

<https://doi.org/10.1038/s42005-025-01984-8>

# A neural network-based synthetic diagnostic of laser-accelerated proton energy spectra

Check for updates

Christopher J. G. McQueen<sup>1,2</sup>, Robbie Wilson<sup>1</sup>, Timothy P. Frazer<sup>1</sup>, Martin King<sup>1,2</sup>, Matthew Alderton<sup>1</sup>, Ewan F. J. Bacon<sup>1</sup>, Ewan J. Dolier<sup>1</sup>, Thomas Dzelzainis<sup>3</sup>, Jesel K. Patel<sup>1,3</sup>, Maia P. Peat<sup>1</sup>, Ben C. Torrance<sup>1</sup>, Ross J. Gray<sup>1,2</sup> & Paul McKenna<sup>1,2</sup> ✉

Machine learning can revolutionize the development of laser-plasma accelerators by enabling real-time optimization, predictive modeling and experimental automation. Given the broad range of laser and plasma parameters and shot-to-shot variability in laser-driven ion acceleration at present, continuous monitoring with real-time, non-disruptive ion diagnostics is crucial for consistent operation. Machine learning provides effective solutions for this challenge. We present a synthetic diagnostic method using deep neural networks to predict the energy spectrum of laser-accelerated protons. This model combines variational autoencoders for dimensionality reduction with feed-forward networks for predictions based on secondary diagnostics of the laser-plasma interactions. Trained on data from fewer than 700 laser-plasma interactions, the model achieves an error level of 13.5%, and improves with more data. This non-destructive diagnostic enables high-repetition laser operations with the approach extendable to a fully surrogate model for predicting realistic ion beam properties, unlocking potential for diverse applications of these promising sources.

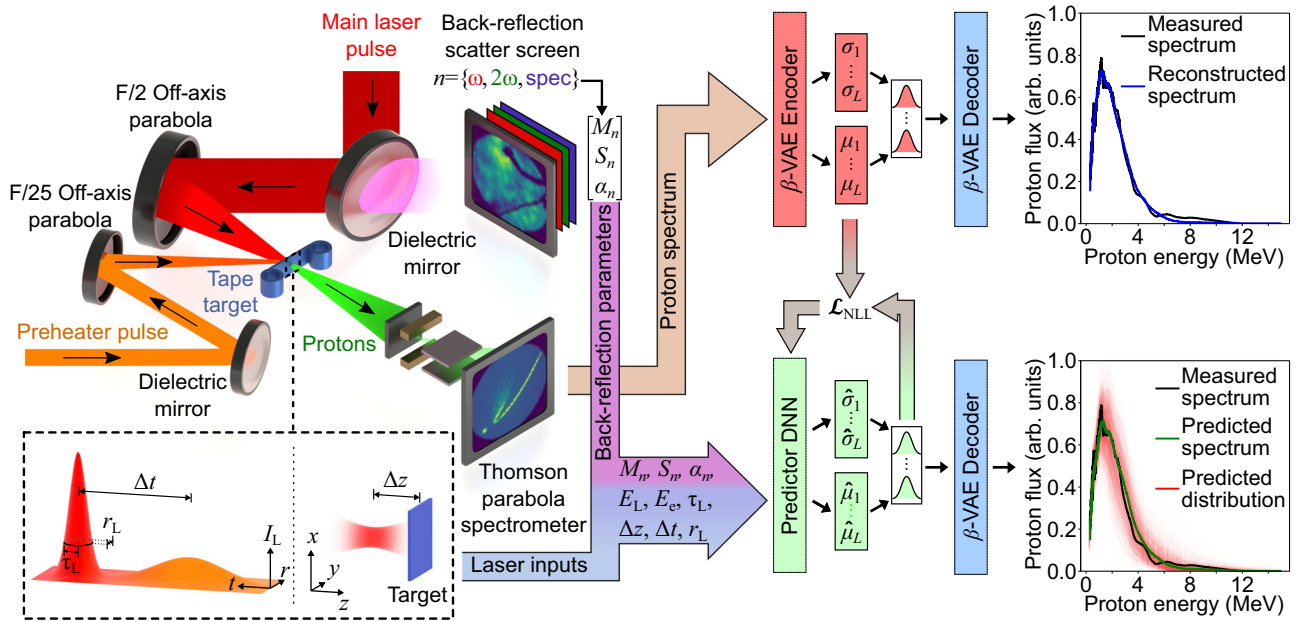
Particle accelerators enable researchers to explore the fundamental structure of matter and develop new technologies for medicine, industry and defense. Laser-plasma-based accelerators generate ultrashort pulses of high-energy particles over short distances, driven by ultra-high field gradients<sup>1,2</sup>. This enables the development of compact particle accelerators capable of delivering ultra-high dose rates. Proton and ion beams produced by laser-driven accelerators have potential applications in medical imaging and FLASH radiotherapy<sup>3–5</sup>. Realising these applications requires new approaches to optimization and control of the acceleration process and beam properties.

The application of machine learning (ML) in laser-plasma accelerator development is an emerging and dynamic area, providing innovative solutions to critical challenges in optimization and control. Early demonstrations include Bayesian optimization approaches in electron and ion acceleration experiments<sup>6–8</sup> and modeling<sup>9,10</sup>, in addition to deep neural networks (DNNs) for rapid deconvolution of laser-plasma diagnostics<sup>11–15</sup>, diagnostics of intense laser focal parameters<sup>16</sup>, modeling laser wakefield acceleration of electrons<sup>17,18</sup>, laser-solid interactions<sup>19</sup> and ion acceleration<sup>20–23</sup>. While DNNs have previously been used to analyze experimental laser-driven ion diagnostics, their application to predicting ion beam properties has thus far been restricted to models trained solely on simulation data. Despite advancements in laser-plasma ion source

development, challenges such as optimization, stabilization and real-time diagnosis remain, where ML can play a transformative role. The growing availability of experimental data from high-repetition rate systems is accelerating greater integration of DNNs in this field.

In the realm of laser-plasma-based ion acceleration, Target-Normal Sheath Acceleration (TNSA) is the most reproducible mechanism, producing short ion pulses with energies reaching tens-of-MeV/nucleon<sup>3,4</sup>. This method involves irradiating a thin foil target with relativistic laser intensities ( $> 10^{18} \text{ W cm}^{-2}$  for  $\sim 1 \mu\text{m}$  light) to produce fast electrons, generating a strong electric field ( $\text{TV m}^{-1}$ ) for ion acceleration. Stabilizing these ion sources and consistently reproducing desired ion beam properties is inherently challenging due to their sensitivity to shot-to-shot variations in laser pulse parameters. Correlating output ion beam properties with input parameters is further complicated by the interplay of numerous interrelated factors. Direct experimental measurement of beam parameters is typically possible only when the ion beam is not being used for irradiation applications or when only a small portion of the beam is needed for that purpose, as direct characterization can disrupt the beam properties. Key applications in radiobiology<sup>24,25</sup> and radiation damage studies often require capturing and focusing the entire ion beam to deliver the required dose. This requirement complicates simultaneous measurement and application, highlighting the

<sup>1</sup>SUPA Department of Physics, University of Strathclyde, Glasgow, UK. <sup>2</sup>The Cockcroft Institute, Sci-Tech Daresbury, Warrington, UK. <sup>3</sup>Central Laser facility, Rutherford Appleton Laboratory, Didcot, UK. ✉e-mail: paul.mckenna@strath.ac.uk



**Fig. 1 | Experimental schematic and model process.** The main laser pulse is transported via multiple stages of the laser beamline and finally turned with a dielectric mirror onto a F/2 parabolic mirror that focuses the pulse onto a tape target, producing a plasma and driving ion acceleration. Laser light back-reflected from the plasma, together with second harmonic light generated in the plasma, passes through the dielectric mirror and onto a scatter screen, where the spatial, intensity and spectral properties are measured. The spectrum of the beam of protons accelerated from the plasma is measured using a Thomson parabola ion spectrometer. As shown in the dashed inset, a lower intensity preheater laser pulse is focused onto the target at time  $\Delta t$  prior to the arrival of the main pulse. The target displacement with respect to the main laser pulse focus,  $\Delta z$ , is also varied. The model training and

prediction processes, and example spectra, are shown on the right. The  $\beta$ -VAE is trained to encode and reconstruct (blue) the measured (black) proton spectra, after which the encoder is used to generate a mean latent space representation for all spectra in the training set. The predictor neural network is trained with the laser parameters and back-reflection moments as inputs to predict the mean and standard deviation of the encoded proton spectra, by minimizing the negative log-likelihood loss function ( $L_{NLL}$ ). The mean latent space prediction is decoded by the  $\beta$ -VAE decoder to produce the predicted spectrum (green), with 1000 samples from the predicted latent space distribution decoded to generate a predicted distribution (each sample shown in a shade of red with reduced opacity to visualise the distribution of overlaid data).

need for reliable modeling approaches. While high-fidelity particle-in-cell (PIC) simulations can model these interactions, they are time-consuming, necessitating simplified empirical models.

Recent advancements in high-power laser repetition rates and rapidly replaceable targets—such as tape-driven foils, droplets, clusters and cryogenic or liquid sheets—enable thousands of laser shots in short experimental campaigns<sup>26</sup>. This capability facilitates extensive parameter scanning, generating large datasets for developing empirical models and training DNNs, which are particularly adept at modeling complex, nonlinear problems.

Here, we introduce a DNN-based synthetic diagnostic of the TNSA-ion energy spectrum, trained and validated with experimental data. A dense feed-forward network, combined with a  $\beta$ -variational autoencoder<sup>27</sup> ( $\beta$ -VAE) predicts the ion beam spectrum and its uncertainty using the laser input parameters and secondary laser-plasma interaction diagnostics. The  $\beta$ -VAE excels in dimensionality reduction to a smooth, interpolatable latent space, enhancing prediction accuracy even with limited, high-dimensional training data. This synthetic diagnostic approach offers real-time insights into ion beam properties without disrupting the beam, facilitating improved operational control.

## Results

### Model input data

The experimental data was obtained from 662 laser-target interactions using the Gemini high power laser<sup>28</sup>, with 80% of the dataset used for hyperparameter tuning and model training, and 20% reserved as a holdout test set for performance estimation. The experiment, depicted schematically in Fig. 1, is described in detail in Methods.

The objective of the DNN is to predict the energy spectrum of the beam of protons accelerated during the laser-target interaction. Therefore, the energy spectrum was measured for each interaction. The DNN inputs

include the on-shot laser energy ( $E_L$ ), pulse duration ( $\tau_L$ ), focal spot radius ( $r_L$ ), the percentage of laser energy within the focal area ( $E_f$ ), and the target distance from the focal point ( $\Delta z$ ). In addition, the temporal offset of a ‘preheater’ laser pulse ( $\Delta t$ ) is included. This preheater laser pulse generates a preplasma at the target front surface, which can significantly influence the evolution of the subsequent main pulse interaction<sup>29,30</sup>.

As an in-situ diagnostic of the interaction, we measured the on-shot backreflected light from the laser-plasma interaction using a spectrometer, along with cameras filtered for both the laser frequency ( $\omega$ ) and the second harmonic ( $2\omega$ ). This reflected light inversely correlates with the energy absorbed during the interaction and is expected to encode information about the proton energy spectrum<sup>31</sup>.

The processed backreflection images and spectra for each shot consist of over  $300 \times 300$  and 1700 pixels, respectively. The high dimensionality of this input data presents a significant challenge for a DNN, known as the *curse of dimensionality*<sup>32</sup>, especially given the limited training data available. To address this, we reduce the dimensionality by calculating the mean ( $M$ ), standard deviation ( $S$ ) and skewness ( $\alpha$ ) of the pixel intensity distributions in the images and spectra, which removes explicit spatial information. Both higher-order moments and image moments—where the images (or spectra) are treated as distributions to preserve spatial or spectral information<sup>33</sup>—were tested during hyperparameter tuning. These methods led to poorer model performance, suggesting that these details do not contribute effectively to the model’s predictive capability.

The nine backreflection moments—comprising the mean, variance, and skewness of the backreflection spectra, laser-frequency-filtered images, and second-harmonic-filtered images—along with the six laser-target parameters, serve as the lower-dimensional inputs to the network. Since the proton spectra are high-dimensional arrays of size 2057, both dimensionality reduction and an inverse reconstruction process are required; these are discussed in the next section.

## Model pipeline

Following the methodology in references<sup>18,34–36</sup>, we employed a  $\beta$ -VAE for the dimensionality reduction of proton spectra. Autoencoders are composed of an encoder and a decoder. The encoder maps the input data to a lower-dimensional ‘latent space’ representation (LSR), while the decoder reconstructs the input data from this encoded representation. The  $\beta$ -VAE enhances this framework by having the encoder generate parameters of a Gaussian probability distribution instead of deterministic values, as illustrated in Fig. 1. By sampling from this distribution and incorporating a loss term that constrains it to approximate a standard normal distribution, a smooth latent space is learned<sup>37</sup>. This process is regulated by the hyperparameter  $\beta$ , which weights the importance of latent space smoothness to the reconstruction error (see the  $\beta$ -VAE architecture section in Methods).

The  $\beta$ -VAE’s encoder reduces the dimensionality of the proton spectra by generating a mean LSR ( $\mu$ ) for each spectrum in the training dataset. A feed-forward DNN, or ‘predictor’, is then trained to predict the LSR from the laser parameters and backreflection moments. The decoder subsequently converts these predicted encodings into proton spectra, as illustrated in Fig. 1. The smoothness of the latent space imposed by the  $\beta$ -VAE architecture allows for reasonable reconstructions, even for slightly misaligned latent space predictions. This predictor-decoder pair therefore forms a synthetic diagnostic for laser-driven protons based purely on input laser parameters and backreflection measurement.

## Uncertainty quantification

Quantifying uncertainties is an essential aspect of applying ML models to practical applications<sup>38</sup>. ‘Aleatoric’ uncertainty stems from inherent randomness or unaccounted variations in the input and output data, such as uncorrected drifts in laser parameters during operation. In contrast, ‘epistemic’ uncertainty originates from limitations in the model design or the size and diversity of the training dataset<sup>39</sup>. Aleatoric uncertainty of predicted spectra is estimated by the predictor network under two assumptions. First, the LSR uncertainty is a multivariate normal distribution. Second, each latent space dimension is statistically independent. The use of the  $\beta$ -VAE architecture supports these assumptions, and they are validated by the model’s performance in the Uncertainty prediction accuracy section. The predictor network outputs a mean value and the natural logarithm of the standard deviation (to ensure positive values) for each latent space dimension. The network is trained using a negative log-likelihood loss function, optimizing the mean and standard deviation prediction simultaneously<sup>40</sup>. Sampling multiple times from the predicted distribution and decoding with the  $\beta$ -VAE decoder generates an empirical distribution of spectra, representing an estimate of aleatoric uncertainty.

For the ‘average’ prediction of a predictor-decoder pair, the predicted mean LSR ( $\mu$  in Fig. 1) can be decoded directly or multiple samples from the predicted latent space distribution can be decoded and averaged. While the latter approach offers slightly better prediction accuracy, decoding the predicted mean directly is faster and thus more suitable for the high repetition rates of modern laser-driven ion sources. Therefore, the average proton spectrum is obtained by directly decoding the predicted mean LSR.

A widely used technique for estimating epistemic uncertainty and enhancing prediction accuracy is ensembling<sup>41</sup>. This involves training multiple neural networks and combining their outputs, ideally with diversity introduced during training, e.g., via varied hyperparameters or training data. During hyperparameter tuning, we found that  $\beta$ -VAE performance was not a limiting factor, so ensembling efforts focused on the predictor networks.

An ensemble of 100 predictor-decoder pairs was trained, consisting of 4  $\beta$ -VAE decoders, each paired with 25 predictors. Diversity was introduced by training each member on random subsets of the training data, as detailed in the Ensembling section in Methods.

This ensemble, hereinafter referred to as the synthetic diagnostic ensemble (SDE), generates an uncertainty distribution for a predicted proton spectrum, incorporating both aleatoric and epistemic uncertainty. Multiple samples are taken from each predictor’s latent space distribution and decoded with the paired decoder, producing a distribution of spectra

from which statistical information, such as confidence intervals, can be derived.

We use the median of the average predicted spectra from all ensemble members as the SDE prediction. This approach is computationally efficient, with only a 0.1% higher error than the best performing method, enabling rapid generation of average spectra while retaining uncertainty estimation through sampling.

Occasionally, the SDE generates spectra which contain non-physical negative values, particularly in low signal spectra. These values are set to zero, as they likely stem from poorly trained regions of the latent space, leading to unphysical output spectra after decoding. When producing the SDE average prediction, it was found that eliminating negative values either before or after calculating the median made negligible difference. In the implemented code, this clipping is performed after the median calculation.

## Proton spectrum and total flux prediction accuracy

Since the ion spectrometer used in this experiment was not calibrated for absolute proton number, absolute error metrics are not particularly insightful for assessing the SDE performance. Instead, we use the percentage of unexplained variance (PUV), a relative error metric, calculated as the ratio of the mean squared error of the model predictions to the variance of the dataset, indicating the percentage of variance not explained by predictions (see the Metrics section in Methods). A naive model that always predicts the mean spectrum of the test dataset, irrespective of the inputs, would have a PUV of 100%. The SDE predictions have an average PUV of 13.5%, indicating high predictive power.

We also evaluated the reconstruction performance of the 4  $\beta$ -VAEs in the SDE, finding an average PUV of 1.9%. The smaller reconstruction error compared to prediction errors indicates that most inaccuracies arise from the predictors’ ability to estimate the LSRs, rather than from the  $\beta$ -VAE reconstruction process.

To visualize model performance, Fig. 2a shows the measured spectra, the mean reconstructions by the  $\beta$ -VAEs and the SDE predictions for the test dataset (20% of the total dataset). These plots demonstrate that while  $\beta$ -VAE reconstructions are closer to the actual spectra, SDE predictions are accurate across most of the dataset.

Figure 2b displays individual measured spectra, SDE predictions, and 1000 random samples from the predicted SDE distributions for the test dataset. This distribution estimates both aleatoric and epistemic uncertainty.

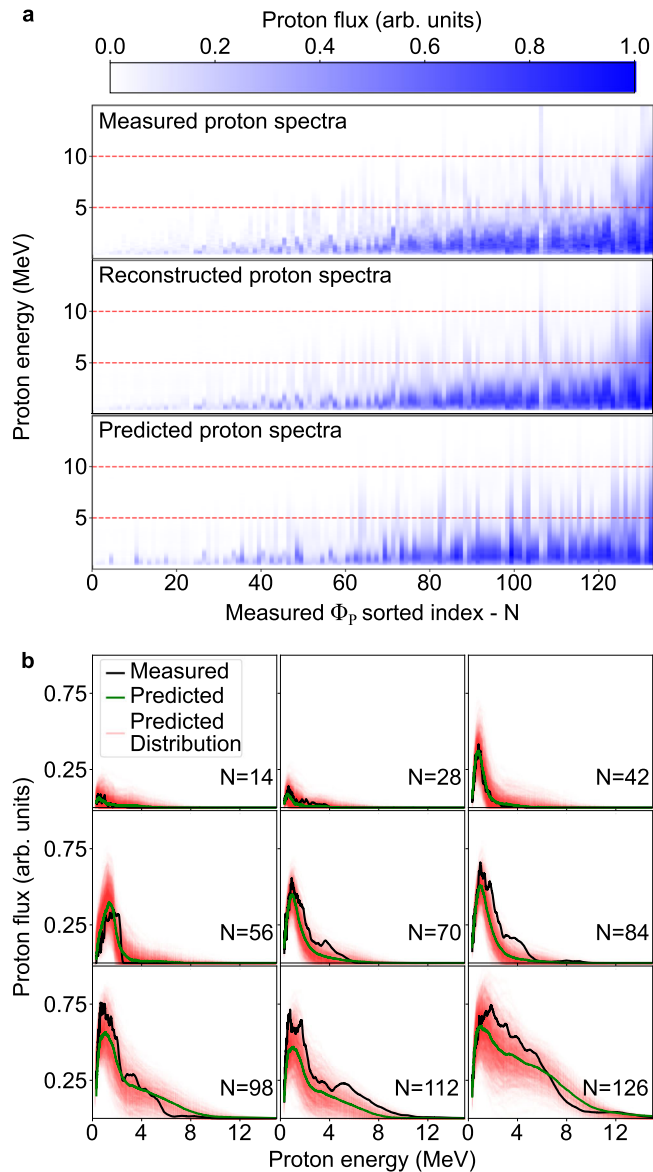
Interpreting the PUV of spectra is challenging due to nonlinear separation of proton energies in the ion spectrometer, which causes resolution to decrease as energy increases. Consequently, the mean squared error, and by extension, the PUV, is biased towards performance at low proton energies. To address this, we also evaluate the performance of the model in predicting the total flux of the spectra,  $\Phi_p$  (the integral over energy), which appropriately weights energy bins based on their width. The PUV of these predictions is 16.4%, indicating low error across the spectral energy range.

For each spectrum, we sampled 1000 times from the SDE’s predicted uncertainty distribution to generate a  $\Phi_p$  uncertainty distribution. The 68% confidence interval (16th to 84th percentile) is shown in Fig. 3a, plotted as a function of the measured  $\Phi_p$ .

Figures 2 and 3a show good agreement between the average predictions and the actual values, within the predicted uncertainty range. The next section will explore the validity and practical value of these uncertainty predictions.

## Uncertainty prediction accuracy

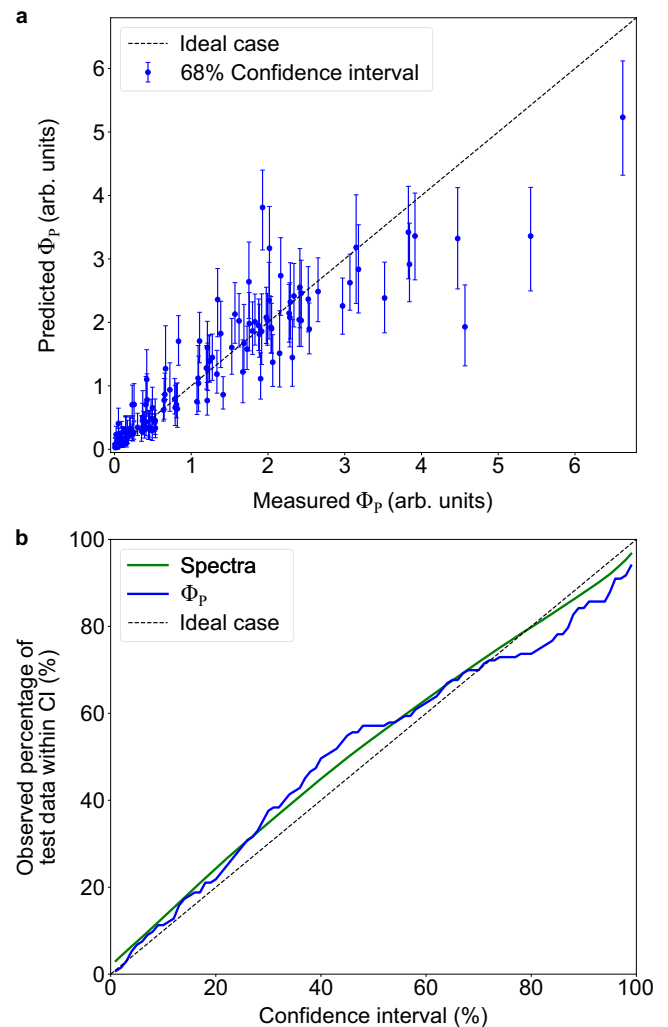
Whilst the uncertainty predictions of the SDE are plausible, their validity needs assessment. Calibration is a crucial aspect of uncertainty prediction, testing whether predicted confidence intervals (CIs) align with observed data. To evaluate this, we use a calibration curve, plotting the percentage of the test dataset (i.e. all energy bins of all spectra or all  $\Phi_p$  values) that fall within a given predicted CI, against the corresponding CI size. Perfect calibration is represented by the line  $y=x$ , with values below the ideal line indicating overconfidence and values above indicating underconfidence<sup>42,43</sup>.



**Fig. 2 | Spectral comparison.** **a** Proton spectra measured experimentally (top), with corresponding average reconstructions from the  $\beta$ -VAEs (middle) and predictions from the predictor-decoder pair in the SDE (bottom). The spectra are sorted by increasing measured  $\Phi_p$ , indexed by  $N$ . Red dashed lines are included as visual guides to assist in identifying specific energy levels. **b** Examples from evenly-spaced,  $\Phi_p$ -sorted indices  $N$ , showing measured spectra (black), average predictions from the SDE (green), and 1,000 overlaid samples from the predicted distribution generated by the SDE (each sample shown in a shade of red with reduced opacity to visualise the distribution of overlaid data).

We computed the percentage of test data within predicted CIs, centred on the median, from 1% to 99% in 1% steps. This is shown for both spectra and  $\Phi_p$  predictions in Fig. 3b, displaying only small deviations from perfect calibration. Calibration error (CE), defined in the Metrics subsection in Methods, quantifies calibration quality, measured as the root mean square (RMS) deviation between the values of the CIs and the observed data within them. Our model achieves a CE of 3.35% for spectra and 5.8% for  $\Phi_p$  predictions. We also calculate signed CE<sup>44,45</sup>, which is positive for predominantly underconfident predictions. The signed CE is 2.1% for spectra and 1.8% for  $\Phi_p$  predictions, suggesting that incorrect uncertainty estimates are generally conservative.

Additionally, we measure the sharpness of predicted distributions, defined in the Metrics subsection in Methods as the average RMS width of the predicted distributions. Lower values indicate narrower distributions, which



**Fig. 3 | Confidence test results.** **a** 68% confidence intervals (centered on the median) of the predicted  $\Phi_p$  distributions plotted as a function of  $\Phi_p$  from the measured spectra (blue markers with error bars), alongside the ideal case of perfect accuracy (black dashed line). **b** Reliability diagram showing the fraction of measured data within predicted confidence intervals for spectra (green) and  $\Phi_p$  (blue). The dashed diagonal line at  $y=x$  represents perfect calibration.

are more insightful, assuming comparable calibration<sup>46</sup>. The sharpness across predicted spectra distributions is 0.065, much narrower than the dataset’s average standard deviation of 0.19. Similarly, the predicted  $\Phi_p$  distributions have a sharpness of 0.4 compared to a standard deviation of 1.2.

The high degree of calibration and sharpness of the predicted distributions highlights the power of this model architecture to be utilized as a synthetic diagnostic, capable of generating accurate and precise uncertainty distributions of on-shot spectra with training on a small dataset.

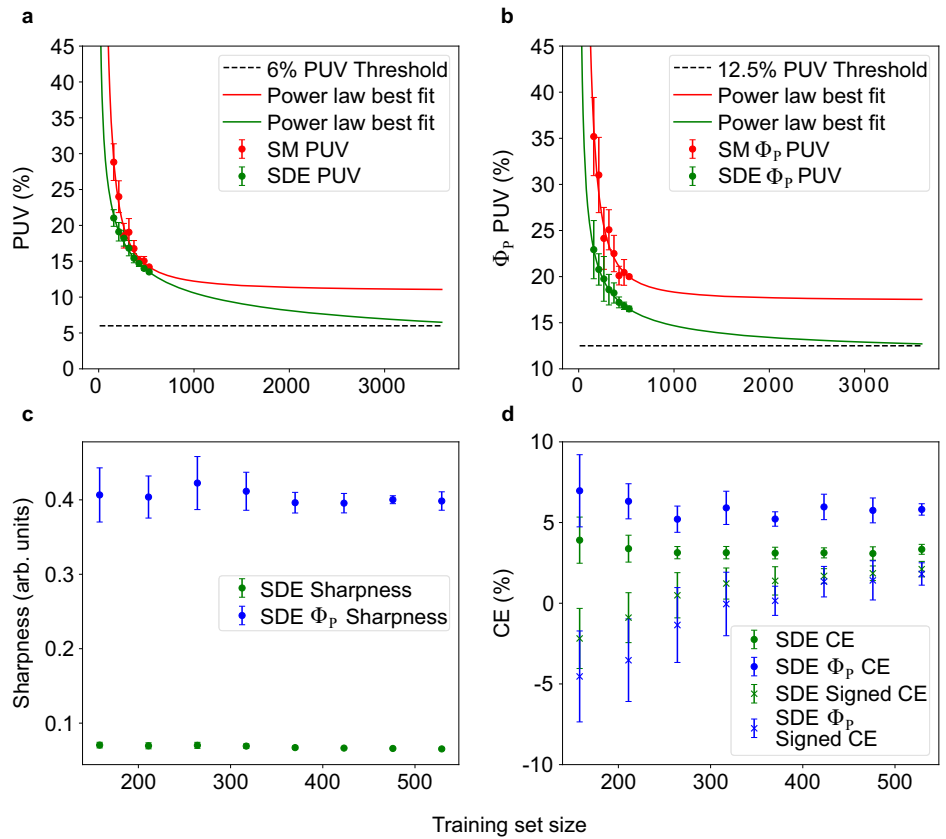
### Training set size performance scaling

The SDE demonstrates promising performance with fewer than 700 training samples. DNNs typically improve with larger datasets, following a power law trend<sup>47–49</sup>. This suggests potential error reduction with larger datasets, as planned for future experiments.

To extrapolate future performance, we retrained the SDE on random sub-samples of the full training dataset of varying sizes, repeating the process ten times to ensure robustness against data distribution changes and training variability. The spectra and  $\Phi_p$  PUV, sharpness, and CE of the SDE predictions were evaluated after each training session.

The results, presented in Fig. 4a, show a substantial reduction in error as more data is added, following a power law trend. For example, with

**Fig. 4 | Effect of training set size on various metrics.** **(a)** Synthetic diagnostic ensemble proton spectra prediction PUV (green), **(b)**  $\Phi_p$  prediction PUV (green), **(c)** sharpness of predicted spectrum (green) and  $\Phi_p$  (blue) distributions, and **(d)** calibration error (CE) and signed CE, all as a function of training set size. Error bars represent variability from repeated resampling and retraining. Plots **(a, b)** include power-law curve fits extrapolated to a training set size of 3600, and include the performance and power-law fit of the surrogate model (red).



3600 shots of training data, achievable in about an hour with a Hz-rate laser system, the PUV for predicted spectra is projected to reach ~6.5%.

The PUV of the  $\Phi_p$  predictions, shown in Fig. 4b, also follows a power-law trend but improves less significantly, reaching about 12.7% after 3600 training shots.

The sharpness of predicted spectra distributions decreases slightly with more data, from 0.07 with < 160 training samples to 0.065 with the full dataset, while the predicted  $\Phi_p$  distribution sharpness remains almost unchanged, as shown in Fig. 4c. This indicates that adding more training data does not significantly sharpen the distributions. However, as previously noted, the distributions are already considerably sharper than the overall variation within the dataset.

The gradual improvement in the SDE  $\Phi_p$  prediction error and negligible change of the  $\Phi_p$  distribution sharpness can be attributed to a small number of, predominantly high-energy, high-flux, spectra where the model underperforms. In contrast, greater improvement is observed in overall metrics due to the bias towards low energy regions of the spectra, where predictions improve significantly. This is potentially driven by the low density of high-energy spectra within the dataset (as can be observed from Figs. 2a, 3a), limiting learning. Alternatively, it could indicate that the backreflection diagnostics do not encode enough information for precise prediction of these high-energy spectra, which may exhibit larger shot-to-shot fluctuations.

These findings indicate significant performance gains are possible with larger datasets and highlight steps for improved performance across a wide parameter range, particularly the need for a balanced dataset or additional diagnostics, supporting advancements in laser-driven proton experiments. Importantly, these results indicate that the scale of the datasets required for significant improvements are within reach on current and next generation Hz repetition rate laser systems.

### Surrogate model

We also performed a dataset size scaling test on a full surrogate model (SM), which is analogous to the SDE but without on-shot backreflection, or other

secondary diagnostic, measurements. This SM would enable both on-line, real-time output ion beam parameter predictions based on measured experimental input laser and plasma parameters, and off-line input parameter space exploration, facilitating rapid optimization.

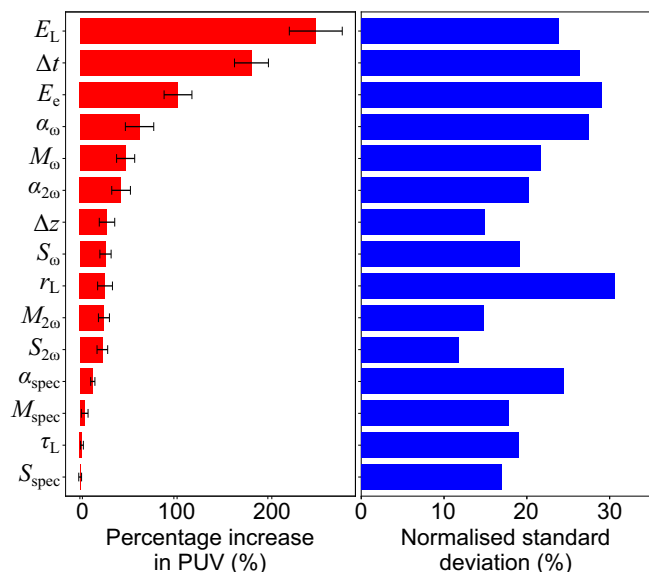
As shown in Fig. 4a, b, the performance of the SM plateaus much more quickly than that of the SDE with the current model architecture. However, a more sophisticated model could potentially achieve enhanced performance with additional data.

### Feature importance

In addition to the SDE's ability to provide accurate, on-shot predictions of the proton spectra, we can also gain insights into the underlying physics through the interpretation of its outputs. While the black-box nature of DNNs poses significant challenges to interpretability<sup>50–52</sup>, Permutation Feature Importance (PFI) provides a way to gauge the relative importance of specific inputs to the model performance. Originally developed for Random Forests<sup>53</sup>, PFI is now widely used for neural networks<sup>54–56</sup>. This method involves shuffling an input in the test set and measuring the change in prediction error, with large increases indicating strong dependency.

We applied PFI to the SDE ten times to account for random shuffling variation, as shown in Fig. 5, along with the standard deviation normalized to the range of each input parameter. The results highlight that laser pulse energy is the most critical feature for determining proton spectra, consistent with experimental studies<sup>57</sup>. Related to this, encircled energy is also important, reflecting its role in determining the total energy transferred to protons<sup>58</sup>.

The preheater offset time emerges as the second most important factor. It determines the density profile of the expanding preplasma, which in turn influences the interaction between the main laser pulse and the plasma, and specifically the energy coupling to fast electrons<sup>29</sup>. While existing analytical models of laser-driven ion acceleration do not fully account for the impact of the plasma density profile, our findings show it to be important, underscoring the advantages of this approach over more limited models. Focal



**Fig. 5 | Influence of input parameters.** The red bars correspond to the percentage increase in the SDE PUV during application of Permutation Feature Importance, with the error bars representing variability from repeated measurements. The blue bars correspond to the standard deviation normalized to the range of each input. The subscripts  $\omega$  and  $2\omega$  identify moments of back-reflected light at the laser frequency and at twice the laser frequency, respectively, and  $spec$  indicates the back-reflected spectrum.

spot size and  $\Delta z$  have some effect, though less significant, while pulse duration exhibits minimal influence within the tested range. As shown, these parameters are less critical compared to energy and preheater offset time, despite being varied similarly during the experiment over their tested ranges.

Applying PFI to the SDE motivates an approach to defining system performance requirements and designing datasets for future ML-based laser-driven ion acceleration experiments. These findings highlight the substantial benefits of ensuring detailed sampling and accurate measurement of laser energy compared to other parameters, such as pulse duration. Additionally, they highlight the necessity of developing a more precise diagnostic of the front surface plasma density profile, given its relative importance as indicated by the model.

## Discussion

We have developed a DNN-based model that accurately predicts proton spectra from high-intensity laser-solid interactions while providing well-calibrated uncertainty estimates. The model employs an ensemble of predictor networks that use laser conditions and secondary backreflection measurements to predict reduced-dimensional spectra representations, which are subsequently decoded using  $\beta$ -VAE networks. By predicting Gaussian distribution parameters, the model generates robust uncertainty estimates that enhance prediction accuracy. The ensemble approach improves both uncertainty estimation and prediction error by combining diverse predictions from multiple predictor-decoder pairs, each trained on distinct subsets of the dataset.

This study presents a reliable framework for non-destructively determining the properties of laser-accelerated proton beams, enabling real-time diagnostics and optimization of future experiments. Traditional laser-plasma setups are constrained by bulky in-vacuum diagnostic systems, requiring large chambers and complex controls. Synthetic diagnostics offer a solution, significantly reducing system size and unlocking new applications (e.g., non-destructive testing and inspection), where source mobility and live diagnostics are essential<sup>59</sup>.

Additionally, our methodology, incorporating techniques such as PFI, enhances our understanding of laser-driven ion acceleration and laser-solid

interactions, providing valuable insights for guiding future experimental design and parameter optimization. The performance of the SDE demonstrates favorable scaling with increasing dataset size; however, the gradual improvement observed in the SDE  $\Phi_p$  predictions highlights the importance of achieving a balanced distribution of spectra in future experiments to maximize performance with limited data. This observation also suggests the potential need for different diagnostics that capture a broader range of interaction information, enabling more precise predictions.

The occasional negative output values in SDE predictions have a minimal impact on model performance. However, this could be addressed more rigorously in future models by applying the rectified linear unit (ReLU) activation function to the decoder output instead of the current linear function. Implementing this change would necessitate retuning the model hyperparameters due to the significant change in the architecture.

The lack of increased prediction accuracy when applying alternative dimensionality reduction techniques to the backreflection data, including the preservation of spatial and spectral information through the use of image moments (as discussed in the Model input data section), may reflect unique characteristics of the backreflection diagnostic itself. A diagnostic more strongly correlated with the proton spectrum could potentially benefit from a dimensionality reduction approach that better preserves such correlations. Furthermore, potential enhancements to the model architecture could include adopting a non-parametric approach to predicting aleatoric uncertainty, removing the assumption that latent space representations must follow a Gaussian uncertainty distribution. This could result in improvements in calibration and sharpness<sup>60</sup>.

Although the SM exhibits limited improvement with additional data, this may stem from model capacity constraints. Re-tuning hyperparameters or modifying the model architecture could enhance the predictor DNN's accuracy, offering a method for efficiently and accurately scanning large parameter spaces and designing future experiments.

Since the model inputs consist of in-situ backreflection measurements from the interaction, the model is unlikely to generalize immediately to datasets from other experimental setups or laser facilities. However, transfer learning presents a promising avenue to address this limitation. By using a pre-trained SDE from one system, it may be possible to learn the complex correlations of a new system with reduced training time and smaller dataset requirements, enabling rapid adaption for future applications<sup>61,62</sup>. This approach holds significant potential for current and emerging high-power laser facilities operating at 1 Hz repetition rate or greater, where an equivalent dataset size to that used in this work could be generated within tens of minutes. Consequently, training an SDE or surrogate model to diagnose and guide experiments could become a routine part of standard operating procedures.

In conclusion, the DNN architecture presented here provides accurate and insightful predictions of the spectral properties of laser-accelerated protons. This approach represents a significant advancement in non-destructive ion beam measurements and highlights the effectiveness of applying DNN models to laser-solid interactions. The technique has broad potential applications, ranging from diagnosing other radiation types (e.g., electrons, neutrons and x-rays) in high-power laser-plasma interactions to predicting outcomes and quantifying uncertainty in fields like nuclear fusion research and astrophysics, where high dimensionality and low data volumes present substantial challenges.

## Methods

### Experimental details

As illustrated in Fig. 1, pulses from the Gemini high power laser were focused to intensities up to  $4.4 \times 10^{20} \text{ Wcm}^{-2}$  onto a  $15 \mu\text{m}$ -thick copper tape-driven target. Prior to focusing, the pulses passed through a double plasma mirror system (not shown in the schematic). The timing of the arrival of a preheater laser pulse (focused to  $1 \times 10^{14} \text{ Wcm}^{-2}$ ) prior to the main laser pulse,  $\Delta t$ , controlled the degree of plasma expansion. The parameter  $\Delta t$  was varied within the range [0.4, 3.68] ns (noting that when the preheater is not used, it is labelled as 0 ns in the dataset) using a motorized

timing slide. Meanwhile,  $\Delta z$ , the spatial offset between the laser focus and the target front surface, was adjusted using motorized drives for the parabola, varying within the range of  $[-100, 100] \mu\text{m}$ . The energy of the preheater beam could not be measured for every shot during the experiment. Instead, it was characterized daily when the system was brought online. The energy variability is anticipated to be comparable to that of the main drive laser pulse, as both are derived from the same source. Since the preheater beam is not tightly focused, any potential wavefront changes are expected to have a minimal impact on the interaction. The wide range over which the preheater offset time was varied is likely the most significant parameter influencing changes in the preplasma.

Measurements of the laser pulse energy and resulting proton and optical radiation were made using a suite of diagnostic tools. The on-target laser energy, measured every shot using a calibrated calorimeter sampling energy from a laser pick-off, was within the range  $[0.65, 4.8] \text{ J}$ . The laser focal spot was periodically imaged with the laser in low-power mode at 10 Hz. The focal spot radius (defined as the half-width at half-maximum) and the encircled energy within the focal radius were calculated for each image and found to be within the range  $[1.35, 2.11] \mu\text{m}$  and  $[21.5, 25.7]\%$ , respectively (see the Data processing section). The pulse duration was adjusted within the range  $[36, 95] \text{ fs}$  via an Acousto-Optic Programmable Dispersive Filter (Dazzler by Fastlite). Daily pulse duration measurements were performed over five shots using a GRENOUILLE<sup>63</sup>, enabling both optimization and characterization of the pulse. The use of a double plasma mirror system required operations to be paused every 50 shots for plasma mirror replacement. Following each replacement, the laser focal spot was reoptimized and imaged to ensure consistency. This process effectively minimized the effects of drift in the laser wavefront. However, any unaccounted-for variations in the laser focus or spectrum could introduce noise into the focal spot or pulse duration input parameters.

The energy spectrum of the beam of accelerated protons was measured using a Thomson parabola ion spectrometer<sup>64–67</sup>, which directed the ions onto a multi-channel plate with phosphor screen. The resulting light output was imaged with an Andor Neo CMOS camera. Backreflected laser light from the target front surface was measured by directing it onto a diffuse scatter screen. The scattered light was imaged at both the laser frequency ( $\omega$ ) and the second harmonic ( $2\omega$ ), in addition to the spectrum being measured across a range from 175 to 1169 nm.

### Data processing

To extract the radii and encircled energies from the focal spot images, background reference images were subtracted and Gaussian curves fit in both x and y dimensions, centered on the image centroid. The focal spot radius was defined as the average half-width at half-maximum (HWHM) of these Gaussian distributions, and the encircled energy was calculated as the ratio of pixel intensity inside the area defined by the HWHM radius to the total intensity across the focal spot.

The accelerated ions are dispersed by the parallel electric and magnetic fields of the Thomson parabola spectrometer in parabolic traces, depending on their charge-to-mass ratio and energy. To reduce X-ray-induced noise, the images of the phosphor light were median-filtered, and custom code was used to extract the proton spectra from the ion traces by performing background subtraction and modeling the energy spread induced by the fields.

Images of the backreflected light were median-filtered, and a background level was estimated from regions outside the near-field, which was then subtracted; any negative pixel values set to zero. The images were cropped to match the near-field beam diameter and masked to eliminate pixels outside this area. The processed backreflection  $\omega$  and  $2\omega$  images were saved as  $320 \times 328$  and  $326 \times 334$  arrays, respectively.

Backreflection spectra, stored as 1714-element arrays, were also median-filtered. Each spectrum had the mean of the lowest 10% subtracted to remove background noise and was clipped to ensure non-negative values.

### Metrics

Several metrics are used to evaluate the performance of the model including the PUV, CE, and sharpness. These are defined as follows:

$$\begin{aligned} \text{PUV} &= 100(1 - R^2) \\ &= 100 \frac{\sum_{i=1}^N \sum_{j=1}^L (y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^L (y_{ij} - \bar{y}_j)^2} \end{aligned} \quad (1)$$

where  $R^2$  is the coefficient of determination and  $N$  is the number of spectra in the test dataset, each with length  $L$ .  $y_{ij}$  and  $\bar{y}_j$  denote the  $j$ th element of the  $i$ th spectrum and the mean spectrum (defined as the mean value for each energy bin over all spectra in the test dataset), respectively.  $\hat{y}_{ij}$  indicates the corresponding predicted value. The PUV, a commonly used performance metric in regression problems, is the ratio of the mean squared error of the model to the variance of the dataset, representing the percentage of the variance in the data not explained by the model. Alternatively, it may be interpreted as the error of the model relative to the error of a naive model that always predicts the dataset mean<sup>68–72</sup>. For  $\Phi_p$  predictions, the same formula applies, with  $L = 1$  and  $\bar{y}_j$  replaced with the mean  $\Phi_p$  of the test dataset.

$$\text{CE} = \sqrt{\frac{1}{99} \sum_{p=1}^{99} (O_p - P)^2} \quad (2)$$

$$\text{Signed CE} = \frac{1}{99} \sum_{p=1}^{99} (O_p - P) \quad (3)$$

are the CE and signed CE respectively, where  $O_p$  is the percentage of the measured data (i.e., all measured energy bins of all spectra, or all  $\Phi_p$ , in the test set) observed to be within a  $P\%$  confidence interval<sup>42,44,45</sup>. We have included an additional square root operation on the CE from the definition in ref. 42 to retain the units.

For a predicted empirical distribution with  $S$  samples, where each sample is given by  $\tilde{y}_{ijs}$ , representing the uncertainty in the  $j$ th element of the  $i$ th predicted spectrum, the sharpness can be defined as

$$\text{sharpness} = \sqrt{\frac{1}{NL} \sum_{i=1}^N \sum_{j=1}^L \frac{1}{S} \sum_{s=1}^S (\tilde{y}_{ijs} - \bar{y}_{ij})^2} \quad (4)$$

Here,  $\bar{y}_{ij}$  is the mean over  $s$  of  $\tilde{y}_{ijs}$ <sup>42,73</sup>.

During hyperparameter tuning, we used the Continuous Ranked Probability Score (CRPS) to evaluate model performance. CRPS integrates both calibration and sharpness into a single metric, making it effective for comparing the overall accuracy of predicted distributions<sup>46,74</sup>. We estimated CRPS by calculating the average of the ‘pinball loss’<sup>75</sup> from the 1st to 99th quantiles of the predicted distributions<sup>74</sup>.

### Data splitting

To accurately evaluate the final model performance, a random 20% of the total data was set aside as a holdout test set, which was not used for data normalization, hyperparameter tuning, or model training. This approach ensures that the final performance metrics are not overly optimistic due to overfitting.

Each DNN model was trained on 90% of the remaining data, with the remaining 10% reserved as a validation set. This configuration enabled the use of the Keras EarlyStopping callback, which halts training if the validation loss does not improve after a specified number of epochs (the ‘patience’ hyperparameter), thereby preventing unnecessary training and overfitting<sup>76</sup>. In addition, we used the ‘restore\_best\_weights’ parameter of EarlyStopping,

restoring the model weights to those found on the training epoch with the lowest validation loss.

### Data normalization

Data normalization, which has been shown to enhance model performance, especially in data-limited scenarios<sup>77,78</sup>, was applied by mapping the training dataset inputs and outputs to the [0, 1] interval. The validation and test sets were scaled using the parameters obtained from scaling the training set, avoiding data leakage. Each  $\beta$ -VAE and predictor in the ensemble was trained on a different random sub-sample of the training data, necessitating unique scaling parameters for each ensemble member's inputs and outputs. Predicted data was then scaled back to the original range to ensure consistent performance evaluation.

### $\beta$ -VAE architecture

The encoder and decoder of the  $\beta$ -VAEs are fully connected dense neural networks. All layers use rectified linear unit (ReLU) activation functions, except for the input and output layers, which are linear. During hyperparameter tuning, the hidden layers of both the encoder and decoder are constrained to be symmetric, with a maximum of 2 hidden layers each.

The  $\beta$ -VAE loss function for each proton spectrum during training is defined as follows<sup>79</sup>:

$$\mathcal{L}_{\beta\text{-VAE}} = \frac{1}{2} \sum_{i=1}^L (y_i - \hat{y}_i)^2 + \frac{\beta}{2} \sum_{j=1}^{L_z} (\mu_j^2 + \sigma_j^2 - \log(\sigma_j^2) - 1) \quad (5)$$

where  $L$  denotes the 'length' of the proton spectrum,  $y_i$  is the  $i$ th element of the ground truth spectrum, and  $\hat{y}_i$  is the corresponding element of the reconstructed spectrum.  $L_z$  represents the length of the latent space, while  $\mu_j$  and  $\sigma_j$  is the mean and standard deviation of the latent space generated by the encoder, respectively.

The first term on the right-hand side of equation (5) captures the reconstruction error from the decoder, whereas the second term represents the Kullback-Leibler (KL) divergence between the latent space distribution and the standard normal distribution.

The parameter  $\beta$  is a hyperparameter that regulates the trade-off between the reconstruction error and the KL divergence. Increasing  $\beta$  results in a smoother and more regularized latent space but may also increase the reconstruction error. During hyperparameter tuning,  $\beta$  was permitted to be set to 0, effectively disabling the KL divergence term. In this scenario, the encoder outputs only the mean LSR, reverting the  $\beta$ -VAE to a standard autoencoder.

### Predictor architecture

The predictor network is a fully connected dense network that uses ReLU activation functions across all hidden layers, with the architecture allowing for up to three hidden layers during hyperparameter tuning.

As discussed in the Uncertainty quantification section, the predictor DNNs use a negative log-likelihood loss function to estimate the parameters of an independent multivariate normal distribution, reflecting the assumed distribution of latent space variables due to experimental noise. The loss function for each encoded spectrum is given by<sup>40</sup>:

$$\mathcal{L}_{\text{NLL}} = \frac{1}{2} \sum_{j=1}^{L_z} \left( \log(\hat{\sigma}_j^2) + \frac{(\mu_j - \hat{\mu}_j)^2}{\hat{\sigma}_j^2} \right) \quad (6)$$

where  $\mu_j$  is the  $j$ th element of the encoded ground truth spectrum generated by a  $\beta$ -VAE encoder, and  $\hat{\mu}_j$  and  $\hat{\sigma}_j$  are the corresponding predicted mean and standard deviation of the predictor DNN, respectively.

### Hyperparameter tuning

The combined  $\beta$ -VAE-predictor model has numerous hyperparameters that need to be tuned for optimal performance. Given the computational

expense of training each DNN, we used Bayesian optimization via the scikit-optimize gp\_minimize package<sup>80-82</sup>.

Initially, we performed Bayesian optimization over broad parameter ranges for 221 iterations to identify regions of high performance. This was followed by a more focused optimization over narrower ranges for an additional 123 iterations. Hyperparameter tuning was conducted using  $k$ -fold cross-validation ( $k=10$ )<sup>83-85</sup>, with nine folds used for training and one used for performance evaluation on every permutation of folds. The nine training folds incurred a further 90/10 random split into training and validation datasets for EarlyStopping. The  $\beta$ -VAE and predictor networks were trained for a maximum of 15,000 epochs, with EarlyStopping patience parameters tuned independently.

During tuning, the batch sizes for training the  $\beta$ -VAEs and predictors were set to 16, 32 or 64. To ensure uniform batch sizes, the training data for each fold was truncated to a multiple of 64. Additionally, the maximum moment order (MMO) of the backreflection diagnostics was treated as a hyperparameter, recognizing that not all moments may be essential for predictions. Including too many unimportant inputs could potentially reduce performance.

The full list of tuned hyperparameters along with their allowed ranges and optimal values found are shown in Table 1.

We selected stochastic gradient descent as the optimizer for the networks during training because of its proven ability to achieve superior generalization performance compared to adaptive algorithms such as ADAM<sup>86-88</sup>.

Given that the model predicts an uncertainty distribution rather than a single average spectrum, we needed a metric that evaluates the error across the entire distribution, not just the average. To achieve this, we generate 1000 samples from the predicted distribution for each ground truth spectrum and use an approximation of the CRPS. Minimizing this approximation enables us to fine-tune the model hyperparameters effectively, ensuring the model accurately captures the overall uncertainty distribution of the data.

### Ensembling

During hyperparameter tuning, we found that the prediction performance of the predictor-decoder pairs was not constrained by the reconstruction error of the  $\beta$ -VAEs, as detailed in the section Proton spectrum and total flux prediction accuracy. Consequently, ensembling the  $\beta$ -VAEs was not essential, and we trained only four  $\beta$ -VAEs, each using a different 10% subset of the training data for validation and EarlyStopping.

For the predictor ensemble, each member is again trained on a random 90% subset of the training data, with the remaining 10% used as validation

**Table 1 | Optimal DNN hyperparameters and tuning ranges**

Hyperparameter	Optimal Value	Tuning Range
$\beta$ -VAE LR	0.000601	[ $1 \times 10^{-4}$ , $1.5 \times 10^{-3}$ ]
$\beta$	0.0265	[0, 2]
$\beta$ -VAE Encoder Layer 1	505	[10, 1000]
$\beta$ -VAE Encoder Layer 2	633	[0, 1000]
Latent Space Size	8	[1, 15]
$\beta$ -VAE Patience	333	[10, 500]
Predictor LR	0.0024	[ $1 \times 10^{-4}$ , $5 \times 10^{-2}$ ]
Predictor Layer 1	1198	[10, 1500]
Predictor Layer 2	1435	[0, 1500]
Predictor Layer 3	247	[0, 1500]
Backreflection MMO	3	[0, 5]
Predictor Patience	145	[10, 500]
Batch Size	32	[16, 64]

Both the model learning rates (LRs) and  $\beta$  were trained on a logarithmic scale. To accommodate  $\beta$  values of zero, a small constant ( $1 \times 10^{-4}$ ) was added during training and subtracted before applying the value in the model.



for EarlyStopping. In addition, a scan was performed to test the CE, sharpness and PUV of the ensemble predictions as a function of different techniques to induce variation between ensemble members. The two methods involved either randomly truncating or padding the training data for each member network to multiples of the batch size (after the validation split) and sampling these resized datasets with or without replacement. When sampling without replacement, padding was achieved by randomly appending samples from the training data, also in multiples of the batch size. This step was unnecessary when sampling with replacement, as it allows direct sampling from a dataset larger than the original. Consequently, both the fraction of the data used for training each network and its distribution were varied.

The optimal performance, in terms of prediction error, CE and sharpness, was achieved when each member was trained with the full training dataset, padded to a multiple of the batch size, and sampled without replacement. It remains unclear which method of padding or truncating is most effective in edge cases where the training data (after the validation split) is already a multiple of the batch size, as no padding or truncating would be applied in this scenario with the current design.

Furthermore, various definitions of the ‘average’ prediction were tested, including the mean and median of all ensemble members, as well as the mean and median of an empirical distribution generated from the ensemble. It was found that the median of the predicted distribution yielded the highest accuracy by a small margin. However, as noted in the Uncertainty quantification section, the median of the average prediction from each ensemble member is computationally more efficient while maintaining a similar error rate.

### Data availability

Data associated with research published in this paper can be accessed at <https://doi.org/10.15129/cee473ca-3d2d-4150-b46b-6f964e8ce9d1>.

### Code availability

The developed code is available at [https://gitlab.cis.strath.ac.uk/mckennagroup/2024\\_synthdiag\\_commsphysics#](https://gitlab.cis.strath.ac.uk/mckennagroup/2024_synthdiag_commsphysics#).

Received: 9 October 2024; Accepted: 30 January 2025;  
Published online: 12 February 2025

### References

- Malka, V. Laser plasma accelerators. *Phys. Plasmas*. **19**, 055501 (2012).
- Hooker, S. Developments in laser-driven plasma accelerators. *Nat. Photonics* **7**, 775–782 (2013).
- Daido, H., Nishiuchi, M. & Pirozhkov, A. S. Review of laser-driven ion sources and their applications. *Rep. Prog. Phys.* **75**, 056401 (2012).
- Macchi, A., Borghesi, M. & Passoni, M. Ion acceleration by superintense laser-plasma interaction. *Rev. Mod. Phys.* **85**, 751–793 (2013).
- Chaudhary, P. et al. Cellular irradiations with laser-driven carbon ions at ultra-high dose rates. *Phys. Med. Biol.* **68**, 025015 (2023).
- Shaloo, R. J. et al. Automation and control of laser wakefield accelerators using Bayesian optimization. *Nat. Commun.* **11**, 6355 (2020).
- Jalas, S. et al. Bayesian Optimization of a Laser-Plasma Accelerator. *Phys. Rev. Lett.* **126**, 104801 (2021).
- Loughran, B. et al. Automated control and optimization of laser-driven ion acceleration. *High Power Laser Sci. Eng.* **11**, e35 (2023).
- Dolier, E. J., King, M., Wilson, R., Gray, R. J. & McKenna, P. Multi-parameter Bayesian optimisation of laser-driven ion acceleration in particle-in-cell simulations. *New J. Phys.* **24**, 073025 (2022).
- Goodman, J. et al. Optimization and control of synchrotron emission in ultraintense laser-solid interactions using machine learning. *High Power Laser Sci. Eng.* **11**, e34 (2023).
- Mariscal, D. A. et al. A flexible proton beam imaging energy spectrometer (PROBIES) for high repetition rate or single-shot high energy density (HED) experiments (invited). *Rev. Sci. Instrum.* **94**, 023507 (2023).
- Mariscal, D. A. et al. Design of flexible proton beam imaging energy spectrometers (PROBIES). *Plasma Phys. Control. Fusion* **63**, 114003 (2021).
- Miyatake, T. et al. Denoising application for electron spectrometer in laser-driven ion acceleration using a Simulation-supervised Learning based CDAE. *Nucl. Instrum. Methods Phys. Res. A* **999**, 165227 (2021).
- Swanson, K. K. et al. Applications of machine learning to a compact magnetic spectrometer for high repetition rate, laser-driven particle acceleration. *Rev. Sci. Instrum.* **93**, 103547 (2022).
- Kuramitsu, Y. et al. Optimization of laser-driven quantum beam generation and the applications with artificial intelligence. *Phys. Plasmas* **31**, 056108 (2024).
- Bukharskii, N. D., Vais, O. E., Korneev, Ph. A. & Bychenkov, V. Y. Restoration of the focal parameters for an extreme-power laser pulse with ponderomotively scattered proton spectra by using a neural network algorithm. *Matter Radiat. Extrem.* **8**, 014404 (2023).
- Kirchen, M. et al. Optimal Beam Loading in a Laser-Plasma Accelerator. *Phys. Rev. Lett.* **126**, 174801 (2021).
- Streeter, M. J. V. et al. Laser wakefield accelerator modelling with variational neural networks. *High Power Laser Sci. Eng.* **11**, e9 (2023).
- Mariscal, D. A. et al. Toward machine-learning-assisted PW-class high-repetition-rate experiments with solid targets. *Phys. Plasmas* **31**, 073105 (2024).
- Djordjević, B. Z. et al. Modeling laser-driven ion acceleration with deep learning. *Phys. Plasmas* **28**, 043105 (2021).
- Djordjević, B. Z. et al. Characterizing the acceleration time of laser-driven ion acceleration with data-informed neural networks. *Plasma Phys. Control. Fusion* **63**, 094005 (2021).
- Liu, Y. et al. Deep learning approaches for modeling laser-driven proton beams via phase-stable acceleration. *Phys. Plasmas* **31**, 013106 (2024).
- Schmitz, B., Kreuter, D., Boine-Frankenheim, O. & Margarone, D. Modeling of a Liquid Leaf Target TNSA Experiment Using Particle-In-Cell Simulations and Deep Learning. *Laser Part. Beams* **2023**, e3 (2023).
- Kroll, F. et al. Tumour irradiation in mice with a laser-accelerated proton beam. *Nat. Phys.* **18**, 316–322 (2022).
- Bin, J. et al. A new platform for ultra-high dose rate radiobiological research using the BELLA PW laser proton beamline. *Sci. Rep.* **12**, 1484 (2022).
- Prencipe, I. et al. Targets for high repetition rate laser facilities: needs, challenges and perspectives. *High Power Laser Sci. Eng.* **5**, e17 (2017).
- Higgins, I. et al. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. *5th Int. Conf. Learn. Represent.* (2017).
- Hooker, C. J. et al. The Astra Gemini project-A dual-beam petawatt Ti: Sapphire laser system. *J. Phys. IV France* **133**, 673–677 (2006).
- McKenna, P. et al. Effects of front surface plasma expansion on proton acceleration in ultraintense laser irradiation of foil targets. *Laser Part. Beams* **26**, 591–596 (2008).
- Gray, R. J. et al. Laser pulse propagation and enhanced energy coupling to fast electrons in dense plasma gradients. *New J. Phys.* **16**, 113075 (2014).
- Ter-Avetisyan, S. et al. Surface modulation and back reflection from foil targets irradiated by a Petawatt femtosecond laser pulse at oblique incidence. *Opt. Express* **24**, 28104–28112 (2016).
- Liu, B., Wei, Y., Zhang, Y. & Yang, Q. Deep Neural Networks for High Dimension, Low Sample Size Data. *Proceedings Twenty-Sixth*

- International Joint Conference Artificial Intelligence* **17**, 2287–2293 (2017).
33. Teague, M. R. Image analysis via the general theory of moments. *J. Opt. Soc. Am.* **70**, 920–930 (1980).
  34. Wu, J. F. & Peek, J. E. G. Predicting galaxy spectra from images with hybrid convolutional neural networks. Preprint at <https://arxiv.org/abs/2009.12318> (2020).
  35. Zhu, B. et al. Data-driven model for divertor plasma detachment prediction. *J. Plasma Phys.* **88**, (2022).
  36. Solera-Rico, A. et al.  $\beta$ -Variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nat. Commun.* **15**, 1361 (2024).
  37. Burgess, C. P. et al. Understanding disentangling in  $\beta$ -VAE. Preprint at <https://arxiv.org/abs/1804.03599> (2018).
  38. Kompa, B., Snoek, J. & Beam, A. L. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digit. Med.* **4**, 4 (2021).
  39. Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach. Learn.* **110**, 457–506 (2021).
  40. Nix, D. A. & Weigend, A. S. Estimating the mean and variance of the target probability distribution. *Proceedings 1994 IEEE International Conference Neural Networks (ICNN'94)* **1**, 55–60 (1994).
  41. Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Adv. Neural Inf. Process. Syst.* **30** (2017). Preprint at <https://arxiv.org/abs/1612.01474>.
  42. Kuleshov, V., Fenner, N. & Ermon, S. Accurate Uncertainties for Deep Learning Using Calibrated Regression. *Proceedings Machine Learning Research* **80**, 2801–2809 (2018).
  43. Gawlikowski, J. et al. A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.* **56**, 1513–1589 (2023).
  44. Verhaeghe, J. et al. Generalizable calibrated machine learning models for real-time atrial fibrillation risk prediction in ICU patients. *Int. J. Med. Inform.* **175**, 105086 (2023).
  45. Seligmann, F., Becker, P., Volpp, M. & Neumann, G. Beyond Deep Ensembles: A Large-Scale Evaluation of Bayesian Deep Learning under Distribution Shift. Preprint at <https://arxiv.org/abs/2306.12306> (2023).
  46. Gneiting, T., Balabdaoui, F. & Raftery, A. E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B* **69**, 243–268 (2007).
  47. Rosenfeld, J. S., Rosenfeld, A., Belinkov, Y. & Shavit, N. A. Constructive Prediction of the Generalization Error Across Scales. Preprint at <https://arxiv.org/abs/1909.12673> (2019).
  48. Hestness, J. et al. Deep Learning Scaling is Predictable, Empirically. Preprint at <https://arxiv.org/abs/1712.00409> (2017).
  49. Sorscher, B., Geirhos, R., Shekhar, S., Ganguli, S. & Morcos, A. S. Beyond neural scaling laws: beating power law scaling via data pruning. Preprint at <https://arxiv.org/abs/2206.14486> (2023).
  50. Fan, F. L., Xiong, J., Li, M. & Wang, G. On Interpretability of Artificial Neural Networks: A Survey. *IEEE Trans. Radiat. Plasma Med. Sci.* **5**, 741–760 (2021).
  51. Zhang, Y., Tiño, P., Leonardis, A. & Tang, K. A Survey on Neural Network Interpretability. *IEEE Trans. Emerg. Top. Comput. Intell.* **5**, 726–742 (2021).
  52. Dong, Y., Su, H., Zhu, J. & Zhang, B. Improving Interpretability of Deep Neural Networks with Semantic Information. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 975–983 (2017).
  53. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
  54. Mandler, H. & Weigand, B. Feature importance in neural networks as a means of interpretation for data-driven turbulence models. *Comput. Fluids* **265**, 105993 (2023).
  55. Lee, S. & Yoo, S. InterDILL: interpretable prediction of drug-induced liver injury through permutation feature importance and attention mechanism. *J. Cheminform.* **16**, 1 (2024).
  56. Fumagalli, F., Muschalik, M., Hüllermeier, E. & Hammer, B. Incremental permutation feature importance (iPFI): towards online explanations on data streams. *Mach. Learn.* **112**, 4863–4903 (2023).
  57. Robson, L. et al. Scaling of proton acceleration driven by petawatt-laser-plasma interactions. *Nat. Phys.* **3**, 58–62 (2007).
  58. Wilson, R. et al. Influence of spatial-intensity contrast in ultraintense laser-plasma interactions. *Sci. Rep.* **12**, 1910 (2022).
  59. Zimmer, M. et al. Demonstration of non-destructive and isotope-sensitive material analysis using a short-pulsed laser-driven epithermal neutron source. *Nat. Commun.* **13**, 1173 (2022).
  60. Tagasovska, N. & Lopez-Paz, D. Single-Model Uncertainties for Deep Learning. Preprint at <https://arxiv.org/abs/1811.00908>.
  61. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? Preprint at <https://arxiv.org/abs/1411.1792>.
  62. Hosny, K. M. M., Kassem, M. A. & Foad, M. M. Classification of skin lesions using transfer learning and augmentation with Alex-net. *PLoS One* **14**, e0217293 (2019).
  63. O’Shea, P., Kimmel, M., Gu, X. & Trebino, R. Highly simplified device for ultrashort-pulse measurement. *Opt. Lett.* **26**, 932–934 (2001).
  64. Carroll, D. C. et al. A modified Thomson parabola spectrometer for high resolution multi-MeV ion measurements—Application to laser-driven ion acceleration. *Nucl. Instrum. Methods Phys. Res. A* **620**, 23–27 (2010).
  65. Thomson, J. J. X. V. I. Rays of positive electricity. *Lond. Edinb. Phil. Mag* **21**, 225–249 (1911).
  66. Mróz, W. et al. Method of processing ion energy distributions using a Thomson parabola ion spectrograph with a microchannelplate image converter camera. *Rev. Sci. Instrum.* **71**, 1417–1420 (2000).
  67. Harres, K. et al. Development and calibration of a Thomson parabola with microchannel plate for the detection of laser-accelerated MeV ions. *Rev. Sci. Instrum.* **79**, 093306 (2008).
  68. Kumar, S., Lohiya, K. S. & Patil, N. Forecasting the Fury: A Deep Learning Approach to Predicting Cyclone Intensity. *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 617–622 (2023).
  69. Goller, M. *Human Agency at Work: An Active Approach towards Expertise Development* Ch. 7 (Springer 2017).
  70. Chapuis, O. & Dragicevic, P. Effects of motor scale, visual scale, and quantization on small target acquisition difficulty. *ACM Trans. Comput.-Hum. Interact.* **18**, 13 (2011).
  71. Hindle, A., Godfrey, M. W. & Holt, R. C. Reading beside the lines: Using indentation to rank revisions by complexity. *Sci. Comput. Program.* **74**, 414–429 (2009).
  72. Schmitz, G. P. J. & Aldrich, C. Neurofuzzy modeling of chemical process systems with ellipsoidal radial basis function neural networks and genetic algorithms. *Comput. Chem. Eng.* **22**, S1001–S1004 (1998).
  73. Tran, K. et al. Methods for comparing uncertainty quantifications for material property predictions. *Mach. Learn.: Sci. Technol.* **1**, 025006 (2020).
  74. Nowotarski, J. & Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew. Sustain. Energy Rev.* **81**, 1548–1568 (2018).
  75. Hong, T. et al. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **32**, 896–913 (2016).
  76. Yarvu, I. B. & Gündüz, S. Y. Predicting Myocardial Infarction Complications And Outcomes With Deep Learning. *Eskiteh. tech. univ. j. sci. tech. appl. sci. eng.* **23**, 184–194 (2022).
  77. Shanker, M., Hu, M. Y. & Hung, M. S. Effect of data standardization on neural network training. *Omega* **24**, 385–397 (1996).
  78. Sola, J. & Sevilla, J. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* **44**, 1464–1468 (1997).

79. Rybkin, O., Daniilidis, K. & Levine, S. Simple and Effective VAE Training with Calibrated Decoders. *Proceedings 38th International Conference Machine Learning Proceedings Machine Learning Research* **139**, 9179–9189 (2021).
80. Victoria, A. H. & Maragatham, G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol. Syst.* **12**, 217–223 (2021).
81. Joy, T. T., Rana, S., Gupta, S. & Venkatesh, S. et al. zHyperparameter tuning for big data using Bayesian optimisation. *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2574–2579 (2016).
82. Wu, J. et al. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *J. Electron. Sci. Technol.* **17**, 26–40 (2019).
83. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*, Ch. 7 (Springer 2009).
84. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2 (IJCAI'95)*, 1137–1143 (1995).
85. Breiman, L. & Spector, P. Submodel Selection and Evaluation in Regression. The X-Random Case. *Int. Stat. Rev.* **60**, 291–319 (1992).
86. Wilson, A. C., Roelofs, R., Stern, M., Srebro, N. & Recht, B. The Marginal Value of Adaptive Gradient Methods in Machine Learning. Preprint at <https://arxiv.org/abs/1705.08292> (2017).
87. Keskar, N. S. & Socher, R. Improving Generalization Performance by Switching from Adam to SGD. Preprint at <https://arxiv.org/abs/1712.07628> (2017).
88. Luo, L., Xiong, Y., Liu, Y. & Sun, X. Adaptive Gradient Methods with Dynamic Bound of Learning Rate. Preprint at <https://arxiv.org/abs/1902.09843> (2019).

### Acknowledgements

We acknowledge the support of staff at the Central Laser Facility at the Rutherford Appleton Laboratory, UK. This research is financially supported by EPSRC (grant numbers EP/V049232/1 and EP/P020607/1) and STFC (grant numbers ST/V001612/1 and ST/X005895/1). It involved the use of the ARCHER2 high-performance computer with access provided via the Plasma Physics HEC Consortia, Grant No. EP/X035336/1. Doctoral funding from EPSRC (EP/T517938/1) and STFC (ST/X508500/1) is gratefully acknowledged.

### Author contributions

C.J.G.M. and R.J.G. processed and analysed the experimental data, in addition to generating and testing the neural network architecture used. R.W., R.J.G. and P.M. conceived the experiment and C.J.G.M., M.K., M.A., E.F.J.B., E.J.D., T.D., J.K.P., M.P.P., B.C.T. participated in various aspects of it, with R.W. and T.P.F. leading the experiment; P.M. and R.J.G. provided overall supervision of the work; the manuscript was prepared by P.M., C.J.G.M., R.J.G. and M.K. with contributions from all authors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Paul McKenna.

**Peer review information** *Communications Physics* thanks Matthew Hill and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025