UNIVERSITY *of* York

This is a repository copy of *Building robust surrogate models of laser-plasma interactions using large scale PIC simulation*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/223889/

Version: Published Version

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

Building robust surrogate models of laser-plasma interactions using large scale PIC simulation

View the article online for updates and enhancements.

Building robust surrogate models of laser-plasma interactions using large scale PIC simulation

Nathan Smith[1],*⬤**, Kate Lancaster**[1]**, Stuart Morris**[2]**, Chris Arran**[1]⬤ **and Chris Ridgers**[1]⬤

[1] York Plasma Institute, University of York, York, United Kingdom
[2] Department of Physics, University of Warwick, Coventry, United Kingdom

E-mail: Nathan.Smith@york.ac.uk

CrossMark

Abstract
As the repetition rates of ultra-high intensity lasers increase, simulations used for the prediction of experimental results may need to be augmented with machine learning to keep up. In this paper, the usage of Gaussian process regression in producing surrogate models of laser-plasma interactions from particle-in-cell (PIC) simulations is investigated. Such a model retains the characteristic behaviour of the simulations but allows for faster on-demand results and estimation of statistical noise. A demonstrative model of Bremsstrahlung emission by hot electrons from a femtosecond timescale laser pulse in the $10^{20} - 10^{23}$ Wcm^{-2} intensity range is produced using 800 simulations of such a laser-solid interaction from 1D hybrid-PIC. While the simulations required 84 000 CPU-hours to generate, subsequent training occurs on the order of a minute on a single core and prediction takes only a fraction of a second. The model trained on this data is then compared against analytical expectations. The efficiency of training the model and its subsequent ability to distinguish types of noise within the data are analysed, and as a result error bounds on the model are defined.

Keywords: laser-plasma interactions, laser-solid interactions, Bremsstrahlung, machine learning, Gaussian process regression

1. Introduction & Background

With next-generation high-intensity, high-repetition rate laser facilities set to begin operation soon, multiple new avenues for research in high intensity laser-matter interactions and available secondary sources are beginning to open [1]. High repetition rates allow for vastly larger scale explorations of the available parameter spaces [2]. With both high intensity and high repetition rates, large parameter scans can be done to study unexplored regimes such as strong field QED [3–5], sophisticated manufacturing methods can be used in industry [6, 7], and production of secondary sources allows for imaging of other systems [8]. Alongside this, it would be beneficial to have quick, efficient models to both predict the outcomes of experiments using these lasers and find optimum sets of parameters for them.

The standard method of modelling high-intensity laser-plasma interactions is through particle-in-cell (PIC) and hybrid-PIC simulations, however these are typically too slow for *in-situ* modelling and have inherent variability due to statistical noise. The latter can be improved through convergence testing, however this further increases the computational cost of performing simulations. Because of this, we are interested in developing a surrogate model, where simulations are done across the relevant parameter space up-front, from which a model is created that can be used to estimate the result of a PIC

simulation, as well as an associated uncertainty on the result, at any set of parameters within that space. Important to note is that this is an interpolative system, and as such is limited to the region simulated, quickly giving results with large error bars as the input moves outside the region. Surrogate models such as this can be made with many different approaches, and are used in the many fields that require extensive use of computationally-intense simulation, namely engineering [9]. In the case of laser-plasma interactions, a surrogate model should (1) be far quicker than an equivalent PIC simulation of the system, (2) reliably interpolate a sparse dataset to produce accurate models and (3) produce reliable estimates of the error, both due to the statistical nature of the underlying simulations and due to the sparse sampling of the parameter space.

Previous work in this field has looked at using such methods for optimization and control of experiments, for instance in maximising the generation of synchrotron radiation in laser-solid interactions [10, 11], modelling the properties of electron bunches in laser wakefield acceleration [12] and calculating emission probabilities for the Breit–Wheeler process [13]. Our work focuses on generalising this approach using Gaussian processes, and from it developing a methodology to use in the future for approximate modelling of laser-plasma systems. We believe this methodology, especially following planned further development, is an elegant solution to the problems posed above, notably in the treatment of noise and uncertainty in the system.

In this paper we produce a surrogate model based on the results of large scale hybrid-PIC simulations of a laser-solid interaction, and evaluate this based on its ability to reproduce known results, the quantification of statistical uncertainty across the parameter space, and how these qualities vary with hyperparameters of the underlying simulations, namely resolution. To test the efficacy of the method, we looked at the physics of bremsstrahlung emission in a plastic target. This was chosen due to this setup having simple scaling laws for the total bremsstrahlung radiation generated allowing for easy comparison, as well as the comparative ease of simulating the system.

The rest of this section, 1.1, covers the physics behind bremsstrahlung emission and introduces some simple scaling laws which we expect our model to be able to reproduce. In section 2 we cover specifics of the system being looked at and how it is simulated, followed by how the resulting surrogate model is created through a Gaussian process regression (GPR) method. In section 3 we cover the overall features of the resultant dataset and the surrogate model produced from it. We also cover variations of this surrogate model resulting from changes to the simulation set-up, to analyze how robust this method to such changes and the statistical nature of PIC simulation. Finally in section 4 we look at the utility and efficiency of this approach to modelling laser-plasma interactions, and on how future work could expand on this to make use of the scheme in experiments of a similar nature.

1.1. Bremsstrahlung

Bremsstrahlung is the radiation produced by charged particles interacting with the electric fields of nuclei, and a full semi-classical description of the process can be found in [14]. In our case, bremsstrahlung comes in the form of x-rays emitted by accelerated electrons as they travel through and are decelerated by the atoms and ions in the target material. A fraction of the electrons at the front surface of the target will be accelerated by the laser pulse to much higher energies than the rest of the electron population and are known as 'hot' electrons. Each of these electrons, with mass m_e, have an energy ϵ that approximately follows an Boltzmann distribution with a mean given by

$$\langle \epsilon \rangle = a_0 m_e c^2, \quad a_0 = \frac{eE_0}{m_e c \omega_L}, \tag{1}$$

where a_0 is the normalized laser amplitude, E_0 is the peak electric field strength and ω_L is the angular frequency of the laser. Assuming a constant laser-to-electron conversion efficiency $\eta_{l \to e}$(this parameter generally varies slowly with the system parameters as well as surface geometry, typically reaching values on the order of $10\% - 30\%$ [15], but is kept constant at $\eta_{l \to e} = 0.3$ in our simulations), this means that the total number of hot electrons injected by a constant laser intensity I_0 incident on an area δA for a period τ is given by

$$N_e = \frac{I_0 \tau \eta_{l \to e} \delta A}{\langle \epsilon \rangle}. \tag{2}$$

At the laser intensities looked at in this study $a_0 \gg 1$ and so our hot electron population is relativistic. In such a situation, the differential radiation cross section for an electron travelling within a material with atomic number Z is given in CGS units by

$$\frac{d\chi}{d\omega} = \frac{16Z^2 e^6}{3m_e^2 c^5} \log \left(\frac{233}{Z^{1/3}} \right), \tag{3}$$

where ω is the frequency of the emitted photon and $\chi(\omega)$ is the sum of all possible cross sections for emitting a photon of frequency up to ω, multiplied by their respective radiated energy. In equation (3) we have assumed the screening of the electric field of the nuclei by their surrounding electrons to be complete, as in [14], and thus the cross section becomes constant with the frequency, which occurs for very high electron energies. It should be noted that this only holds for frequencies up to a certain value, ω_{max}, which approximately corresponds to the total energy of the electron, $\hbar \omega_{max} \simeq \epsilon = \gamma m_e c^2$.

We can integrate this cross section for all frequencies up to the maximum frequency and multiply the material number density n_i to find the energy radiated per unit distance travelled by the electron as

$$\frac{dE_{brems}}{dx} = n_i \int_0^{\omega_{max}} \frac{d\chi}{d\omega} d\omega. \tag{4}$$

In the case of both screened and unscreened cross sections, the result scales with the total energy of the electron. Finally, we can see that this emission will continue for as long as the electron travels through the material, and so we expect more Bremsstrahlung emission for large target depths. From this we can infer some simple scaling laws we expect any model we produce to follow.

$$\eta \propto n_i I_0^{1/2} d. \tag{5}$$

This simple model can be used to benchmark our final model to see if it matches expectations. Note that this does not take in to account other energy loss effects and simplifies some elements of the physics (such as the assumed total screening), and so the the actual scaling of our model with these parameters will not be exact. Additionally, equating the electron path length with target depth is not entirely accurate, as refluxing at the target edges will increase the electron path length, albeit with some energy loss at each scattering.

2. Methodology

Our study was performed in two stages; first, 800 simulations of a laser solid interaction were performed using the 1D version of Hybrid-EPOCH, a PIC code where the material and its cold electrons are treated as a background field rather than as a set of ion and electron macro-particles. This significantly speeds up computation times by allowing a lower resolution to be used and only needing to track the hot electron population. This code is described further in [16] and is an extension to the full-PIC code EPOCH described in [17]. This set was repeated with differing grid resolution. Secondly, these datasets were used with a Gaussian Process regression framework to create surrogate models, which are then evaluated based on the bremsstrahlung physics laid out in (1.1).

2.1. Simulation set-up

Our simulation setup consists of a 1D space filled with plastic, modelled by uniformly distributed Carbon and Hydrogen in a $1:2$ ratio, respectively. The size of this space is the depth of the target d, while the densities of the two material components sum to the total number density n_i. The incident laser is not directly simulated, instead electrons are injected into the simulation space directly with properties that are based on the given laser parameters. These electrons are initialized with energies following a Boltzmann distribution such that the probability density function of the electron energies, $p(\epsilon)$, is given by

$$p(\epsilon) = \frac{1}{\langle \epsilon \rangle} \exp\left(-\frac{\epsilon}{\langle \epsilon \rangle}\right), \tag{6}$$

where the mean electron energy $\langle \epsilon \rangle$ is equal to that given in equation (1). The definition of the normalized field amplitude can be rewritten as $a_0 = 8.5 \times 10^{-6} \sqrt{I_0 \lambda^2}$, where I_0 is the laser intensity, and λ is the laser wavelength. The incoming laser has a 800 nm wavelength. The number of electrons

Table 1. Variable simulation parameters, the limits used for them and the units of each.

Parameter	Minimum	Maximum	Units
Intensity	10^{20}	10^{23}	Wcm^{-2}
Pulse length	1	1000	fs
Target depth	1	100	μm
Number density	10^{28}	10^{30}	m^{-3}

injected per time step matches the laser's Gaussian-shaped temporal intensity profile, with starting and end cut-offs two standard deviations in both cases from the intensity peak. This standard deviation (i.e. the pulse timescale) is set as the random parameter τ, and so the laser time profile has a FWHM of 2.355τ. Due to the 1D nature of the simulations, no spatial profile is used for the laser/electron injection, with simulation properties assumed to be uniform in the unsimulated transverse dimensions by the code. Macroparticles are therefore weighted assuming that each 1D cell represents a 3D cell with transverse extents of 1 m.

In a typical laser plasma interaction, the hot electrons are ejected out of the rear of the target in greater numbers than ions. This results in a charge imbalance and thus creates an electric field at the rear of the target (called the sheath field) that accelerates ions and reflects many of the hot electrons back in to the target with some energy loss and scattering [18]. As this requires ions to be simulated, this effect can not be directly simulated in hybrid-EPOCH, so instead its effects are approximated by use of TNSA boundary conditions. These have three defined parameters; an escape energy, scattering energy loss, and scattering angle. The electric field strength, and as such the energy required to escape the target, is proportional to the hot electron temperature, and so the escape energy parameter is calculated as proportional to the mean electron energy, up to a constant κ_{esc}, which is defined as 1.5 in these simulations. The scattering energy loss is calculated in the same manner, but with a proportionality constant of $\kappa_{\mathrm{TNSA}} = 0.0027$. The scattering angle is kept at a constant 20°. These choices are based on previously done standard EPOCH simulations [16] and are discussed further later.

There are four parameters of interest that are varied from simulation to simulation. These are listed above in table 1. Each parameter is sampled from a log-uniform distribution between their listed maxima and minima. The hot electrons in each simulation emit high energy photons, the dynamics of which are frozen so that they do not leave the simulation. Additionally, while the simulation is only 1D, the photons and electrons have momentum calculated for all three spatial dimensions. The electron dynamics and Bremsstrahlung emission are simulated for 0.5 ns in all simulations.

Additionally, at each point sampled in parameter space the simulations were performed twice, once with 40 nm grid spacing and once with 100 nm spacing, in order to investigate the effect on the resulting surrogate model when using simulations with lower resolution and as such increased noise. The timestep used in these simulations is calculated as the time taken for light to traverse a single cell multiplied by 0.8,

and is kept constant throughout the simulation. As such the 40 nm and 100 nm simulations have timesteps of approximately 0.107 fs and 0.267 fs, respectively.

2.2. Modelling the results

The collected simulation data is used to fit a model via GPR. Gaussian processes are a generalisation of the Gaussian distribution, where functions themselves are sampled. This can be viewed as the limit of a multivariate gaussian distribution as the number of variables is taken to infinity, where each variable reflects the value of the function at some point. This effectively produces a probability distribution of possible functions. In this generalisation the covariance matrix of the distribution becomes what is known as a kernel function, $k(\mathbf{x}_i, \mathbf{x}_j)$, relating how correlated the value of a sampled function at a point $\mathbf{x}_i$, is to that at another point $\mathbf{x}_j$. The choice of this function will determine the form of any function sampled from the resultant Gaussian process. When datapoints are observed, a conditional distribution is calculated from the overall distribution (called the prior distribution in the manner of Bayesian inference) using the observed values, giving a new Gaussian process (i.e. the posterior) that accounts for the new observation. Furthermore, the hyperparameters of the kernel function are optimized using this observed data, varying them such that the log marginal likelihood of observing the set of datapoints is maximised.

In this study, Gaussian processes were used to model the variation of laser-to-Bremsstrahlung conversion efficiency across the parameter space for a simulated laser-solid interaction. We expect this to be a slowly varying function where values are locally correlated but with some amount of noise present. Because of these requirements our model made use of a square exponential function kernel with an added white noise kernel to account for the noise present in the simulations. The resulting kernel function between two points in parameter space $\mathbf{x_i}, \mathbf{x_j}$ is given by

$$k(\mathbf{x_i}, \mathbf{x_j}) = a^2 \exp\left(-\frac{|\mathbf{x_i} - \mathbf{x_j}|^2}{2l^2}\right) + w\delta(\mathbf{x_i}, \mathbf{x_j}), \quad (7)$$

where $\delta(a, b)$ is a function that equals unity when $a = b$ and zero otherwise, and a, l and w are the kernel hyperparameters, representing function magnitude, length scale and noise respectively. These hyperparameters are those optimized during fitting to find the Gaussian process which produces functions that best match the data. In addition to the kernel hyperparameters, an error on the training data can be defined with the parameter α. This acts similarly to the white kernel described above but is added solely for the correlation at training points, and is used in cases where target data has some uncertainty but the underlying function is not expected to be noisy.

The conversion efficiency, i.e. the total energy of Bremsstrahlung photons produced divided by the total input energy, is used as the output of the model. The model itself is logarithmic in nature, taking the logarithm of the input parameters and giving the total energy and relevant uncertainty in logarithmic form. This is due to the widely varying values of the parameters and resultant amount of Bremsstrahlung emission. The assumed uncertainty, α, on data points for fitting (as the true uncertainty is something we wish the model to produce) is tuned to prevent over fitting of the Gaussian process to the data. Below a critical value there is a trade off between uncertainty in the data points and that from the white kernel, while above the critical value the model overestimates the error in the simulations. Following training of the model, expectation values and the variance of our parameter of interest can be found at any point within the parameter space. Further information on the exact details of Gaussian processes can be found in [19].

Separate models are fitted for differing numbers of grid cells, differing numbers of data points and across many values of α in order to investigate the efficacy of the method.

3. Results

The Simulations were run over multiple parallel jobs on the Viking2 cluster. The number of processing cores used by each simulation varied based on the simulation parameters. In total the 40 nm set required 84 000 CPU-hours whilst the 100 nm set took 14 000-CPU hours. Once all simulations were completed, the total energy emitted as bremsstrahlung is calculated from the sum of photon energies multiplied by the weights of their respective particles. The total Laser energy is calculated assuming a uniform 1×1m spatial distribution for the laser pulse for consistency with how macroparticles are weighted by the 1D code. The ratio between these two values is then calculated as the Laser to Bremsstrahlung conversion efficiency. Figure 1(a) shows how this conversion efficiency varies across a reduced form of the parameter space.

3.1. Gaussian Process model

The dataset summarised in figure 1(a) was used to fit a GPR model. With a point uncertainty of $\alpha = 10^{-10}$ (effectively certain conversion efficiencies) and a training set of 800 simulation data points, the parameters of the optimized kernel function were found to be $a = 1.014, l = 2.377$ and $w = 1.21 \times 10^{-3}$. This training process took 82 s, with subsequent prediction of 10 000 datapoints requiring 0.7 s, both on a single CPU core. The convergence of these parameters as the number of simulation data points increases is shown in figure 2.

Additionally, how the accuracy of the model improves in this scenario has been investigated using an additional set of 80 simulations. These are within the same parameter space as the training data, and have been generated in the same manner using a cell size of 40 nm. The predictions of the Gaussian process at the parameter values of this test data are compared against the calculated simulation values using the coefficient of determination,

$$R^2 = 1 - \sum_i \frac{(\eta_i - \eta_i^{\text{true}})^2}{\sigma_{\text{true}}^2}, \quad (8)$$
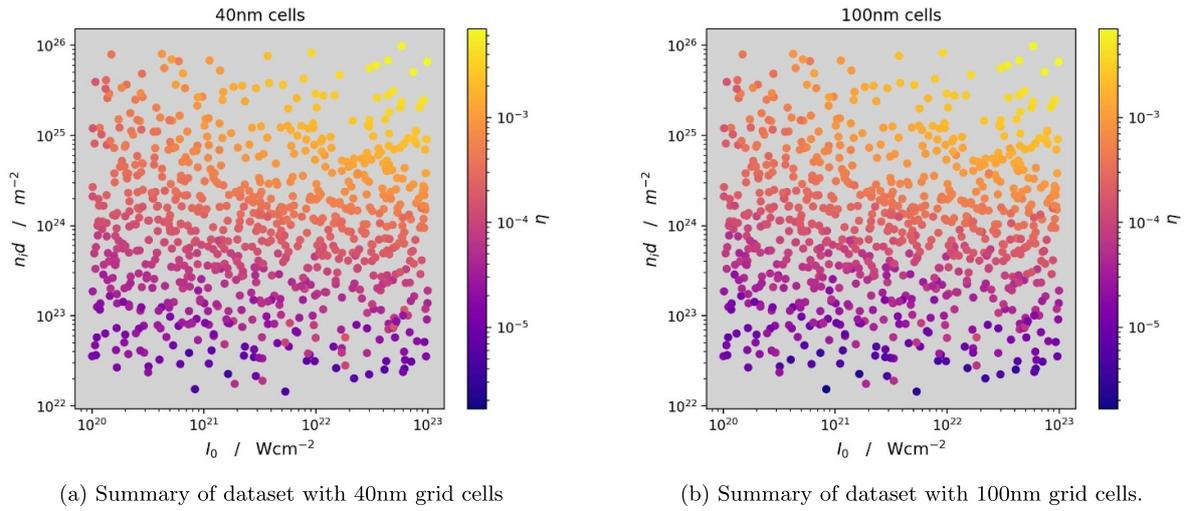
4

(a) Summary of dataset with 40nm grid cells   (b) Summary of dataset with 100nm grid cells.

Figure 1. Plots of the conversion efficiencies calculated at randomly sampled points across the simulation parameter space, for the same parameters in simulation with (a) 40 nm grid cells and (b) 100 nm grid cells. Pulse length is omitted from the plot axes as the conversion efficiency was found to not vary with this. Peak laser intensity is across the *x*-axis while the product of the target depth and target density is across the *y*-axis.
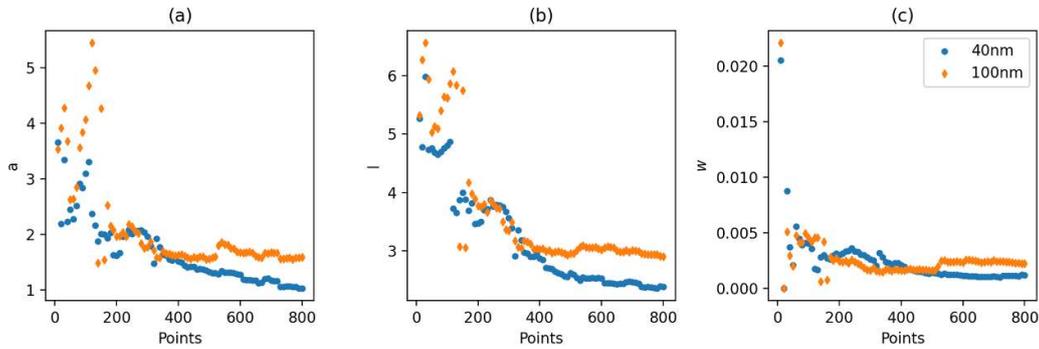

Figure 2. Convergence of the three kernel hyperparameters a, l and w in plots (a), (b) and (c), respectively.

where η_i and η_i^{true} are the predicted and simulation calculated conversion efficiencies for set of parameters i, respectively, and σ_{true}^2 is the variance of the simulation calculated conversion efficiencies. A result of $R^2 = 1$ indicates a model that perfectly predicts the simulation results, while $R^2 = 0$ indicates one which exclusively returns the mean value of the true data, with $R^2 < 0$ in the case of a model performing worse than this. The results as the number of training samples increases, given in the form of the proximity to 1 as a measure of the model error, are given in figure 3. It should be noted that this coefficient of determination is calculated directly from the model outputs, i.e. in logarithmic space, and so the values calculated may not completely reflect those in real space.

As the target uncertainty α is an undetermined parameter of the dataset and will have an impact on the calculated white noise of the model, a logarithmic scan of α is done to investigate how the resulting surrogate model changes. The optimum white noise kernel for each value of alpha is shown in figure 4. From this we can see that below a critical value, α and the noise present in the surrogate mode, δ, will sum to this critical value, which can as such be considered to be the 'true' noise of the dataset used. Above this value, the model could be considered

to be underfitting the data, however the mean prediction does not change significantly.

The full model outlined at the start of this section has been used to produce predictions of conversion efficiency for full scans of each of the four parameters. These are shown in figure 5, where each parameter is scanned over while the other three parameters in each instance are kept fixed. Without accounting for any other effects, the maximum possible conversion efficiency in the simulation set up is $\eta = 0.3$, i.e. the laser-to-electron efficiency, corresponding to all electrons emitting the entirety of their energy as bremsstrahlung. This will not be reached however due to multiple other effects. The nature of the TNSA boundaries used will mean approximately 22% of electrons will escape the simulation quickly, and some energy will be lost by all electrons upon reflection by the TNSA boundaries. Additionally, other energy loss mechanisms are dominant for low energies and so we expect much lower conversion efficiencies, especially for a small Z value.

The variation of conversion efficiency with the different parameters is quantified by fitting scaling laws to the mean prediction across a scan of the parameter of interest. This can only be done for the low-emission end of the scale, as

for three of the four parameters the conversion efficiency begins to tend to a maximum as the parameter increases. The calculated fits are shown alongside the sample scans in figure 5.

The behaviour of model with changing pulse length matches expectations; a longer pulse length just increases the number of electrons and total energy, without changing the electron dynamics greatly. Because of this, the conversion efficiency is mostly flat. For very small pulse lengths, the total number of injected electron macroparticles is comparatively low due to the pulse length approaching a similar scale to the simulation time step. As such statistical noise will increase, resulting in the high error in this region.

3.2. Variation with simulation hyperparameters

The model above is trained on simulations using 40 nm wide cells. To test how differing simulation resolutions change the results, the simulations were repeated at the same points in parameter space with grid cell sizes of 100 nm. The conversion efficiencies calculated for this are shown in figure 1(b). These are, generally, very similar to the efficiencies calculated for a smaller cell size. When a GPR model is fitted to this data, the kernel hyperparameters are found to have an optimum of $a = 1.260, l = 2.900$ and $w = 2.23 \times 10^{-3}$. From this we can see a relatively small change in the main shape parameters, with a more significant change in the noise parameter, with it increasing by a factor of approximately 2.34. The overall trend seen in the previous dataset as α is varied, that being that the white noise parameter and alpha sum to a constant value, is also seen here but with an appropriately larger 'true noise', as shown in figure 4. Convergence of the model however occurs at approximately the same number of total simulations, with differing parameters following convergence, as shown in figure 2. This seems to indicate that despite the increased noise, a consistent model can be still be reached quickly with lower resolution. Finally, as can be seen in figure 3, model accuracy experiences the same increase after approximately 10 simulations, however following this the calculated error in the model decreases much slower than in the higher resolution simulations.

4. Discussion

In the study of laser-plasma interactions, PIC codes are a ubiquitous tool for modelling the physics involved, with use in many different regimes and setups. However, they do not come without issues, namely the long simulation times required. This expensive computational aspect makes parameter scans impractical; A grid search of a four-dimensional parameter space, with 10 values across each dimension, results in 10 000 data points to simulate. This would be impractical, even in the case of 1D simulations. One other major issue is the noise of simulations. The number of simulated particles is vastly smaller than in a real interaction, and so due to the random nature of
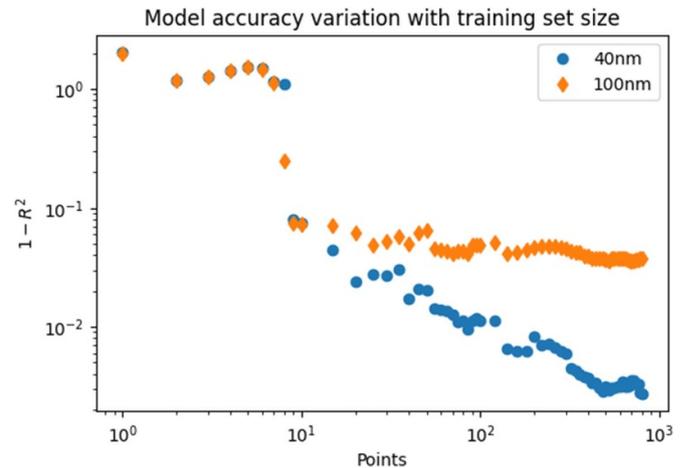

Figure 3. Variation of the model error, calculated as the difference between the coefficient of determination R^2 and its highest possible value, 1. The results for models from both datasets are included and it can be seen that, following an initial drop in error at approximately 10 points, the error decreases following a scaling law.

emission and other processes the electrons undergo, there will be some amount of noise in the results. This variation means that often simulations will need to be repeated multiple times in order to check if the simulation results are not just random noise, further increasing the computation time required.

As shown here however, it is possible to reduce the required data by a huge margin. This is due to the bayesian nature of the GPR model; an appropriate choice of prior vastly simplifies the model fitting process, and the model is optimized during this. Figure 2 show's that for the setup used in this work, this optimization process converges after approximately 400 data points. Additionally, accurate predictions can be made even early than this, as the models have high accuracy after very little data, as can be seen in figure 3. Furthermore, this is through a crude method; the data points are randomly sampled from the parameter space via a log-uniform distribution. This can be greatly improved upon within an 'active learning' scheme, where the next point to be sampled is chosen by finding the maximum value of an acquisition function. In our case where we are interested in exploring a parameter space, this function would simply be the model noise, such that uncertainty is minimized. In cases where optimization of a system is desired, the mean prediction can be used alongside the noise, with the latter weighted depending on how strongly exploration is wished for compared to exploitation.

The time savings of using a model such as this are significant; both datasets required thousands of CPU-hours to compute, while model training took on the order of a minute and prediction a fraction of a second. Because of this extreme time saving, models such as this are useful in cases where prediction is desired for a very large number of data points. One could potentially increase cell size further in order to reduce simulation time, however this can only be done so much before cell sizes begin to approach the size of the simulation space,
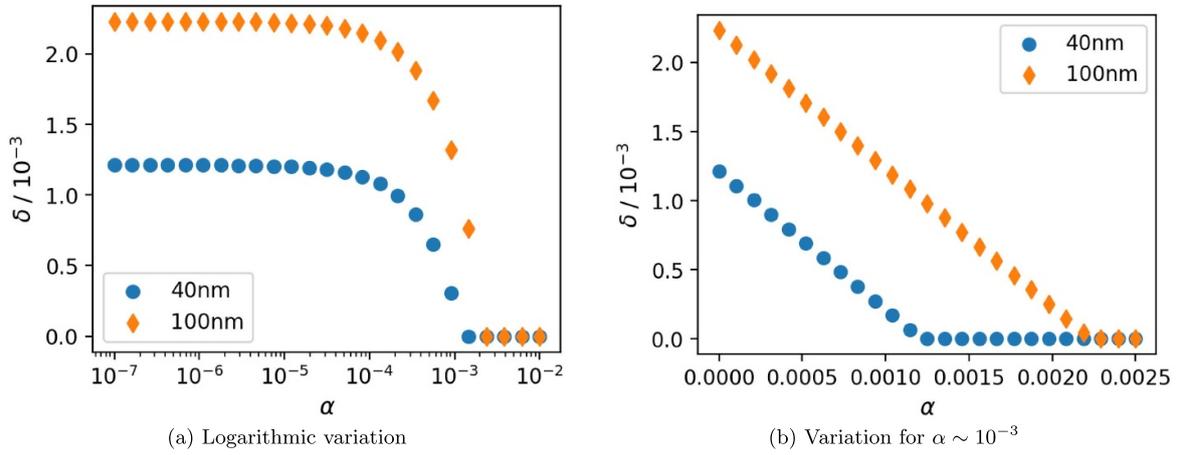
(a) Logarithmic variation

(b) Variation for $\alpha \sim 10^{-3}$

Figure 4. Variation of the optimal value of the white noise hyperparameter δ as the target data uncertainty is varied. The variation over different orders of magnitude is shown in (a), while in (b), which shows the variation across a linear axis, it can be seen that α and δ sum to a constant value, the 'true' noise of the model.

(a) Intensity

(b) Pulse duration

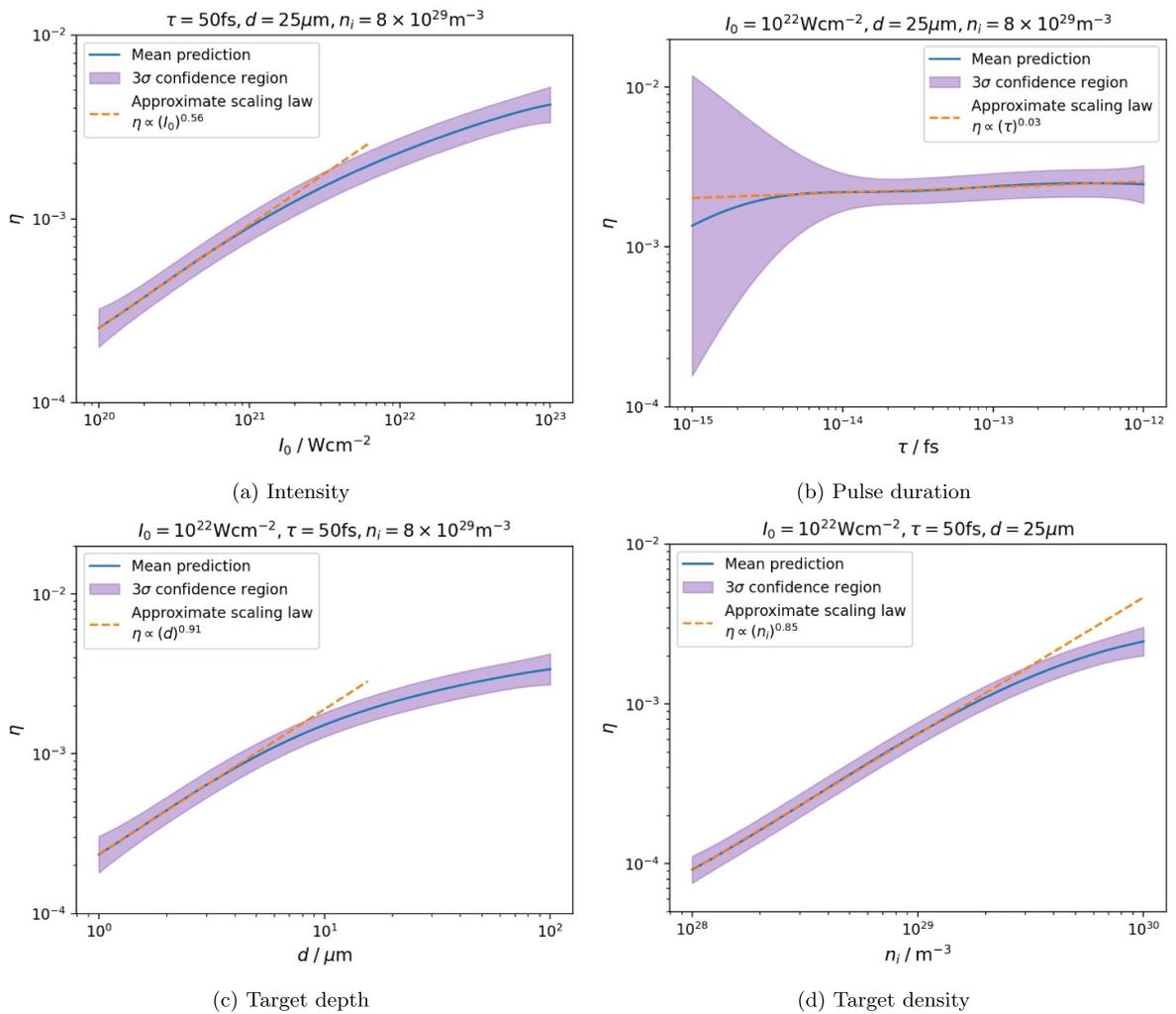(c) Target depth

(d) Target density

Figure 5. Example scans across the parameter space using the trained model, for each of the four variable parameters. Scaling laws are fitted to these scans to investigate how close the model is to the approximate analytical model.

or the time step approaches the timespan of electron injection. In both cases this will result in a poor approximation of the underlying system, and will still require significantly more computation time than our model, as even with a cell size of 1 micron, an average simulation in our dataset would take 10 CPU-minutes while poorly resolving the system.

The advantages of GPR also allow easy accounting for the noise of the simulations. As in our demonstration, the noise of the simulations can be directly quantified by including it within the GPR kernel. This additionally allows us to further refine our exploration method by subtracting the noise from the white kernel from the noise on a prediction, allowing a separation between the statistical noise of the simulation and the epistemic uncertainty resulting from a lack of sampling in a region. One matter to address with this however is that the current implementation considers the noise constant across parameter space. Already we can see this is not entirely accurate, as for small pulse lengths the uncertainty in results increases dramatically. This is likely due to the pulse length beginning to approach the value of the timestep, and as such the injection of electrons is poorly resolved. Variation of noise across the parameter space such as this could be accounted for in the model by using a white noise kernel that varies, preferably with this variation determined by hyperparameters that can be optimized.

Although the system looked at in this paper is relatively simple, it is a step towards a more complex framework for creating surrogate models from PIC data. Already it accomplishes the three requirements outlined in the introduction, those being quick evaluation, accurate interpolation and robustness to noise. It is a relatively simple matter to implement additional variable parameters in future models, especially in the case of using 2D PIC, where additional variables such as target width, spot size and incident angle become available. In these cases, the need for an active learning approach becomes necessary as computation times are increased even further in 2D and beyond. The computational efficiency of such methods can be improved further by using a more sophisticated method of choosing the initial data points than uniform random sampling, such as Latin hypercube sampling [20].

Some other major limitations of the method are also imposed as consequences of the simulations used. Some parts of the simulation may not match the behaviour of real physics, whether this be due to approximation or for a lack of empirical data on the physics in that region. In the former case we have the implementation of the TNSA boundaries, where a specific escape energy and energy loss is defined for each simulation as being proportional to the mean hot electron energy. The proportionality constant of this relation is the same in all simulations, however in previous work such as [16], this has been found to not be the case. Not including this variation changes the collective dynamics of the hot electrons and so will affect the bremsstrahlung result. One example of the latter issue is the physics of electron injection at the high end of our intensity range, where it hasn't been shown that the hot electron temperature follows the scaling used in our simulations and as such phenomena such as relativistic transparency could change this

behaviour. Once again, this will change the overall dynamics of the electrons and so our results at high intensities may not match real behaviour.

Finally, this method shows promise in improving some of the limitations of PIC codes, in particular the treatment of boundary behaviour in hybrid-PIC. While the escape energy and energy loss are parameterized as proportional to the mean energy of the injected electrons, the proportionality constant calculated from traditional PIC is dependent on the system parameters. The exact relationship however is not well quantified, and often, for instance in this investigation, the values of κ_{esc} and κ_{TNSA} are chosen somewhat arbitrarily. GPR can be used as a tool to better choose such values in the future, by using the methods outlined above to produce a surrogate model for the two values across the parameter space. This can then be used alongside the hybrid-PIC code in the future to better model the physics of electrons at the boundary.

5. Conclusion

In this paper, an investigation in to the use of GPR for producing surrogate models of laser-plasma interactions has been performed. As a simple example of this, simulations were performed in 1D hybrid-PIC of the emission of bremsstrahlung by hot electrons in a plastic target. The data from these simulations was used to train a surrogate model using GPR, using a square exponential kernel with added white noise. After fitting using the training data, this was then evaluated on how well it matches an analytical approximation of the bremsstrahlung emission. Other measures of the efficacy of the method were looked at, such as statistical noise and convergence of the kernel hyperparameters during training. Further improvements that will be made to the model are discussed, namely that of implementing active learning via an acquisition function, as well as potential use cases, such as building a surrogate model for the process of Target Normal Sheath Acceleration.

ORCID iDs

Nathan Smith ⬛ https://orcid.org/0000-0002-8991-8509
Chris Arran ⬛ https://orcid.org/0000-0002-8644-8118
Chris Ridgers ⬛ https://orcid.org/0000-0002-4078-0887

References

[1] Gales S and team E-N 2015 Laser driven nuclear science and applications: The need of high efficiency, high power and high repetition rate laser beams *Eur. Phys. J. Spec. Top.* **224** 2631

[2] Furch F J, Witting T, Osolodkov M, Schell F, Schulz C P and Vrakking M J J 2022 High power, high repetition rate laser-based sources for attosecond science *J. Phys. Photon.* **4** 032001

[3] Baumann C and Pukhov A 2019 Laser-solid interaction and its potential for probing radiative corrections in strong-field quantum electrodynamics *Plasma Phys. Control. Fusion* **61** 074010

[4] Fedeli L, Sainte-Marie A, Zaim N, Thévenet M, Vay J L, Myers A, Quéré F and Vincenti H 2021 Probing strong-field qed with doppler-boosted petawatt-class lasers *Phys. Rev. Lett.* **127** 114801

[5] Macleod A J 2022 From theory to precision modelling of strong-field qed in the transition regime *J. Phys.: Conf. Ser.* **2249** 012022

[6] Račiukaitis G 2021 Ultra-short pulse lasers for microfabrication: A review *IEEE J. Sel. Top. Quantum Electron.* **27** 1

[7] Kalaycioğlu H, Elahi P, Akçaalan I and i Ilday F 2018 High-repetition-rate ultrafast fiber lasers for material processing *IEEE J. Sel. Top. Quantum Electron.* **24** 8800312

[8] Hädrich S, Krebs M, Hoffmann A, Klenke A, Rothhardt J, Limpert J and Tünnermann A 2015 Exploring new avenues in high repetition rate table-top coherent extreme ultraviolet sources *Light Sci. Appl.* **4** e320

[9] Alizadeh R, Allen J K and Mistree F 2020 Managing computational complexity using surrogate models: a critical review *Res. Eng. Des.* **31** 275

[10] Goodman J, King M, Dolier E J, Wilson R, Gray R J and McKenna P 2023 Optimization and control of synchrotron emission in ultraintense laser-solid interactions using machine learning *High Power Laser Sci. Eng.* **11** e34

[11] Goodman J, King M, Dolier E J, Wilson R, Gray R J and McKenna P 2024 Optimization and control of synchrotron emission in ultraintense laser-solid interactions using machine learning - corrigendum *High Power Laser Sci. Eng.* **12** e12

[12] Streeter M J V *et al* 2023 Laser wakefield accelerator modelling with variational neural networks *High Power Laser Sci. Eng.* **11** e9

[13] Watt R A, Rose S J, Kettle B and Mangles S P D 2023 Monte Carlo modeling of the linear Breit-Wheeler process within the geant4 framework *Phys. Rev. Accel. Beams* **26** 054601

[14] Jackson J D 1999 *Classical Electrodynamics* 3rd edn (Wiley)

[15] Nilson P M *et al* 2010 Scaling hot-electron generation to high-power, kilojoule-class laser-solid interactions *Phys. Rev. Lett.* **105** 235001

[16] Morris S, Robinson A and Ridgers C 2021 Highly efficient conversion of laser energy to hard x-rays in high-intensity laser-solid simulations *Phys. Plasmas* **28** 103304

[17] Arber T D *et al* 2015 Contemporary particle-in-cell approach to laser-plasma modelling *Plasma Phys. Control. Fusion* **57** 113001

[18] Rusby D R, Armstrong C D, Scott G G, King M, McKenna P and Neely D 2019 Effect of rear surface fields on hot, refluxing and escaping electron populations via numerical simulations *High Power Laser Sci. Eng.* **7** e45

[19] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (*Adaptive Computation and Machine Learning*) (MIT Press)

[20] Loh W-L 1996 On Latin hypercube sampling *Ann. Stat.* **24** 2058