Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography

Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography

P J Doolan[1], M Testa[2], G Sharp[2], E H Bentefour[3], G Royle[1] and H-M Lu[2]

[1] Department of Medical Physics and Bioengineering, University College London, London, UK

[2] Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA

[3] Ion Beam Applications (IBA), 3 Chemin du Cyclotron, Louvain la Neuve, B-1348, Belgium

E-mail: paul.doolan.09@ucl.ac.uk

CrossMark

Abstract

A simple robust optimizer has been developed that can produce patient-specific calibration curves to convert x-ray computed tomography (CT) numbers to relative stopping powers (HU-RSPs) for proton therapy treatment planning. The difference between a digitally reconstructed radiograph water-equivalent path length (DRR$_{WEPL}$) map through the x-ray CT dataset and a proton radiograph (set as the ground truth) is minimized by optimizing the HU-RSP calibration curve. The function of the optimizer is validated with synthetic datasets that contain no noise and its robustness is shown against CT noise. Application of the procedure is then demonstrated on a plastic and a real tissue phantom, with proton radiographs produced using a single detector. The mean errors using generic/optimized calibration curves between the DRR$_{WEPL}$ map and the proton radiograph were 1.8/0.4% for a plastic phantom and $-2.1/-0.2\%$ for a real tissue phantom. It was then demonstrated that these optimized calibration curves offer a better prediction of the water equivalent path length at a therapeutic depth. We believe that these promising results are suggestive that a single proton radiograph could be used to generate a patient-specific calibration curve as part of the current proton treatment planning workflow.

1. Introduction

Proton therapy has the potential to deliver better dose distributions than radiotherapy by photon/electron beams, but there are a large number of uncertainties that make predicting the beam range in the patient difficult. Currently, these uncertainties are accounted for with increased planning margins, resulting in the irradiation of more tissue than is necessary. One of the largest uncertainties comes from the conversion of the patient anatomical information from one of photon attenuation to one of proton energy loss, requiring x-ray computed tomography (CT) numbers to be converted into relative stopping powers (RSPs). This process is thought to contribute 0.5–1.8% of the total proton beam range uncertainty, dependent on whether the uncertainty in the mean ionization energy is considered (Schaffner and Pedroni 1998, Paganetti 2012).

Currently most centres use a single function (referred to in the remainder of this document as the HU-RSP curve) for this conversion, although it is known that there is not a one-to-one correspondence between CT numbers and RSPs (Yang *et al* 2012). The stoichiometric approach first proposed by Schneider *et al* (1996) is widely accepted as the most accurate method for producing the calibration curve for this conversion. The final step of this procedure involves calculation of the theoretical RSPs of human biological tissues, using composition data for average, healthy, adults (Woodard and White 1986, White *et al* 1987, ICRU 1989). As such, the stoichiometric calibration is not specific to the patient being imaged. Yang *et al* (2010) showed that neglecting typical patient-to-patient variations in density (4%), calcium content (2%) or hydrogen content (1%) causes inaccuracies in the RSP prediction of 2.2%, 2.0% and 1.3%, respectively.

Proton radiography offers a potential solution to this discrepancy as it produces maps in units of water-equivalent path length (WEPL). Reconstruction of individual proton radiographic projections allows a 3D map of the patient RSP to be generated, in what is referred to as 'proton CT'. Proton radiography is not used routinely, for this or any other purpose, however, largely because the most common approach requires high speed tracking of the protons before and after the patient and measurement of the residual energy of the protons exiting the patient (Schneider and Pedroni 1995, Pemler *et al* 1999, Schneider *et al* 2004, Schulte *et al* 2004, Shinoda *et al* 2006, Talamonti *et al* 2010, Sipala *et al* 2011). The majority of these prototype designs are bulky, expensive and difficult to incorporate into the clinical environment.

Using only a single detector positioned beyond the patient offers a more applicable solution for proton radiography and as such there has been much research into this field recently (Gelover-Reyes *et al* 2011, Seco and Depauw 2011, Telsemeyer *et al* 2012, Testa *et al* 2013, Poludniowski *et al* 2014). The general concept is to measure a bulk dose. However, while convenient, not tracking protons means the individual paths cannot be reconstructed and the bulk dose becomes mixed with protons that have taken different paths through the patient. Methods therefore have to be introduced to detect and handle this effect, known as 'range mixing' (Bentefour *et al* 2012).

Optimizing the calibration curve with proton radiography offers the potential to combine the mapping of RSPs from radiography with the high spatial resolution of the x-ray CT. In a single previous effort to produce patient-specific calibration curves using proton tracking radiography, Schneider *et al* (2005) split the calibration curve in the soft tissue region

(between −200 and 135 HU) into five parts and randomly varied the curve 20 000 times until the difference between the proton radiograph and a digitally reconstructed radiograph WEPL (DRR$_{\text{WEPL}}$, generated from the x-ray CT dataset) is minimized. Only protons that travelled straight through the proton tracking modules were considered for analysis. Their work showed promising results, with an improvement in the mean range prediction from 3.6 to 0.4 mm for a dog patient. However, no further investigation of the technique and its introduction to routine clinical practice has since been reported by this or any other groups. In this work we present a method for optimizing the calibration curve using single-detector proton radiography. Specifically, we:

(a) Devise and implement an optimization scheme that minimizes the difference between the measured proton radiograph (known WEPL map) and DRR$_{\text{WEPL}}$ map generated from the x-ray CT dataset, by varying the HU-RSP calibration curve.
(b) Validate the optimizer against a set of synthetic datasets, in which the CT number and RSP are controlled and known.
(c) Demonstrate the full approach on phantoms, using WEPL maps measured by single-detector time-resolved proton radiography.
(d) Confirm that the estimate of the WEPL map at the treatment depth is improved using the optimized calibration curve, by comparing with a WEPL map measured by the method of dose extinction.

2. Methods and materials

2.1. Optimization scheme

The structure of the optimization procedure is shown schematically in figure 1. The DRR$_{\text{WEPL}}$ map is created using the Siddon transform implemented in Plastimatch (plastimatch.org), and the length of intersection of each ray trace with every voxel, together with the voxel CT HU value, is stored in a matrix file for future computations. A single ray corresponds to a single pixel in the DRR$_{\text{WEPL}}$ map, so the density of rays matches the resolution of the DRR$_{\text{WEPL}}$. The CT HU values are then converted into RSPs using the given HU-RSP calibration curve. For a single ray trace i, the individual intersection lengths l_j multiplied by the RSP of the voxel, summed across all the voxels j the ray passes through, gives the calculated WEPL, w_c^i, of the ray trace:

$$w_c^i = \sum_j l_j F\left(\text{HU}_j\right).$$

(1)

Combining all the ray traces together gives the DRR$_{\text{WEPL}}$ map. In equation (1) $F(\text{HU}_j)$ is a function that describes the calibration conversion and it is typically a single piecewise linear function:

$$F = \begin{cases} 0.001 & \text{for} & \text{HU} \leq -1000 \\ a + b\,\text{HU} & \text{for} & -1000 < \text{HU} \leq 0 \\ a' + b'\,\text{HU} & \text{for} & 0 < \text{HU} \leq 40. \\ \vdots & & \vdots \end{cases}$$

(2)

The division of the linear sections is a choice of the user. In this work all results are compared to a 'generic' calibration curve used for general treatment planning at our institution, formed of three linear portions (−1000 to 40 HU with a gradient of 0.001, 40 to 2990 HU with a gradient of 0.0005, and a line with zero gradient between 2995 and 9960 for titanium). For
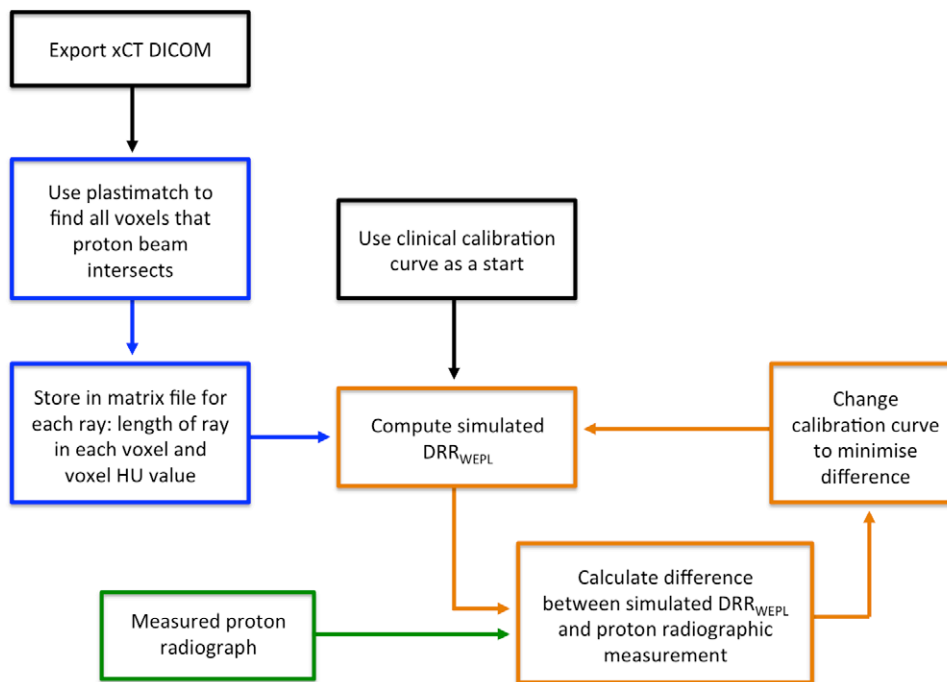
Figure 1. Schematic of the DRR optimization procedure. Plastimatch operations are shown in blue, measurements in green and optimization processes in orange.

optimization, this curve was split into 25 linear spline points, based on the material binning suggested for Monte Carlo simulations by Schneider *et al* (2000), with an additional point at water. The 26 points were [−1000; −950; −120; −83; −53; −23; 0; 7; 18; 80; 120; 200; 300; 400; 500; 600; 700; 800; 900; 1000; 1100; 1200; 1300; 1400; 1500; 2990] (the titanium correction was excluded from the optimization procedure as this would ordinarily be overridden as part of treatment planning).

For the optimization, the parameters *a*, *b*, *a'*, *b'*… in equation (2) were varied until the difference between the DRR$_{WEPL}$ and the proton radiographic WEPL reached a minimum. The cost function describing this difference *Δ*, which must be minimized, was thus defined:

$$\Delta = \sum_i \frac{(w_m^i - w_c^i)^2}{(w_m^i)^2},$$ (3)

where w_m^i is the proton radiography (measured) value of the ith pixel and w_c^i is the WEPL (calculated) value of the ith pixel.

The optimization was performed using Matlab's built-in Nelder–Mead optimization function, 'fminsearch'. In addition to the convenience of a Matlab built-in function, Nelder–Mead optimizations are more robust to local minima and do not require an equation to be provided for the derivative of the cost function (Lagarias *et al* 1998). Details of the input variables used in optimization scheme can be found in table 1. When optimizing, 24 spline points (two points were fixed, see below) were allowed to vary in the y-direction only (HU stays fixed but RSP changes). The number of variables, *n*, therefore was equal to 24. We also imposed other restrictions, namely that: (i) the curve must be monotonic; and (ii) that the curve must pass

Table 1. Input variables for the optimization.

Variable	Details	Value used
MaxFunEvals	Maximum number of evaluations of the cost function (where n is the number of variables = 24)	$200 \times n$
TolFun	Absolute tolerance on function value	1×10^{-4}
Linear sections	Number of individual lines that make up the piecewise linear calibration curve	25
Constraints	Must pass through the points of air and water Must be monotonic	Air: HU = −1000, RSP = 0.001 Water: HU = 0, RSP = 1

through the points of air (HU = −1000, RSP = 0.001) and water (HU = 0, RSP = 1); by introducing a high cost penalty in the optimization if these constraints were not met.

For the synthetic datasets, for which the desired solution was known, repeats ran on a while loop until the solution converged. For the real measured datasets we ran two repeats, each of 2400 iterations ($100 \times n$), using the output of the first iteration as the input for the second. This allows the optimizer to form another simplex, which has been shown ensures the Nelder–Mead algorithm can produce a locally optimal solution (Simon 2011).

2.2. Optimizer validation against simulated datasets

To validate the function of the optimizer we worked with synthetic data. Real CT datasets contain noise and real proton radiographs are subject to multiple Coulomb scattering (MCS), so it would be otherwise difficult to discern whether the optimizer had reduced the errors to the theoretical minimum. Using literature tissue compositions from Woodard and White (1986), White *et al* (1987) and ICRU (1989), theoretical CT numbers were calculated using the parameterization of our scanner and the standard procedure described by Yohannes *et al* (2012), while the RSPs were computed using the approximation of the Bethe-Bloch formula given by Schneider *et al* (1996). Simulated phantoms were constructed and perfect, known WEPL maps were generated based on the geometry. DRR$_{WEPL}$ maps were created with no divergence in the ray tracing to isolate the performance of the optimizer. The calibration curves were optimized against these known WEPL maps, for a set of synthetic datasets shown in figure 2.

2.2.1. Single materials, multiple materials and heterogeneous phantom.
The simplest task for the optimizer is to alter the calibration curve at a single point only, which is the case for homogeneous objects composed of one material. To test this, 30 individual spherical phantoms were simulated, each with a radius of 15 cm and composed of a single homogeneous material (tissues from table 2 of Yohannes *et al* (2012)). An example for the breast tissue is shown in figure 2(*a*). To test the ability of the optimizer to simultaneously reduce the error for multiple materials at once, a synthetic phantom was created (figure 2(*b*)) of $5 \times 5 \times 5$ cm³ blocks of 28 different materials across the whole HU range (same tissues as for the homogeneous spheres, excluding 'blood' and 'cell nucleus'). To verify that the optimizer works through heterogeneous materials, a synthetic phantom was composed of two of the multiple material phantoms (figure 2(*b*)) in succession with the rows of the second phantom offset. In both the multiple material and heterogeneous phantoms the voxels were $5 \times 5 \times 5$ mm³ in size, so each material block contained $10 \times 10 \times 10$ pixels.

2.2.2. Anthropomorphic phantom and real patient CT data.
An anthropomorphic pelvic phantom (custom made by The Phantom Laboratory, Greenwich, NY) was CT scanned
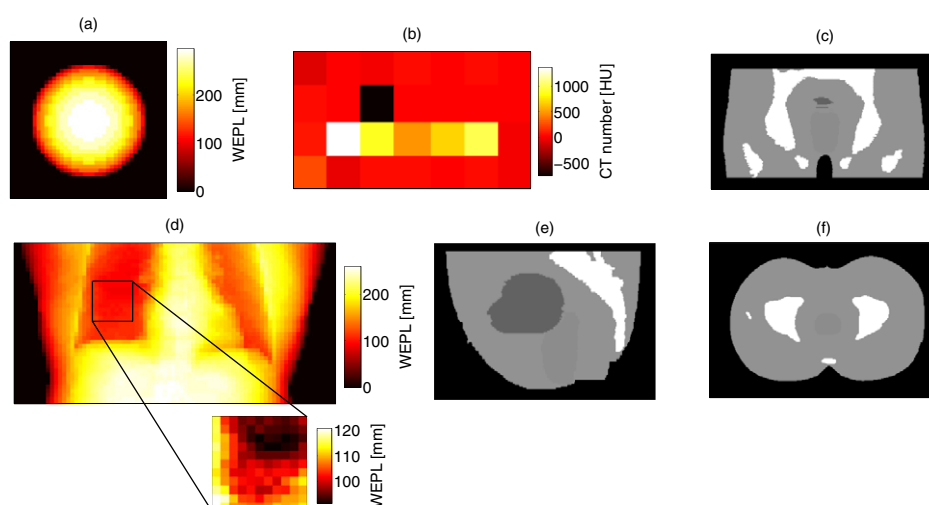
Figure 2. Synthetic phantoms used to validate the optimizer: (*a*) single, homogeneous material phantom; (*b*) multiple material phantom; anthropomorphic pelvic phantom in (*c*) coronal, (*e*) sagittal and (*f*) axial views; (*d*) lung patient, with region optimization region shown by the inset.

(figures 2(*c*), (*e*) and (*f*)). To avoid discrepancies due to CT artifacts and to allow a known WEPL map to be constructed, the CT numbers of each organ were overridden to a single value and assigned a single RSP. Although currently not used clinically because of the potential dose to the rectum, better proton range prediction could allow for the use of anterior proton fields for prostate treatments, which would offer a significant dose sparing advantage over photon treatments (Tang *et al* 2011). As such, optimization was applied over an anterior field. As a final test, the CT numbers of each organ in a lung patient were overridden to a single value (figure 2(*d*)). For real patients we only need to know how the curve should be optimized for the field to be treated, so for speed the optimization was performed on a smaller cropped region, as shown in the inset of figure 2(*d*).

2.3. Impact of CT noise

In all previous validation tests, optimization involved minimizing the difference between two perfect images: (i) a DRR_{WEPL} formed through a CT with no noise; and (ii) a simulated known WEPL map. In real patient cases the CT will contain noise, which is a stochastic phenomenon, usually assumed to have a Gaussian distribution (Chvetsov and Paige 2010). Therefore to investigate the impact on the optimization, the CT numbers of each organ in the pelvic phantom (figures 2(*c*), (*e*) and (*f*)) were overridden to a Gaussian distribution of values. Optimization was performed on datasets in which a range of noise levels (standard deviations of 1, 2, 3, 5 and 10%) was applied to the scaled CT numbers (+1000 HU compared to usual values) of each organ.

2.4. Proton radiographic and dose extinction measurements

All radiographic measurements were made in the proton gantry at Massachusetts General Hospital, using a spread-out-Bragg peak with a range of 20 cm and a modulation width of

18 cm (allowing us to capture all WEPLs within the range 2–20 cm). Proton radiographs were acquired using the time-resolved proton radiographic method, the full details of which can be found elsewhere (Lu 2008, Gottschalk *et al* 2011, Testa *et al* 2013). In summary, rotation of the range modulator (RM) wheel, at a typical well-defined 600 rpm, delivers a series of Bragg peaks spread out in depth. Using the RM wheel essentially as a precise clock, at each depth there is a unique time-varying dose pattern which can be converted to a water-equivalent depth based on a calibration in water.

Our calibration was based on the root-mean-square (RMS) deviation of each modulator cycle (Gottschalk *et al* 2011) and the detector used was an array of 249 semiconductor diodes arranged on an octagonal matrix (see figure 5(*b*)), with a diagonal pitch of 7.07 mm and a linear pitch of 10 mm separating diodes on the same row. The data acquisition module has a sampling time of 2 ms, which is sufficiently small to capture the 100 ms cycle of the RM wheel.

Optimization was conducted as detailed in section 2.1, with DRR_{WEPL} maps replicating the geometry of the proton radiographic measurements. In our gantry protons are assumed to originate from a single point 227 cm upstream of the isocentre, so the beams are divergent in the ray tracing. The DRR_{WEPL} map was produced at the same location as our diode array in the measurements, which was placed directly below the phantom (with no air gap). The set up for the proton radiographic measurements is shown in figure 3(*c*).

To determine if the optimization procedure results in a more accurate WEPL map at the therapeutic depth, we compared with the simple, robust method of dose extinction. In this method a controlled dose (10 MU) was delivered and increasing thicknesses of water-equivalent material were placed between the proton beam and the phantom until the dose to all of the detectors is zero. Depth dose curves for each diode were then formed, from which the thickness that stops 50% of our protons, also known as the R80, was computed. These R80 values were subtracted from the beam range (20 cm) to compute the WEPL as measured by dose extinction (DE_{WEPL}).

2.5. Plastic and real tissue phantoms

To test the optimization procedure we imaged the phantoms shown in figure 3 (and explained in the subsequent paragraphs). They are both composed of two halves. The reason for this is that radiographs are made through the entire patient, but the desired quantity for treatment is the WEPL to a therapeutic depth. Assuming the therapeutic depth is the distal edge of the top half of the phantom, the procedure consists of the following three steps: (i) acquire time-resolved proton radiograph through the entire phantom; (ii) optimize the calibration curve by minimizing the difference between the DRR_{WEPL} and the proton radiographic WEPL; and (iii) predict the WEPL to the distal edge of the top half using the optimized calibration curve. To validate that the optimization improves the WEPL estimate, we split apart the phantoms and conducted dose extinction measurements (section 2.4) on the top halves only.

In the top half of the plastic phantom (figure 3(*a*)), there are eight tightly packed inserts from the Gammex 467 phantom (Gammex Inc., Middleton, WI), arranged in a 12 cm PMMA box (1 cm thick walls) with a central divider. Water was used to fill in the gaps, up to the insert height (7 cm). The insert materials ranged from RSPs of 0.94 (Adipose) to 1.60 (Cortical bone), which together with the thickness of the PMMA box gave a range of WEPLs from 7.6 to 12.2 cm. The bottom half of the plastic phantom was a piece of homogeneous solid water, equivalent to 5.12 cm of water.

For the real tissue phantom, the top half (same 12 cm PMMA box as for the plastic phantom) was filled with a single piece of boneless beef chuck (fat left on), a non-smoothed pork
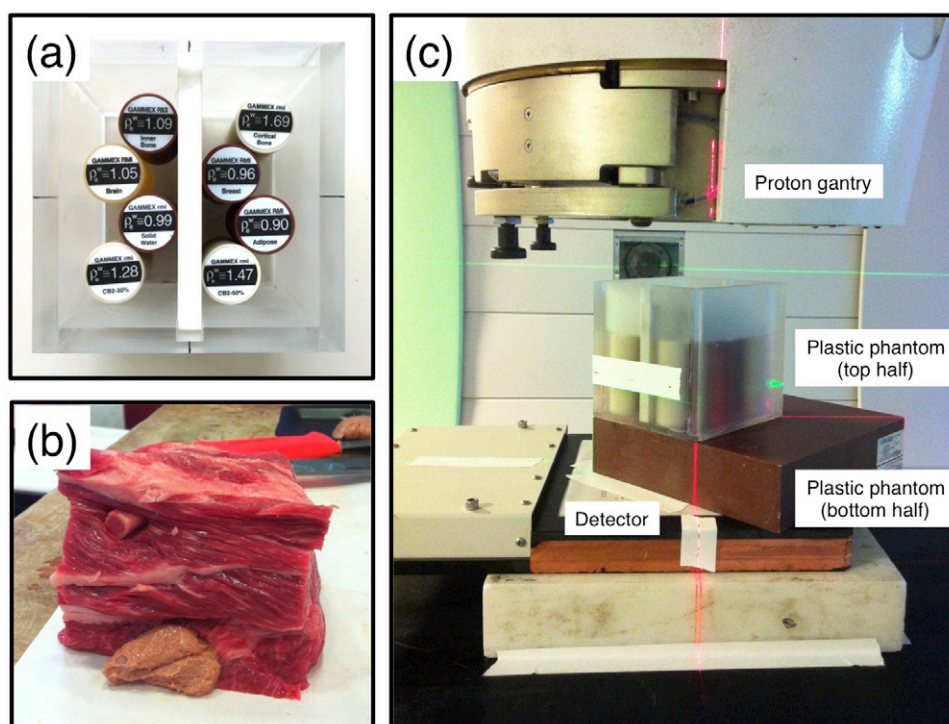
Figure 3. Phantoms imaged with proton radiography. (*a*) Plastic phantom, top-down view showing insert locations. (*b*) Top half of the real tissue phantom. (*c*) Set up for proton radiographic measurements.

bone and a strip of veal liver along the base (figure 3(*b*)). The bottom half was composed of a simple commercial lunchbox filled with another single piece of boneless beef chuck and another non-smoothed pork bone. The bones in the top and bottom halves were oriented 90° to each other, with some overlap. Given that the real tissue was not frozen or fixed during the experiment, CT scans were acquired approximately 1 h before and 30 min after the proton radiographic measurements to allow for quantification of any changes in tissue shape. The latter volume was warped to the former volume using a b-spline deformable registration (with the software Plastimatch) and tissue changes were computed along each direction to assess the stability of the tissue during the experiment.

2.6. Handling range mixing

The inability to be able to reconstruct individual proton tracks is one of the major obstacles that needs to be addressed for the routine use of single detector proton radiography. Since protons undergo MCS, they take a curved trajectory through the body and thus there can be multiple paths to a given point. If all of the most probable paths pass through the same amount of water-equivalent material, then all the protons at the given point will have a well-defined energy distribution (assuming the protons entering are mono-energetic) and it is possible to define the WEPL. If, however, protons can get to the point of interest via two (or more) equally probable paths with different energy losses, then the energy distribution is not well defined and the final signal becomes mixed. Since single detector proton radiography does not allow

Table 2. Differences between DRR$_{\text{WEPL}}$ and known WEPL maps, before and after optimization, for the set of synthetic datasets.

Test	Generic RMSE (mm)	Optimized RMSE (mm)	Reduction (%)
1. Homogeneous	2.72	2.6×10^{-6}	100.0
2. Multiple	0.69	0.09	87.0
3. Heterogeneous	1.44	9.7×10^{-6}	100.0
4. Anthropomorphic	4.13	0.10	97.6
5. Lung patient	3.92	0.01	99.7

us to reconstruct the proton paths, it is imperative that instances of range mixing are identified as the WEPL value is no longer reliable and the particular diode should not be used as part of the optimization.

For our purposes we applied a simple method to remove potentially range mixed diodes, by assuming that range mixing occurs at the boundaries of two materials. To identify the boundaries, we converted a 1 mm resolution DRR$_{\text{WEPL}}$ map into a map of local standard deviation by assessing the standard deviation of all surrounding pixels (on a 3×3 square). Any pixels having a standard deviation of >1% were defined to be close to a boundary and could thus be subject to range mixing. Any diode locations that overlapped with these pixels were then masked out during optimization.

3. Results

3.1. Optimizer validation against simulated datasets

The results for all the synthetic dataset validation tests are detailed in table 2. Tests 1 and 2 were completed on a 32 bit 4 GB CPU architecture, while tests 3, 4 and 5 were completed on a 64 bit 8 GB CPU architecture. The optimizer was able to either completely eradicate or significantly reduce the root-mean-square-error (RMSE) across the WEPL map in all tests, with a wide range of optimization times (18 s to 4.1 h) dependent on the complexity of the phantom and the CPU architecture. Figure 4 shows an example result for the anthropomorphic pelvic phantom: (*a*) shows the known total WEPL map; (*b*) shows the calibration curves in the soft tissue region before and after optimization. Before optimization, the DRR$_{\text{WEPL}}$ difference map and histogram of differences compared to the known WEPL map are shown in (*c*) and (*d*) respectively. The same results using the optimized calibration curve are shown in (*e*) and (*f*).

3.2. Impact of CT noise

The optimizer was shown to handle noise in the CT number well, with results before and after optimization detailed in table 3.

3.3. Optimization against time-resolved proton radiographs

Example proton radiographs of the real tissue phantom, acquired with the time-resolved technique, are shown in figure 5: a 3D map is shown in (*a*), a 2D map (showing the diode locations) is shown in (*b*), and the 2D map after masking of the pixels potentially subject to range mixing is shown in (*c*). As explained in the methods, optimization was only conducted on the unmasked diodes. The results following 120 and 125 min of optimization for the plastic and
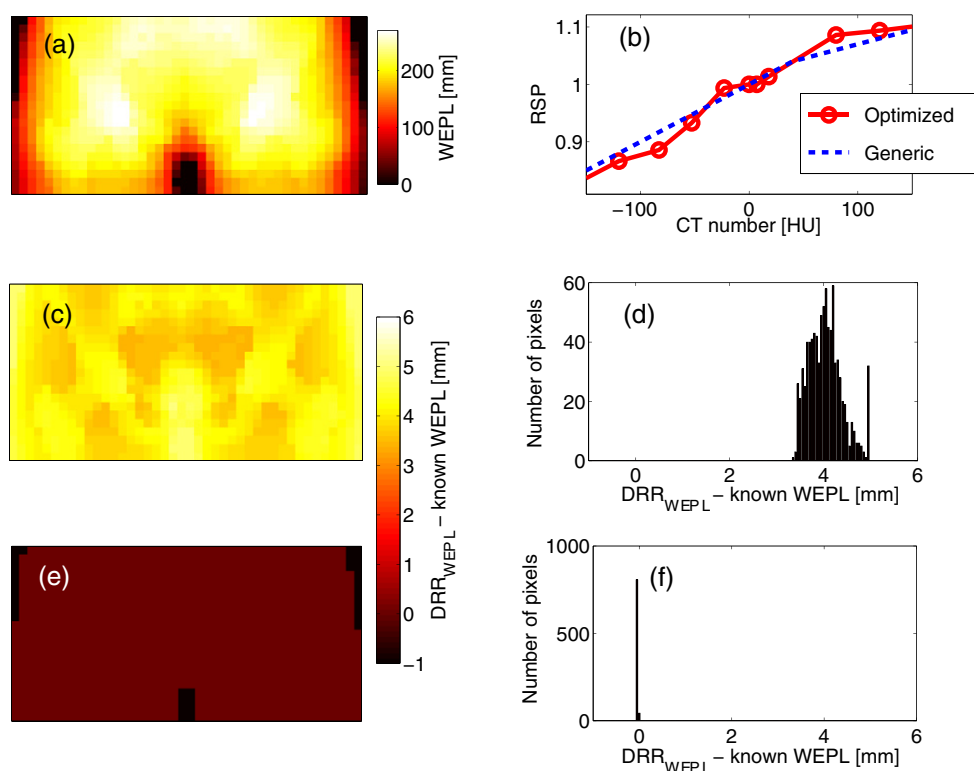
Figure 4. Optimization results for the anthropomorphic pelvic phantom. (*a*) Known total WEPL map. (*b*) Optimized and generic calibration curves in the soft tissue region. (*c*) DRR_WEPL difference map and (*d*) histogram of differences compared to the known WEPL map using the generic calibration curve. (*e,f*) Same as (*c,d*), but for the optimized calibration curve.

Table 3. Differences between DRR_WEPL and known WEPL maps, before (generic) and after (optimized) optimization, for the different robustness tests.

Test	Generic RMSE (%)	Generic μ/σ (%)	Optimized RMSE (%)	Optimized μ/σ (%)
CT noise (1%)	0.84	−0.84 / 0.10	0.03	0.00 / 0.03
CT noise (2%)	0.90	−0.89 / 0.08	0.03	0.00 / 0.03
CT noise (3%)	0.85	−0.85 / 0.07	0.02	0.00 / 0.02
CT noise (5%)	0.72	−0.72 / 0.07	0.01	0.00 / 0.01
CT noise (10%)	0.56	−0.55 / 0.09	0.02	0.00 / 0.02

real tissue phantoms, respectively, can be seen in figure 6. The relevant statistics before and after optimization are detailed in table 4.

From the registration it was found that the tissue remained quite static during the course of the 3 h experiment. With reference to the directions in figure 3, it was found the mean/max deformations were: −0.53/ −1.48 mm in the LR direction; −0.46/ −1.99 mm in the AP direction; and 0.07/ −1.51 mm in the SI direction.
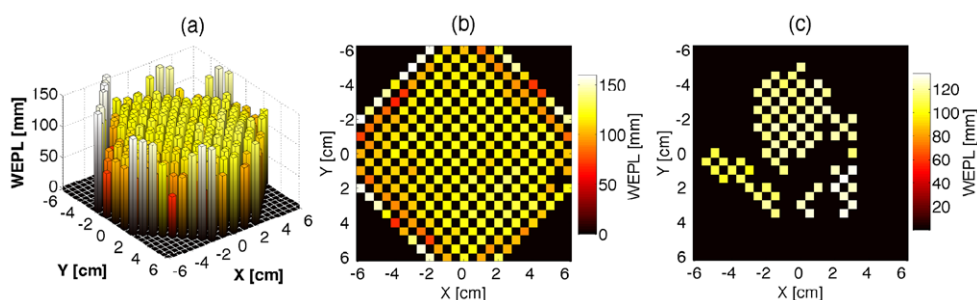
Figure 5. Proton radiographs of the real tissue phantom. (*a*) 3D WEPL map. (*b*) 2D WEPL map. (*c*) 2D WEPL map with range mixing mask applied.

3.4. Estimation of the WEPL at the therapeutic depth

To validate that the optimized calibration curve offers an improvement in the WEPL estimate at a therapeutic depth, independent measurements of the top half of the phantom were made using dose extinction. Integrated doses were recorded by the imager behind the top halves of the phantom, for a number of thicknesses of solid water on top. A comparison of the DRR_{WEPL} maps at the therapeutic depth (constructed using the generic and optimized calibration curves) with the DE_{WEPL} maps are shown in figure 7, with the relevant statistics detailed in table 4.

4. Discussion

4.1. Validation against synthetic datasets

Synthetic datasets were used to validate the optimizer. Based on the assigned RSPs for the HU values in the synthetic dataset, perfect WEPL maps could be created against which to optimize. In all cases the error defining the cost function was either completely eradicated or significantly reduced. The only cases where the optimizer failed to completely minimize the error were for the multiple material and heterogeneous phantoms. In these phantoms there were 28 different materials, but the optimized calibration curve was only split up into 25 linear sections. As such, it would have been very unlikely for the optimizer to completely eradicate the errors. The fact that there was a significant reduction demonstrates the optimizer's robustness to potentially conflicting requirements. The results also showed that the complexity of the dataset does not impact on the results, with complete eradication of the error for the anthropomorphic and patient geometries.

The perfect datasets were critical in constructing the optimizer, but it was also important to validate that the optimizer was robust in more realistic cases. It was found to function as expected with varying CT noise levels, with an almost identical final error following optimization. As to be expected, there was an intrinsic spread in the errors that the optimizer could not remove due to the noise.

4.2. Optimization against time-resolved proton radiographs

The results in figure 6 demonstrate that the procedure works when optimizing against real proton radiographs. For both the plastic and real tissue phantoms all computed statistics were improved. Following optimization on the plastic phantom the RMSE difference was reduced from 3.4 to 2.6% and the mean was reduced from 1.8 to 0.4%. For the real tissue phantom the
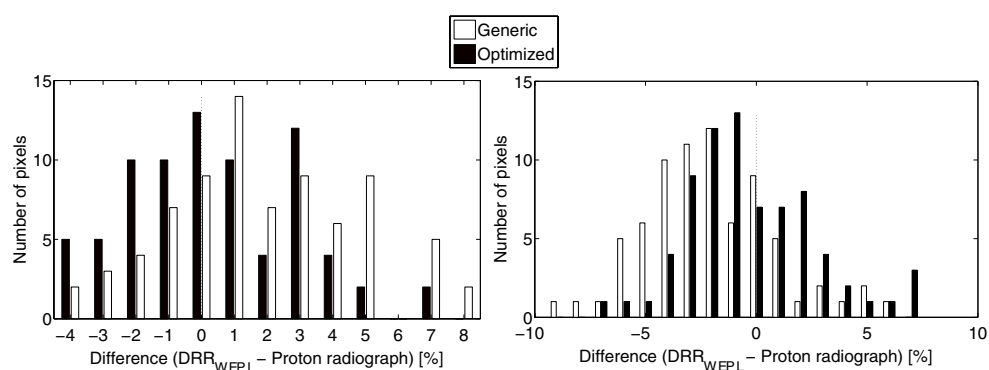
Figure 6. Histogram of differences for the generic and optimized calibration curves for (left) the plastic phantom and (right) the real tissue phantom.

Table 4. Differences between DRR_{WEPL} and proton radiographic WEPL or DE_{WEPL} maps, before (generic) and after (optimized) optimization, for the plastic and real tissue phantoms. Maximum and minimum values are the biggest differences in the positive and negative directions, respectively.

Test	Plastic RMSE	Plastic μ/σ	Plastic max/min	Real RMSE	Real μ/σ	Real max/min
Proton radiograph						
Generic	3.38%	1.83/2.86%	7.85/ −3.89%	3.57%	−2.06/2.93%	6.06/ −8.69%
Optimized	2.58%	0.36/2.57%	6.70/ −4.66%	2.93%	−0.23/2.94%	8.08/ −7.07%
DE_{WEPL}						
Optimized	3.36%	0.67/3.32%	7.62/ −7.14%	1.96%	−1.55/1.21%	2.66/ −4.25%
Generic	2.53%	−0.82/2.41%	4.01/ −4.37%	1.39%	0.56/1.28%	4.40/ −2.57%

RMSE was reduced from 3.6 to 2.9% and the mean error was reduced from −2.1 to −0.2%. As expected the improvements are more modest than for the synthetic datasets that had a single solution, but they are sufficiently large to warrant the clinical value of the technique.

For the real tissue phantom the tissue shape changes were small during the course of the experiment. Maximum changes of −1.48/ − 1.99/ −1.51 mm in the LR/AP/SI directions were found between the pre- and post-radiographic CT images, which were acquired 4 h apart in time. The time difference between the pre-radiographic CT image, upon which the optimization was performed, and the proton radiographic measurement was around 1 h. Assuming a constant rate of shape change, differences in the real tissue between the CT image and the proton radiograph would be less than 0.5 mm in all directions. Given the pitch between adjacent diodes was never less than 7 mm, we assume that such changes did not impact on our results.

4.3. Estimation of the WEPL at a therapeutic depth

Splitting apart the phantoms, dose extinction measurements allowed for the generation of WEPL maps at therapeutic depths. Comparing the DE_{WEPL} maps with DRR_{WEPL} maps at the same depth, it was clear that for both phantoms the optimized curve matched the DE_{WEPL} better than the generic curve. The RMSEs for the plastic phantom were 3.4/2.5% for generic/ optimized calibration curves. The same statistics for the real tissue phantom were 2.0/1.4%.
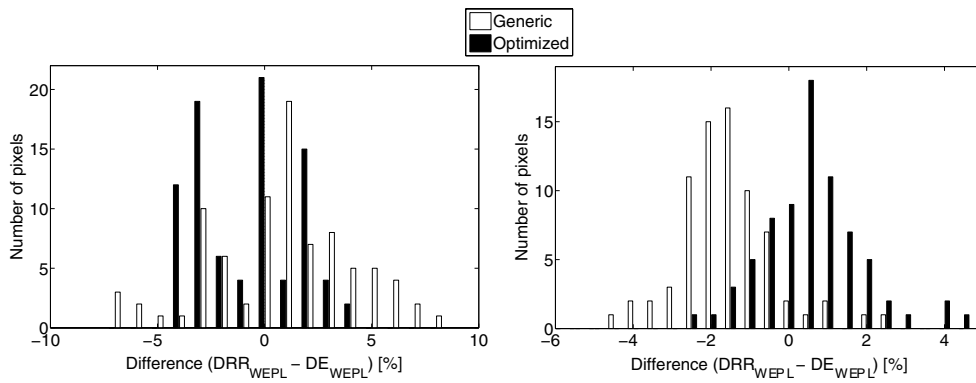
Figure 7. Comparison at therapeutic depth. Histogram of differences between the DRR$_{WEPL}$ and the DE$_{WEPL}$, before and after optimization, for (left) the plastic phantom and (right) the real tissue phantom.

These improvements are similar to those of the previous section, suggesting that optimization of the calibration curve using proton radiography across the whole patient allows for a more accurate estimate of the WEPL at a therapeutic depth.

4.4. Technique limitations

The technique proposed here relies on the ability to produce an accurate proton radiographic image. Optimizing against an inaccurate proton radiographic image would result in an inaccurate calibration curve being produced and an inaccurate estimate of the RSPs in the patient. We are confident, however, that the single-detector time-resolved approach we utilized in this study was able to produce accurate WEPL values as in a previous study the same equipment was able to achieve millimeter accuracy in a pure water phantom (Testa *et al* 2013).

Another source of uncertainty is the applicability of straight-line DRR ray tracing to proton paths. It is well known that protons do not take straight line paths due to the effect of MCS, and thus a straight line approximation would underestimate the WEPL. One potential solution would be to use the optimized calibration curve as input for a Monte Carlo simulation (by binning into materials and assigning compositions and densities) so that the effects of MCS are accounted for. The optimized calibration curve could then be fine-tuned, using the results of the simulation. We believe this underestimation to be a second order effect in our results, however. The maximum WEPL was 17.32 cm for the plastic phantom and 13.31 cm for the real tissue phantom. The Highland (1975) formula to compute the root mean square projected angle of scattering σ_s (radians) is given by,

$$\sigma_s = S_2 \frac{\sqrt{X/X_0}}{p\beta c} \left[1 + \varepsilon \log_{10}(X/X_0) \right], \tag{4}$$

where $S_2 = 14.1$ MeV and $\varepsilon = 1/9$ are empirical factors; X is the thickness of the scatterer (cm); X_0 is the radiation length (36.08 cm for water in our case as we use WEPLs); p is the momentum (MeV/c); and β is the particle velocity as a fraction of the speed of light c. At the energy

used (approximately 175 MeV) and at the largest thicknesses of our phantoms, $\sigma_s = 0.0292$ rad for the plastic phantom and $\sigma_s = 0.0253$ rad for the real tissue phantom. An angular deflection leads to a projected lateral deflection D as described by Leroy and Rancoita (2009):

$$D = \frac{\sigma_s}{\sqrt{3}} X. \tag{5}$$

In our experiments the average lateral deflections were thus $D = 0.29$ cm for the plastic phantom and $D = 0.19$ cm for the real tissue phantom. Given the closest pitch between adjacent diodes (0.7 cm) was larger than both of these values, we believe that a straight line approximation was valid. If the protons travel through material with a composition different from water, such as bone, it is possible these deflections will increase. The deflections will also increase if the distance between the patient and the detector is increased.

It is likely more accurate results will be produced by assuming a cubic spline path (Li *et al* 2006) or a most likely path (Williams 2004), however this requires the tracking of individual particles (which is not possible with a single detector). Additionally, we are not aware of any software that can currently easily compute the length of intersection of such paths with the CT dataset. This is suggested as an area of future work.

In the current implementation the optimization times are probably too long. This time is dependent on three main factors. (i) The voxel size and number of voxels in the CT dataset determines the number of intersections stored in the ray tracing matrix file, which affects the time to calculate the DRR$_{WEPL}$ map. This has to be recomputed for each iteration. (ii) The number of pixels in the two images, between which the difference is trying to be minimized, affects the time to calculate the cost function. (iii) The stopping tolerance of the optimization directly affects how long the optimization takes to converge. The first of these, computation of the DRR$_{WEPL}$ map using the ray tracing details, was the major factor for slow optimization times in our implementation. This could be significantly reduced, however, by producing field-specific calibration curves over much smaller volumes. The proton radiograph would be acquired in the same direction as the intended treatment beam, so only those tissues in the field of view would need to be optimized. It is also possible that an alternative optimization scheme, such as gradient descent, could also speed up the minimization.

Another limitation of our specific set of measurements was the spatial resolution (pitch between diodes up to 1 cm) of the proton imager. This could be easily improved by using one of the various promising technologies currently available such as fluorescent screens coupled to charge-coupled device cameras (Ryu *et al* 2008, Muraishi *et al* 2009, Bentefour *et al* 2013); commercial flat-panel detectors (Telsemeyer *et al* 2012); or complementary metal oxide semiconductor (CMOS) active pixel sensors (CMOS APS) (Gelover-Reyes *et al* 2011, Seco and Depauw 2011, Poludniowski *et al* 2014).

Additionally, production of the DRR$_{WEPL}$ map is reliant on the accuracy of the CT. As such, removal or reduction of artifacts such as beam hardening or streaking (with improved reconstruction algorithms or dual-energy CT) must be considered before producing the DRR.

4.5. Clinical application

The optimization scheme developed in this work is insensitive to the method of production of the proton radiograph. Although we used the time-resolved single detector proton radiographic method, the optimization scheme could, in theory, be used for proton tracking radiographic devices. Provided the ray tracing of the DRR can be effectively modeled (i.e. the

source-to-isocenter and source-to-imager distances are known), patient-specific curves could be generated.

We envisage the proton treatment planning workflow being updated as follows: (i) acquire an x-ray CT for treatment planning; (ii) acquire a proton radiograph in the intended treatment beam direction; (iii) generate a patient-specific HU-RSP calibration curve by optimization; and (iv) create a treatment plan using the patient-specific calibration curve.

5. Summary and conclusions

In this work we have shown that it is possible to construct a robust optimizer that can produce patient-specific calibration curves, using the information from simple single-detector proton radiographs. Following validation with a set of perfect (containing no noise) synthetic datasets, the robustness of the optimizer was confirmed using CT datasets containing noise.

We then demonstrated the technique with real proton radiographs produced using a simple diode array and with the time-resolved radiographic technique. For a phantom composed of real tissues, the optimization procedure was able to reduce the mean error between a DRR_{WEPL} map and a proton radiograph from -2.1% using a generic calibration curve to 0.2% using an optimized calibration curve. Using this optimized calibration curve, produced using proton radiography through the entire patient, it was demonstrated that a more accurate estimate of the WEPL at a therapeutic depth could also be made. By splitting the phantom in half, the RMSE difference between the DRR_{WEPL} map and an independent DE_{WEPL} map decreased from 2.0 to 1.4% using the optimized calibration curve.

With further development, we believe that this procedure could be introduced as part of the proton treatment planning workflow. A proton radiograph would be acquired at the time of the planning x-ray CT and calibration curves would be generated on a patient-by-patient basis.

References

Bentefour E H *et al* 2012 Effect of tissue heterogeneity on an *in vivo* range verification technique for proton therapy *Phys. Med. Biol.* **57** 5473–84

Bentefour E H *et al* 2013 Methods and device for dose based proton radiography *American Association of Physicists in Medicine 55th Annual Meeting and Exhibition* (Indianapolis, IN, 4–8 August 2013) p 308

Chvetsov A V and Paige S L 2010 The influence of CT image noise on proton range calculation in radiotherapy planning *Phys. Med. Biol.* **55** N141–9

Gelover-Reyes E *et al* 2011 Proton radiography with silicon active sensors *ESTRO Proc. 99 (London)* p S531

Gottschalk B *et al* 2011 Water equivalent path length measurement in proton radiotherapy using time resolved diode dosimetry *Med. Phys.* **38** 2282–8

Highland V 1975 Some practical remarks on multiple scattering *Nucl. Instrum. Methods Phys. Res.* A **29** 497–9

ICRU 1989 Tissue substitutes in radiation dosimetry and measurement *ICRU Report 44* (Bethesda, MD: ICRU)

Lagarias J *et al* 1998 Convergence properties of the Nelder–Mead simplex method in low dimensions *SIAM J. Optim.* **9** 112–47

Leroy C and Rancoita P G 2009 *Principles of Radiation Interaction in Matter and Detection* (Singapore: World Scientific)

Li T *et al* 2006 Reconstruction for proton computed tomography by tracing proton trajectories: a Monte Carlo study *Med. Phys.* **33** 699–706

Lu H-M 2008 A point dose method for in vivo range verification in proton therapy *Phys. Med. Biol.* **53** N415–22

Lu H-M, Mann G and Cascio E 2010 Investigation of an implantable dosimeter for single-point water equivalent path length verification in proton therapy *Med. Phys.* **37** 5858

Muraishi H *et al* 2009 Evaluation of spatial resolution for heavy ion CT system based on the measurement of residual range distribution with HIMAC *IEEE Trans. Nucl. Sci.* **56** 2714–21

Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99–117

Pemler P *et al* 1999 A detector system for proton radiography on the gantry of the Paul-Scherrer-Institute *Nucl. Instrum. Methods Phys. Res.* A **432** 483–95

Poludniowski G *et al* 2014 Proton-counting radiography for proton therapy: a proof of principle using CMOS APS technology *Phys. Med. Biol.* **59** 2569–81

Ryu H *et al* 2008 Density and spatial resolutions of proton radiography using a range modulation technique *Phys. Med. Biol.* **53** 5461–8

Schaffner B and Pedroni E 1998 The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power *Phys. Med. Biol.* **43** 1579–92

Schneider U *et al* 2004 First proton radiography of an animal patient *Med. Phys.* **31** 1046–51

Schneider U *et al* 2005 Patient specific optimization of the relation between CT-Hounsfield units and proton stopping power with proton radiography *Med. Phys.* **32** 195–9

Schneider U and Pedroni E 1995 Proton radiography as a tool for quality control in proton therapy *Med. Phys.* **22** 353–63

Schneider U, Pedroni E and Lomax A 1996 The calibration of CT Hounsfield units for radiotherapy treatment planning *Phys. Med. Biol.* **41** 111–124

Schneider W, Bortfeld T and Schlegel W 2000 Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions *Phys. Med. Biol.* **45** 459–78

Schulte R *et al* 2004 Conceptual design of a proton computed tomography system for applications in proton radiation therapy *IEEE Trans. Nucl. Sci.* **51** 866–72

Seco J and Depauw N 2011 Proof of principle study of the use of a CMOS active pixel sensor for proton radiography *Med. Phys.* **38** 622–3

Shinoda H, Kanai T and Kohno T 2006 Application of heavy-ion CT *Phys. Med. Biol.* **51** 4073–81

Simon E 2011 Improving the convergence of Nelder–Mead (and so fminsearch) *Matlab Central* http://uk.mathworks.com/matlabcentral/fileexchange/33328-improving-the-convergence-of-nelder-mead--and-so-fminsearch- [Accessed 17 November 2014]

Sipala V *et al* 2011 PRIMA: an apparatus for medical application *Nucl. Instrum. Methods Phys. Res.* A **658** 73–7

Talamonti C *et al* 2010 Proton radiography for clinical applications *Nucl. Instrum. Methods Phys. Res.* A **612** 571–5

Tang S *et al* 2011 Improvement of prostate treatment by anterior proton fields *Int. J. Radiat. Oncol. Biol. Phys.* **83** 408–18

Telsemeyer J, Jäkel O and Martišíková M 2012 Quantitative carbon ion beam radiography and tomography with a flat-panel detector *Phys. Med. Biol.* **57** 7957–71

Testa M *et al* 2013 Proton radiography and proton computed tomography based on time-resolved dose measurements *Phys. Med. Biol.* **58** 8215–33

White D, Woodard H and Hammond S 1987 Average soft-tissue and bone models for use in radiation dosimetry *Br. J. Radiol.* **60** 907–13

Williams D 2004 The most likely path of an energetic charged particle through a uniform medium *Phys. Med. Biol.* **49** 2899–911

Woodard H and White D 1986 The composition of body tissues *Br. J. Radiol.* **59** 1209–18

Yang M *et al* 2010 Theoretical variance analysis of single-and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues *Phys. Med. Biol.* **55** 1343–62

Yang M *et al* 2012 Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration *Phys. Med. Biol.* **57** 4095–115

Yohannes I *et al* 2012 A formulation of tissue- and water-equivalent materials using the stoichiometric analysis method for CT-number calibration in radiotherapy treatment planning *Phys. Med. Biol.* **57** 1173–90