

# Machine learning for the prediction of stopping powers

William A. Parfitt, Richard B. Jackman\*

London Centre for Nanotechnology, and Department of Electronic and Electrical Engineering, University College London, 17-19 Gordon Street, London WC1H 0AH, UK

## ARTICLE INFO

**Keywords:**  
Machine learning  
Random forest  
Stopping power  
SRIM

## ABSTRACT

The stopping power of a material upon interaction with an energetic ion is the key measure of how far that ion will travel. The implications of accurate particle range calculations are tremendous, affecting every single application in which particle radiation is involved, from nuclear power to medicine. An approach is presented which attempts to overcome current shortcomings in the theoretical understanding of stopping power, as well as the methods used to interpret and exploit measured data. This is a considerable challenge, however the use of a novel machine learning methodology is shown to hold great promise in this endeavour: the ultimate aim being the ability to correctly predict the stopping value for any energy, ion and target combination, having no pre-existing experimental data.

A random forest regression algorithm is trained using over 34,000 experimental measurements, representing stopping power values for 522 ion-target combinations across the energy range  $10^{-3}$  to  $10^2$  MeV/amu, and ion and target atomic masses 1 to  $> 240$ . Evaluation is carried out using several fundamental error metrics, over the whole dataset as well as for individual combinations, to provide the most comprehensive understanding of performance when tested under strict cross-validation criteria. The resulting model is shown to yield predicted stopping power curves corresponding closely to those of the true experimental values, with an ability to generalise across target elements, compounds, mixtures, alloys and polymers, irrespective of phase, and for a wide range of ion masses.

## 1. Introduction

The understanding of particle interactions in matter has profound implications in physics and beyond. Particularly useful is the ability to calculate the range of energetic particles, which lose energy as they pass through a material. The energy loss  $dE$  of an ion traversing path length  $dx$  through a material is known as its **stopping power**  $S(E)$ ,

$$S(E) = -\frac{dE}{dx} \quad (1)$$

which may generally be resolved into three contributing components,

$$S(E) = S_{el}(E) + S_{rad}(E) + S_{nuc}(E) \quad (2)$$

the electronic stopping power,  $S_{el}(E)$ , due to inelastic collisions with atomic electrons causing ionisation and excitation; the radiative stopping power,  $S_{rad}(E)$ , attributed to the emission of Bremsstrahlung radiation; and the nuclear stopping power,  $S_{nuc}(E)$ , resulting from elastic electromagnetic collisions with nuclei in which recoil energy is lost to the atoms [1].

The radiative stopping power is proportional to  $(m_e/M)^2$  where  $m_e$  and  $M$  are the electron and ion rest masses, respectively, and is

considered negligible for particles other than electrons and positrons. The electronic stopping power is the dominant component for 'light' ions ( $Z \leq 10$ ) and may be approximately described by the Bethe theory [2], which itself is derived upon the assumptions of the first-order Born approximation. Deviations from this theory may be accounted for in a number of correction terms, including the Barkas correction [3,4], the Bloch correction [5], the shell correction [6], and the density effect correction [7,8].

For large particle energies the corrections are minimal, and the Bethe equation provides a sufficiently accurate prediction of stopping power. However, in the low energy regime (e.g. below  $\sim 2$  MeV for  $^4\text{He}$  [9]) the interpretation breaks down. The nuclear stopping contribution and adjustments such as the shell correction and charge shrouding effects become more significant.

There are methods for attempting to describe the stopping power where theoretical models are incomplete. A widely accepted approach is to use semi-empirical fitting formulas that provide adequate fits to existing experimental stopping power data for various incident ions and target materials; such as that proposed by Varelas and Biersack [10]. In particular, the principal parameters to this formula are provided by Ziegler [11–13], with other coefficients proposed by Watt [14]. These

\* Corresponding author.

E-mail addresses: [william.parfitt.13@ucl.ac.uk](mailto:william.parfitt.13@ucl.ac.uk) (W.A. Parfitt), [r.jackman@ucl.ac.uk](mailto:r.jackman@ucl.ac.uk) (R.B. Jackman).

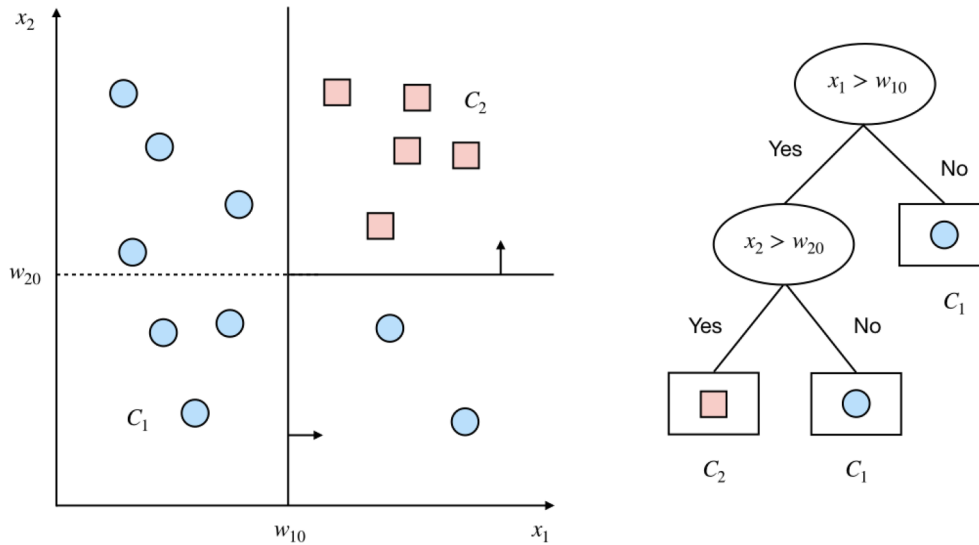


Fig. 1. An example of a simple decision tree. Ovals represent decision nodes, rectangles represent leaf nodes (adapted from [20]).

coefficients are compiled and published by the International Commission on Radiation Units and Measurements (ICRU) [9]. Additionally Powers [15] publishes some fitting parameters for a limited number of compound targets. However the majority of existing experimental data is provided only for elements, restricting the viability of fitting formulas for more complex material compositions. In this case, values may be approximated as a weighted linear combination of the stopping powers for their atomic constituents – known as the Bragg additivity rule. This suffers further errors arising from fact that stopping power is strongly influenced by chemical bonding of the atomic constituents, the phase of the material, and channelling effects. It is generally acknowledged [9] that improving the accuracy of the Bragg additivity rule by semi-empirical methods is better applied to interpreting existing data than predicting the stopping power of other materials.

The reality is that accurate stopping power data is crucial for applications in medical radiotherapy, particle research, space technology, nuclear energy generation and wider industry. Inaccuracies may generate large errors in particle range calculations, which are derived by integration of the stopping power with respect to energy. Whilst the present understanding of particle energy loss in matter is vast, it still does not allow for an accurate analytical solution to the stopping power function at all energies. There is a concurrent incompleteness of experimental data, and the parameterised approximations currently used to solve these problems are burdened with errors and the inability to generalise well to other ions and target materials. Ideally, data is required for a great number of elements, mixtures and compounds; for a range of incident ionising particles; and across a complete energy spectrum.

Machine learning (ML) methods are therefore perfectly placed to provide an alternative solution; using the current incomplete experimental datasets to predict unknown values. A novel approach to the calculation of stopping power is presented in this paper whereby an ML algorithm is trained and evaluated with the goal of achieving a complete model – capable of accurately predicting the stopping power values for any feasible incident ion, energy and target material.

## 2. Machine learning

The field of machine learning – using computers to learn from data, improve their algorithms and make intelligent decisions – has seen a significant uplift recently. This is the result of four main driving forces: faster and more affordable processing hardware; increased connectivity and communication between computers; greater quantities, storage and

access to data; and well-funded, extensive research efforts resulting in better and more efficient ML algorithms [16]. Consequently machine learning technology is now present in almost every aspect of modern life, from suggesting a movie to watch [17], diagnosing skin cancers in photographs [18], to making instantaneous stock market predictions [19].

Conventional computer programs require a user to have a deep understanding of a system in order to write a set of instructions that compute the desired outputs from a set of inputs. Instead, ML relies on the principle that what may be lacking in knowledge of a system can be made up for in data [20]. As such, if it is possible to compile a large number of examples, a ML algorithm may be ‘taught’ to detect statistical patterns and regularities in these examples to the extent that it is able to accurately make its own predictions for unseen data.

ML methods can be broadly classified by the type of data passed to the computer during training and the expected outputs of the trained model. *Supervised learning* occurs when a training sample consists of a set of input variables, also known as the *feature vector*, and the corresponding output to be predicted, known as the *label*. Given a dataset with several samples the algorithm learns to map the features to the labels, such that new labels can be predicted for unlabelled data. Other learning approaches include *unsupervised learning*, where labels are not provided, and the computer is instead tasked with clustering patterns or trends within the input data; and *reinforcement learning*, where the computer seeks to take actions within a defined environment to maximise a cumulative ‘reward’. In the case of supervised learning, the method may be further separated into the types of labels to be produced. These might be *classification* problems or *regression* problems, where the labels are qualitative or quantitative, respectively.

The type of problem informs the learning algorithms that are applicable to it. One leading supervised learning model used for both classification and regression problems is a decision tree ensemble, sometimes called a random forest (RF). A decision tree is a hierarchical model in which the output labels are determined through a series of recursive splits [20]. Each tree consists of several *decision nodes*. At each node  $m$  a test function  $f_m(x)$  is evaluated, where the discrete outcomes of the test form each of the resulting branches. For a given input vector, these tests are applied at each node until one of the *leaf nodes* is reached, at which point the value of this leaf is equivalent to the predicted output of the model, fig. 1. The test function aims to discriminate such that the error (however it may be defined) is minimised to the greatest possible extent at each split. A leaf node therefore defines a region of the input vector space where instances all have the same label

**Table 1**

A description of the terminology and quantities within the final ‘cleaned’ data.

Term	Meaning	Number
<i>dataset</i>	All cleaned experimental data	1
<i>ion</i>	Incident particle	32
<i>target</i>	Material under irradiation by the <i>ion</i>	97
<i>combination</i>	A unique <i>ion-target</i> pair	522
<i>sample</i>	Each observation (row) in the <i>dataset</i> , consisting of a single stopping power value given for a particular energy, <i>ion</i> and <i>target</i>	34,237

(for a classification problem) or similar numeric value (for regression problems) and the error is lowest. Individual decision trees, however, are very prone to overfitting the training data and do not generalise very well. To avoid this problem, multiple decision trees known as a random forest (RF) can be used [21,22]. RF relies on the principle of *bootstrap aggregation* in which a random sample of the training data is taken with replacement and used to fit each tree in the forest. The output prediction of each individual tree can then be averaged to produce the output of the ensemble overall, reducing susceptibility of the model to noise in the data. RF is further developed such that for each split in a tree a random subset of the input features are chosen, in order to de-correlate the trees.

### 3. Methods

#### 3.1. Objective

The objective of this work is to train and optimise a machine learning model that is capable of predicting mass stopping power (the stopping power per unit density). It should take inputs describing any incident ion: from protons  $Z = 1$  and alpha particles  $Z = 2$ , to heavy ions such as uranium  $Z = 92$ ; for any target material: pure elemental materials, complex crystalline, amorphous or polymeric compounds, and mixtures of varying composition; in any state: solid, liquid or gas. The model should then be able to generate predictions for stopping power both individually and across a continuous incident ion energy range.

Key to the utility of the method is the ability to make fully ‘blind’ predictions. That is, for a given ion-target combination the model should never have seen any prior experimental data when generating predictions. The model will be trained on a dataset comprising experimental data for numerous ion-target combinations and use complex learned feature relationships to infer the stopping power function for any new unseen combination. Second to this, the model should be generalisable to achieve consistent accuracy across the full range of ion-target combinations. Care must be taken to avoid overfitting to a particular subset or feature of the data. Weaknesses and limits, instances where the model fails to perform well or is biased, should be evaluated.

All data cleaning, pre-processing, modelling and evaluation is performed in Python 3.7 as part of the Anaconda 5.3 distribution.

#### 3.2. Dataset

The training data is comprised of distinct experimental observations in the form:

Ion, target, energy  $\rightarrow$  measured stopping power

Over several decades these measurements have been made by numerous independent authors and subsequently published in the literature. The data is, in general, publicly-available in the form of journal publications and industry reports. Nevertheless, in the first instance, it is presented in a sparse and non-standardised format: embedded within text documents, written in multiple languages, quoted in different units,

varying error levels and precision. A unique and free-to-access database is available on the website of the International Atomic Energy Agency (IAEA), Nuclear Data Services [23]. This prodigious resource was pioneered by Professor Helmut Paul, University of Linz, transferred to the IAEA in 2015 and is currently under the collaborative management of Dr. Claudia Montanari [24]. The database makes available compilations of experimental measurements starting from 1928 to present. These cover a broad energy spectrum from eV to GeV, for ions in the range  $Z = 1$  to  $Z = 92$ , and approximately one hundred target elements, compounds and mixtures. The data is accessible in various forms, of which the file ‘SCSDData.zip’ was downloaded for this work and contains historical measurements taken up to September 2015.

However the data in this form is, to a degree, still non-standard and inconsistent – particularly for ML purposes. The files are therefore prepared, both manually and using automatic scripts such that the final dataset structure is self-consistent, and where energy and mass stopping power are in units of [MeV/amu] and [MeV. cm<sup>2</sup>. g<sup>-1</sup>], respectively. The process is considered to be an efficient solution for handling the large database but has the inevitable downside in that a portion of potentially useful training data is lost, including more recent observations. This could be extracted and processed to improve the future performance of the model. The ‘cleaned’ dataset is as described in Table 1, with associated terminology to be used here on.

#### 3.3. Features

This data is then further processed to make it suitable for training a ML model. This takes the form of feature extraction and pre-processing. Feature extraction involves extending the dataset to contain relevant input variables that may be predictive of the stopping power. By composing a supplementary dataset of fundamental ion and target material parameters, the following features are considered: the atomic mass of the incident ion; the atomic number of the incident ion; the relative atomic (or molecular) mass of the target material; the state of the target material *one-hot* encoded as three binary vectors for solid, liquid or gas; the normalised fractional composition of the target material, encoded as a sparse matrix where each column represents an element in the periodic table and the values contained in the columns indicate the relative fractional composition of the target materials (eg. for pure aluminium, the column representing ‘Al’ will contain a value of 1.0 – whereas for aluminium oxide, Al<sub>2</sub>O<sub>3</sub>, the column ‘Al’ will contain a value of 0.4). Other features generated include the mean ionisation energy of the target material, the mean density of the target material, and a binary vector indicating whether the target material is polymeric or not. The final predictor of the stopping power is the incident ion energy per unit mass [MeV/amu] given by the experimental dataset. Equivalently, the value (label) to be predicted is the experimental stopping power.

#### 3.4. Model

The decision tree machine learning model – ‘RandomForestRegressor’ – is provided as part of the Scikit-Learn library in Python, version 0.19.1 [25–27]. During training, each decision tree is built from a number of samples of the input training data drawn with replacement, to produce a bootstrapped ensemble. A random subset of the features are used at each split – according to the principle of a random forest. The sizes of the trees are left unrestricted, that is: the minimum number of samples required for a split to occur is set to 2; the trees are fully expanded until all the leaves are ‘pure’, i.e. they cannot be split any further, regardless of tree depth; and the minimum number of samples required at each terminal leaf node is set to 1. The number of trees used is 100; any increase in this number showed only a marginal improvement in model performance at the expense of a much greater computing time. All other hyperparameters are set to their default values including the splitting criterion, which is set to optimise

for mean squared error (MSE). Training is only performed on a single processing thread, although this may be parallelised for speed improvements if required. The predicted output is calculated as the mean of the outputs of each tree in the ensemble. Through training on the features provided in the dataset and obtaining the best splits in each tree, the model learns to calculate outputs for any new set of input features.

### 3.5. Error metrics

Four error metrics are used to quantify the performance of the model in terms of the predicted values  $y_{pred}$  against the true experimental values,  $y_{true}$ . The first is the coefficient of determination,  $R^2$ ,

$$R^2 = 1 - \frac{\sum (y_{true} - y_{pred})^2}{\sum (y_{true} - \bar{y}_{true})^2} \quad (3)$$

where the numerator of the fraction describes the size of the residuals (difference between predicted and true values) produced by the model; and the denominator describes the size of the residuals for a naive model in which the regression predicts the average value,  $\bar{y}_{true}$  [28]. The value of  $R^2$  is a widely used metric to describe the ‘goodness of fit’ of a regression model to the observed data, in which a perfect model achieves a score of  $R^2 = 1$ . An arguably better metric for the usefulness of a model is the root-mean-squared-error (RMSE),

$$RMSE = \sqrt{\frac{\sum (y_{true} - y_{pred})^2}{n}} \quad (4)$$

where the square root is taken of the sum of squared residuals, divided by the number of samples,  $n$ . The value of the RMSE is therefore expressed in the units of the quantity being estimated, in this case the stopping power, and therefore provides a more tangible accuracy metric. The RMSE represents an estimate for the standard deviation of the model (for an unbiased model it is equal to the standard deviation), which scales with the magnitude of the stopping power, and hence smaller values indicate a better model. The fact that the square of the residuals is summed means that it penalises large errors to a much greater extent. To avoid this, in the case where independent large errors are not relatively much worse than the overall errors, the mean-absolute-error (MAE) may be evaluated,

$$MAE = \frac{1}{n} \sum |y_{true} - y_{pred}| \quad (5)$$

which measures the absolute (unbiased) sum of residuals. Finally, the most intuitive – but also fundamentally limited – metric of error is the mean-absolute-percentage-error (MAPE),

$$MAPE = \frac{100}{n} \sum \left| \frac{y_{true} - y_{pred}}{y_{true}} \right| \quad (6)$$

in which the percentage deviation of the residuals is calculated. This metric is easily comparable with other work that quotes errors in terms of percentages, such as the ICRU reports [1,9], however suffers from several drawbacks. The first is that a ‘divide-by-zero’ error may be encountered if the true value is zero (not an issue in this instance). The most significant, however, is that there is no upper bound on the percentage error for forecasts that are too high. Hence model optimisation based upon the MAPE metric will systematically favour models that make too-low predictions over those that are too high.

### 3.6. Model evaluation

Two evaluations of model performance are sought. The first is the ability of the model to fit to the *training data* – i.e. the samples in the experimental dataset that are used to train the model. This is achieved by training the model on the full dataset and then using the trained model to make predictions for each sample in the full dataset. The error

metrics described above are calculated to indicate how well the predicted and true experimental values align for each combination, in addition to how well the model fits onto the dataset as a whole.

This evaluation is an assessment of the fundamental internal performance of the model but not indicative of its external predictivity on new data that has not been used in training. Hence the second evaluation is on that of the predictions against new and ‘unseen’ data. In lieu of acquiring new experimental observations in a laboratory, a more practical (and statistically meticulous) method is to remove, or *hold-out*, a portion of the dataset known as the *test data*. The model is trained on the remainder of the data, and then asked to make blind predictions on the test set.

A rigorous method by which this can be implemented is *k*-fold cross-validation (CV). This a process in which the dataset is divided into *k* partitions. Of these partitions, *k*–1 are used to train the model with the remaining partition held-out for testing the model. This process is repeated *k* times as the hold-out sample is rotated through the full dataset, with the model retrained each time on the remainder, until all the data has been used both for both training and testing. The total error of the model is given as the mean average of the error metric evaluated for each hold-out sample.

The ultimate benefit of the model lies in its ability to predict the full energy range stopping power for ion-target combinations with no existing experimental data. Hence to comply with this strict condition, the dataset is divided into folds that each contain all samples for a unique combination, such that  $k = 522$ .

## 4. Results

### 4.1. Dataset analysis

A fundamental problem when attempting to build a generalisable machine learning model is unbalanced training data. Training data that under or over-represents certain samples and features will, in the best case, result in a model that is better at some prediction tasks than others – and in the worst case, produce a model that heavily overfits to the nuances of the training data and is unable to generalise at all. It is therefore important to understand the dominant patterns within the dataset before training such that these imbalances can be accounted for. Unfortunately, any database that is made up of real-world observations will contain inherent sampling biases. In the case of the stopping power dataset, experimental physicists are incentivised to make measurements of ions and targets that are useful to their specific field, and less so to improve the continuity of the existing knowledge base which may otherwise be scientifically irrelevant. Certain combinations will attract intense study and hence be indulged with many hundreds of observations by many authors. Others will be less so and may only contain a single observation by a lone author. Fig. 2 is a representation of the distribution of samples within the dataset. The black squares represent a unique ion-target combination, in which the ion has atomic mass  $A_{ion}$  and the target has relative atomic mass  $A_{target}$ . The histograms represent the relative distribution of samples within the dataset for both ions and for targets, where the total area under each histogram is equal to the total number of samples, 34,237.

There is a noticeably greater density of combinations containing low mass ions: data exists for every ion up to  $Z = 22$ . Above that, the dataset represents only a fraction of the heavier ions up to  $Z = 92$ . The number of samples is heavily weighted towards the lighter ions, with hydrogen ( $Z = 1$ ) and helium ( $Z = 2$ ) the most commonly observed. The range of target materials favours lower relative atomic masses and lighter incident ions, although the distribution of samples is more continuous. Notably, the largest number of samples are recorded for gold ( $A \sim 197$ ), aluminium ( $A \sim 27$ ) and silver ( $A \sim 108$ ). The bias of the dataset towards the lighter combinations is unsurprising considering the relative abundance and practicality of lighter ions and target materials. On average each combination contains  $\sim 66$  observations. However, the



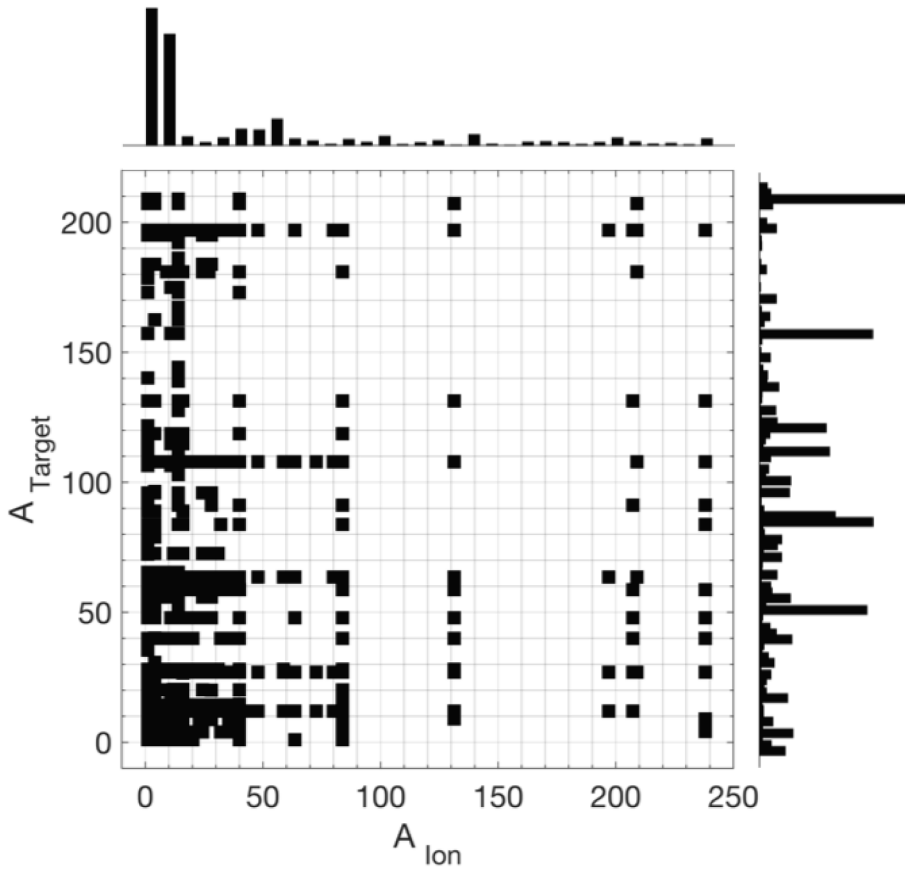


Fig. 2. Distribution of combinations within the dataset by (relative) atomic mass, one square represents one unique combination out of 522. There are 32 ions along the x-axis and 97 targets on the y-axis. The histograms represent the total number of samples for each ion or target mass, a larger bar indicates more samples at that mass. For clarity, histogram x-axis not to scale.

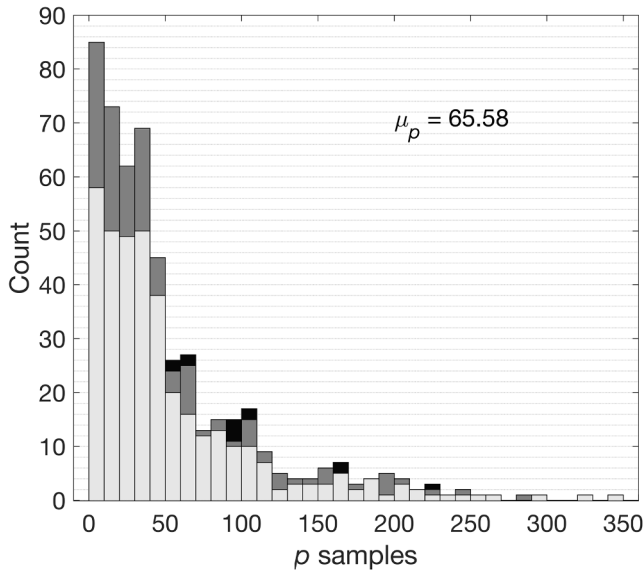


Fig. 3. Number of samples,  $p$ , per combination. Of which solid (light grey), gaseous (dark grey), liquid (black) target materials. Mean number of samples for any single combination is  $\sim 66$ . For clarity x-axis is truncated – not shown are individual bin counts for 10 combinations ranging up to 1078 samples.

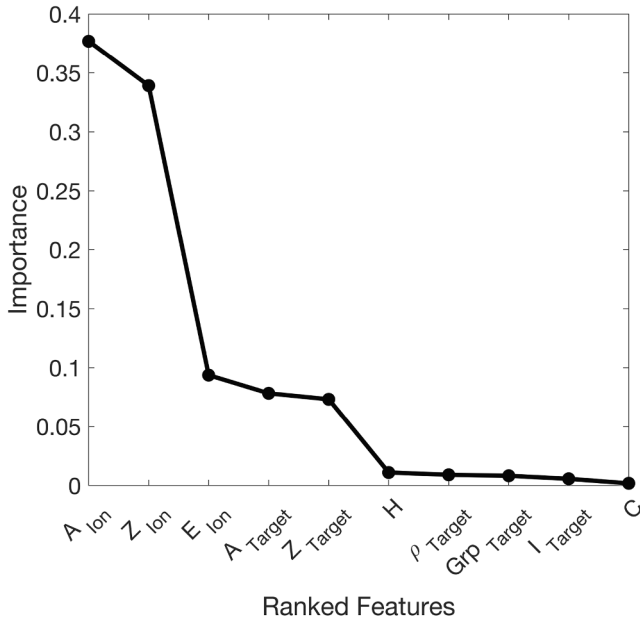
distribution of the number of samples  $p$ , Fig. 3, is heavily skewed towards small numbers of observations for each combination. The modal value lies between 1 and 10 stopping power observations, likely contributed from a single study and not covering the full energy range. There also exist some examples with much greater numbers of measurements, including 10 combinations that are not shown in Fig. 3 with stopping power observations taken across a wide energy range totalling

between 350 to over 1000 samples. 76.6% of the dataset is comprised of solid target materials (light grey bars), with the remaining 21.0% gases (dark grey bars) and 2.4% liquids (black bars).

#### 4.2. Feature importance

The relative importance of the features is calculated by the mean decrease in impurity (MDI) [29] – defined as the total decrease in node impurity, weighted by the number of samples reaching the node, averaged over all the trees in the forest. MDI gives an understanding of how much each feature influences the successful performance of the model and can allow those most useful to be selected for. As a secondary consequence, feature importance may also yield interesting insights into real-world, physical, correlations. However, it is important to note that the objective of the model is that of optimising predictions, and not of drawing meaningful inferential conclusions.

Fig. 4 shows the ranked importance for the top ten features, out of a total of 121. The first five have an overwhelming influence on the model, namely, the atomic numbers and masses of the ion and target, as well as the ion energy per unit mass. On the other hand, target density has little impact. These relationships might be expected when considering the form of theoretical descriptions of the mass stopping power. Interestingly, the mean ionisation energy is much less important, as is the physical state of the target material, and whether it is polymeric. The fractional target composition, comprising 109 elemental vectors, is also mostly irrelevant. The presence of these features apply individually to each target material – hence they are not key predictors overall. It should be noted that both hydrogen and carbon score higher than all other elements, which may either signify that the presence of these elements – or their various chemical groups – may more strongly influence the stopping power. More likely is that there is a selectivity attributed to the model through a higher number of hydrocarbon observations listed in the dataset. This would indeed mean that the model



**Fig. 4.** The relative importance of features listed for the top ten features: ion atomic mass,  $A_{ion}$ ; ion atomic number,  $Z_{ion}$ ; ion energy,  $E_{ion}$ ; target (relative) atomic mass,  $A_{target}$ ; target (relative) atomic number,  $Z_{target}$ ; hydrogen fraction,  $H$ ; target density,  $\rho_{target}$ ; target periodic group,  $Grp_{target}$ ; target mean ionisation energy,  $I_{target}$ ; carbon fraction,  $C$ .

is performing more successfully on these materials but it has no physical correlation to the stopping power.

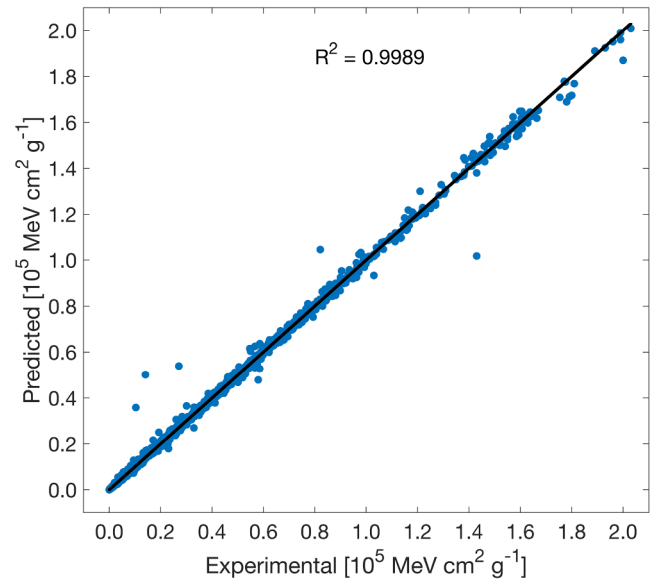
#### 4.3. Performance on training data

The random forest regression model is trained by supervised learning on the full dataset. Each sample is represented by the vector of features described above and mapped to its associated experimental stopping power value. After training, predictions are made for each sample of the dataset to assess how well the model fits to the training data it has seen. The quality of a model can be assessed by comparing the model predictions against the original experimental stopping power values, Fig. 5. Very good correlation is observed, and nearly all points align with the black line that represents a perfect predicted outcome. The calculated value of  $R^2 = 0.9989$  indicates a near-ideal fit to the 34,237 samples in the training data across the entire range of stopping powers. The calculated error metrics are displayed in Table 2.

Errors are also calculated individually for each combination. The frequency distributions of these errors are shown in Fig. 6, along with their mean values. These values are different to those presented in Table 2 because of their differing methodologies in derivation. Both the dataset as a whole, and the individual combinations, display low RMSE and MAE metrics in relation to the total variation in the stopping power values – from 0 to over 200,000 MeV. cm<sup>2</sup>. g<sup>-1</sup>. Additionally MAPE is slightly greater than 2% across the full range of values. It can therefore be assumed that both the quantity of data and the architecture of the model is sufficient to identify accurate relationships between the input features and stopping power values. This is indicative of a low bias model that is not underfitting the training data.

#### 4.4. Performance on test data

The true test of a model, and essential to its ultimate advantage, is to generate predictions for unseen test data that has not been trained on. This is carried out using  $k$ -fold CV for each ion-target combination. Fig. 7 plots the predicted values against experimental observations where each combination, in turn, has been held-out for test data. The



**Fig. 5.** Predicted values for all samples plotted against their ‘true’ experimental values. Predictions are made by model on the training data. High  $R^2$  value and small deviations from the ‘ideal’ model (black line) are observed, indicating that the model fits very well to the dataset.

**Table 2**

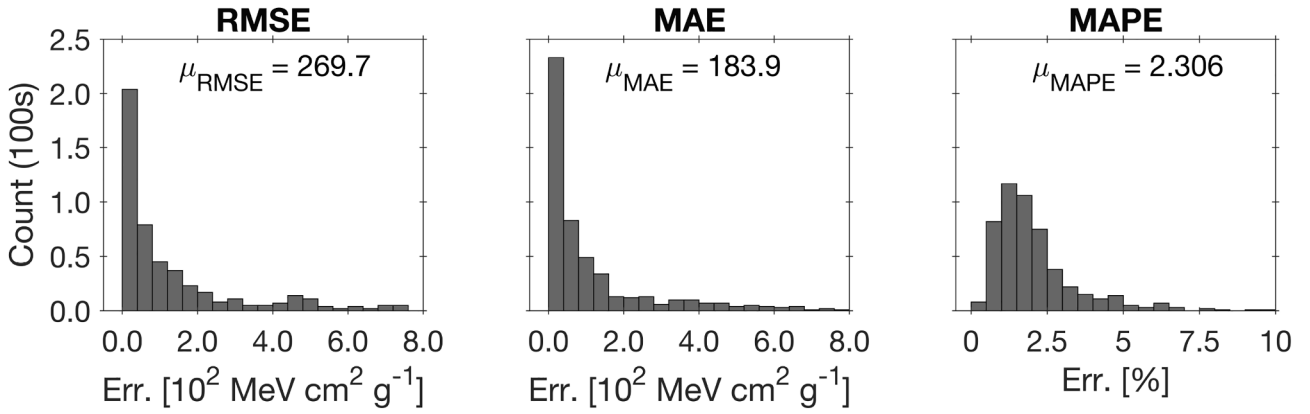
Error metrics calculated on all samples in dataset, to 4 significant figures.

$R^2$	RMSE [MeV. cm <sup>2</sup> . g <sup>-1</sup> ]	MAE [MeV. cm <sup>2</sup> . g <sup>-1</sup> ]	MAPE [%]
0.9989	492.7	86.94	2.273

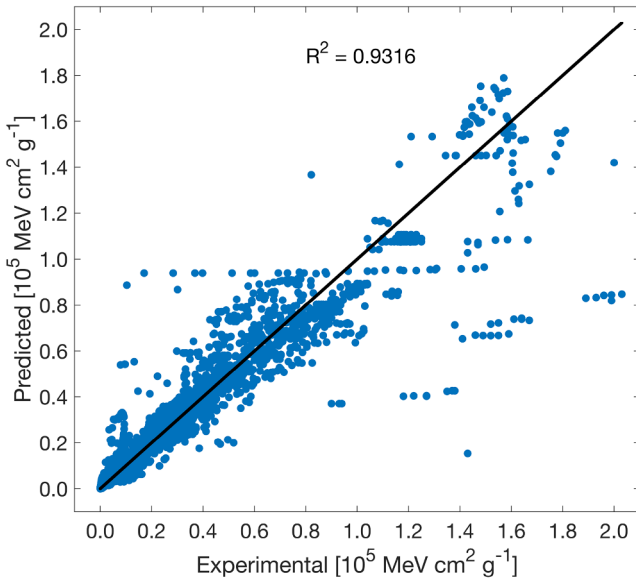
predictions are therefore ‘blind’ and rely on the ability of the model to generalise well to new data in order to achieve low error. Overall, the predictions align with the true experimental values.  $R^2 = 0.9316$  suggests that the model is predicting the test data well. It is immediately clear that the majority of stopping power values within the dataset are distributed towards the lower end of the range. In this region, representing stopping powers for light ions incident on heavy targets, the model appears to make predictions very near to the ‘ideal’ values (black line). For higher stopping powers, associated with heavy ions incident on light target materials, the model is unable to make sufficiently good predictions and the spread increases accordingly. This is possibly a consequence of the lesser quantity training data for higher stopping power values. Interestingly, a line of identical predicted values can be resolved at approximately 100,000 MeV. cm<sup>2</sup>. g<sup>-1</sup>, suggesting a possible overfit to a feature in the training data. There are a few additional clusters of high-error predictions that may also be attributed to this.

The  $R^2$  value can be considered a measure of the proportion of the variation in the predicted values that can be described by the model [28]. However the value of the model lies in its overall error – not in how well it explains dataset variability. Hence RMSE, although not a metric for the success of the dataset fitting process itself, is a better indicator of the usefulness of the model. Table 3 gives the calculated error metrics on all the predictions as a whole. Given the broad range of stopping powers, up to 200,000 MeV. cm<sup>2</sup>. g<sup>-1</sup>, an RMSE = 3,944 MeV. cm<sup>2</sup>. g<sup>-1</sup> is acceptable. The MAE = 814.8 MeV. cm<sup>2</sup>. g<sup>-1</sup> is several times lower and would suggest that the higher RMSE is due to a small number of very large prediction errors – possibly caused by overfitting or by the outliers at high stopping power values. The MAPE could intuitively be interpreted as an overall 85% ‘accuracy’, with this accuracy expected to be greater for smaller stopping powers.

Analysis of the error metrics for individual ion-target combinations can provide greater insight into the limits of the model. Fig. 8 shows the distribution of errors calculated for each combination, along with the



**Fig. 6.** Error metrics calculated for each individual combination. Histograms show distribution of error magnitudes across all combinations in the dataset. The mean error values,  $\mu$ , are calculated across all combinations, and therefore differ from those values calculated for all samples.



**Fig. 7.** Predicted values for all samples plotted against their ‘true’ experimental values. Predictions are made by model on the unseen test data. An imbalance of data distribution can be observed. Better predictions are made for lower stopping power values, which are better represented in the dataset. Additionally, linear clustering, possibly caused by overfitting, reduces the predictive quality of the model.

**Table 3**

Error metrics calculated on all samples in dataset, to 4 significant figures.

$R^2$	RMSE [MeV. cm <sup>2</sup> . g <sup>-1</sup> ]	MAE [MeV. cm <sup>2</sup> . g <sup>-1</sup> ]	MAPE [%]
0.9316	3944	814.8	15.03

mean values. Only the lower end of the distributions are displayed for clarity, obviating the need to show a long tail of single-count bins. Since the RMSE and MAE scale in proportion to the magnitude of the stopping power, it is expected that the distribution of these errors will be in proportion with the stopping powers in the dataset. Correspondingly, the greatest number of counts is seen in the smallest bin (0–100 MeV. cm<sup>2</sup>. g<sup>-1</sup>), and thereafter the majority of combinations have an RMSE and MAE below 1000 MeV. cm<sup>2</sup>. g<sup>-1</sup>. The MAPE has a modal value between 4–8%, with a long tail (truncated for clarity) consisting of outliers with much higher errors.

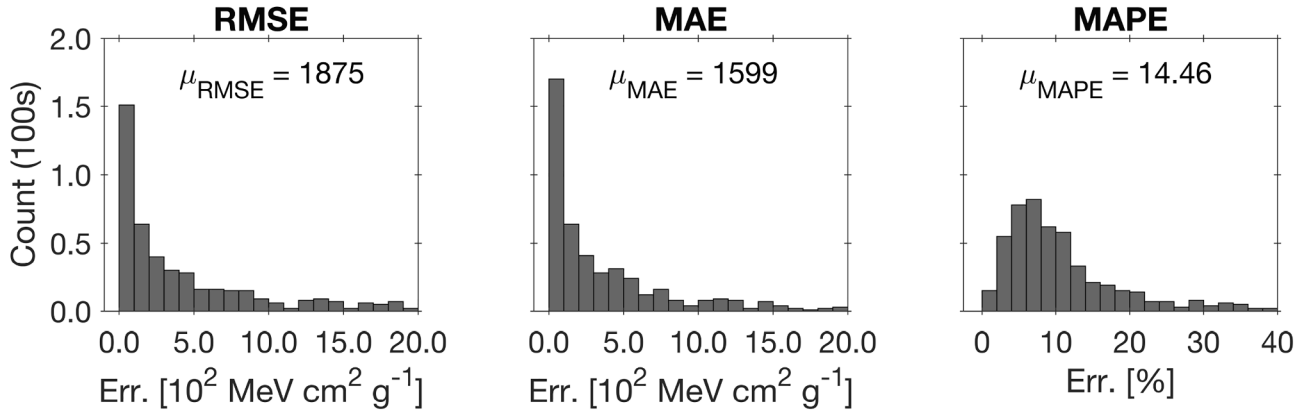
It is useful to visualise the source of these outlier values and identify boundaries within which the model performs the best. These boundaries should inform how the model may be improved to generalise

better. Fig. 9 shows a plot of the ion and target masses for each combination,  $A_{Ion}$  and  $A_{Target}$ , respectively, against their RMSE error. Colours represent RMSE errors below 5,000 MeV. cm<sup>2</sup>. g<sup>-1</sup> (green), between 5,000 and 15,000 MeV. cm<sup>2</sup>. g<sup>-1</sup> (yellow) and above 15,000 MeV. cm<sup>2</sup>. g<sup>-1</sup> (red). Several patterns can be identified. Firstly, the model performs best on the high number of combinations that contain lower-mass ions. This would undoubtedly be a consequence of a greater amount of training data in this regime, although this may perhaps not be the only reason. Secondly, the highest errors occur along the rear, right-hand face of the graph corresponding to light target materials. In particular, hydrogen and helium targets are very poorly predicted, especially at higher ion masses where the stopping powers are larger. However, this is not directly attributable to a lack of data in this region – the dataset consists of 801 observations for a helium target (8th highest target material) and hydrogen 616 (12th highest). Hence the poor generalisability of the model for lighter targets may be due to more complex non-linear relationships that are not present for heavier targets.

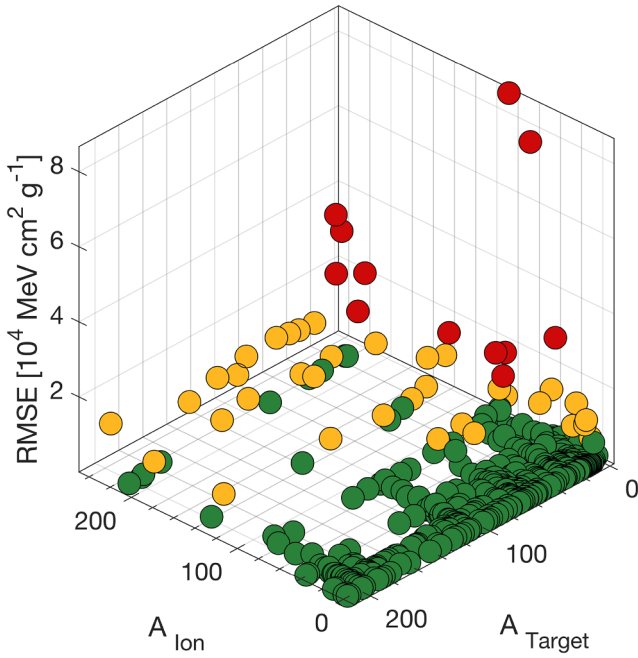
The analysis of the model allows two empirical boundary conditions to be proposed:  $A_{Ion} < 50$  and  $A_{Target} > 4$ . These describe the regime in which optimal predictive performance is achieved. Beyond this the model struggles to generalise and produce good predictions, due to a lack of training data, but also possibly as a result of more complex physical relationships in these regions. Table 4 shows the result of re-training of the model with various boundary conditions. Both the RMSE and MAE are observed to decrease by approximately 75%, which is to be expected considering the reduced number of samples with high stopping powers. The model is, however, observed to improve as a result of the added boundaries both in terms of greater  $R^2$  and lower MAPE. Despite these conditions the dataset still comprises nearly 30,000 samples and therefore remains generalised, with the improvements not falsely gained from overfitting to a small subset of the data.

#### 4.5. Individual combination examples

The performance of the model is further evaluated on the predicted stopping power functions for each individual combination. A subset of combinations is presented here, describing the general features observed across the full dataset of 522 combinations. In the following figures the predicted stopping power values are plotted in blue, as a function of energy in the range  $10^3$  to  $10^2$  MeV/amu. These predictions are generated for 200 energies equally spaced along the logarithmic axis. The experimental data that exists, which doesn’t necessarily cover the full energy range, is plotted in red. The predictions are made without the model having ‘seen’ any of the experimental values and are therefore solely generated by the patterns and relationships determined by the random forest within the remainder of the dataset. As such, the following figures represent the true power of the model to predict the



**Fig. 8.** Error metrics calculated for each individual combination. Histograms show distribution of error magnitudes across all combinations in the dataset (these are truncated for clarity along the x-axis and omit some individual combinations with very large errors). The mean error values,  $\mu$ , are calculated across all combinations, and therefore differ from those values calculated for all samples.



**Fig. 9.** The distribution of RMSE with ion and target masses,  $A_{ion}$  and  $A_{target}$  respectively. Higher errors are concentrated along the rear right-hand face of the graph – suggesting that the model does not perform well on very light targets, and heavier incident ions. The largest errors occur for a hydrogen target.

**Table 4**

Error values (evaluated on the whole dataset) are greatly reduced when the model is re-trained and cross-validated with the application of boundary conditions on the training data. Only 14% of samples are lost, suggesting that the error metrics are still representative of a large dataset and well-generalised model.

Boundary Conditions	Samples	$R^2$	RMSE	MAE	MAPE
None	34,237	0.9316	3944	814.8	15.03
$A_{target} > 4$	32,457	0.9627	2817	666.5	11.62
$A_{ion} < 50$	31,553	0.8920	1523	331.9	14.76
$A_{target} > 4, A_{ion} < 50$	29,849	<b>0.9378</b>	<b>1053</b>	<b>235.2</b>	<b>11.08</b>

stopping power function across a broad energy range, for materials where experimental data is incomplete or does not exist.

Fig. 10(a) presents a typical output for a hydrogen ion (proton)

incident on a lead target. It can be seen that the model is able to replicate the shape of the stopping power function to a high degree, with minimal discontinuities or noise. There is a discontinuity seen at higher energies, which may be symptomatic of a shortage of training data in this region. The scatter within the blue curve is due to the fact that the model is generating the predictions individually for each energy value. A smooth predicted stopping power function therefore should not be expected, nor should it be desired – the experimental training data itself is not perfect and therefore contains a component of irreducible error that the model is successfully not overfitting to. In general, the predicted values are seen to align extremely well with the experimental observations and achieve correspondingly low error metrics. The remaining low-energy portion of the curve that is not measured against the experimental values looks like a sensible extrapolation of the curve, and given the low errors on the higher energy data, could be considered a well-founded prediction. Fig. 10(b) shows a typical response for an elemental gas, argon. Similarly, good alignment with experimental values is seen, notably in the low-energy region, and the model produces a consistent stopping power function curve – boasting the ability of the model to generalise to gaseous targets.

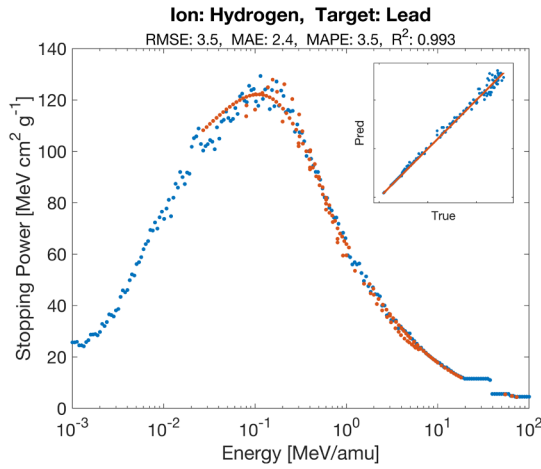
More complex targets also perform well. Namely, mixtures such as air, Fig. 11(a); alloys such as havar, Fig. 11(b); polymeric resins such as formvar, Fig. 11(c); and compounds such as silicon dioxide, Fig. 11(d).

It is interesting to assess the performance of the model on very ‘noisy’ experimental data, which has been accumulated through many different experiments, such as for iron, Fig. 12(a); nickel, Fig. 12(b); silicon, Fig. 12(c); and zinc, Fig. 12(d). Many of these combinations contain vastly different sets of observations, particularly around the peak of the function. Hence one unavoidable source of error, that cannot be attributed to the quality of the model, is simply disagreement of the predictions against a large number of inconsistent experimental observations. However, it is notable that the blind predictions on these noisy combinations tend to favour a single trend in the experimental data values. This may conversely lend credibility to the favoured set of experimental values, as it would suggest that they are statistically more representative than the others.

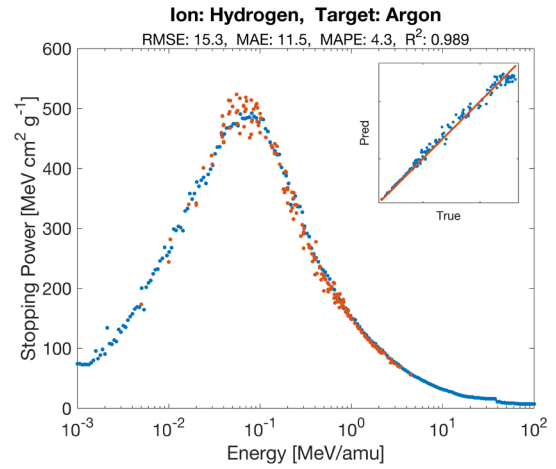
The fact that the model may be trained on numerous combinations of highly noisy and self-contradictory data, and yet is able to predict clean and continuous stopping power curves is a strong indication that overfitting is not occurring, and that the model is a robust and well generalised. Furthermore, it suggests large potential performance gains if the training data can be cleaned and rationalised.

Predictions generated for heavier ions are presented in Fig. 13. The model is largely able to produce sensible stopping power functions that agree well with the experimental values. However, the continuity of the curve degrades with increasing ion mass. The relative deficiency of





(a) Hydrogen ion on lead target

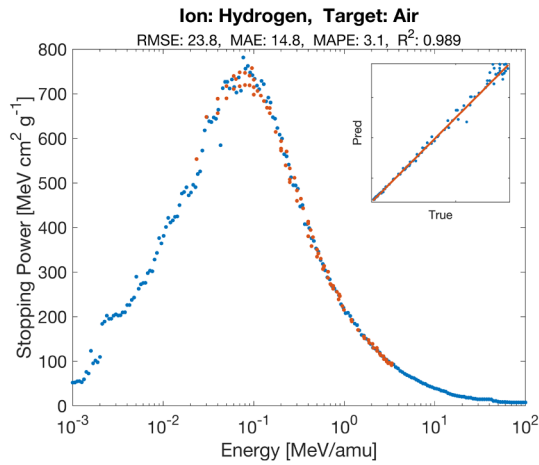


(b) Hydrogen ion on argon target

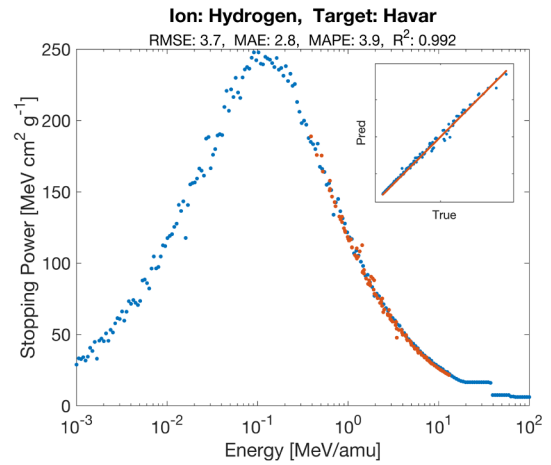
**Fig. 10.** Stopping power function: experimental observations,  $y_{true}$  (red); predictions,  $y_{pred}$  (blue).

experimental observations for higher ion masses is immediately apparent and is an attributable cause for this effect. The poorest congruity is observed in the high energy region: plateaus and discontinuities can be seen, which may be the source of the linear clusters identified in

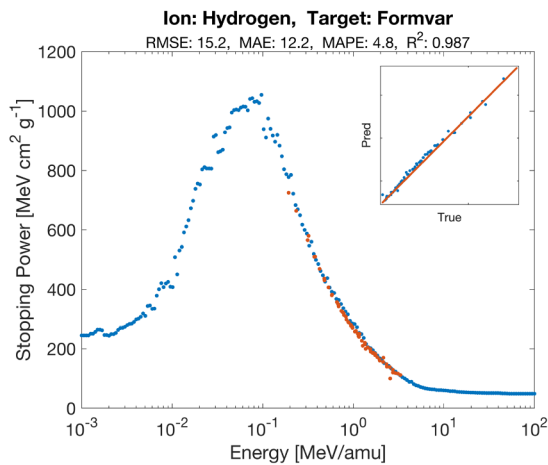
Fig. 7. RMSE and MAE are observed to scale approximately in proportion with the magnitude of the stopping power, as expected. Higher ion masses experience greater stopping powers. In relation to the magnitude these errors can be considered relatively small, accounting



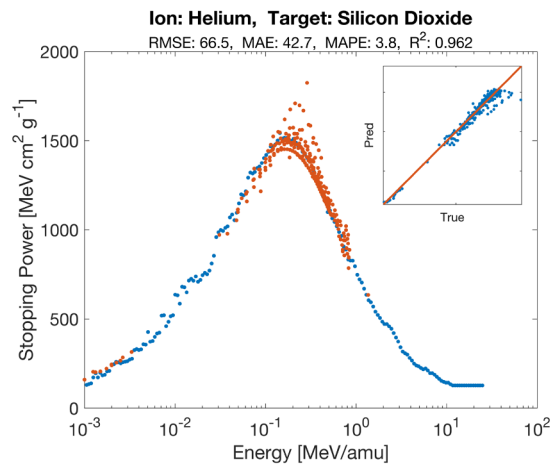
(a) Hydrogen ion on air target



(b) Hydrogen ion on havar target



(c) Hydrogen ion on formvar target



(d) Helium ion on silicon dioxide target

**Fig. 11.** Stopping power function for various complex target materials: experimental observations,  $y_{true}$  (red); predictions,  $y_{pred}$  (blue).

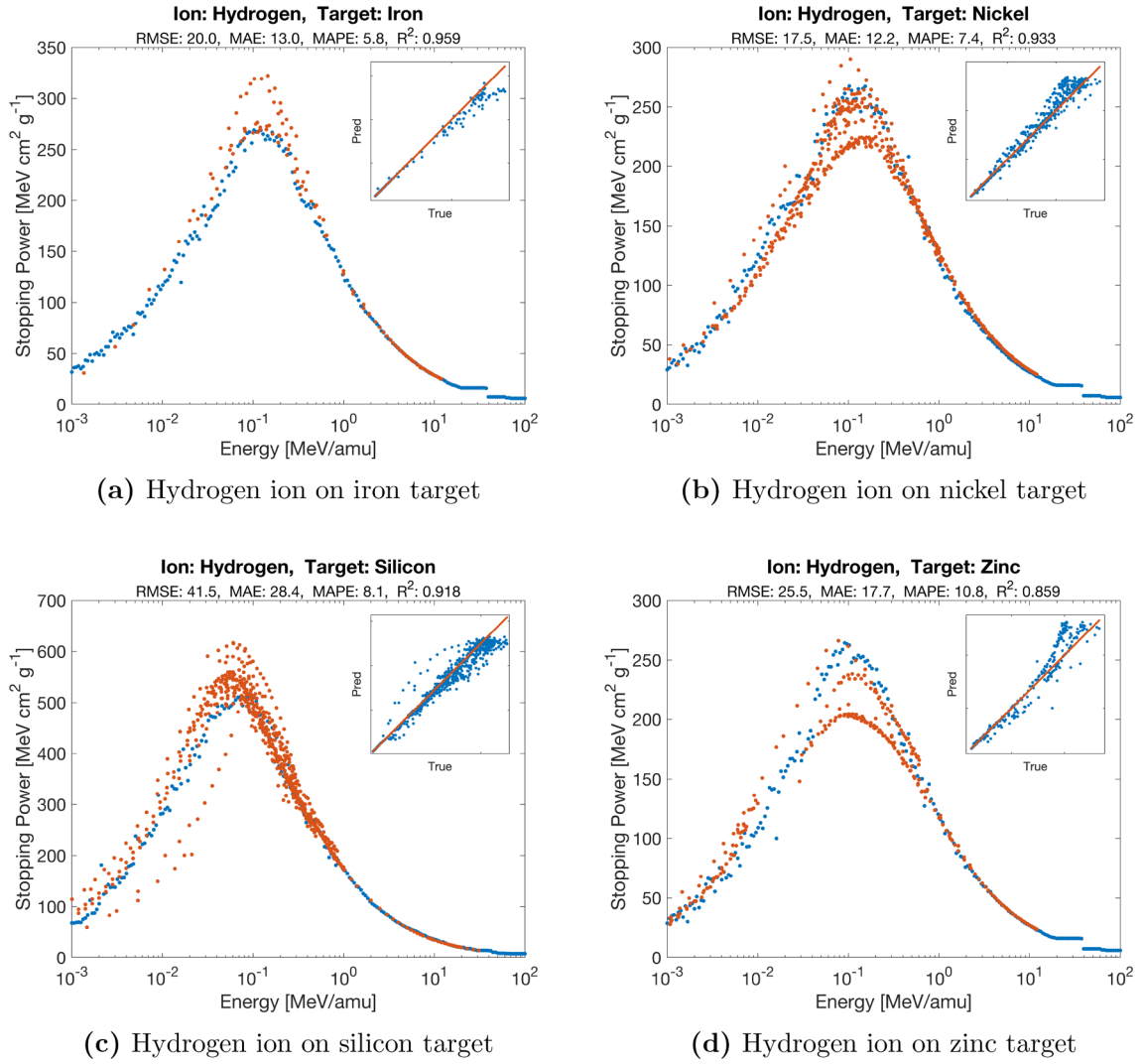


Fig. 12. Stopping power function for examples of noisy or incongruous experimental samples: experimental observations,  $y_{true}$  (red); predictions,  $y_{pred}$  (blue).

for between 4 and 10% of the peak stopping power in Fig. 13.

## 5. Discussion

### 5.1. Model assessment

The initial assessment of any ML method should be on the bias and variance of the model. A high bias is consistent with underfitting, in which the model does not exhibit a representative relationship between the input features and the predicted output. This is commonly judged by the performance on the training data, for which large errors would indicate a large bias. It is promising that the errors on the training set given in Table 2 would indicate that the bias is minimised. High variance would suggest the opposite problem – in which the model is overfitting to the data. Low training data errors combined with high errors on test data would indicate poor generalisation and hence a large variance. A random forest model has the benefit of reducing overfitting by design, in that each decision tree is fitted on a different subset of features and samples. Additionally, testing through cross-validation should provide a comprehensive assessment of the model generalisability. A few artefacts of possible overfitting have been identified in the predictions, but overall the model appears to generate consistent and sensible predictions. The errors in Table 3 are not unreasonable given other factors involved and would not suggest an intolerably high variance. There is also a significant degree of scatter, or irreducible

error, within the experimental data – in some cases this is very severe. This means that even a perfect model will suffer a minimum level of error.

Given this, the second assessment should be whether the model is fit for purpose. The direct use of the model is to be able to generate a full stopping power function,  $S(E)$ , for any ion-target combination, without any knowledge of experimental data. Figs. 10–13 display this capability is achieved satisfactorily for a broad diversity of ions and targets. A smooth curve fit is produced and is observed to align well with experimental data without appreciably high errors. Predicted values may be generated across the full energy range to form a functional curve from which ion ranges may be calculated. Two boundary conditions have been identified that define ions and targets that the model is better at predicting. These conditions are invariably the result of a lack of experimental training data outside these regimes; however may also point to a wider physical difference in the behaviour of light targets and heavy ions. This would need to be corroborated by further experimental work, and the supplementation of this data in future models.

### 5.2. Comparison to other methods

Considering that the work is novel in its particular implementation, it is difficult to make a direct comparison with previous work. The method is similar to those employed in quantitative structure–property relationship (QSPR) studies, in which ML models are used to generate

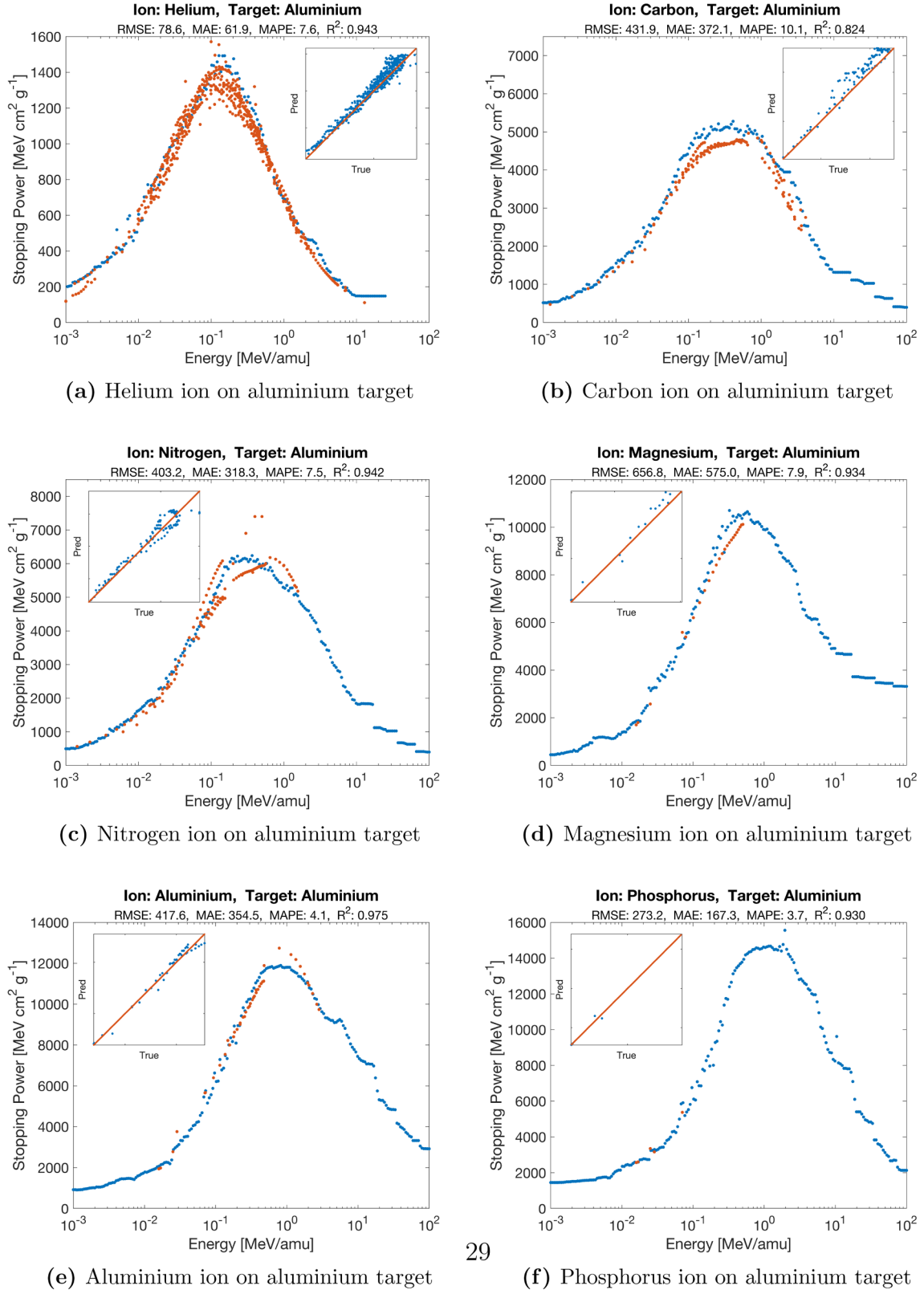


Fig. 13. Stopping power function for heavier ions on an aluminium target: experimental observations,  $y_{true}$  (red); predictions,  $y_{pred}$  (blue).

regression or classification predictions of unknown material properties based upon structural and molecular information [30–32]. These tend to be used in chemistry, for example, to predict aqueous solubility [33] or ionisation of organic molecules [34]; biology, to predict protein interactions [35]; or medicine, to identify chemical structures that may

perform well as drugs [36]. These use cases clearly do not allow quantitative comparison, however the methods used for the model development and validation in this work are very similar and of equally high standard [37–39].

The machine learning approach may also be compared with the

**Table 5**

A comparison between the alternative MAPE values published for SRIM-2010 and those derived for the random forest model. Values are calculated for various incident ions: hydrogen, helium, lithium and heavier ions. For each row:  $\bar{\epsilon}$  = the average MAPE calculated for all samples in that row.

Ions	SRIM-2010		RF Model		
	Samples	$\bar{\epsilon}$	Samples	Train $\bar{\epsilon}$	Test $\bar{\epsilon}$
H	8,300	4.0	11,380	2.4	11.9
He	6,500	3.9	9,255	2.1	9.1
Li	1,400	4.8	712	2.1	24.5
Be - U	9,000	5.8	12,879	2.0	12.4
All	25,200	4.6	34,237	2.1	11.6

current semi-empirical methods, such as SRIM [40]. The methods of Ziegler et al. differ substantially, but the aim is similar. SRIM appears to use an alternative definition of MAPE to define the error of their model,

$$MAPE^* = \frac{100}{n} \sum \left| \frac{y_{true} - y_{pred}}{y_{pred}} \right| \quad (7)$$

They publish these values compared against 25,200 experimental data points, for one of their latest updates: SRIM-2010, Table 5, in which the objective is to provide the best possible curve fit to experimental data.

For comparison, values in Table 5 are quoted for both training data, where the ML method may be considered to be performing a similar ‘curve-fitting’ task to SRIM; and the test data, in which the ML method is uniquely able to produce ‘blind’ predictions for unseen data. It can be concluded that the performance of SRIM, as evaluated solely on MAPE, is better than that of the blind test predictions of the random forest. However, it is inferior in comparison to the predictions made on the training data. Ziegler et. al. state that if erroneous experimental outliers are omitted (those that differ by >25%) the overall error of their model drops to 4.0%. Similar analysis of the RF model achieves 2.0% on the training data and 6.6% on the test data.

### 5.3. Model improvements

The current state of the model is such that it remains under-optimised in several aspects. It represents the most pessimistic result achievable by such a method. Primarily, it is disadvantaged by the quality and distribution of the training data. This may ultimately prove to be an impenetrable issue without collection of large amounts of additional experimental data, however it would be expected that substantial improvements may be achieved by simple processing of the current dataset. Such treatment may include: balancing of the dataset such that all ions and targets are represented more equally, training only on combinations that contain experimental data across the full energy range, applying further boundary conditions such as on the state of the target material to produce a more accurate but narrower model. Secondly, the large noise component in the experimental data must be addressed. The dataset should be selectively processed to minimise the scatter and remove outliers, particularly in the case where the experimental data consists of several contradictory stopping power curves. Finally, the large number of samples discarded in the processing of the original database should be extracted and added to the dataset, including those more recent experimental measurements – which may additionally benefit from more modern and refined techniques.

The feature generation process should be developed to improve the quantity of potentially useful information for training. New features could be added that convey more detail on the structure of the target materials, such as molecular fragments, chemical groups, crystallinity or bonding states – which has been shown to improve the performance of SRIM [41].

In terms of the model it would be expected that there may be

algorithms that are better suited to this kind of problem within the multitude of machine learning tools available. Promising candidates would be certain linear models or neural networks; or even the use of separate models for each energy regime. Further improvement could be attained with random forests by optimisation of the model hyperparameters (number of trees, depth, samples at each split, etc.) which can often be determined by an extensive gridsearch across the parameter space.

## 6. Conclusions

The work in this paper introduces a new paradigm for determining the stopping power of ion-target combinations. It is an area that suffers from a number of shortcomings not only in its theoretical underpinning, but also in the methods used to interpret and exploit experimental data. The implications of accurate stopping power calculations are of considerable importance and hence there is a great need to make improvements.

It has been shown that the novel use of machine learning holds substantial promise in fulfilling this demand. A random forest regression algorithm is trained using over 34,000 stopping power observations, representing 522 ion-target combinations across the energy range  $10^{-3}$  to  $10^2$  MeV/amu, and ion and target atomic masses 1 to > 240. Based on extensive evaluation through  $k$ -fold cross-validation against several error metrics it has been demonstrated that the model is able to fit to training data with superior accuracy. However, crucially, it also is able to make low error predictions on unseen test data. This signifies the extraordinary power of the model to make fully blind predictions for ion-target combinations that do not have experimental data, with the ability to generalise across elements, compounds, mixtures, alloys and polymers; in solid, liquid and gaseous states; and for a wide range of ion masses. These results lay the groundwork for further investigation and significant future improvements to the model.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors wish to thank the Engineering and Physical Sciences Research Council (EPSRC) and BAE Systems (Maritime) for the award of a Cooperative Awards in Science and Technology (CASE) studentship to fund this work; as well as those involved in the compilation and maintenance of the IAEA stopping power database, which is a tremendously valuable resource.

## References

- [1] S.M. Seltzer, J.M. Fernandez-Verea, P. Andreo, P.M. Bergstrom Jr, D.T. Burns, I. Krajcar Bronic, C.K. Ross, F. Salvat, Key Data for Ionizing-Radiation Dosimetry: Measurement Standards and Applications (Report No. 90), Technical Report, International Commission on Radiation Units and Measurements (ICRU), 2014.
- [2] H.A. Bethe, Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie, Ann. Phys. 5 (1930) 325–400.
- [3] J.C. Ashley, R.H. Ritchie, W. Brandt, Z3 effect in the stopping power of matter for charged particles, Phys. Rev. B 5 (1972) 2393–2397.
- [4] J.C. Ashley, R.H. Ritchie, W. Brandt, Z3 dependent stopping power and range contributions, Phys. Rev. A 8 (1973) 2402–2408.
- [5] F. Bloch, Zur Bremsung rasch bewegter Teilchen beim Durchgang durch die Materie, Ann. Phys. 16 (1933) 285–320.
- [6] H. Bichsel, Stopping Power of Fast Charged Particles in Heavy Elements (Report IR-4550), Technical Report, National Institute of Standards and Technology, 1991.
- [7] R.M. Sternheimer, S.M. Seltzer, M. Berger, Density effect for the ionization loss of charged particles, Phys. Rev. B 26 (1982) 6067–6076.
- [8] R.M. Sternheimer, The density effect for the ionization loss in various materials, Phys. Rev. 88 (1952) 851–859.
- [9] M.J. Berger, M. Inokuti, H.H. Andersen, H. Bichsel, D. Powers, S.M. Seltzer, D.



- Thwaites, D.E. Watt, Stopping Powers and Ranges for Protons and Alpha Particles (Report No. 49), Technical Report, International Commission on Radiation Units and Measurements (ICRU), 1992.
- [10] C. Varelas, J. Biersack, Reflection of energetic particles from atomic or ionic chains in single crystals, *Nucl. Instrum. Methods* 79 (1970) 213–218.
- [11] H.H. Andersen, J.F. Ziegler, Hydrogen: Stopping Powers and Ranges in All Elements, The Stopping and Ranges of Ions in Matter, vol. 3, Pergamon Press, Elmsford, New York, 1977.
- [12] W. Donahue, W.D. Newhauser, J.F. Ziegler, Analytical model for ion stopping power and range in the therapeutic energy interval for beams of hydrogen and heavier ions, *Phys. Med. Biol.* 61 (2016) 6570–6584.
- [13] J.F. Ziegler, Helium: Stopping Powers and Ranges in all Elemental Matter, The Stopping and Ranges of Ions in Matter, vol. 4, Pergamon Press, Elmsford, New York, 1977.
- [14] D. Watt, Stopping Cross Sections, Mass Stopping Powers and Ranges in 30 Elements for Alpha Particles (1 keV to 100 MeV) (Ph.D. thesis), University of St. Andrews, Scotland, 1988.
- [15] D. Powers, Stopping Cross, Sections and Ranges for Alpha Particles in Matter (Report prepared for the ICRU Stopping Power Committee), Technical Report, Department of Physics, Baylor University, Waco, Texas, 1978.
- [16] D. Hill, 7 Key Factors Driving the Artificial Intelligence Revolution, available at <https://singularityhub.com/2016/08/29/7-factors-driving-the-artificial-intelligence-revolution/> (accessed on February 22nd, 2018).
- [17] P. Lops, M. de Gemmis, G. Semeraro, Content-based Recommender Systems: State of the Art and Trends, *Recommender Systems Handbook*, Springer, Boston, MA, 2011.
- [18] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (2017) 115–118.
- [19] S. Shen, H. Jiang, T. Zhang, Stock market forecasting using machine learning algorithms, Technical Report, Department of Electrical Engineering, Stanford University, 2012.
- [20] E. Alpaydin, Introduction to Machine Learning, third ed., The MIT Press, Cambridge, MA, 2014.
- [21] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [22] T. Ho, Random Decision Forests, in: 3rd International Conference on Document Analysis and Recognition, Montreal, Canada, pp. 278–282.
- [23] H. Paul, Stopping Power of Matter for Ions: Graphs, Data, Comments and Programs, available at <https://www-nds.iaea.org/stopping/> (accessed on September 5th, 2018).
- [24] C.C. Montanari, P. Dimitriou, The IAEA stopping power database, following the trends in stopping power of ions in matter, *Nucl. Instrum. Methods Phys. Res., Section B: Beam Interactions Materials Atoms* 408 (2017) 50–55.
- [25] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, G. Varoquaux, API design for machine learning software: experiences from the scikit-learn project, Technical Report, 2013.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, Scikit-learn: machine learning in python, *J. Mach. Learn. Res.* 12 (2012) 2825–2830.
- [27] Scikit-learn User Guide: 1.11. Ensemble Methods, available at <http://scikit-learn.org/stable/modules/ensemble.html#forest> (accessed on February 27th, 2018).
- [28] D.L. Alexander, A. Tropsha, D.A. Winkler, Beware of R2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models, *J. Chem. Inf. Model.* 55 (2015) 1316–1322.
- [29] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Chapman and Hall, London, 1984.
- [30] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna, V. Prachayasittikul, A practical overview of quantitative structure-activity relationship, *EXCLI J.* 8 (2009) 74–88.
- [31] P. Liu, W. Long, Current mathematical methods used in QSAR/QSPR studies, *Int. J. Mol. Sci.* 10 (2009) 1978–1998.
- [32] K. Roy, S. Kar, R. Das, What is QSAR? Definitions and Formulism, A primer on QSAR/QSPR modeling: Fundamental Concepts, Springer-Verlag Inc., New York, NY, 2015, pp. 2–6.
- [33] D.S. Palmer, N.M. O'Boyle, R.C. Glen, J.B.O. Mitchell, Random forest models to predict aqueous solubility, *J. Chem. Inf. Model.* 47 (2007) 150–158.
- [34] R. Fraczekiewicz, In silico prediction of ionization, *Comprehensive Med. Chem. II* 5 (2007) 603–626.
- [35] E. Freyhult, K. Andersson, M. Gustafsson, Structural modeling extends QSAR analysis of antibody-lysozyme interactions to 3D-QSAR, *Biophys. J.* 84 (2003) 2264–2272.
- [36] C. Nantasenamat, I. Chartchalern, V. Prachayasittikul, Advances in computational methods to predict the biological activity of compounds, *Expert Opin. Drug Discov.* 5 (2010) 633–654.
- [37] P. Gramatica, Principles of QSAR models validation: Internal and external, *QSAR Combinatorial Sci.* 26 (2007) 694–701.
- [38] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, *Mol. Inform.* 29 (2010) 476–488.
- [39] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, *Mol. Inform.* 22 (2003) 69–77.
- [40] J.F. Ziegler, J.P. Biersack, U. Littmark, The Stopping and Range of Ions in Solids, The Stopping and Ranges of Ions in Matter, vol. 1, Pergamon Press, Elmsford, New York, 1985.
- [41] J.F. Ziegler, J.M. Manoyan, The stopping of ions in compounds, *Nucl. Instrum. Methods Phys. Res., Section B: Beam Interactions with Materials and Atoms* 35 (1988) 215–228.