This is the Accepted Manuscript version of an article accepted for publication in Physics in Medicine & Biology. IOP Publishing Ltd is not responsible for any errors or omissions in this version or any version deried from it. The Version of Record is available online at https://doi.org/10.1088/1361-6560/abf1fd .

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

ACCEPTED MANUSCRIPT

Machine learning for proton path tracking in proton computed tomography

To cite this article before publication: Dimitrios Lazos *et al* 2021 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/abf1fd

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2021 Institute of Physics and Engineering in Medicine**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 86.134.173.73 on 08/04/2021 at 14:08

TITLE: **Machine Learning for proton path tracking in proton Computed Tomography**

Dimitrios Lazos[1], Charles-Antoine Collins-Fekete [2], Miroslaw Bober[1], Philip Evans[1, 3] and Nikolaos Dikaios[1, 4]

[1] Centre for Vision, Speech and Signal Processing, Department of Electrical and Electronic Engineering, University of Surrey, Guildford, GU2 7XH, United Kingdom

[2] Department of Medical Physics and Biomedical Engineering, University College London, London, WC1E 6BT, United Kingdom

[3] Chemical, Medical and Environmental Science, National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, United Kingdom

[4] Research Centre of Mathematics, Academy of Athens, Athens, 11527, Greece

Accepted Manuscript

Abstract

A Machine Learning approach to the problem of calculating the proton paths inside a scanned object in proton Computed Tomography is presented. The method is developed in order to mitigate the loss in both spatial resolution and quantitative integrity of the reconstructed images caused by multiple Coulomb scattering of protons traversing the matter. Two Machine Learning models were used: a forward neural network and the XGBoost method. A heuristic approach, based on track averaging was also implemented in order to evaluate the accuracy limits on track calculation, imposed by the statistical nature of the scattering. Synthetic data from anthropomorphic voxelized phantoms, generated by the Monte Carlo Geant4 code, were utilised to train the models and evaluate their accuracy, in comparison to a widely used analytical method that is based on likelihood maximization and Fermi-Eyges scattering model. Both neural network and XGBoost model were found to perform very close or at the accuracy limit, further improving the accuracy of the analytical method (by 12% in the typical case of 200MeV protons on 20 cm of water object), especially for protons scattered at large angles. Inclusion of the material information along the path in terms of radiation length did not show improvement in accuracy for the phantoms simulated in the study. A neural network was also constructed to predict the error in path calculation, thus enabling a criterion to filter out proton events that may have a negative effect on the quality of the reconstructed image. By parametrizing a large set of synthetic data, the Machine Learning models were proved capable to bring - in an indirect and time efficient way - the accuracy of the Monte Carlo method into the problem of proton tracking.

Keywords: Machine Learning, proton Computed Tomography, proton path tracking, multiple Coulomb scattering, Geant4 Monte Carlo, proton stopping power

Accepted Manuscript

1. Introduction

Proton therapy takes advantage of the property of protons that they release a significant part of their energy towards the end of the path, creating compact and well-defined energy deposition profiles. The sharp dose fall-off of the proton beams allows the planning of highly conformal dose distributions to the treatment target, while at the same time sparing the surrounding healthy tissue. However, these desirable features of proton therapy make it vulnerable to uncertainties in the proton treatment delivery, including those in set up, patient motion and proton range, the latter being related to miscalculations on tissue inhomogeneity and patient stopping power estimation.

Although proposed as early as x-ray CT, proton CT (pCT) is a relatively unexplored field, especially in its clinical aspects, with its own practical/theoretical challenges and image artefacts. Multiple Coulomb scattering is the major factor limiting the spatial resolution of pCT images. Passing through matter, protons undergo a large number of small angle deflections, ending up after exiting the imaged object with a lateral displacement of the order of few millimetres. While proton trackers are capable of recording, with satisfactory accuracy, the position and orientation before and after traversing the patient, the proton path inside the body is unknown, thus compromising clinically needed spatial resolution and tissue quantitative information.

Theoretical approaches, based on Fermi-Eyges (Eyges 1948) and Molière (1947) multiple scattering theories, as well as empirical/phenomenological ones have been proposed to improve incomplete proton track information. The methods are based on the information that is recorded during the acquisition for each proton, regarding the position, motion direction and energy, before and after passing through the imaged object. Most of the methods attempt to fit the proton track to the measurements by means of maximizing likelihood and/or employing Bayesian inference schemes (Williams 2004, Schulte *et al* 2008, Wang *et al* 2010, Collins-Fekete *et al* 2017, Lazos *et al* 2020). The least action-based approach, which involves solving the Euler-Lagrange equation, has also been proposed (Erdelyi 2009, Krah *et al* 2019). Another category of solutions on a simpler theoretical base is the use of parametric curves. Collins-Fekete *et al* (2015) proposed an empirical cubic spline path model that is calibrated by parametrizing Monte Carlo data, without using any multiple scattering theory. Williams (2004), Li *et al* (2006) and Erdelyi (2009) proposed /utilised cubic parametric curves for the case where energy loss is not accounted for.

Such theoretical tools however are based on simplifying assumptions, necessary for the analytical solution of typically complicated integral-differential equations. The Fermi-Eyges model for example, accounts for statistical correlation between angular and spatial coordinates, through a joint Gaussian distribution, but it underestimates large angle scattering. These sources of inaccuracy have a negative effect on the path estimation and subsequently on the pCT image numerical consistency. In addition, computational time in these typically inverse methods is a concern, since pCT reconstruction through an iterative method is by itself a time-consuming procedure. The stochastic nature of particle scattering and energy loss seems to set a barrier to the achievable accuracy, which is translated into pCT inherent spatial resolution limits.

To overcome the inaccuracies of the underlying physics models and the computational burden of analytical solutions, we propose in this article a machine learning approach for the problem of proton path tracking inside the imaged object. Machine learning algorithms bring new possibilities to this problem due to their capability to model complex data dependencies and to learn from large sets of high-dimensional data through training. Machine learning tools such as artificial neural networks (NNs) have shown promising results in the problem of particle track reconstruction from accelerator data (Tsaris *et al* 2018) and due to their ability for parallelization have greatly reduced computation times.

The proton tracking problem in pCT is characterized by smaller degree of complexity and lower dimensionality. In our approach we calculate the sequence of the proton positions by employing a feed forward NN and an extreme gradient boosting (XGB) model (Chen and Guestrin 2016), trained over a large set of Monte Carlo simulated proton tracks in matter.

Accepted Manuscript

The input of the network consists of the position, direction and energy of each proton before and after the medium. The proposed NN is compared against another state of the art machine learning algorithm such as boosted decision trees.

A ~~critical~~ question in path reconstruction in pCT is if the knowledge of the materials along the path can improve the accuracy of the estimation. This is because the theoretical cross sections of multiple scattering, such as those in Fermi–Eyges or Molière theories, depend on the material through both density and atomic number, and the same is valid for compounds and mixtures. Such information can be extracted, with some degree of uncertainty, from the maps of the relative stopping power (RSP) that are given by pCT, thus potentially creating a positive feedback in accuracy. To test this hypothesis, we constructed a NN where the input, besides the measured variables (position, direction and energy), includes a number of radiation length values along the path. Being capable to detect and reproduce any observable dependency, artificial NNs can be used to test the validity of this hypothesis. Another NN is also constructed that attempts to predict the error of the reconstructed path, enabling a criterion to filter out proton tracks that may have a negative effect on the quality of the reconstructed image.

In the following, we present a solution of the path reconstruction problem based on machine learning models (namely artificial NN and boosted decision tree) and compare their accuracy against a widely accepted theoretical approach, based on likelihood maximisation (Williams 2004, Schulte *et al* 2008). Geant4 Monte Carlo (MC) data from voxelized phantoms and simple phantoms have been used for training the models and evaluate their accuracy. We also utilise the same models to check if the knowledge of the radiation length of the material(s) of the imaged object can enhance the method accuracy.

Accepted Manuscript

2. Methods

2.1. Problem definition

We assume the proton enters and leaves the medium at points $\mathbf{M}_0 = (x_0, y_0, z_0)$ and $\mathbf{M}_1 = (x_1, y_1, z_1)$, respectively (figure 1(a)). Direction vectors at these points are $\mathbf{P_0}$= (u_0,v_0,w_0) and $\mathbf{P_1}$=(u_1,v_1,w_1), assuming $||\mathbf{P}_0||=||\mathbf{P}_1||=1$. The direction vectors are measured by two pairs of trackers, one positioned at each side of the imaged object. The same trackers record the proton position, while the points $\mathbf{M}_0$ and $\mathbf{M}_1$ are calculated by projecting these positions on the object surface, the latter can be given by a simple filtered backprojection reconstruction, even without elaborated correction of proton path. The initial energy is known from the accelerator settings, while the final one is measured by the range telescope. In this context, the problem of calculation of the path $\boldsymbol{S}(t) = (x(t), y(t), z(t))$ can be written as a function of the known variables:

$$\boldsymbol{S}(t) = F(t, \mathbf{M}_0, \mathbf{M}_1, \mathbf{P}_0, \mathbf{P}_1, E_0, E_1), \tag{1}$$

where E_0 and E_1 the proton energy before and after passing through the medium, while t is a temporal variable so that $0 \le t \le 1$, $\boldsymbol{S}(0) = \mathbf{M}_0$ and $\boldsymbol{S}(1) = \mathbf{M}_1$. Without loss of generality, for each proton we select a coordinate system so that $\mathbf{P}_0$ = (1, 0, 0) and $\mathbf{M}_0$ = (0, 0, 0). Similar to the methodology followed in other path calculation approaches (Williams 2004, Schulte *et al* 2008), we assume that the proton moves independently along the *y*- and the *z*-axis, and so the motion along each axis can be treated separately. This simplifies the problem from 3D to 2D. Equation 1 can thus be rewritten for the *y*-axis:

$$y(t) = F_{\mathrm{y}}(t, x_1, y_1, \theta_{y1}, E_0, E_1), \tag{2}$$

where θ_{y1} is the plane exit angle (on *xy* plane), while the position variable t has been set equal to x/x_1. Similar equation exists for *z*-axis. In this study, we assume that this equation can be approximated by two ML models: a feed forward NN (denoted as N_{2D}) and an extreme gradient boosting (XGB) decision tree. Proton track data for training and qualitative assessment of the proposed models was obtained by Geant4 MC simulation. It should be noted that for measured data $(\mathbf{M}_i, \mathbf{P}_i, E_i, i = 0, 1)$ of each proton there is an infinite multitude of paths $\boldsymbol{S}(t)$ that can fit the measured values. Indeed, according to the Fermi-Eyges theory the proton position at each depth follows a Gaussian distribution, with energy- and depth-dependant spread. The resulting path $\boldsymbol{S}(t)$ is estimated by our method, in the same way the Fermi-Eyges model gives the path that maximises likelihood. In our case the estimated path may be defined as the path that minimises the L_2 norm between the NN predictions and the actual paths, produced by MC simulation.

Another NN (denoted as N_{Err}) is constructed to predict the error of the track given by the N_{2D}, relative to the actual MC simulation path. The function this NN models is:

$$\mathrm{N_{Err}}: \quad \left|y_{\mathrm{N_{2D}}}(t) - y_{\mathrm{MC}}(t)\right| = F_{\mathrm{Err}}(t, x_1, y_1, \theta_{y1}, E_0, E_1), \quad 0 \le t \le 1 \tag{3}$$

where $y_{\mathrm{N_{2D}}}(t)$ is the track estimated by the N_{2D} and $y_{\mathrm{MC}}(t)$ is the path given by Geant4 simulation. In this modelling the absolute value of the error is set as the target value.

While the functional form of the equation (2) is the most simple and transparent form of the problem, in practice it may be inefficient, since it requires the initial proton direction to coincide with the *x*-axis. This is achieved by 2D/3D coordinate system rotation. To avoid this computational burden, we attempt to approximate directly the function of equation (1) through a NN. In this way the 2D/3D rotations are essentially incorporated in the functionality of the NN and performed in a much faster linear fashion. To simplify equation (1) and reduce the dimensionality of the problem we apply the

Accepted Manuscript

computationally inexpensive parallel translation of the coordinate system, so that $\mathbf{M_1}' = \mathbf{M}_1 - \mathbf{M}_0$ and $\mathbf{M_0'} = \mathbf{M}_0 - \mathbf{M}_0 = (0,0,0)$. In this way the entry point is omitted from the set of input variables. Another NN is constructed (N_{3D}) to approximate equation (1) that now becomes:

$$N_{3D}\colon\ \boldsymbol{S}(t) = F(t, \mathbf{M}_1', \mathbf{P}_0, \mathbf{P}_1, E_0, E_1), 0 \le t \le 1. \quad (1a)$$

The N_{3D} has three outputs, one for each direction. The variable t for a path point $\mathbf{M}_t$ is defined by $t = D_{\mathbf{M}_t}/D_{\mathbf{M}_1}$, the ratio of the distances $D_{\mathbf{M}_t}$ and $D_{\mathbf{M}_1}$ from the origin of the points $\mathbf{M}_t$ and $\mathbf{M}_1$ projected on the line of the initial proton direction. To produce training and evaluation data we rotate the data used for N_{2D}, at random directions, uniformly distributed over the 4π steradians. The vectors $\mathbf{M}_1'$, $\mathbf{P_0}$ and $\mathbf{P_1}$ and all the path points are rotated around y-axis by an angle a, determined by sampling uniformly for its cosine in $(-1, 1)$ and subsequently around x-axis, by an angle uniformly sampled in $[0, 2\pi)$.

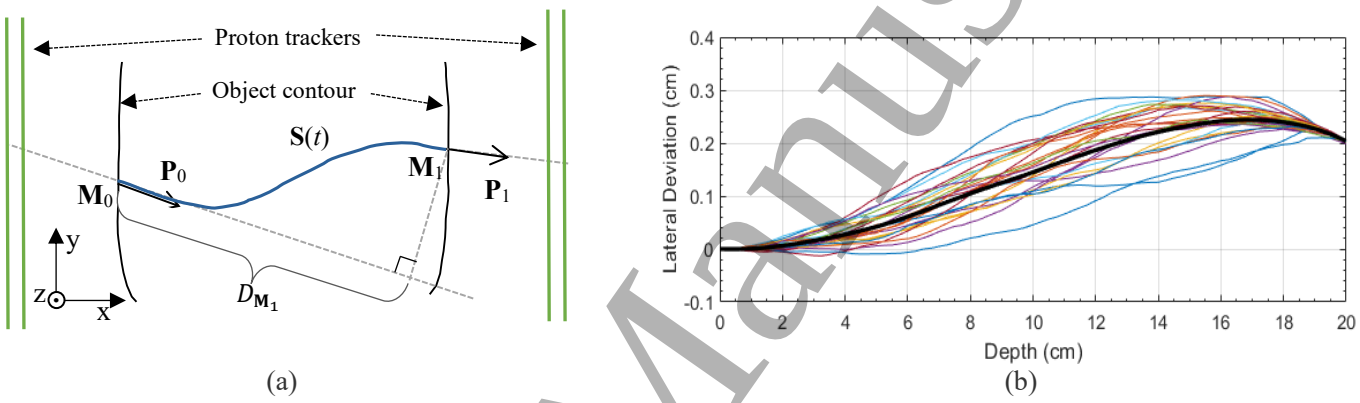


Figure 1. a) Diagram showing vectors $\mathbf{M}_0$, $\mathbf{M}_1$, $\mathbf{P}_0$ and $\mathbf{P}_1$. b) Demonstration of the concept of path averaging. Axes are not in scale.

Accepted Manuscript

2.2. Machine Learning models

The developed algorithms were based on feed forward NNs and XGB which is a form of boosted decision trees (Chen and Guestrin 2016). A NN parametrises a function Y=Y(X), (not to be confused with coordinates x and y) by means of a grid of neurons, simple parametric functions that combine linear and non-linear features in their response. In a graphical representation of the feed forward NN (figure 2), the neurons lie in successive, hierarchically positioned layers, where the neurons input comes from the lower layers, including the first layer that contains the input variables. The connections between neurons are parameters called weights. The set $\boldsymbol{\Omega}$ of the network parameters is optimised for a minimal error between model predictions $Y(X_i; \boldsymbol{\Omega})$ and training data points Y_i:

$$\boldsymbol{\Omega} = \arg\min_{\boldsymbol{\Omega}} \sum_{i=1}^{N} [Y(X_i; \boldsymbol{\Omega}) - Y_i]^2 \,. \quad (4)$$

In the above a quadratic penalty function has been assumed. In our implementation, a feed forward NN with 2 hidden layers with 19 and 2 neurons respectively was assumed. The rectified linear unit ("ReLU") activation function was used, and the "softmax" at the final layer (Gao and Pavel 2018). Training was performed by minimizing the L_2 norm, using adaptive learning rate optimization (ADAM) (Kingma and Ba 2014). The learning rate was 0.001, weight decay of 0.0005 and training batch of 256. Training and evaluation of the NNs was performed in "Matlab" programming language (The MathWorks, Inc.), on a typical desktop computer (2.4GHz, 6GB of RAM).

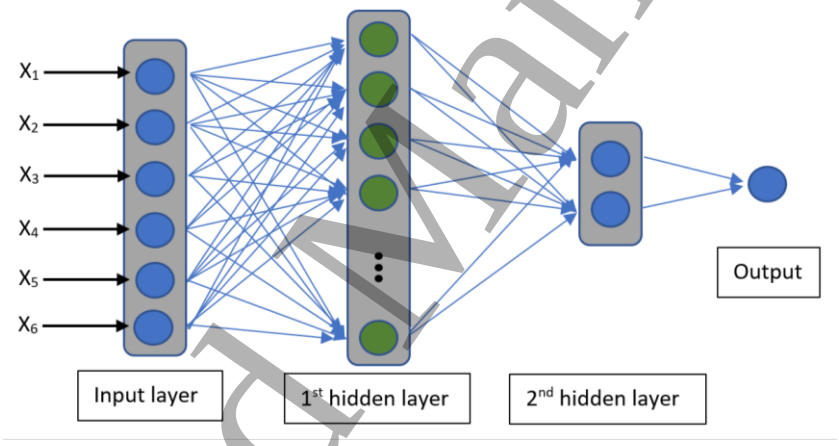


Figure 2. Schematic diagram of the 19×2 NN used in this study.

XGB is a fast implementation of a gradient boosted decision tree. Decision trees are regression models in the form of a tree structure; however, they are prone to bias and overfitting. Boosting is a method of sequentially training weak classifiers (decision trees) to produce a strong classifier where each classifier tries to correct its predecessor to prevent bias-related errors. Gradient boosting tries to fit the errors made in the initial fit and correct the corresponding errors in further training. XGB run with maximum tree depth of 12, boosting learning rate 0.1, number of boosted trees to fit 1000, subsampling parameter of 0.9 and sampling level of columns by tree of 0.8. A "softmax" objective function was used, and L_2 norm as an evaluation metric. Training and evaluation were performed by "sklearn" software machine learning library in "Python" programming language (Pedregosa *et al* 2011), and run on the "Kaggle" computation platform (www.kaggle.com).

Accepted Manuscript

2.3. Path averaging

A more straightforward approach based on path averaging over a dictionary of Monte Carlo proton tracks was also implemented for comparison. The track for every proton was calculated by averaging a subset of tracks that share the same value with the proton in question, in terms of the following measured quantities: path length projected on *x*-axis (x_1), lateral displacement (y_1), plane angular deflection (θ_{y1}), entry and exit energy (figure 1(b)). The path $\boldsymbol{S}_i$ of the *i*-th proton with measured data equal to (x_{1i}, y_{1i}, θ_{y1i}, E_{0i}, E_{1i}) was given by:

$$S_i(x) = \overline{S_k}(x\,|[x_1, y_1, \theta_{y1}, E_0, E_1]_i), k \in Q_i, \tag{5}$$

where Q_i is the subset of the library protons that match the proton in question. Practically this subset is populated with protons paths $S_k(x)$ that are in the vicinity of the *i*-th proton in the 5-fold space, defined by the input variables, by means of the tolerance values:

$$k \in Q_i \;\text{ if }\; \Delta x_1 \le |x_{1k} - x_{1i}|, \Delta y_1 \le |y_{1k} - y_{1i}|, \Delta\theta_{y1} \le \left|\theta_{y1k} - \theta_{y1i}\right|,\; \Delta E_0 \le |E_{0k} - E_{0i}|, \Delta E_1 \le |E_{1k} - E_{1i}|, \tag{6}$$

where the tolerances $\Delta x_1, \Delta y_1, \Delta\theta_{y1}, \Delta E_0$ and ΔE_1 are set to 5 mm, 0.05 mm, 0.05 deg, 5 MeV and 2MeV respectively. The method is sensitive to these tolerance values, since large values reduce accuracy, while small ones affect statistics by diminishing sample size. The values above provided the best trade-off between accuracy and statistics in voxelized phantoms. To reduce the probability of solving for unmatched paths, we applied this method to paths after 5% filtering of the path library. The filtering cuts the 5% of the paths with the higher exit plane angle θ_{y1}, thus reducing the appearances of extreme values of lateral displacement also, since the latter is highly correlated to the angle θ_{y1}. A minimum sample size of 70 tracks is set. However even after filtering, there are cases that the sample size is less than 70 paths. In such cases, the most restrictive tolerance is repetitively increased by 5%, until the sample size exceeds the minimum value. The root mean square error (RMSE) measured by the method is not sensitive on the selection of this limit (the difference in error between sample sizes of 70 and 10 is 5 μm for 200 MeV protons incident on 20 cm of water). Since the training data was not enough to support the required accuracy, a set of at least 2×10^6 tracks was produced by MC simulation for every phantom and energy the method was evaluated. To further reduce the size requirements, the following property that comes out of space symmetry was applied:

$$\overline{S_k}(x\,|[x_1, y_1, \theta_{y1}, E_0, E_1]_i), = -\overline{S_k}(x\,|[x_1, -y_1, -\theta_{y1}, E_0, E_1]_i). \tag{7}$$

No other form of symmetry (such that between the *y*- and *z*-axis) were utilised in this method, or elsewhere in the study for the purpose of data augmentation.

Accepted Manuscript

2.4. Inclusion of material information

To test the hypothesis that the accuracy of the path calculation can be improved if material information along the proton path is added in the input, the N_{RL} was constructed that incorporated such information in terms of inverse radiation length L_R. With this addition, the function the NN approximates becomes:

$$N_{RL}:\ y(t) = F_y(t, x_1, y_1, \theta_{y1}, E_0, E_1, \{1/L_R(t_k)\}). \quad (8)$$

The input variables $L_R(t_k)$ are given at an equidistant grid of K points, $t_k = \Delta t/(k + 1/2)$, where k = 0, 1, ..., K-1 and $\Delta t = x_1/K$. Radiation length information of the materials along the path is provided by the Geant4 output at distances of 0.2mm.

Accepted Manuscript

2.5. Detector effects

The performance of the proposed method is evaluated in more realistic scenarios where the detector limited spatial and energy resolution reduces the accuracy of the track measurements (Bodd *et al* (2014)). The case of a pencil beam perpendicularly incident on 20 cm water is considered (figure 3). As shown in the figure, passing through the detectors, protons are detected with errors Δy_{0A}, Δy_{0B}, Δy_{1C} and Δy_{1D}, respectively.

Except from the positional errors in determining the points $\mathbf{M}_0^{"}$ and $\mathbf{M}_1^{"}$, these errors result also in parallax errors in the entry and exit angles θ_0 and θ_1. We study the effect of these detector errors on the path prediction of the N_{3D}. Assuming that the detector errors follow Gaussian distribution, we sample their values from such distribution (the values of 0.10 mm, 0.15 mm and 0.20 mm as the standard deviation $\sigma_{\Delta y}$ of the distribution were simulated). A second variable in the study is the distance d between adjacent detectors, while the gap between the inner detectors and the object is assumed constant at 10 cm. After applying the detector error on 400k MC proton tracks, the proton paths are calculated by the N_{3D} network:

$$\boldsymbol{S}(t) = \mathrm{N}_{3\mathrm{D}}(t, \mathbf{M}_1'' - \mathbf{M}_0'', \mathbf{P'}_0, \mathbf{P'}_1, E_0, E_1), 0 \le t \le 1, \tag{9}$$

where $\mathbf{P'}_0$ and $\mathbf{P'}_1$ the normalized direction vectors, determined after the errors on the detection points are applied. Also studied is the case of error in the measurement of the exit energy E_1, assumed it Gaussian with a standard deviation σ_E between 0.05 MeV and 10MeV. In this case, paths are calculated by the N_{2D} network, without assuming positional errors.

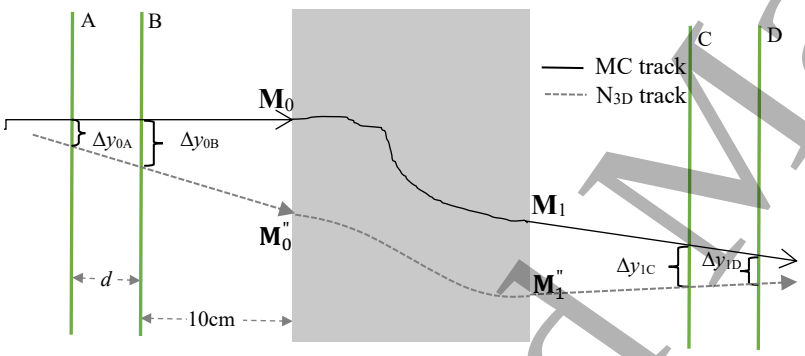


Figure 3. Diagram showing the MC track and the track estimated by the N_{3D} network, after applying the Gaussian –assumed detector error.

Accepted Manuscript

2.6. Training data acquisition and performance evaluation

2.6.1. Monte Carlo simulation

Geant4 Monte Carlo code, version 10.2.1 (Agostineli *et al* 2003) has been utilized to create a large number of proton tracks in anthropomorphic voxelized phantoms and homogenous water phantoms, for initial proton energies ranging from 180 MeV to 350 MeV. The training dataset was composed from proton tracks in water layers with thicknesses varying for 15 cm to 35 cm, as well as voxelized ones (CIRS Head, CIRS Pelvic and CIRS Lung, Table 1). Evaluation of the accuracy of the proposed models was performed on water phantoms as well as on a paediatric head phantom, the CIRS pelvic and an abdomen phantom. A pencil beam, perpendicular to the surface was assumed in the simulation of water phantoms and a parallel beam for the voxelized ones, covering the entire phantom area (table 1). Proton position, direction and energy, before and after passing through the media were recorded (similar to experimental pCT measurements), to serve as the input parameters of the proposed models. Positions of the protons along their path were also recorded every 0.2 mm to be used as the output training data. The following physics lists were enabled in Geant4: the standard electromagnetic option 3 for electrons and ions tracking in the absence of magnetic field, the ions elastic model (G4HadronElasticPhysics) and both the QMD (G4IonQMDPhysics) and Bertini/Ion cascade (G4HadronInelasticQBBC) models for elastic and inelastic scatter. The simulation of multiple Coulomb scattering was based on the Lewis model (Goudsmit and Saunderson 1940) along with the Urban model (Urban 2006).

Table 1. Composition of the training dataset.

Phantom	**Initial Proton Energy** (MeV)	**Number of proton tracks**	**Beam Dimensions** (cm × cm)
CIRS Head	180-350	1.4 M	17 x 35
CIRS Pelvic	>>	1.5 M	23 x 44
CIRS Thorax	>>	1.2 M	23 x 44
Water 15, 19, 23, 27, 31, 35 cm	>>	6 x 150 k	Pencil Beam

Accepted Manuscript

2.6.2. Training data and accuracy index

The training set of the N_{2D}, N_{3D}, N_{RL} and XGB consisted of 5×10^6 proton tracks. For every path the target variable (lateral displacement) was evaluated at one value of the temporal variable t randomly selected in the interval $(0, 1)$. The training set of N_{Err} consisted of 4.2×10^6 tracks, different from those used for the training of N_{2D}. The target value in this case was the absolute value of the error in displacement of the path reconstructed by the N_{2D} network. The proposed models calculated the paths on an equidistant grid of 100 points, determined by the t_k variable. The performance of the 2D models and path averaging was evaluated in terms of the RMS error in lateral deviation on *y*-axis, according to the following:

$$\mathrm{RMSE}(t) = \left\{\frac{1}{N}\sum_{i=1}^{N}\left[y_i(t) - y_{\mathrm{MC},i}(t)\right]^2\right\}^{1/2}, \tag{10}$$

The error of N_{3D} was evaluated in similar fashion in 3D space. The maximum - along the path - RMS error is presented, as accuracy index, occurring after the middle point of the path. The evaluation set of the protons tracks created by the simulation, consisted of at least 100k tracks for every case statistics are presented.

The proposed ML models were compared against a likelihood maximisation method, based on the Fermi-Eyges multiple scattering model and commonly referred as most-likely-path (MLP) (Williams 2004, Schulte *et al* 2008). We used the proposed (Schulte *et al* 2008) polynomial fit to evaluate the factor $1/\beta^2p^2$ as a function of depth (up to 20 cm) for a beam of 200 MeV in 20 cm water medium. We applied the same technique to calculate this factor for proton beams of 230 MeV and 250 MeV on water media of 30 cm and 20 cm respectively (in the case of 30 cm a 9th degree polynomial was necessary for accurate fitting). For application of the MLP method on inhomogeneous media, such as the voxelized phantoms, the proton energy *E* at depth *x* is approximated by:

$$E(x) = E_1 - \int_0^x \left(\frac{\mathrm{d}E}{\mathrm{d}x'}\right)\mathrm{d}x', \tag{11}$$

where $\mathrm{d}E/\mathrm{d}x'$ the energy- and material-dependant stopping power.

2.6.3. Data filtering

In pCT reconstruction many recorded events are not considered for calculation, since they display large deviation from the expected averages and may result in large position errors on the calculated path. We present statistics for the case of filtering out the 5% of the tracks of the evaluation set with the larger angular deviation and when no such filtration is applied. For the N_{2D}, N_{RL} and XGB models and path averaging method the filtering criterion is the absolute value of angular deflection $|\theta_{y1}|$ projected on *xy*-plane (plane angle). Another filtering scheme is applied to remove 5% of protons with larger positional error, predicted by the N_{Err} network. No filtration has been applied on the training data.

2.6.4. Effect on pCT image resolution

An estimation of the modulation transfer function (MTF) was also performed to quantify the effect of the limited path accuracy in pCT image resolution. Calculation was done for the center of a cylindrical water phantom, 20 cm in diameter, and was based on the statistics of the residuals for the 20 cm water layer, presented in figure 5(a). We used the following equation proposed by Plautz *et al* (2016):

$$\mathrm{MTF}(f) = \exp(-2\pi\sigma_0^2 f^2), \tag{12}$$

where f is the spatial frequency in units of lp/mm and σ_0^2 the error in lateral displacement, assumed Gaussian-distributed, at the depth of 10 cm on the 20 cm water phantom for 200 MeV protons. Values of $\mathrm{MTF}_{10\%}$ and $\mathrm{MTF}_{50\%}$ for the four path reconstruction methods were calculated for the case of 5% filtration on proton paths.

Accepted Manuscript

3. Results

Training times ranged between 57 min (N_{2D}, 306 iterations) and 160 min (XGB, 3843 iterations), on the platforms specified in section 2.2. Computation times per 10^6 paths were 2sec and 10 sec for N_{2D} and XGB respectively. Table 2 presents the maximum - along the path - RMS error for 200 MeV, 230 MeV and 250 MeV protons passing through the paediatric head, the abdomen and the pelvis phantoms, respectively. Because this error depends on the path length inside the object, errors are presented for three values of such lengths (17±0.5cm for head phantom, 22±0.5 cm for pelvis and 31±0.5cm for abdomen phantom). When 5% of the paths with the highest angular deflection are filtered out, N_{2D} and XGB present errors, ranging from 0.35 mm for the head phantom for proton initial energy of 200 MeV, up to 0.9 mm for abdomen phantom, for 230 MeV protons. Statistics are presented for sets of 100k proton paths, in each class, identical for all methods. Figure 4 plots the 5%, 25%, 75% and 95% percentiles of the (near half-normal) distribution of the absolute value of the path error at mid depth, for 200 MeV and 250 MeV protons passing through the pelvis phantom (unfiltered data). Except of the short path lengths where all four methods seem to converge in error, both N_{2D} and XGB perform better than the analytical method that is based on likelihood maximisation. The figure also shows that the error increases with the path length and can reach or exceed 1.5 mm.

Table 2 shows also that when radiation length, evaluated on 100 points along the path is included in the input of the N_{RL} model, the RMS error is the same as N_{2D} (which does not include such information). The same is observed for different numbers of evaluation points. Table 2 shows the error of the network N_{3D} that gives the coordinates of the path points directly in the coordinate system assumed in the reconstruction. For the case of no filtering these 3D error values are divided by $\sqrt{2}$ in order to be comparable with the 2D errors given by the other methods, assuming motion independence along *y*- and *z*-axes. In the case of 5% filtering the N_{3D} is applied to the same data as N_{2D} (with initial direction along the *x*-axis), but reported is the error along *y*-axis.

Table 2. Maximum, along the path, RMS error in lateral displacement (mm) for protons incident on paediatric head phantom, abdomen phantom and pelvis phantom.

Track Reconstruction Method	**Data filtering**	**Pead. Head** 200MeV Path length: 17±0.5cm	**Pelvis** 200MeV Path length: 22±0.5cm	**Abdomen** 230MeV Path length: 31±0.5cm
N_{2D}	0%	0.42	0.73	1.03
	5%	0.35	0.63 (0.60)*	0.88
N_{RL}, k = 100	0%	0.45	0.73	1.06
	5%	0.35	0.63	0.90
N_{3D} **	0%	0.43	0.74	1.05
	5%	0.36	0.64	0.90
XGB	0%	0.41	0.72	1.00
	5%	0.35	0.62 (0.60)*	0.86
Path Averaging	0%	0.43	0.76	1.05
	5%	0.35	0.63 (0.60)*	0.88
Likelihood Maximization (Schulte *et al*, 2008)	0%	0.49	0.90	1.44
	5%	0.37	0.71 (0.67)*	1.03

( )* : values after 5% filtering based on N_{Err} error prediction.

Accepted Manuscript

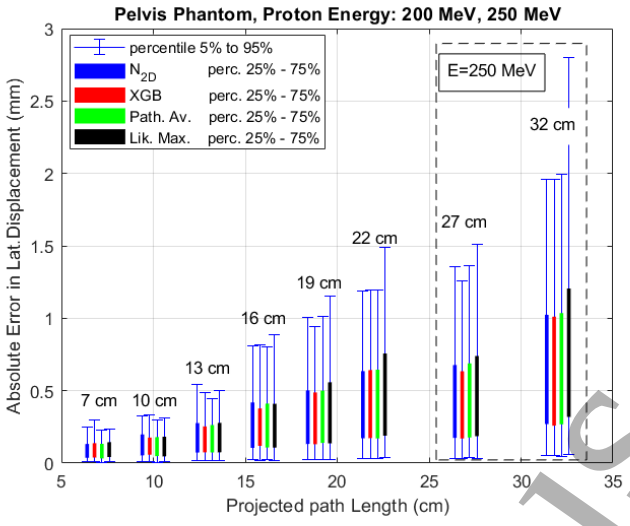


Figure 4. Diagram comparing the 5%, 25%, 75% and 95% percentiles of the maximum along the path absolute error, for the four methods considered in this study. No data filtration is applied.

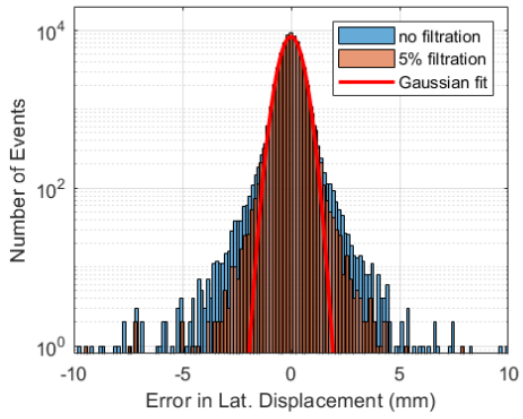
Table 3 presents the maximum along the path RMS error in lateral displacement for 200 MeV and 250 MeV protons travelling through a 20 cm in thickness water layer. It shows a significant advantage of the N_{2D} and the XGB over the likelihood maximization method. For 200 MeV protons paths after 5% filtering the accuracy improvement is 12.3% for both. Maximum - along the path - RMS errors by path averaging method are slightly higher than those of N_{2D} and XGB when a set of 2×10^6 paths is used for path searching. Figure 5(a) presents the histogram of errors in lateral displacement at depth of 10 cm, for paths calculated by the XGB model, showing a deviation from the Gaussian distribution, even after 5% filtering. Figure 5(b) shows the correlation between the XGB– predicted and actual MC values at depth of 10 cm, while figure 5(c) compares the error distributions of the studied methods in terms of cumulative density function.

Table 3. Maximum (along the path) RMS errors (mm) for 200 MeV, 230 MeV and 250 MeV protons incident on water layers.

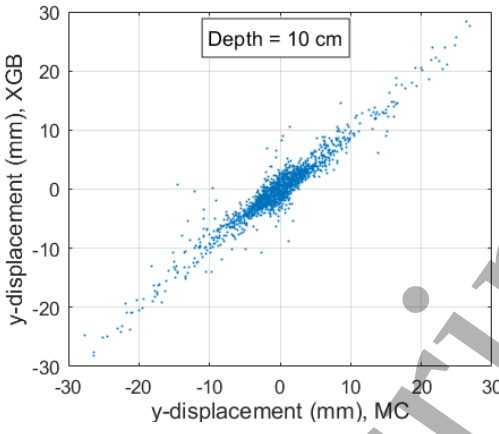
Track Reconstruction Method	Data filtering	200 MeV 20 cm	250 MeV 20 cm	230 MeV 30 cm
N_{2D}	0%	0.58	0.43	1.05
	5%	0.50 (0.46)*	0.36	0.95
XGB	0%	0.57	0.43	1.02
	5%	0.50 (0.46)*	0.35	0.93
Path averaging	5%	0.50 (0.46)*	0.37	0.96
Likelihood Maximization (Schulte *et al*, 2008)	0%	0.97	0.75	1.52
	5%	0.57 (0.53)*	0.44	1.13

$(\,)^*$: values after 5% filtering based on N_{Err} error prediction.

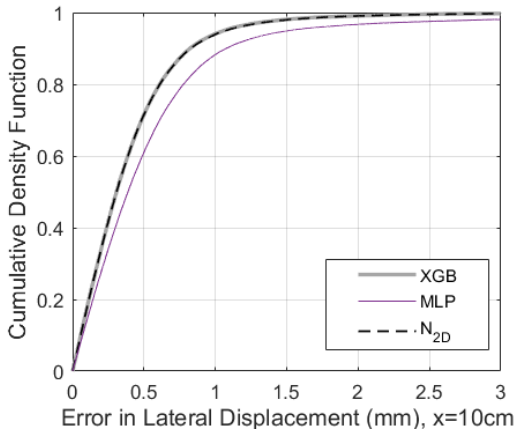
Accepted Manuscript



(a)



(b)



(c)

Figure 5. a) Distribution of the XGB- path error in lateral displacement at depth of 10 cm for 200 MeV protons on 20 cm of water. (b) Correlation between the XGB- predicted lateral displacement at depth of 10 cm vs the actual MC value. Correlation coefficient is 0.95. (c) Cumulative density function of the maximum along the path error distributions of paths reconstructed by XGB, N_{2D}, path averaging and likelihood maximization.

Table 4 presents the $MTF_{10\%}$ and $MTF_{50\%}$ given by the four path calculation methods, at the center of a cylindrical water phantom, 20 cm in diameter, scanned by 200 MeV protons. Filtration of 5% on angular deflection has been applied.

Table 4. Values of $MTF_{10\%}$ and $MTF_{50\%}$ at the center of a cylindrical water phantom, 20 cm in diameter. Proton energy: 200 MeV, data filtration on angle: 5%.

Track Rec. Method	$MTF_{10\%}$ (lp/mm)	$MTF_{50\%}$ (lp/mm)
N_{2D}	0.74	0.41
XGB	0.74	0.41
Path averaging	0.74	0.41
Likelihood Max/ion	0.65	0.36

Figure 6 compares absolute values of the actual path error (of the paths reconstructed by N_{2D}) with the error predicted by the N_{Err} for 200 MeV protons passing through the pelvis phantom. The RMS discrepancy between the above errors is 0.34

Accepted Manuscript

mm, with correlation coefficient of 0.49 (figure 6). Despite the suboptimal correlation, the N_{Err} offers a better filtering criterion than the deflection angle. We applied this (based on N_{Err}) criterion to filter out the 5% of the tracks that are expected to give the larger errors in position in the cases of pelvic phantom and 20 cm thick water object, for proton energy of 200 MeV. As Table 2 and 3 show (values in parentheses), the remaining tracks, reconstructed by N_{2D} give a maximum along the path error that is about 4.8% and 8.0% respectively lower than that given by filtration based on the angle. This accuracy improvement is also observed for the other three methods of path calculation.

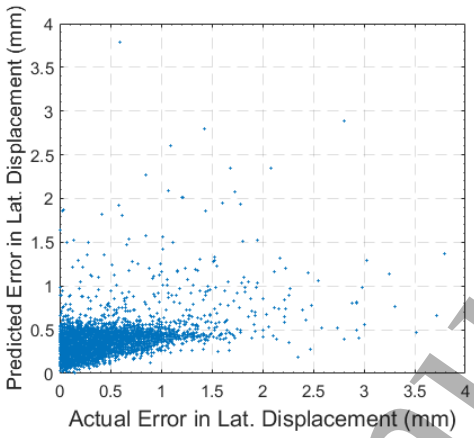
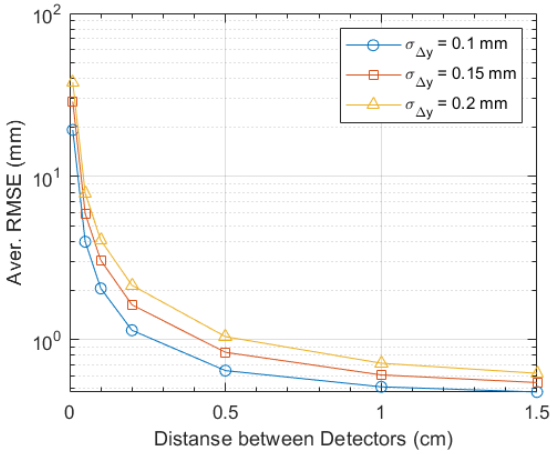


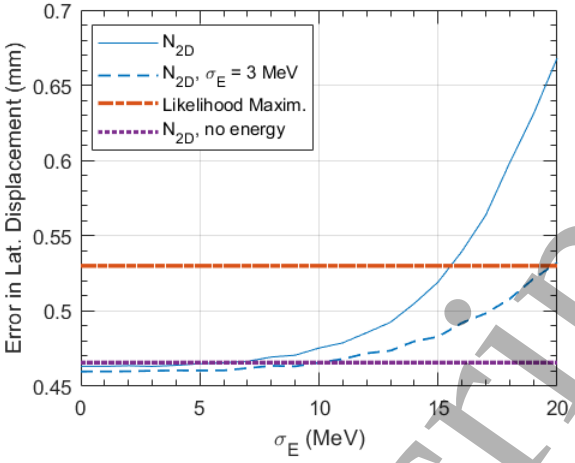
Figure 6. Correlation between the N_{Err} path predicted error in lateral displacement at the middle of the path and the actual error, equal to the difference between the N_{2D} -calculated path and the MC path. Data are from pelvis phantom, for proton energy of 200 MeV.

Figure 7(a) shows the dependence of the average- along the path- RMSE on the distance between detectors d for three values of standard deviation of the Gaussian – assumed detector position error. Average RMSE increases with the detector error $\sigma_{\Delta y}$ and is maximised for small gaps between detectors, due to the increase of the angular error. Figure 7(b) shows the error in lateral displacement at the middle of the path ($x = 10$ cm), calculated by the N_{2D}, as a function of the standard deviation σ_E of the Gaussian error in measurement of the proton exit energy E_1. The error is almost constant for $\sigma_E < 5$ MeV, and rapidly increases for $\sigma_E > 10$ MeV. The path error is smaller when an energy measurement error of σ_E=3 MeV has been applied on the training dataset. Also marked in the figure is the error when exit energy E_1 is not included in the training dataset, as well as the error of the path calculated by likelihood maximization, which does not depend on the energy error since the model is based on average energy profile along the path.

Accepted Manuscript



(a)



(b)

Figure 7. a) Dependence of the average- along the path- RMSE on distance between detectors for three values of standard deviation of the Gaussian – assumed detector position error. B) Escalation of the RSME at depth of 10 cm (middle point of the 20 cm water layer) as a function of the Gaussian- assumed error in measurement of the exit energy, for noiseless and noisy data.

Accepted Manuscript

4. Discussion

Our work shows the ability of ML to provide a simple, easy to implement, time effective and accurate solution for proton tracking in pCT, a problem characterised by the typically complicated mathematics of the transport equation. Analytical models provide solutions of adequate accuracy, leaving however a margin for improvement. In the typical case of 200 MeV protons on 20 cm of water, likelihood maximisation based on Fermi-Eyges gives RMS error of 0.57 mm (table 3, column of 5% filtration), while that based on Molière's model gives RMS error of 0.53 mm (Lazos *et al* 2020). These values are 14% and 6% respectively, higher than what the error of the N_{2D} (0.50 mm). On the other hand, ML does not rely on an analytical model and its respective assumptions and is able to reproduce the accuracy of the training data. ML also reduces the computational time, as analytically estimating the proton path is very time consuming but computing it from a trained model is rapid. As table 4 shows, the improved track accuracy has a direct effect on the resolution in terms of MTF. Our results also show that the N_{Err} predicts better the path error, enabling a criterion for filtering out paths that may affect the quality of the reconstructed image. Combining deflection angle, lateral displacement and exit energy and incorporating a more accurate high angle single scattering model, this criterion is more comprehensive than one based only on the deflection angle, or build upon Fermi-Eyges model (Williams 2004) that (criterion) does not account for deflection angle and displacement. Figure 7 shows that the path error significantly increases with the measurement error on the trackers, while energy error affects calculations to a lesser degree. The latter error reduces when random errors in energy are incorporated in the training data.

Even though average track is conceptually different with the most likely path, given by other methods that are based on likelihood maximization, the approach of path averaging provides paths completely in line with the requirements of pCT. Despite the lack of practicality, since a sufficient number of matching paths must be available for every path in question, thus preventing its use in practice, the method is found useful in the framework of this study, particularly for providing a baseline for evaluation. This way also of calculation does not require any assumption regarding the shape or the form of the solution, however its accuracy maybe affected by the non-zero tolerance values regarding the selection of the subset of the paths, that match the one in question.

Our study shows that there is no benefit in accuracy from including the along-the-path material information in the NN-based model for the inhomogeneities found in the head, pelvis and abdomen phantoms, as long as the exit energy is included in the input variables. Since the phenomenological base the NNs work on, makes it difficult to generalize on findings, it should be mentioned that this conclusion may not be valid for conditions other than the specific ones assumed in the study. These conditions include the inhomogeneities of the simulated phantoms (thus excluding metal implants, and high- or low-density materials, often used in clinical practice in contact with the patient), the specific material parameter (radiation length) along the actual MC track and the accuracy that the studied NNs can achieve. Our study does not exclude the possibility of improved accuracy if such material information, but lateral to the path, is included in a ML model (Kellaf *et al* 2019). This finding may seem controversial, since in scattering problems, the depth of inhomogenities is expected to have an effect on the lateral spread of scattered particles. As Eyges mentioned (Eyges 1948), this comes out directly from the mathematics of the Fermi-Eyges model. However our findings are in agreement with the literature. In a likelihood maximisation approach, Lazos *et al* (2020) found no accuracy improvement when RSP values from the reconstructed pCT images were included in the likelihood calculation. On the same theoretical concept, Collins-Fekete *et al* (2017) found also no or little benefit from utilization of information (along the track) from pre-existing x-ray CT images, while Brook and Penfold (2020) found no improvement for a head phantom. It should be also noted that although the exit energy E_1 summarises the effect of the SPR along the path, there is not redundancy by including SPR values in the input variables, since the energy loss is a statistical process and subsequently deviations from the average energy loss are expected.

Accepted Manuscript

Even the proton tracks have fine details, because of the stochastic nature of the scattering, paths reconstructed by heuristic methods, such as NN, XBG and path averaging, without making any assumption regarding their shape, are found to be smooth curves. This supports the evidence that 3rd degree parametric curves such as cubic splines or cubic Bezier curves can be used to represent paths, without reduction in accuracy because of their limited flexibility.

We conclude this discussion by reviewing the major limitations of our approach. Being an ad hoc solution based on synthetic data, the ML models are subjected on any limitation, inaccuracy and systematic error of the training data. The accuracy of the proposed ML models is questionable outside the area they are trained, due to incomplete formation of the mapping function. This is particularly important in the proton tracking problem because three of the input variables (angular deflection, lateral displacement and exit energy) follow uneven density distributions that also allow rare occurrences of extreme values. Considering that the number of proton events, collected in a pCT scan is in the order of 10^8, far exceeding the about 5×10^6 size of the training set, it is certain that extreme values outside the limits of the training set will appear. In this case the accuracy of the ML models will be challenged without any data filtering. We selected the size of 5×10^6 tracks to ensure diversity and adequately represent anatomical sites, energy values and path deviations. However, increasing the training set to address the above concern is not a practical solution.

Another concern is that we use the same source of data (Geant4 simulation) to train and evaluate the ML models, as well as to compare their performance against other methods, committing the so called “inverse crime” (Wirgin 2004). If there is a systematic error in our implementation of the MC simulation, the proposed ML models will reproduce it and in comparison with a third method - likelihood maximization in this case - it will be attributed as error of the latter. However, accuracy of the Geant4 is widely manifested, while the code has been extensively used for testing methods in radiation transport, including the ones in proton tracking in pCT. In this context, our approach can be seen as an efficient parametrization of the output of the Geant4 code in the setting of the pCT that attempts to bring - under the limitations discussed above - the MC accuracy into the problem.

Accepted Manuscript

5. Conclusions

Our study succeeded to show that ML can provide us with accurate, time effective, practical and easy to implement solutions to the proton tracking problem in pCT, thus giving the possibility of further improving spatial resolution, which closely depends on the precision proton tracks are determined. By incorporating a comprehensive MC–based stochastic multiple scattering model, the proposed methods are proved more accurate than the likelihood maximization that utilizes either Fermi-Eyges model or Molière theory to analytically describe proton scatter. In agreement with the literature, our study also finds that material information along the proton tracks, in terms of either RSP or radiation length does not improve the accuracy of the path calculation, for the inhomogeneities found in a head and a pelvis phantoms. Furthermore, ML enables the prediction of the track error, thus providing a more comprehensive and sensitive criterion for removing tracks that may harm the reconstructed image.

Accepted Manuscript

Acknowledgements

Dimitrios Lazos and Nikolaos Dikaios are supported by an EPSRC research grant (EP/R009988/1). Nikolaos Dikaios is also supported by a fellowship in Royal Society (INF/R1/19103). Charles-Antoine Collins-Fekete is supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) No. 502962 – 2017.

Accepted Manuscript

References

Agostinelli S, Allison J, Amako K, Apostolakis J, Araujo H, Arce P, Asai M, Axen D, Banerjee S, and Barrand G 2003 Geant4–a simulation toolkit *Nucl. Instrum. Methods Phys. Res*. A **506** 250–303

Bopp C, Rescigno R, Rousseau M and Brasse D 2014 The impact of tracking system properties on the most likely path estimation in proton CT. *Phys Med Biol.* **59** N197–210

Brooke M D and Penfold S N 2020 An inhomogeneous most likely path formalism for proton computed tomography. *Phys. Med.* **70** 184–95

Chen T and Guestrin C 2016 XGBoost: A Scalable Tree Boosting System (arXiv: 1603.02754)

Collins-Fekete C A, Doolan P, Dias M F, Beaulieu L and Seco J 2015 Developing a phenomenological model of the proton trajectory within a heterogeneous medium required for proton imaging *Phys. Med. Biol.* **60** 5071–82

Collins-Fekete C A, Bär E, Volz L, Bouchard H, Beaulieu L and Seco J 2017 Extension of the Fermi–Eyges most-likely path in heterogeneous medium with prior knowledge information *Phys. Med. Biol.* **62** 9207–19

Kingma D P and Ba J L 2014 ADAM: A method for stochastic optimization ICLR (arXiv:1412.6980)

Erdelyi B 2009 A comprehensive study of the most likely path formalism for proton-computed tomography *Phys. Med. Biol*. **54** 6095–122

Eyges L 1948 Multiple scattering with energy loss *Phys. Rev.* **74** 1534–5

Gao B and Pavel L 2018 On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning (arXiv:1704.00805)

Goudsmit S and Saunderson J L 1940 Multiple scattering of electrons *Phys. Rev.* **57** 24–9

Khellaf N *et al* 2019 Effects of transverse heterogeneities on the most likely path of protons *Phys. Med. Biol.* **64** 065003

Krah N, Létang J M and Rit S 2019 Polynomial Modelling of Proton Trajectories in Homogeneous Media for Fast Most Likely Path Estimation and Trajectory Simulation *Phys. Med. Biol.* **64** 195014

Lazos D, Collins-Fekete C-A, Evans P and Dikaios N 2020 Molière maximum likelihood proton path estimation approximated by cubic Bézier curve for scatter corrected proton CT reconstruction *Phys. Med. Biol*. **65** 175003

Li L, Liang Z, Singanallur J V, Satogata T J, Williams D C and Schulte R W 2006 Reconstruction for proton computed tomography by tracing proton trajectories: A Monte Carlo study *Med. Phys.* **33** 699 -706

Molière G 1948 Theorie der Streuung schneller geladener Teilchen II. Mehrfach- und Vielfachstreuung *Z. Nat. Forsch.* A **3** 78–97

Pedregosa *et al* 2011 Scikit-learn: Machine Learning in Python *J Mach Learn Res* **12** 2825–30

Plautz TE, Bashkirov V, Giacometti V, Hurley RF, Johnson RP, Piersimoni P, Sadrozinski HF, Schulte RW and Zatserklyaniy A 2016 An evaluation of spatial resolution of a prototype proton CT scanner *Med. Phys.***43** 6291–300

Schulte R W, Penfold S N, Tafas J T and Schubert K E 2008 A maximum likelihood proton path formalism for application in proton computed tomography *Med. Phys*. **35** 4849–56

Tsaris A *et al* The HEP.TrkX Project: Deep Learning for Particle Tracking 2018 *J. Phys.: Conf. Ser.* **1085** 042023

Urban L 2006 A multiple scattering model in Geant4 *CERNOPEN-2006-077* (Geneva: CERN) pp 1–14 (https://cds.cern.ch/record/1004190?ln=fr)

Wang D, Mackie T R and Tomé W A 2010 On the use of a proton path probability map for proton computed tomography reconstruction *Med. Phys.* **37** 4138–45

Williams D C 2004 The most likely path of an energetic charged particle through a uniform medium *Phys. Med. Biol.* **49** 2899–912

Wirgin A 2004 The inverse crime (ArXiv preprint math-ph/0401050)

Accepted Manuscript