}$ into $[0,1]^{d_{\text{in}}}$. The function f operates on the measure space $(\mathbb{R}_+^{d_{\text{in}}}, \mathcal{B}, \lambda)$ with the Borel-σ-algebra $\mathcal{B}$ and the Lebesgue measure λ. It is hence measureable in the mathematical sense. In addition, we expect f to be a continuous function based on the physical background of the problem. We can hence conclude that f is Lebesgue integrable and suitable to be approximated for example by a width bounded ReLU network.

To validate this approach, we setup a simulation study based on the simple beamline layout shown in Fig. 1. The main simulation study is split into three parts:

1. Building a large number of data sets (training, validation and test),
2. Training the FCNN and evaluation of the performance using the test data sets,
3. Comparison of the FCNN performance to the traditional fit method, as discussed above.

The first step of the simulation study is to build a large number of data sets. Creating a data set consists of two steps:

- Numerical tracking of the beam from the cathode to the location of the solenoid,
- Numerical simulation of the solenoid scan.

First, the emittance at the solenoid position is determined by numerical tracking of the particles. In this step the solenoid field is set to zero. In addition to the emittance, other beam parameters, such as the beam size, divergence, or bunch length can be recorded as well. Then, the simulation domain is extended up to the position of the screen, which is used in the experiment to record the beam size vs. the solenoid focusing strength. The experiment is then simulated for M focusing strength settings. It is important to setup the scan range such that the resulting data includes the beam size minimum, i.e. the focus, as it carries most of the information about the emittance at the solenoid position [11]. We use the well established code ASTRA [33], which takes space charge effects into account. The beam size vs. focusing strength scan data functions as the data set to be interpreted by the FCNN. Over the course of the study, a specific way to prepare the input data turned out to yield the best results. For each scan, $M/2$ scan points centered around the minimum beam size are interleaved with the relative focusing strength difference $\delta B_i = (B_i - B_{\text{foc}})/B_{\text{foc}}$, where B_{foc} is the setting corresponding to the minimal beam size. The data set S_{in} is then of the form

$$S_{\text{in}} = [\delta B_1, \sigma_1, \delta B_2, \sigma_2, ..., \delta B_{(M/2)}, \sigma_{(M/2)}], \qquad (17)$$

where σ_i is the ith rms beam size. Each of these data sets is labeled with a set of important beam and simulation input parameters. These labels are then used to perform so-called supervised training of the FCNN. After the learning process, the FCNN is able to predict each of these parameters from given scan data, which is prepared according to Eq. 17. Figure. 2 summarizes how the data sets are created and what types of data they contain.

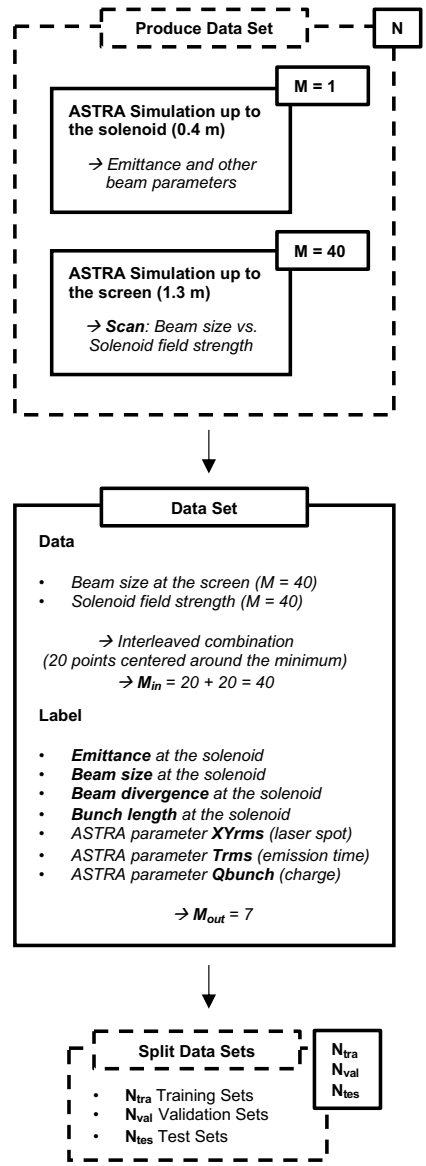

FIG. 2. Diagram showing how the FCNN training, validation and test data is produced. N is the total number of data sets. M is the number of ASTRA simulations. M_{in} is the final number of data points to be fed to the FCNN. M_{out} is the length of the output vector, i.e. the number of predicted beam and simulation parameters. The data sets are split into three categories, where in this study the common split of $N_{\text{tra}} = 0.6 \cdot N$, $N_{\text{val}} = 0.2 \cdot N$ and $N_{\text{tes}} = 0.2 \cdot N$ is used.

For this particular study $N = 16066$ random data sets with $M = 40$ were produced. Each data set differs in the three key ASTRA input parameters *total charge*, *laser spot size* and *cathode emission time*. Table I summarizes the parameter ranges used for this study, which are losely based on typical settings at the ARES linac at the time. The parameters are varied according to a uniform distribution. The ARES S-band gun was simulated with an, at the time available, peak gradient of 65 MV/m, resulting in a final $\gamma = 6.8$. Based on the parameter ranges shown in Table I, a convergence study in terms of required macro particles in the numerical simulation was performed. For the highest possible charge density, 10000 particles were found to be sufficient.

TABLE I. ASTRA input parameter ranges for the data set production.

Parameter	Value
Bunch charge	$[0.01, 2.1]$ pC
Laser spot size (flat top diameter)	$[240, 400]$ µm
Cathode emission time (rms)	$[60, 100]$ fs

The neural network was implemented using the TensorFlow framework [34]. The input layer has M neurons, corresponding to the length of S_{in}. Then one hidden layer with M and two hidden layers with $M/2$ neurons are added in order to capture non-linearities in the system. The system is then coupled to the output layer of size M_{out}, which corresponds to the number of ASTRA input and simulated beam parameters to be predicted by the network. The overall layout is hence

$$M \to [M] - [M/2] - [M/2] \to M_{\text{out}}. \tag{18}$$

This particular layout was determined empirically. Each neuron is coupled to every neuron of the following layer, or in other words the layers are fully connected. The neurons are activated using the well established rectified linear activation function (ReLU) [35, 36]. Training of the network is performed using a combination of the ADAM and ADAGRAD gradient decent algorithm [37] with the mean squared error (MSE) as the loss function. The available N data sets are split into three categories. N_{tra} training sets, N_{val} validation sets and N_{tes} test sets. The training sets are used to adjust the neuron weights during the training procedure, while the performance of the network is judged after each so-called epoch based on the validation sets, which are not used during training. This is done to avoid overfitting the training data. An epoch refers to one forward and backward pass of the entire training data. Finally, the performance of the resulting model is determined using the test sets, which have not been part of the learning procedure at all. We use the common split of $N_{\text{tra}} = 0.6 \cdot N$, $N_{\text{val}} = 0.2 \cdot N$ and $N_{\text{tes}} = 0.2 \cdot N$. The network was trained for ~ 10000 epochs using ADAM and another ~ 10000 epochs using ADAGRAD [38].

IV. SIMULATION STUDY - DATA SET

Before evaluating the prediction performance of the neural network, it is useful to inspect the training data set. Since we are especially interested in analyzing phase advance scan data for space charge dominated beams, the laminarity parameter ρ at the solenoid position was calculated for each data set (cf. Eq. 2). Figure 3 shows ρ vs. the bunch charge. The color scale indicates the laser spot size on the cathode used in the particular simulation. In addition, the distribution of ρ across the whole data set is shown. It can be seen that all of the data sets lie in the $\rho \gg 1$, i.e. space charge dominated, regime ($\rho_{\text{min}} = 17.8$). Also, the higher the charge, the higher the value for ρ, as expected. In addition, the color scale reveals that the smaller the laser spot size on the cathode, the higher the value for ρ. The sensitivity of ρ on the laser spot size strongly depends on the bunch charge.

As noted above, the traditional fit method only works if Eq. 15 is satisfied. Figure 4 shows the fit feasibility criterion for each data set, with the same color code as in Fig. 3. None of the data sets satisfies the criterion, which leads to the expectation that the traditional fit method should not work well on the training data (and with that in reality for the ARES working point, which is the basis for the parameter space shown in Table I).

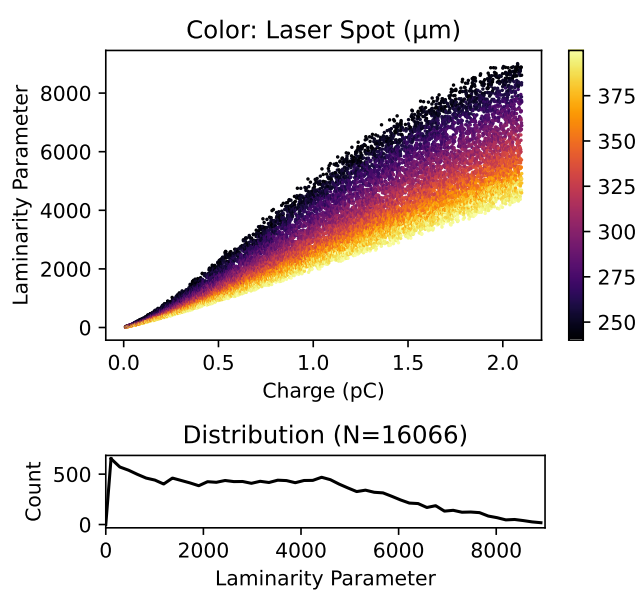

FIG. 3. Top: Scatter plot of the laminarity parameter ρ at the solenoid according to Eq. 2 vs. the bunch charge for all training data sets. $\gamma = 6.8$ for this population. The color indicates the laser spot size on the cathode for each data set. Bottom: Distribution of the laminarity parameter.

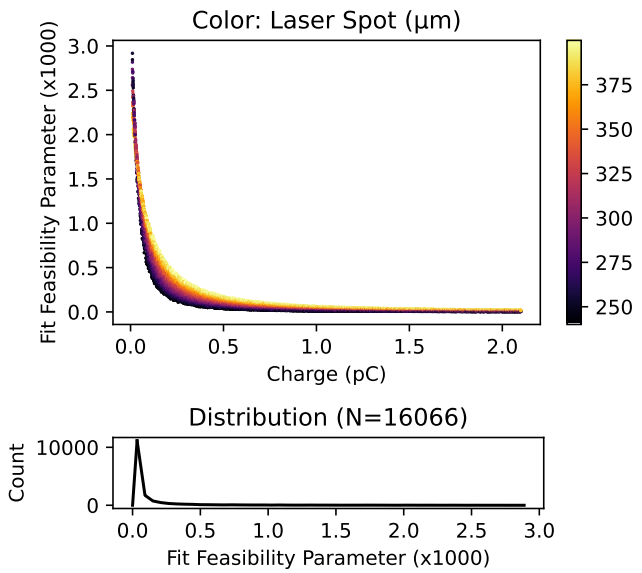

FIG. 4. Top: Scatter plot of the fit feasibility parameter according to Eq. 15 vs. the bunch charge for all training data sets. $\gamma = 6.8$ for this population. The color indicates the laser spot size on the cathode for each data set. Bottom: Distribution of the fit feasibility parameter. *Note that the plot shows the values multiplied by 1000 for readability.*

V. RESULTS AND COMPARISON

In this section the performance of the pre-trained FCNN is presented. We also compare its performance against the traditional fit method discussed above. In order to evaluate the performance of the FCNN, we try to predict the labels of the N_{tes} test data sets, which were not used in the supervised training procedure. The main goal of the study is to predict the transverse emittance, but since the labels include a number of other simulation and beam parameters, the FCNN also provides predictions of these. In order to better quantify the prediction performance for the different label components, the relative error between predicition and truth was calculated for each data set. This is shown in Fig. 5. In addition to the error distributions, a radar plot visualizes the prediction performance in terms of number of data sets in 1 %, 5 % and 10 % relative error intervals respectively (see Table II for the actual percentages). In the ideal case, the heptagon would be filled completely. Inspection of the results reveals that some quantities are predicted much better than others. Specifically, it can be seen that the cathode emission time is predicted particularly bad. This result is somewhat expected, however, because this quantity refers to the longitudinal phase space at emission time, which cannot directly be accessed via a transverse beam size measurement. All quantities, which refer to the transvere phase space at the solenoid show very good prediction performance with $< 5\,\%$ error. The prediction performance of both bunch charge and bunch length at the solenoid needs to be considered in more detail. In case of the bunch charge, values $\lesssim 0.5\,\text{pC}$ are predicted much less accurately. This can be explained by the lack of significant space charge effects, which alter the shape of the beam size vs. focusing strength curve, effectively leading to degeneracy w.r.t. the initial bunch charge. Despite being a quantity of the longitudinal phase space, the bunch length is generally predicted with an error $< 10\,\%$. This is because here the bunch length is directly correlated with the bunch charge and hence space charge effects. The prediction performance decreases towards smaller bunch lengths, which can be explained by the fact that the bunch length at the solenoid actually increases with the initial bunch charge. Therefore the same argument applies as for the bunch charge.

From the prediction results and the ground truth a mean relative error was calculated over the whole test data set, yielding a mean prediction error for the emittance at the solenoid of 1.0 %. Beam size and divergence at the solenoid are predicted very accurately with an error less than 0.1 %. In addition to the beam parameters at the solenoid position, ASTRA input parameters were predicted. The laser spot size is predicted with an error down to 0.7 %. Emission time and bunch charge are predicted with errors of 32.1 % and 27.1 % respectively. The fact that the laser spot size is predicted best out of the three input parameters is due to the fact that it has the strongest effect on the shape of the beam size vs. focusing strength data, especially for low charges (linear dependence of the thermal emittance). These results, as well as the percentage of predictions within a 1 %, 5 % and 10 % relative error interval are summarized in Table II.

The main goal of the study is to find a better way to determine the transverse emittance from phase advance scan data in the $\rho \gg 1$ regime, as well as in regimes where Eq. 15 does not hold. It is hence useful to evaluate the emittance prediction performance in form of the relative error versus these two quantities. Figure 6 shows the result of this analysis using both the FCNN, as well as the traditional fit routine (cf. Sec. II).

As expected from Fig. 3 and Fig. 4, the traditional fit yields inaccurate results across the whole data set. The FCNN, on the other hand, performs much better even for very high values of ρ.

VI. EXTENSION TO MEASURED DATA

So far, the neural network was trained and tested solely with ideal phase advance scan data. This means that the training, validation, as well as the test data sets contain only perfectly evenly spaced data points with flawless beam size values. In addition, all scans were simulated within the same focusing strength range. Although this approach yields very good prediction performance for simulated data, it cannot be applied to experimental data, for several reasons. First, in a measured data set it

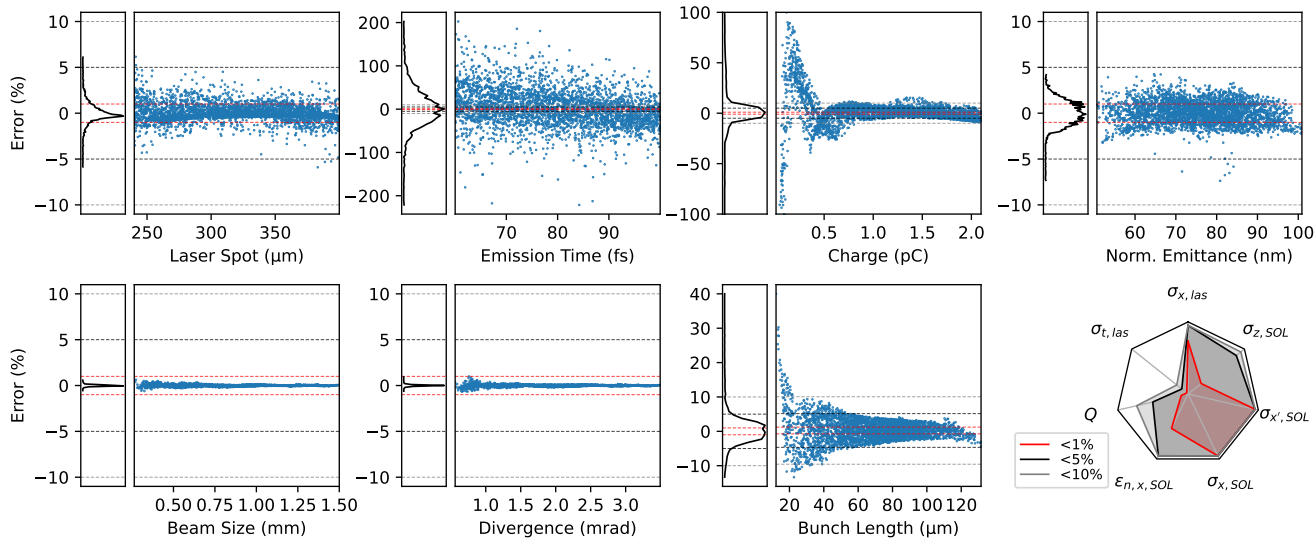

FIG. 5. Relative error distributions for all verification data sets and label components. The last plot is a radar plot showing the percentage of data sets within a 1 % (red), 5 % (black) and 10 % (gray) relative error interval respectively. The dashed lines in the distribution plots refer to the same error intervals. Q : Bunch charge, $\sigma_{t,\mathrm{las}}$: Emission time, $\sigma_{x,\mathrm{las}}$: Laser spot size, $\sigma_{z,\mathrm{SOL}}$: Bunch length at the solenoid, $\sigma_{x',\mathrm{SOL}}$: Divergence at the solenoid, $\sigma_{x,\mathrm{SOL}}$: Beam size at the solenoid, $\varepsilon_{\mathrm{n},x,\mathrm{SOL}}$: Normalized emittance at the solenoid.

TABLE II. Summary of the prediction performance for the different training label components. Mean error refers to the absolute prediction error and is calculated for the whole verification data set. The three remaining columns refer to the fraction of the verification data set within a certain prediction accuracy interval.

	Mean Error (%)	$< 1\,\%$ Error (%)	$< 5\,\%$ Error (%)	$< 10\,\%$ Error (%)
Laser Spot Size	0.71	76.59	99.84	100.00
Emission Time	32.14	2.46	11.67	21.05
Bunch Charge	27.05	10.24	52.74	76.74
Emittance at the Solenoid	1.01	55.67	99.81	100.00
Beamsize at the Solenoid	0.05	100.00	100.00	100.00
Divergence at the Solenoid	0.05	100.00	100.00	100.00
Bunch Length at the Solenoid	2.60	24.35	89.76	98.13

is not guaranteed that all data points are evenly spaced. It is also not guaranteed that the scan range corresponds to the trained one and that the measured values of the focusing strength are correct [39]. Finally, the measured beam sizes are subject to jitter and systematic measurement errors like resolution limitations.

In order to take all of this into account, the training procedure was modified. Each data set is now created with a slightly different focusing strength scan range. Since the number of scan points is kept constant, the spacing between data points is now slightly different every time, which might also be the case in reality. Furthermore, the first and last focusing setting were added to the range of predicted parameters ($\rightarrow M_{\mathrm{out}} = 9$). Measurement errors were taken into account by generating noisy data sets from the ideal sets by adding normally distributed errors. Both relative and absolute errors on focusing strength and beam size were considered with magnitudes based on experience at ARES. From each data set, $N_{\mathrm{err}} = 100$ noisy sets with relative errors and N_{err} noisy sets with absolute errors were generated. In addition, the procedure was repeated, this time enforcing a 10 µm resolution limit on the beam sizes. Including the ideal data, the total number of data sets now increases to $N_{\mathrm{tot}} = 2(2N_{\mathrm{err}} + 1) \cdot N = 6458532$. To visualize the importance of training the FCNN with noisy data, we fed data sets with increasing artificial noise to both the network trained solely on ideal data, as well as one trained on noisy data. The results are shown in Fig. 7. It can be seen that the population with less than 5 % decreases significantly with noise level, if the FCNN is not trained on noisy data.

We performed the same general analysis for the new network, as described above, and saw the same overall behaviour. The results are summarized in Table III. Compared to the network based on ideal data, the performance is slightly worse, but still for the majority of data sets the emittance is predicted with less than 5 % error.

TABLE III. Summary of the prediction performance of the modified network for the different training label components. Mean error refers to the absolute prediction error and is calculated for the whole verification data set. The three remaining columns refer to the fraction of the verification data set within a certain prediction accuracy interval.

	Mean Error (%)	< 1 % Error (%)	< 5 % Error (%)	< 10 % Error (%)
Laser Spot Size	2.76	22.25	87.85	98.83
Emission Time	15.97	4.25	21.08	43.48
Bunch Charge	14.13	6.44	30.89	67.20
Emittance at the Solenoid	2.51	29.43	89.17	97.51
Beamsize at the Solenoid	0.38	96.34	100.00	100.00
Divergence at the Solenoid	0.38	96.63	100.00	100.00
Bunch Length at the Solenoid	4.48	12.74	60.03	95.02

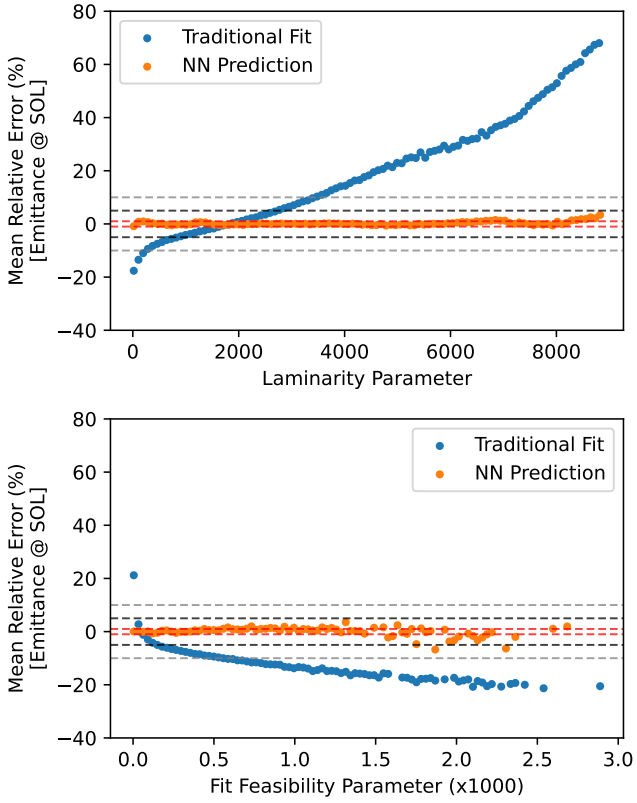

FIG. 6. Mean emittance FCNN prediction performance, as well as traditional fit results vs. the laminarity parameter (top) and the fit feasibility criterion (bottom). The data is calculated for 100 equally sized laminarity parameter slices. $\gamma = 6.8$ for this population. *Note that the bottom plot shows the fit criterion values multiplied by 1000 for readability.*

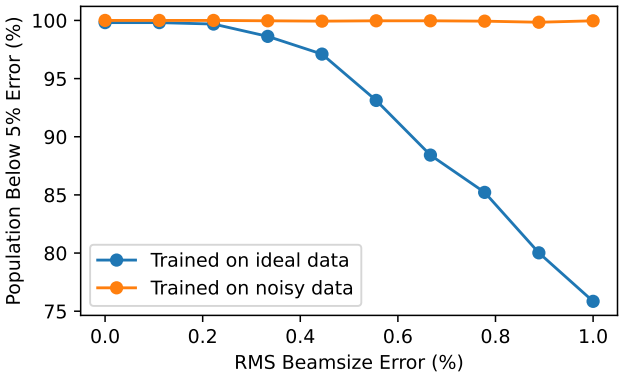

FIG. 7. Population with less than 5 % prediction error on the emittance vs. different artificial noise levels. The FCNN layout is the same in both shown cases. *Note that in order to enable a direct comparison, the FCNNs used here were trained on the data sets with fixed scan range and constant spacing. This explains the generally better performance compared to Table III.*

VII. MEASUREMENTS AT THE ARES LINAC

As a real world test, we conducted emittance measurements using the phase advance scan technique at the ARES linac at DESY. The layout of the measurement setup is shown in Fig. 1. We took data for several bunch charges by adjusting an attenuator in the cathode laser beamline. The charge was measured both with a Faraday cup, which can be inserted into the beamline instead of the scintillating screen and a cavity based charge monitor ~ 0.7 m downstream of the screen [40]. All measurements shown here were performed according to the procedure introduced in Sec. II. Transverse beam sizes were determined from camera images of a scintillating Ce:GAGG (Cerium doped Gadolinium Aluminium Gallium Garnet) screen. The spatial resolution of the system is specified to be $\sim 10\,\mu$m [41].

Figure 8 shows the emittance values obtained from the measurements using both the FCNN, as well as the traditional fit method. The data points are compared to an ASTRA simulation including space charge based on the machine settings at the day of the measurements, including uncertainty. It can be seen that the FCNN results are much closer to the expected values than the results obtained from the envelope equation fit. It is interesting to note that Fig. 8 reproduces the expected behaviour shown in Fig. 6, as the fit underestimates the emittance for charges < 0.5 pC and overestimates them for higher charges. The FCNN result follows the ASTRA curve much closer, mostly staying within the uncertainty of the simulation.

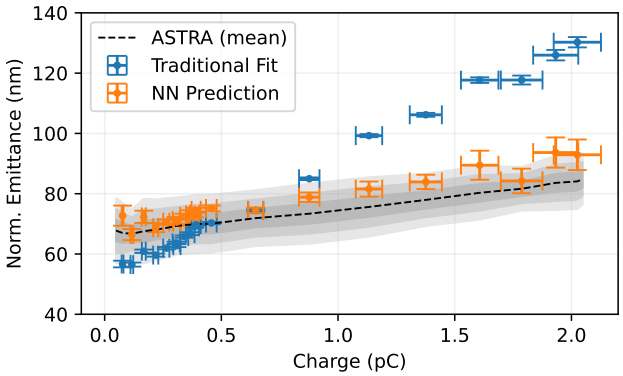

FIG. 8. Emittance measurements conducted at the ARES linac at DESY using the phase advance scan technique for different bunch charges. The black dashed line denotes the mean ASTRA simulation result. The shaded areas correspond to the 1, 2 and 3 σ uncertainty of the simulation, based on 100 simulations with normally distributed input parameters.

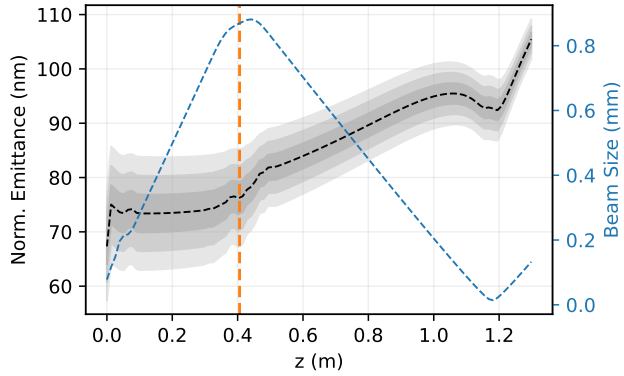

FIG. 9. ASTRA simulation of the evolution of the transverse normalized emittance between gun and grid for a 1 pC bunch with $\gamma = 6.8$. The orange dashed line marks the position of the solenoid, which is used to perform the phase advance scan measurements. The black dashed line denotes the mean ASTRA simulation result. The shaded areas correspond to the 1, 2 and 3 σ uncertainty of the simulation, based on 100 simulations with normally distributed input parameters. The blue dashed line shows the evolution of the beam size – the beam reaches its focus slightly before the grid.

In order to cross-check the obtained results, we performed additional emittance measurements using a grid mask based method, as described in [14], using the same machine setup. Three different grids were used for the measurements: An SPI G200TH TEM grid [42], an SPI G300 TEM grid [43] and a custom made pepper pot [44]. Measurements were performed for two different charge settings, 1 pC and 2 pC. These values were deliberately chosen to be in the strongly space charge dominated regime, where the traditional fit yields particularly bad results. At ARES, the grid masks are installed at the same z-position as the screen used to record the phase advance scan data. This means that the emittance obtained from the grid measurements will always be different from the phase advance scan result, as the phase advance scan yields the emittance at the position of the focusing element. The emittance is furthermore expected to be different, because in order to image the grid, the beam needs to be focused slightly before the grid, which can lead to emittance growth. Nevertheless, it is still possible to compare the measured values to ASTRA simulations, which would show that the ARES setup depicted in Fig. 1 can be well simulated with ASTRA. This would validate the FCNN results indirectly. Figure 9 shows an ASTRA simulation of the grid measurement using a 1 pC bunch with $\gamma = 6.8$. It can be seen that the emittance is strongly affected by focusing the beam down. The measurement results from all three grids, as well as the expected values from the ASTRA simulation are summarized in Table IV. The results are very close to the expected value in the 1 pC case. The measurement of the 2 pC beam shows a slightly higher than expected emittance value, which is in line with the high uncertainty of the high charge results shown in Fig. 8. We hence conclude that ASTRA simulates the ARES beamline shown in Fig. 1 well.

TABLE IV. Summary of the grid based measurements, used to verify the FCNN results. The measurements are compared to the ASTRA simulation result.

Method	ε_n (nm) @ 1 pC	ε_n (nm) @ 2 pC
ASTRA Simulation	$\mathbf{105.5 \pm 1.2}$	$\mathbf{152.9 \pm 3.0}$
SPI G200TH TEM Grid	107.4 ± 5.8	165.4 ± 15.7
SPI G300 TEM Grid	102.7 ± 3.8	156.0 ± 15.6
Pepper Pot	105.6 ± 11.9	156.1 ± 27.4
Mean (Measurements)	$\mathbf{105.2 \pm 8.0}$	$\mathbf{160.5 \pm 20.3}$

A. Prediction of other parameters

As discussed above, the FCNN also predicts fixed machine parameters and other charge dependent beam parameters to varying accuracy (see Table III). Table V summarizes the predicted fixed machine parameters in comparison to what was used in the experiment. It can be seen that both the mean laser spot size and pulse length are predicted to be larger than the expected values. Since measuring the laser spot size on the cathode directly is very difficult in the ARES setup, the predicted values fall within the uncertainty. As discussed in Sec. V, the prediction of the laser pulse length should be treated with caution. The solenoid scan range is predicted well within the uncertainties.

Figure 10 shows the prediction results for the beam charge, as well as the charge dependent beam parameters *bunch length*, *beam size* and *beam divergence* at the solenoid. As in Fig. 8, the data points are compared to an ASTRA simulation including space charge based on the machine settings at the day of the measurements, includ-

TABLE V. Prediction results for fixed machine parameters.

Parameter	Prediction	Experiment
Laser Spot Size (µm)	338.5 ± 0.5	320 ± 30
Laser Pulse Length (fs, rms)	87.34 ± 0.04	76 ± 8
Solenoid Field – Start (mT)	130.516 ± 0.001	130.5 ± 0.1
Solenoid Field – End (mT)	150.783 ± 0.001	150.7 ± 0.1

ing uncertainty. It can be seen that the beam charge prediction fits the measured values well. Both beam size and beam divergence follow the ASTRA curve well, albeit at the lower end of the uncertainty, denoted by the shaded area. The bunch length follows a more linear charge dependence than expected from the ASTRA simulation, which might be attributed to either the not fully known temporal and spatial laser pulse shape at the cathode, as well as the overall prediction performance of parameters of the longitudinal phase space (see Sec. V).

In order to cross-validate the prediction results, an ASTRA simulation using the mean predicted laser spot size and pulse length (see Table V) was performed. The results are are shown in Fig. 10 as the blue dashed line. Indeed, a larger spot size and longer pulse length lead to results closer to the lower end of the uncertainty in all three cases, which can be explained by the reduced charge density. Remaining discrepancies might be explained by the not fully known temporal and spatial laser pulse shape at the cathode.

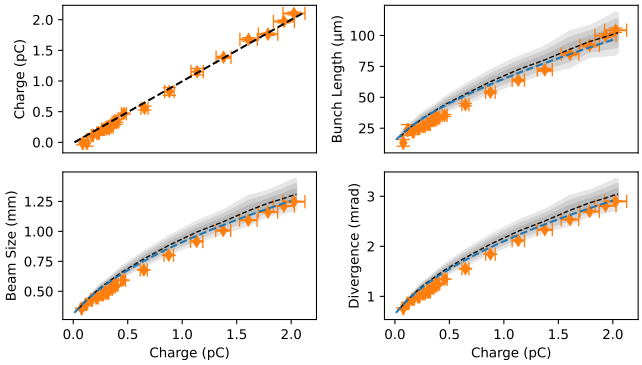

FIG. 10. Prediction results for other charge dependent beam parameters. The black dashed line denotes the mean ASTRA simulation result. The shaded areas correspond to the 1, 2 and $3\,\sigma$ uncertainty of the simulation, based on 100 simulations with normally distributed input parameters. The blue dashed line corresponds to an ASTRA simulation using the mean predicted laser spot size and pulse length (see Table V).

VIII. CONCLUSION AND OUTLOOK

We have shown in simulation that a pre-trained fully connected neural network can be used to predict the transverse emittance from phase advance scan data even in the $\rho \gg 1$ regime and in case a traditional envelope equation based fit is mathematically not feasible. We have optimized the network for real-world measurement data and achieved $< 5\,\%$ error for the majority of the test data set population (89.2 %), resulting in a mean relative error of 2.5 %. We have applied our method to measurements conducted at the ARES linac at DESY and compared the predictions to numerical simulations using the well benchmarked code ASTRA, as well as results obtained from the traditional fit method. As expected from the simulation study, the FCNN predictions are much closer to what is expected from the numerical simulation. We have furthermore cross-validated the results using additional emittance measurements based on a grid mask based method.

In addition to the transverse emittance, the network also predicts other key beam and machine parameters to varying accuracy. While quantities directly tied to the transverse phase space are predicted as accurate or better than the emittance, quantities tied to the longitudinal phase space, such as the bunch length, are predicted less accurate, as expected. It should be noted, that in our study the gun setting is not a variable in the process of training the FCNN. This means that for each gun setting (gradient and phase) a separate FCNN has to be trained. Inclusion of these two parameters could be part of a future study. Furthermore, difficult to directly access parameters, such as the thermal emittance could be added. In conclusion, we have demonstrated that pre-trained FCNNs can be a powerful tool for the analysis of previously difficult to interpret data sets.

IX. MODEL AVAILABILITY

The FCNN models are available from the corresponding author upon request in TensorFlow format.

ACKNOWLEDGMENTS

We acknowledge support from DESY (Hamburg, Germany), a member of the Helmholtz Association HGF. Specifically, we would like to thank the DESY beam diagnostics group for their support during the measurement runs at ARES. We also thank the DESY Maxwell team for providing the compute resources we used to generate the training data sets in a reasonable time. Finally, we thank R. Mayet (Halodi Robotics AS) for NN-related consulting.

[1] J. R. Dwyer, C. T. Hebeisen, R. Ernstorfer, M. Harb, V. B. Deyirmenjian, R. E. Jordan, and R. Dwayne Miller, Femtosecond electron diffraction: Making the molecular movie, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **364**, 741 (2006).

[2] C. Pellegrini, A. Marinelli, and S. Reiche, The physics of x-ray free-electron lasers, Rev. Mod. Phys. **88**, 015006 (2016).

[3] R. J. England, R. J. Noble, K. Bane, D. H. Dowell, C.-K. Ng, J. E. Spencer, S. Tantawi, Z. Wu, R. L. Byer, E. Peralta, K. Soong, C.-M. Chang, B. Montazeri, S. J. Wolf, B. Cowan, J. Dawson, W. Gai, P. Hommelhoff, Y.-C. Huang, C. Jing, C. McGuinness, R. B. Palmer, B. Naranjo, J. Rosenzweig, G. Travish, A. Mizrahi, L. Schachter, C. Sears, G. R. Werner, and R. B. Yoder, Dielectric laser accelerators, Reviews of Modern Physics **86**, 1337 (2014).

[4] F. X. Kärtner, F. Ahr, A. L. Calendron, H. Çankaya, S. Carbajo, G. Chang, G. Cirmi, K. Dörner, U. Dorda, A. Fallahi, A. Hartin, M. Hemmer, R. Hobbs, Y. Hua, W. R. Huang, R. Letrun, N. Matlis, V. Mazalova, O. D. Mücke, E. Nanni, W. Putnam, K. Ravi, F. Reichert, I. Sarrou, X. Wu, A. Yahaghi, H. Ye, L. Zapata, D. Zhang, C. Zhou, R. J. D. Miller, K. K. Berggren, H. Graafsma, A. Meents, R. W. Assmann, H. N. Chapman, and P. Fromme, AXSIS: Exploring the frontiers in attosecond X-ray science, imaging and spectroscopy, Nuclear Inst. and Methods in Physics Research, A **829**, 24 (2016).

[5] F. Giorgianni, M. P. Anania, M. Bellaveglia, A. Biagioni, E. Chiadroni, A. Cianchi, M. Daniele, M. Del Franco, D. Di Giovenale, G. Di Pirro, M. Ferrario, S. Lupi, A. Mostacci, M. Petrarca, S. Pioli, R. Pompili, V. Shpakov, C. Vaccarezza, and F. Villa, Tailoring of Highly Intense THz Radiation Through High Brightness Electron Beams Longitudinal Manipulation, Applied Sciences **6** (2016).

[6] P. A. P. Nghiem, R. Assmann, A. Beck, A. Chancé, E. Chiadroni, B. Cros, M. Ferrario, A. Ferran Pousa, A. Giribono, L. A. Gizzi, B. Hidding, P. Lee, X. Li, A. Marocchino, A. Martinez de la Ossa, F. Massimo, G. Maynard, A. Mosnier, S. Romeo, A. R. Rossi, T. Silva, E. Svystun, P. Tomassini, C. Vaccarezza, J. Vieira, and J. Zhu, Toward a Plasma-based Accelerator at High Beam Energy with High Beam Charge and High Beam Quality, Physical review accelerators and beams **23**, 031301 (2020).

[7] M. J. Rhee, Refined definition of the beam brightness, Physics of Fluids B: Plasma Physics **4**, 1674 (1992), https://doi.org/10.1063/1.860076.

[8] J. Rees and L. Rivkin, SLAC-PUB-3305, *(On measuring emittances and sigma matrices.)* (1984).

[9] M. G. Minty and Z. F, *Measurement and Control of Charged Particles Beams* (Springer-Verlag, Berlin, Heidelberg, New York, 2003).

[10] M. Hachmann and K. Flöttmann, Measurement of ultra low transverse emittance at regae, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **829**, 318 (2016), 2nd European Advanced Accelerator Concepts Workshop - EAAC 2015.

[11] M. Hachmann, *Transverse emittance measurement at REGAE via a solenoid scan*, Master's thesis (2012), universität Hamburg, Diplomarbeit, 2012.

[12] K. Floettmann, Emittance compensation in split photoinjectors, Phys. Rev. Accel. Beams **20**, 013401 (2017).

[13] M. Zhang, Report no. FERMILAB-TM-1988, *Emittance Formula for Slits and Pepper-pot Measurement.* (1996).

[14] D. Marx, J. Giner Navarro, D. Cesar, J. Maxson, B. Marchetti, R. Assmann, and P. Musumeci, Single-shot reconstruction of core 4d phase space of high-brightness electron beams using metal grids, Phys. Rev. Accel. Beams **21**, 102802 (2018).

[15] M. Ferrario, M. Migliorati, and L. Palumbo, Space Charge Effects, , 331 (2016), 26 pages, contribution to the CAS - CERN Accelerator School: Advanced Accelerator Physics Course, Trondheim, Norway, 18-29 Aug 2013.

[16] A. Solopova *et al.*, SRF Cavity Fault Classification Using Machine Learning at CEBAF, in *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019*, International Particle Accelerator Conference No. 10 (JACoW Publishing, Geneva, Switzerland, 2019) pp. 1167–1170, https://doi.org/10.18429/JACoW-IPAC2019-TUXXPLM2.

[17] A. R. Maier, N. M. Delbos, T. Eichner, L. Hübner, S. Jalas, L. Jeppe, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepp, M. Trunk, P. A. Walker, C. Werle, and P. Winkler, Decoding sources of energy variability in a laser-plasma accelerator, Phys. Rev. X **10**, 031039 (2020).

[18] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, Demonstration of model-independent control of the longitudinal phase space of electron beams in the linac-coherent light source with femtosecond resolution, Phys. Rev. Lett. **121**, 044801 (2018).

[19] A. Ivanov and I. Agapov, Physics-based deep neural networks for beam dynamics in charged particle accelerators, Phys. Rev. Accel. Beams **23**, 074601 (2020).

[20] A. Sanchez-Gonzalez, P. Micaelli, C. Olivier, T. R. Barillot, M. Ilchen, A. A. Lutman, A. Marinelli, T. Maxwell, A. Achner, M. Agåker, N. Berrah, C. Bostedt, J. D. Bozek, J. Buck, P. H. Bucksbaum, S. C. Montero, B. Cooper, J. P. Cryan, M. Dong, R. Feifel, L. J. Frasinski, H. Fukuzawa, A. Galler, G. Hartmann, N. Hartmann, W. Helml, A. S. Johnson, A. Knie, A. O. Lindahl, J. Liu, K. Motomura, M. Mucke, C. O'Grady, J.-E. Rubensson, E. R. Simpson, R. J. Squibb, C. Såthe, K. Ueda, M. Vacher, D. J. Walke, V. Zhaunerchyk, R. N. Coffee, and J. P. Marangos, Accurate prediction of x-ray pulse properties from a free-electron laser using machine learning, Nature Communications **8**, 15461 (2017).

[21] C. Emma, A. Edelen, M. J. Hogan, B. O'Shea, G. White, and V. Yakimenko, Machine learning-based longitudinal phase space prediction of particle accelerators, Phys. Rev. Accel. Beams **21**, 112802 (2018).

[22] D. Ratner, F. Christie, J. P. Cryan, A. Edelen, A. Lutman, and X. Zhang, Recovering the phase and amplitude of x-ray fel pulses using neural networks and differentiable models, Opt. Express **29**, 20336 (2021).

[23] M. Kirchen, S. Jalas, P. Messner, P. Winkler, T. Eichner, L. Hübner, T. Hülsenbusch, L. Jeppe, T. Parikh, M. Schnepp, and A. R. Maier, Optimal beam loading in a laser-plasma accelerator, Phys. Rev. Lett. **126**, 174801 (2021).

[24] S. Jalas, M. Kirchen, P. Messner, P. Winkler, L. Hübner, J. Dirkwinkel, M. Schnepp, R. Lehe, and A. R. Maier, Bayesian optimization of a laser-plasma accelerator, Phys. Rev. Lett. **126**, 104801 (2021).

[25] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems, Phys. Rev. Accel. Beams **23**, 044601 (2020).

[26] E. Panofski, R. Assmann, F. Burkart, U. Dorda, L. Genovese, F. Jafarinia, S. Jaster-Merz, M. Kellermeier, W. Kuropka, F. Lemery, B. Marchetti, D. Marx, F. Mayet, T. Vinatier, and S. Yamin, Commissioning results and electron beam characterization with the s-band photoinjector at sinbad-ares, Instruments **5**, 10.3390/instruments5030028 (2021).

[27] V. Kumar, Understanding the focusing of charged particle beams in a solenoid magnetic field, American Journal of Physics **77**, 737 (2009), https://doi.org/10.1119/1.3129242.

[28] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems **2**, 303 (1989).

[29] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[30] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, The expressive power of neural networks: A view from the width, in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017) pp. 6231–6239.

[31] B. Hanin and M. Sellke, Approximating continuous functions by relu nets of minimal width (2018), arXiv:1710.11278 [stat.ML].

[32] B. Hanin, Universal function approximation by deep neural nets with bounded width and relu activations, Mathematics **7**, 992 (2019).

[33] K. Floettmann, ASTRA - A Space Charge Tracking Algorithm, `http://www.desy.de/~mpyflo`.

[34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[35] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks (JMLR Workshop and Conference Proceedings, Fort Lauderdale, FL, USA, 2011) pp. 315–323.

[36] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for activation functions (2017), arXiv:1710.05941 [cs.NE].

[37] S. Ruder, An overview of gradient descent optimization algorithms (2017), arXiv:1609.04747 [cs.LG].

[38] In our study, the training procedure usually took $< 1\,\mathrm{h}$ on an Apple M1 processor. Production of the training data sets, however, can take several days, depending on the available compute infrastructure. In order to speed-up data set production, we used the parallelized version of ASTRA.

[39] This issue is already somewhat alleviated by using δB_i, as described in Sec. III.

[40] D. Lipka, J. Lund-Nielsen, and M. Seebach, Resonator for Charge Measurement at REGAE (2nd International Beam Instrumentation Conference, Oxford (United Kingdom), 16 Sep 2013 - 19 Sep 2013, 2013).

[41] C. Wiebers, M. Holz, G. Kube, D. Nölle, G. Priebe, and H.-C. Schröder, Scintillating Screen Monitors for Transverse Electron Beam Profile Diagnostics at the European XFEL (2nd International Beam Instrumentation Conference, Oxford (United Kingdom), 16 Sep 2013 - 19 Sep 2013, 2013).

[42] SPI G200TH TEM Grid, available at `https://www.2spi.com/item/202hc-xa`, *last access: 22nd April 2022*.

[43] SPI G300 TEM Grid, available at `https://www.2spi.com/item/2030c-xa`, *last access: 22nd April 2022*.

[44] Custom made Tungsten pepper pot (laser drilled). Hole diameter: 15 µm, pitch: 85 µm, thickness: 20 µm.