The use of proton radiography to reduce uncertainties in proton treatment planning

Paul Doolan

Department of Medical Physics and Bioengineering

University College London

A thesis submitted for the degree of

Doctor of Philosophy

December 2014

I, Paul Doolan confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

.....

Abstract

The availability and demand for proton therapy is rapidly expanding across the globe. One of the key decisions that must be made in the procurement process by all new centres is which proton treatment planning system (TPS) to purchase. The first topic of this thesis is therefore to evaluate the performance of three different proton TPSs in the planning of ten meningioma patients. The comparison is built upwards from the beam commissioning and attempts are made to make as many variables as possible consistent between systems. Few statistically significant differences were found between the plans, although differences between the systems (such as layer spacing and spot positioning) are discussed. It is hoped this work will be of general use to the whole proton physics community and will encourage further development of proton TPSs from vendors.

One of the major sources of range uncertainty in current proton treatment planning is due to the necessary conversion of the patient's X-ray computed tomography (CT) dataset from CT numbers to relative stopping powers (RSPs). The remainder of the thesis looks to address this. The stoichiometric procedure is considered the most accurate method to generate the X-ray CT to RSP calibration curve. In the third chapter of the thesis an investigation is made into the errors of this procedure: specifically, the theoretical calculation of the RSP, step four of the process. The impact of these errors on the proton beam range is calculated for both phantom and patient cases.

It has been suggested that proton radiography could offer a solution to the uncertainty in this calibration curve. The fourth chapter of the thesis therefore looks at a novel method of proton radiography, which involves taking the dose ratio of two pristine Bragg peaks. The investigation proceeds with a theoretical analysis of the application limits of the technique, together with an experimental validation of the theoretical approach.

The fifth chapter of the thesis demonstrates an approach that uses proton radiography to improve the calibration curve. Assuming the information

in the proton radiograph to be correct, the calibration curve can be optimised by comparison with a digitally reconstructed radiograph through the X-ray CT. The function of this optimiser is validated on synthetic datasets and its application is demonstrated with real measurements on plastic and real tissue phantoms. The technique is also shown to offer an improvement in the water-equivalent path length prediction at a therapeutic depth.

Acknowledgements

I would firstly like to thank my supervisors, Professor Gary Royle, Professor Adam Gibson (both University College London, UCL) and Dr El Hassane Bentefour (Ion Beam Applications), for their continual guidance and assistance throughout my studies and without whom this work would not have been possible. I am very grateful to Dr Hsiao-Ming Lu for his superb guidance and support during my time at Massachusetts General Hospital (MGH).

I would also like to show my appreciation to Dr Ivan Rosenberg (University College London Hospital, UCLH), with whom I discussed many aspects of my work and analysis. I am also indebted to Dr Chris Ainsley at The Hospital of the University of Pennsylvania (UPenn) for his great support in my attempts to commission their beam and for his specialist expertise throughout the treatment planning project. I would like to acknowledge Derek D'Souza, Chris Stacey and the radiotherapy physics team at UCLH for their support in this work. The guidance of Maura Kirk, also of UPenn, was critical in beam geometry selection and other planning enquiries. I would also like to thank the patience, support and plan verification provided by the support staff of Varian, Philips and Elekta.

From my time at MGH I would first like to thank Dr Mauro Testa, for his enduring support during all the proton radiography measurements (that often extended deep into the night). Dr Gregory Sharp also receives my gratitude for his Plastimatch modifications, his patience and for his expert opinion throughout the project. I would like to thank Dr Joao Seco for his inspiration and direction in the work of assessing range effects. I would like to acknowledge the support of the operators at MGH, without whom the late night measurements would not have been possible.

I would like to gratefully acknowledge the financial support received from the Engineering and Physical Sciences Research Council (UK) and Ion Beam Applications (Louvain-La-Neuve, Belgium), who have allowed me to complete this work. I also owe gratitude to the UCL Graduate School,

who funded the secondment to UPenn, and to the Institute of Physics and Engineering in Medicine and the Institute of Physics for funding trips to conferences.

From a personal aspect, I would also like to pay tribute to the group on Cox 3 who made my time at MGH so enjoyable and got me hooked on American coffee.

I would like to thank the spirit and atmosphere engendered by all the students and staff of the Radiation Physics group at UCL. 11 o'clock will forever be associated with tea time and questionable biscuits.

I could not be more grateful στην αγάπη μου, for the emotional and scientific support throughout the entirety. Η αγάπη σου kept me strong when I needed it most and I feel so lucky to have you by my side.

I would like show my appreciation to my grandparents, Irene and Patrick, who have shown me nothing but encouragement since my first days at school.

Finally, I reserve special praise for my parents, John and Elaine, and brother Liam, for their love and support throughout the (many) years of study. Without your continual positivity I would not be where I am today.

Contents

Chapter 1

Introduction

The physical properties of protons can, in some cases, allow for an improved dose distribution compared with conventional X-ray radiotherapy (Isacsson et al. [1996]; Lee et al. [1994]; Miralbell et al. [1997]). The Bragg peak, characterised by a sharp distal and lateral penumbra and shown in figure 1.1, makes it possible to limit the dose to critical structures in ways that are not possible with X-ray radiotherapy (Wilson [1946]). Additionally, the low entrance doses ensure that the integral dose deposited is always lower for protons than photons (Paganetti [2012]). This sparing of surrounding normal organs from low- and mid-dose levels can help to lower the probability of radiation-induced secondary malignancies (Chung et al. [2008]).

This distinct advantage has led to the use of proton therapy for a variety of indications, including ocular treatments using low proton energies (Damato et al. [2005]; Kacperek [2009]), cranio-spinal irradiations (Archambeau et al. [2000]), lung cancer (Moyers et al. [2001]) and a variety of paediatric treatments (Merchant [2009]), among others. Falling construction costs has led to a significant increase in the interest in, and availability of, proton beam therapy (Peggs [2002]). As of 2010, over 60,000 patients had received part or all of their radiation therapy from proton beams (Bouyon-Monteau et al. [2010]). Since 2010, 9 new centres have opened worldwide, with a further 30 planned to open by the end of 2016 (PTCOG [2013]). Many of these new centres are opening with pencil beam scanning (PBS) rather than the traditional passive scattering techniques, driven by improved efficiency (passive techniques require patient-specific apertures and compensators for each field) and potentially lower neutron dose (Hall [2006]) (because of the absence of physical beam spreading and shaping devices in the nozzle). Further details of the techniques available in proton therapy are detailed in section 2.1.3.

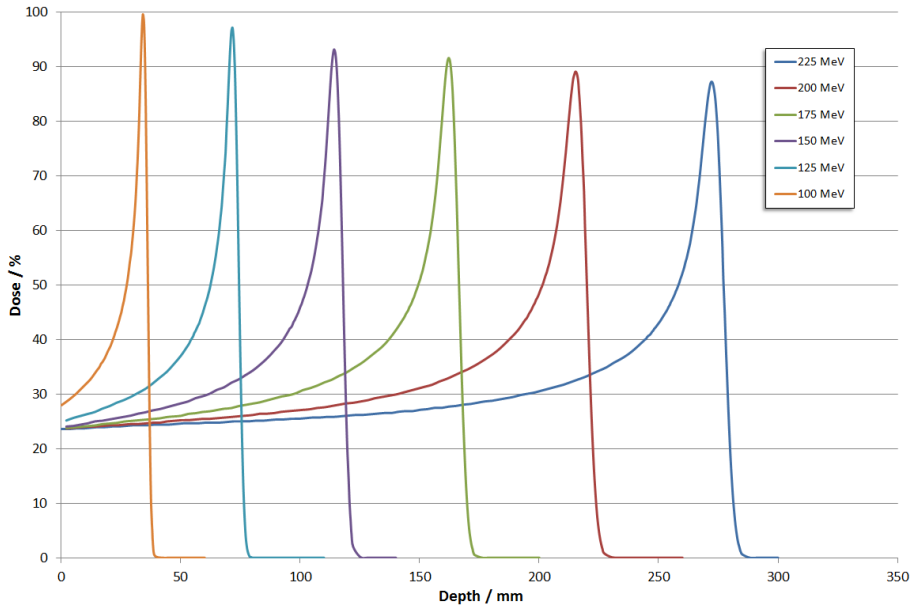


Figure 1.1: Example Bragg peaks from the proton beam at The Hospital of the University of Pennsylvania.

1.1 Range uncertainty

Given that the major advantage of proton therapy over photon therapy is the distinctive Bragg peak, which is relied upon to shape the dose, it is imperative that the proton beam range (the depth in tissue that protons reach) is well known. The uncertainty on this position, known as 'range uncertainty', requires additional margins to be added to the treatment volume (compared to photon therapy), limiting its efficacy. As such there is huge interest in the development of a technique that can accurately predict or measure the proton range. Such a device or technique will lead to further reduction of treatment planning margins, reduced toxicity, better outcomes and thus wider use of proton therapy.

Range uncertainty arises from a number of sources, both systematic and random in nature. A detailed review of the systematic errors, which forms a large part of this thesis, is given in section 3.1.2. Paganetti [2012] demonstrated that one of the largest uncertainties derives from inhomogeneities in the patient volume, leading to a degradation of the Bragg peak shape (Urie et al. [1986]). Assuming inhomogeneities simply shift the Bragg peak position ignores the fact that protons take somewhat curved paths due to multiple Coulomb scattering (MCS). Urie et al. [1986] showed that the deviation of protons from a straight line due to inhomogeneities leads to drastic changes in the Bragg peak shape (if measured beyond the patient) and the dose distribution within the patient.

In addition to patient inhomogeneities, range uncertainty arises from other sources deriving from the patient imaging, i.e. in the measured computed tomography (CT)

number, and uncertainties in calculating the relative stopping power (RSP) of the tissue. Random uncertainties are largely related to the position of the target relative to the beam and are thought to blur out over a typical fractionation scheme. However, for some indications this is not the case as the patient anatomy is known to change slowly and systematically (Knopf and Lomax [2013]).

The Bragg peak shape means that patient positioning errors in proton therapy can result in drastic changes in the dose distribution, unlike photon therapy which usually results in a simple shift of the dose (Knopf and Lomax [2013]). Currently proton range uncertainty is 'managed' by designing plans that are 'robust' against such uncertainties. One such approach is the careful selection of treatment field directions, such that the sparing of a critical structure is reliant on the lateral penumbra and not the distal fall-off. Another approach, suggested by Rietzel et al. [2006] and Thomas [2006] but to our knowledge not currently used in the clinic anywhere, is to design beam-specific planning target volumes (PTVs) that have larger margins at the distal edge to account for both the weighting of the dose at the Bragg peak and the systematic errors.

In general, proton treatment planning uses a generic margin recipe of 3.5% of the range (water-equivalent, to the distal edge) plus an additional absolute value. Paganetti [2012] explains that the historical 3.5% value originates from the addition (in quadrature) of the CT number uncertainty using 1980's CT technology (2%) and the conversion from CT Hounsfield units (HU) to RSPs (1%), multiplied by 1.5 standard deviations. The MD Anderson Proton Therapy Center, the Loma Linda University Medical Center and the Hospital of the University of Pennsylvania ('UPenn', with whom a collaboration was formed for chapter 2 of this thesis) all use a margin recipe of 3.5% + 3 mm. At Massachusetts General Hospital ('MGH', where chapter 5 of this thesis was conducted), a margin of 3.5% + 1 mm is used. The overriding aim of this thesis is to reduce the 3.5% value to a lower level.

There are a multitude of methods to measure the proton beam range within the patient: from implantable dosimeters (Gottschalk et al. [2011]; Lu [2008b]) to prompt gamma imaging (Min et al. [2006]; Verburg et al. [2013]) and positron emission tomography (Parodi et al. [2005]). For a complete list the reader is directed towards the comprehensive review of Knopf and Lomax [2013].

In this thesis we aim to address the problem of range uncertainty, not by measurement of the beam *in vivo*, but by a better characterisation of the patient tissues. An improved estimation of the RSP distribution within the patient will reduce proton range uncertainty and will ultimately allow for a reduction in PTV margins. We investigate the applicability of proton radiography as a tool for this purpose.

1.2 Major results and novelty

The work in this thesis is separated into four chapters:

- All new proton centres must decide which proton treatment planning system (TPS) to purchase. In the first chapter of this thesis the performance of three commercial proton TPSs (Eclipse, Pinnacle^3 and XiO) are evaluated, through the planning of ten meningioma patients. To date there have been no such works anywhere in the literature, largely because there was previously little choice. Following commissioning of the same beam data on each system, the same ten patients were planned with as many variables as possible consistent between systems.

  It was found that both Eclipse and Pinnacle^3 produced beam models within clinically-accepted tolerances (the beam model of XiO was not assessed). In the version tested, Eclipse overestimated the build-up (by up to 2.5%) for energies $E \geq 175\,\text{MeV}$ and Pinnacle^3 consistently overestimated the Bragg peak depth, with a maximum distal R50 (the depth at which the dose falls to 50% of the Bragg peak maximum) error of 0.5 mm.

  Few statistically significant differences were found between the plans, although there were clear differences in the layer spacing and spot positioning. Such information may prove useful to the proton physics community and it is hoped it will encourage further development of proton TPSs from vendors.

As discussed in section 1.1, the conversion of the patient's X-ray CT dataset from CT numbers to RSPs contributes 0.5-1.8% to proton range uncertainty. The overall aim of the work in chapters 3, 4 and 5 in this thesis is to minimise this error.

- There are numerous formulae (Bichsel, Janni, ICRU, Schneider) for computing the theoretical stopping power of tissues, a critical part of the stoichiometric calibration procedure. It is not clear (i) which of these formulae should be used, and (ii) the magnitude of errors each introduces. With respect to measurements in a proton beam, it was found that there was a systematic underestimation of the RSP between -0.6% to -2.1% for all approaches except Janni (which overestimated by 0.5%) for seven Gammex inserts. These errors translated into overestimations in the beam range of 1.1% to 1.5% (-0.3% underestimation for Janni).

  By developing a gold standard that matched our measurements, it was shown that the agreement across 73 human tissues was 0.04±0.37% (mean ± standard deviation) when using the Schneider approximation with the ICRU elemental ionisation energies and a water ionisation energy of 78 eV.

- Proton radiography has been mooted as a potential solution to the range uncertainty introduced by the calibration curve, as it generates a direct map of the patient RSPs (i.e. no need to convert from CT numbers). The limits and applicability of a novel method of single-detector proton radiography, which involves taking the dose ratio of two pristine Bragg peaks, is determined. To date, the applicability of this 'dose ratio' technique has only been investigated on the distal side of the lower energy Bragg peak due to the sharp fall-off. An investigation is therefore made into the use of proximal side, which has the potential to allow for a much wider range of water-equivalent thicknesses (WETs) to be imaged. Comparisons are made with use of the distal side.

  The theoretical approach - using an analytical approximation for the Bragg peak to generate theoretical dose ratio curves - was validated using measurements of a sapphire sphere with a complementary metal oxide semiconductor active pixel sensor (CMOS APS). In the theoretical investigation it was found that, provided the noise level in the dose ratio image was less than 2%, a larger spread of WETs could be imaged using the proximal side of the Bragg peak (max 5.31 cm) compared to the distal side (max 2.42 cm). A square field equivalent size of 7.6 cm^2 could be imaged in a paediatric brain for a required accuracy in the WET of 3 mm and a 1% noise level in the dose ratio image, but otherwise the technique was found to have limited applicability to different patient sites.

- After demonstrating the limited applicability of proton radiography as a single solution in the previous chapter, it is shown that it can be combined with X-ray CT to generate patient-specific calibration curves. Assuming the information in the proton radiograph to be correct, the calibration curve can be optimised by comparison with a digitally reconstructed radiograph water-equivalent path length (DRR_{WEPL}) map through the high resolution X-ray CT. The function of this optimiser is validated on synthetic datasets, with reductions in the root-mean-square difference between the DRR_{WEPL} and a known WEPL, before and after optimisation, of 87-100%. Its application is demonstrated with real measurements on plastic and real tissue phantoms, using the established 'time-resolved' method of single-detector proton radiography. The mean error in the WEPL for the generic/optimised calibration curves was 1.8/0.4% for the plastic phantom and -2.1/-0.2% for the real tissue phantom. The technique was also shown to offer an improvement in the prediction at the therapeutic depth.

Chapter 2

Proton treatment planning system comparison

Work in this chapter can be found in the following journal article and was presented at the following meetings:

- (Accepted) **Doolan P**, Alshaikhi J, Rosenberg I, Ainsley C G, Gibson A, D'Souza D, Bentefour E H, Royle G **2014** A comparison of the dose distributions from three proton treatment planning systems in the planning of meningioma patients with single-field uniform dose pencil beam scanning *Journal of Applied Clinical Medical Physics*

- Alshaikhi J, Amos R, **Doolan P**, D'Souza D, Royle G, Rosenberg I **2014** Comparison of proton treatment planning systems *International Conference on Radiation Medicine (ICRM)*, Riyadh, Saudi Arabia, P48 (Poster)

- **Doolan P**, Rosenberg I, Ainsley C, Gibson A, Royle G **2013** A comparison of the beam configuration modules of two proton treatment planning systems *European Society of Therapeutic Radiology and Oncology (ESTRO) 2nd Forum*, Geneva, Switzerland, PD-570 (Oral presentation)

As can be seen, there were many contributors to this work. J Alshaikhi produced the plans on XiO, under strict guidelines from myself. C G Ainsley and the physics team at UPenn acquired the integrated depth doses and spot profiles and helped with the commissioning of each system. All other work in this chapter was conducted by myself.

2.1 Background

2.1.1 The systems

This work was conducted as part of the procurement process for the new proton centre at University College London Hospital (UCLH). The TPSs under consideration include: Eclipse (Varian Medical Systems), Pinnacle^3 (Philips Radiation Oncology Systems), RayStation (RaySearch) and XiO (Elekta). In this work the Eclipse, Pinnacle^3 and XiO systems (summarised in table 2.1) were assessed by planning a number of different patient cases on each system.

It should be noted that, at the time of the study, Pinnacle^3 was a non-clinical research system. Philips had not yet received regulatory clearance and it was still under development. During the course of the study five different versions of Pinnacle^3 were released, so different parts of the study were completed under different versions. As such, some results found in older versions may improve in later versions.

2.1.2 Previous treatment planning system comparisons

There have been many publications of treatment planning comparisons between photons and protons (a selection: (Cozzi et al. [2001]; St Clair et al. [2004]; Stuschke et al. [2012]; Trofimov et al. [2007]; Yoon et al. [2011]), specifically for meningiomas: (Arvold et al. [2012]; Combs et al. [2012])) and for different proton therapy techniques (Arjomandy et al. [2012]; Noa et al. [2012]; Tang et al. [2013]; Yeung et al. [2014]). A number of publications exist on the theory and accuracy of the dose calculation algorithms for two of the systems (Eclipse (Schaffner [2008]; Schaffner et al. [1999]; Ulmer and Matsinos [2011]; Ulmer and Schaffner [2011]; Zhu et al. [2013]) and XiO (Rana et al. [2013]; Soukup et al. [2005])). For Pinnacle^3, only the theoretical basis is available (Bortfeld [1997]; Gottschalk et al. [1993]; Kooy et al. [2010]; Lynch et al. [1991]; Soukup et al. [2005]), because at the time of the study Philips had yet to receive regulatory clearance for the software. While there have been numerous comparisons of commercial photon TPSs (Ayyangar et al. [2005]; Eldesoky et al. [2012]; Fogliata et al. [2003, 2005, 2007]), we are not aware of any publications comparing different commercial proton TPSs. This is largely because

Manufacturer	System	Version
Varian	Eclipse	10.0.39
Philips	Pinnacle^3	9.100.0
RaySearch	RayStation	3.4.9
Elekta	XiO	4.7.00

Table 2.1: Details of proton TPSs under consideration at UCLH.

proton TPSs were typically built in-house, but the rapid expansion of proton therapy has led the major photon TPS vendors to develop or purchase their own modules. This study compares the plans produced by three proton TPSs, Eclipse (Varian), Pinnacle^3 (Philips) and XiO (Elekta), for a set of ten meningioma patients in which as many variables as possible were kept consistent. It is the hoped that such an assessment will provide additional information to the procurement process for new proton centres and will encourage further proton TPS development.

2.1.3 Delivery modes

Proton treatments currently fall into two modes: scattering or scanning. Passive scattering, the most common current treatment, consists of a beam of fixed energy intercepted by various beam modifiers. After initial beam energy modification upstream, the energy is refined by placing material in the path (variable range shifter, range modulator wheel or ridge filter). The initial pencil-shaped beam is laterally spread to a more useful, larger Gaussian profile using one or two scattering foils (double scattering is more common). The lateral shape of the beam is made to match the lateral shape of the tumour (plus any margin) using a custom (field- and patient-specific) brass aperture. Dose conformation in the depth direction is achieved using custom (field- and patient-specific) compensators of lucite or wax.

Scanning treatments are far more convenient, but until recently have not been used routinely because the technology was not commercially available and/or centres lacked experience. In scanning the beam energy is refined before the nozzle, using range shifters. The pristine pencil beam is deflected using magnetic dipoles to scan the beam across the patient. Beam energy corresponds to a given depth in tissue, so once the cross-section of a tumour has been scanned the energy is altered (usually lowered) and the next cross-section is scanned. Scanning has an advantage over scattering techniques as it allows for more conformal treatment plans, particularly at the proximal tumour edge (which cannot be conformed to with conventional scattering techniques). In addition, apertures and compensators are not required for every patient field, saving both time and money. There are worries over delivery times and interplay with patient motion (Grassberger et al. [2013]), but the general trend is for new centres to open with scanning techniques.

2.1.3.1 Modes of scanning

Scanning can either be delivered discretely on a spot-by-spot basis (known as spot scanning) or in a continuous mode (also known as PBS). In either of these modes the fluence at each position can be constant (known as uniform scanning) or modulated

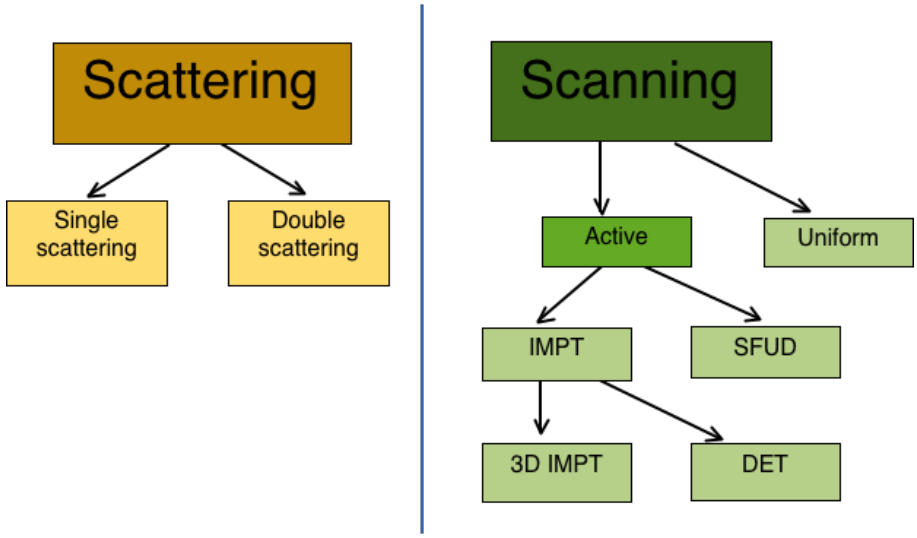


Figure 2.1: Classification of the techniques possible in proton therapy. Abbreviations are given in the text.

(active scanning). In uniform scanning it is possible to have different spot weights using a compensator, as with passive scattering.

In active scanning there are two common techniques: (i) single field uniform dose (SFUD), where each field is individually optimised to deliver a uniform dose distribution to the target; and (ii) intensity modulated proton therapy (IMPT), where the final dose distribution is the result of contributions from multiple fields, each of whose individual contribution to the target dose distribution is non-uniform (Lomax [1999]). In addition, there are two different 'flavours' of IMPT: (i) 3D IMPT, where Bragg peaks from all beams are placed throughout the volume and their weights are optimally adjusted; and (ii) distal edge tracking (DET), in which Bragg peaks are placed only on the distal edge of the target (McGowan et al. [2013]). Liu et al. [2012b] showed that the latter requires many fields and is less robust to uncertainties, but Oelfke and Bortfeld [2000] demonstrated that it offers reduced integral dose, improved lateral fall-off and reduced dynamic range of Bragg peak weights. The classification of all these techniques can be seen in figure 2.1.

2.1.4 Optimisation

Treatment planning is a procedure that takes as its input a model of the radiation beam and patient anatomy and produces two outputs: (i) instructions to the machine to deliver the treatment; and (ii) the expected distribution in the patient. Dedicated TPSs are used for this process.

Active scanning has a large number of degrees of freedom, with the planner being able to control the position, energy and intensity of every pencil beam (which can have a standard deviation (Gaussian in shape) as small as 3 mm in air). Assuming

the spots are spread uniformly across the tumour and the number of energy layers are fixed, the intensity (weight) of each pencil beam must be optimised to achieve the best possible coverage of tumour and sparing of normal tissue. Typically there can be of the order of 10^3 pencil beams per treatment, so this must be an inverse planning process performed by a dedicated TPS.

Inverse optimisation involves the minimisation of a function that quantifies aspects of the dose distribution and specifies trade-offs between target volume coverage and organ at risk (OAR) sparing (Schwarz [2011]). The routine to minimise this cost function is similar between all systems and is thus unlikely to cause differences between plans in itself. However, each software has a different method of pre-weighting and positioning spots prior to the minimisation step, which is likely to affect the final plan quality.

At the time of this work, to the best of the author's knowledge, there have been no publications investigating the performance of proton optimisation algorithms in current commercial TPSs, most likely because there previously was little choice. As recently as 10 years ago most proton TPSs were built in-house. The rapid expansion of proton therapy has however led to the major photon TPS vendors to produce or purchase their own system. For instance the planning system developed at the Paul Scherrer Institute (PSI) forms the basis of the Varian Eclipse proton TPS; the system developed at MGH and Harvard Cyclotron Laboratory has been adapted by Elekta (previously CMS) to form the basis of the proton planning in XiO; and the system built by the University of Florida is currently being adapted by Philips to form their proton module within Pinnacle^3 (Yeung [2009]). With many new centres due to open in the next 5-10 years there is demand for an evaluation of these commercial systems.

2.1.5 Meningioma

One method of comparing TPSs is to plan the same patient on each system. The multitude of settings available in each system makes this a difficult process, but it is possible to isolate some conditions for testing. It is important that the patients selected are suitable for scanning proton therapy and provide a difficult test to each system, so that differences can be observed. In an ideal case many different treatment sites would be tested, however this was not possible due to time constraints. It was concluded, in consultation with an experienced medical physicist, that meningiomas would be appropriate for this purpose.

Meningiomas account for 13-26% of all primary intracranial tumours (Kleihues et al. [2002]). A number of treatment planning intercomparisons have shown that proton beam therapy can substantially decrease the dose to OARs in the brain

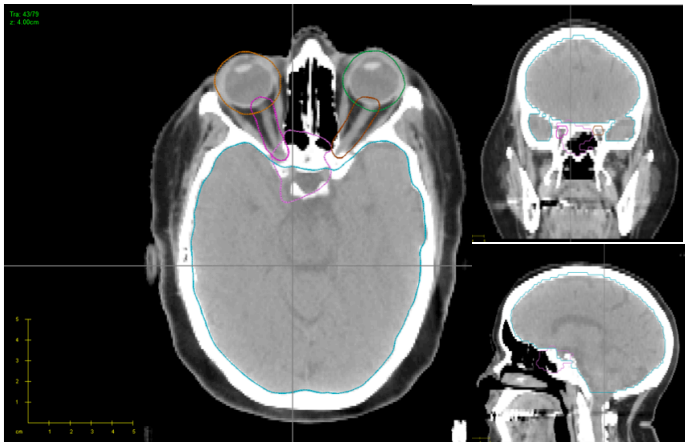


Figure 2.2: Meningioma patient to be used in the TPS comparison. Slices have been selected to best demonstrate the difficulty of the case in the (left) transverse, (top right) coronal and (bottom right) sagittal directions. Organs outlined: light pink = clinical target volume; dark pink = right optic nerve + 3 mm; dark brown = left optic nerve + 3 mm; orange = right orbit + 3 mm; green = left orbit + 3 mm; blue = brain.

(Baumert et al. [2001]; Lin et al. [2000]) and Weber et al. [2004a] confirmed this specifically in meningioma patients. This advantage of proton therapy, together with the difficulty of treating a volume that has a number of overlapping or surrounding critical structures, make these suitable cases on which to conduct a dose distribution comparison of different systems.

An example of one of the patients used in the comparison is shown in figure 2.2. The clinical target volume (CTV) encompasses the visible tumour, with an extension to areas likely to have microscopic spread. The difficulty in this case is that the right optic nerve overlaps with the target volume, a problem that was typical of the ten patients selected. Here the sharp dose fall off of the proton depth dose curve could be used in a superior or superior oblique beam, sparing many of the organs at risk. In photon plans such an arrangement would be ill-advised because the X-rays would continue through the rest of the patient's body, depositing dose along critical structures such as the spine.

2.1.6 Proton beam commissioning

A fair comparison requires all systems to have the same beam data. Therefore, raw PBS beam data from UPenn's horizontal fixed beam line, an IBA machine (Ion Beam Applications, SA, Louvain-La-Neuve, Belgium), were added to each system and commissioned according to the requirements and tools of each TPS.

Most modern TPSs are based on models rather than measurements, i.e. they do not use this measured beam data directly but instead fit a model to the data and perform all calculations using the model. It is possible to create a dose model that accurately characterises the output of a given machine by iteratively adjusting

the dose model parameters and evaluating the quality of the match between the measured and computed data. Even after this manual adjustment the model in each system will be different, so it is important to commission the beam so that any limitations are highlighted. Such commissioning must be completed for all beam data that will be used for subsequent treatment planning. In this comparison study, any differences between the beam model in different systems may lead to different results. As such, it is important any differences at this level are known and well-documented.

Proton beam commissioning is simpler than photon beam commissioning, as far fewer measurements are required:

1. *Integrated depth dose (IDD) curves:* pristine (not modified by scattering plates) Bragg peaks from single, unscanned spots are measured at a range of depths in water for a number of discrete energies. It is important the ionisation chamber is large enough to collect the entire charge from the spot. To avoid inaccuracy in such modelling most systems recommend using curves every 1-2 $\mathrm{g/cm^2}$ (i.e. every 1-2 cm in water), which corresponds to energies approximately every 5-10 MeV.

2. *In-air spot profiles:* lateral profiles (in x- and y-directions) of the spot are measured at a range of positions from the source, for as many energies as available.

3. *General machine parameters:* numerous properties of the machine must be entered into the system.

2.1.6.1 Integrated depth doses

Each system uses different models to calculate the proton dose in a patient or phantom. These are the same models that are used in beam configuration to model the measured depth dose curves input by the user and thus it is important to understand the differences. These models are often a compromise between speed and accuracy.

2.1.6.1.1 Pinnacle^3 (the Bortfeld model)

The analytical model proposed by Bortfeld [1997] is used to approximate the depth dose distribution for protons within Pinnacle^3. The dose D as a function of depth z can be parameterised according to the following equation:

$$D(z) = \Phi_0 \frac{e^{-\xi^2/4}\sigma^{1/p}\Gamma(1/p)}{\sqrt{2\pi}\rho p \alpha^{1/p}(1+\beta R_0)} \left[\frac{1}{\sigma} L_{-1/p} + \left(\frac{\beta}{p} + \gamma\beta + \frac{\epsilon}{R_0}\right) L_{-(1+p)/p}(-\xi)\right] \quad (2.1)$$

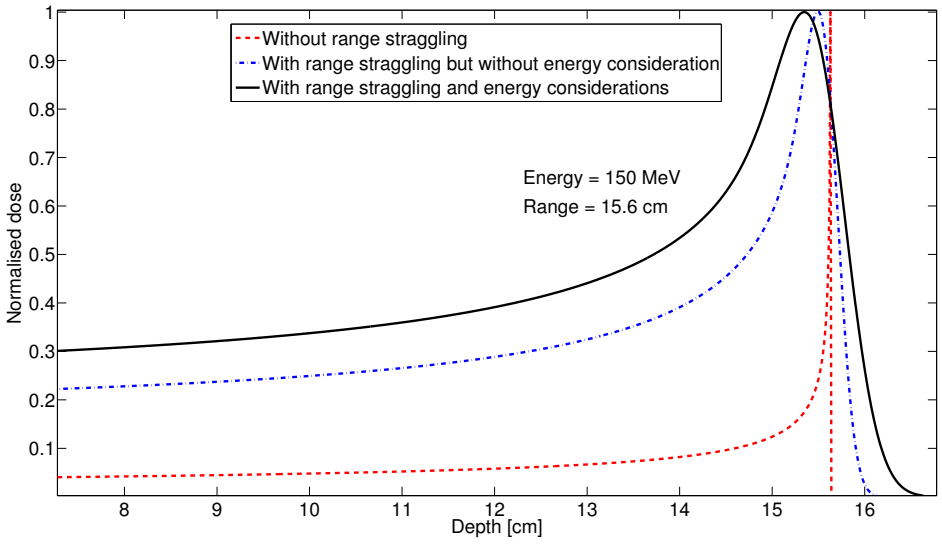


Figure 2.3: Normalised depth dose curves for 150 MeV protons in water, for the three Bortfeld equations.

where Φ_0 is the primary fluence; $\xi = (R_0 - z)/\sigma$, where R_0 is the range of the proton beam (distal 90%), z is the depth along the beam direction and σ is the width of the Gaussian range straggling; Γ is the gamma function; ρ is the density of the medium; and L is a parabolic cylinder function. The other parameters are constants and their values for protons in water are given in table 2.2.

Equation 2.1 takes account of (i) range straggling, a phenomenon in which different protons lose different amounts of energy; and (ii) the energy spectrum of the initial proton beam (assumed to consist of two parts: a peak, approximated by a Gaussian; and a small 'tail' extending toward low energies). There are other Bortfeld equations that do not take account of one or both of these and their different depth doses are shown in figure 2.3. It can be seen that consideration of these two factors not only has a distinct effect on the shape of the Bragg peak, but also its position.

Parameter	Description	Value	Unit
p	Exponent of range-energy relation	1.77	-
α	Proportionality factor	0.0022	cm MeV^{-p}
β	Slope parameter of fluence reduction relation	0.012	cm^{-1}
γ	Fraction of locally absorbed energy released in non-elastic nuclear interactions	0.6	-
ϵ	Fraction of primary fluence contributing to the tail of the energy spectrum	$\approx 0.0 - 0.2$	-

Table 2.2: Summary of constants and parameters used in the theoretical model proposed by Bortfeld [1997]. The values are for protons in water. The last value is machine specific.

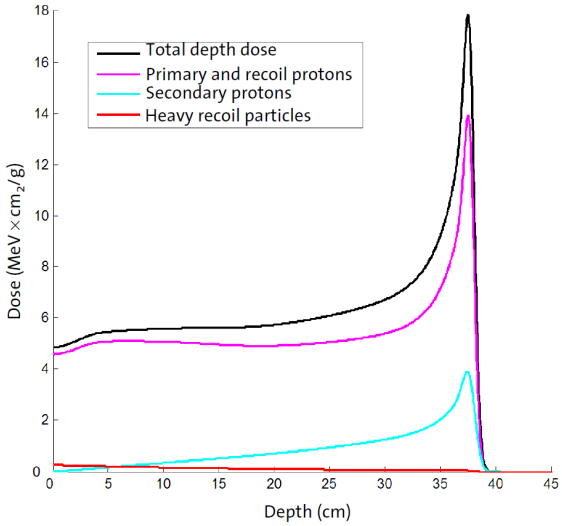


Figure 2.4: Contribution of components to the total depth dose, for a 250 MeV proton beam in water, using the beamlet model. Varian [2011]

Pinnacle[3] uses the distribution that takes account of both factors, as described by equation 2.1. A problem with this model is that it does not account for dose build-up at the entrance (for high energy proton beams) or the transport of secondary protons. Eclipse therefore uses a different model that does take account of the build-up, details of which are provided in section 2.1.6.1.2.

2.1.6.1.2 Eclipse (the Beamlet model)

The dose calculation model, known as 'Proton Convolution Superposition' in Eclipse, is based on the model of proton beamlets in water by Ulmer and Schaffner [2011]. The beamlet is separately modelled in depth and perpendicular to the beam direction.

The depth dose curve is modelled as an infinitely broad, non-divergent beam impinging orthogonally to the surface, so the beamlet dose deposition only depends on the incoming energy spectrum. The dose curve is divided into contributions from primary protons, secondary protons and recoil particles. The lateral dose distribution is modelled by a 2D Gaussian distribution. The 3D dose distribution of a single beamlet can be found by multiplying the number of protons, the normalised lateral distribution and the maximum (central axis) dose on a given plane. The dose from a real proton beam line is composed of a continuum of beamlets with spatially varying intensities in the plane perpendicular to the beam axis. A summary of the contribution of primary, secondary and recoil particles to the total depth dose is shown in figure 2.4, for an incident energy $E_0 = 250\,\text{MeV}$ and a straggling parameter (accounting for the finite energy spread of the incident protons and range straggling) $\tau_{incident} = 0.2\,\text{cm}$.

2.1.6.2 In-air spot profiles

When protons pass through material they interact with the Coulomb and hadronic field of its nuclei, resulting in a small energy loss and a small angle deflection. While the deflection is small, because of the large number of interactions the effect becomes considerable. This process is known as multiple Coulomb scattering (MCS) and the theory was first derived by Molière [1948] and developed by Bethe and Ashkin [1953]. These deflections are random but their cumulative effect can be modelled using a Gaussian angular distribution when viewed on a macroscopic scale. As such, both systems model the lateral source size at any distance from the machine head by a two-dimensional Gaussian distribution.

2.2 Aim

The aim of this part of the thesis was to compare the performance of three proton TPSs, to inform the procurement process for the new UCLH proton centre. This was conducted through the planning of ten meningioma patients on all systems.

A full comparison of the entire TPS of each system would be impractical, due to the size, complexity and number of so many different features on each system. Rather than attempt such a task, that would quickly become out of date as the TPS evolves, our preliminary work focused on a very specific case (SFUD for meningioma patients). Even then, there were restrictions to the approach that could be taken. Firstly, the vendors have proprietary rights to their software and so it was not possible to know the internal details of the algorithms. Secondly, the complexity of the systems makes it very difficult to single out a specific component for testing. As such we took a pragmatic approach of comparing the best plan that each system could produce, with the same clinical constraints. Such an approach has both advantages (e.g. it allows the use of options offered by each TPS) and disadvantages (e.g. there are different implementations between TPSs and planners). The accuracy of each dose calculation algorithm is an important topic, but it is something that we decided not to investigate in this work due to the scale and depth required by the task. Also, phantom studies are not purely a reflection of the inherent accuracies of the dose calculation algorithms themselves, but also of how well the respective beam models have been configured; and the latter can be improved with experience. Given our relative inexperience with each system, we decided to avoid this ambiguity and simply assessed the beam models by their abilities to reproduce the measured input data; all TPSs were found to be capable of doing this within clinical tolerances (Doolan et al. [2013]).

The comparison was formed of two main analyses of 10 patient plans for each

system: (i) the DVH metrics for the target and OARs; and (ii) an assessment of the uniformity and distribution of spot weights. No qualitative features (ease of use, visualisation features etc.) were compared to avoid the possibility of subjectivity. Although there are numerous ways to compare different TPSs, we believe that this is the first study, albeit a preliminary one, to explicitly compare clinical implementations of different TPSs for proton therapy.

2.2.1 Objectives

In order to be able to compare different optimisation algorithms the same patient was planned on each system. To achieve this aim, there were a number of objectives that must be fulfilled:

1. The image data and the outlines of target and organs at risk must be consistent on each system.

2. Protons and X-rays interact differently in tissue and so the HUs from the X-ray planning CT must be converted to RSPs. This requires the production of a calibration curve, which must be made consistent between systems. An investigation into the uncertainties introduced by the use of this single curve, together with efforts to reduce them, form chapters 3, 4 and 5 of this thesis.

3. In a comparison the beam data needs to be consistent between TPSs. It is therefore necessary to acquire beam data from a clinical proton machine, load the data into each system and model the beam.

4. After the most appropriate beam geometry has been determined, the meningioma patients must be planned on each system and compared.

2.3 Materials and methods

This intercomparison study was conducted at a dose planning level (plans were made on the TPS, without measurement validation). SFUD plans, using a horizontal fixed beam geometry, were produced for three TPSs available to us:

1. Pinnacle^3, Philips Radiation Oncology Systems (Fitchburg), version 9.1.00. This was a non-clinical research system, as Philips had not yet received regulatory clearance for this software;

2. Eclipse, Varian Medical Systems, version 10.0.39;

3. XiO, Elekta, version 4.70.00.

Fixed horizontal SFUD fields were employed in the study because, at the time of writing, they are probably the most common implementation of spot scanning proton beams due to their relative robustness over IMPT fields to range uncertainties (Lomax [2008]; Weber et al. [2004b]). Although we allowed couch rotations, by removing the additional degrees of freedom provided by IMPT plans, differences between systems could be more easily identifiable.

2.3.1 Proton beam commissioning

A fair comparison requires all systems to have the same beam data. Therefore, the raw PBS beam data from UPenn's horizontal fixed beam line, an IBA machine, were added to each system (my work involved the commissioning of Eclipse and Pinnacle^3 only, so only these systems are described in this thesis). The beam models were tuned within each system until the differences from the input data were within clinical tolerances (Doolan et al. [2013]). A consistent minimum spot monitor unit (MU) of 0.021 MU was enforced on all systems to ensure the plans were deliverable (so that the statistical error on the spot dose, as dictated by the charge measurement resolution of the monitor chambers in the nozzle, is within 1%). These MU constraints are dealt with during post-processing, with all systems rounding equivalently, the impact of which is discussed in depth by Zhu et al. [2010]. The HU to RSP calibration curve was determined using a stoichiometric calibration (Schneider et al. [1996]) and was identical between all three systems. Many more details on this stoichiometric procedure can be found in the chapters 3, 4 and 5 of the thesis. As one patient had titanium clips inside the target, the HU had to be overridden in contouring because of saturation in the image.

2.3.1.1 Integrated depth doses

At UPenn, proton IDDs were acquired using a fixed horizontal beamline (i.e. the gantry cannot be rotated around the patient) directed through the side of a water tank. The WET of the side of the tank (18 mm WET), an offset from the surface (20 mm - this is required as it is difficult to position the chamber precisely at the surface) and the detector WET (4 mm) meant that 42 mm was the shallowest effective water-equivalent depth that the detector could be placed. All measurements therefore have no values below 42 mm. The lack of data in this build-up region proved to be problematic for one of the planning systems, as discussed in section 2.5.1.1.1. Measurement step sizes were not equidistant and were adapted for different energies: 0.5-1.0 mm in the area of the Bragg peak and 3 mm elsewhere.

The detector used was a parallel plate PTW 34070 Bragg peak chamber, as shown in figure 2.5 (parallel plate chambers are recommended as they have a small

dimension in the depth direction). It has a sensitive volume of 10.5 cm^3, with a reference point 3.47 mm below the chamber surface (which contributes to the detector 4 mm WET mentioned above). In depth doses measurements with spot scanning the dose from a single spot is measured, so it is important that the chamber is large enough so that the whole beam spot is captured in the chamber. To achieve this, Eclipse recommends (Varian [2011]) using a chamber with a radius R_{min} of at least 3 x σ_s, where

$$\sigma_s = \sqrt{\sigma^2 + 2 \times (0.0307 \times R_0)^2}. \tag{2.2}$$

The factor '3' originates from 3 x standard deviations covering approximately 99% of a Gaussian distribution; σ is the standard deviation of the Gaussian-modelled spot; the factor '2' accounts for the lateral distribution of large angle scattered protons and protons from nuclear interactions, which is approximately twice the size of primary protons; the factor '0.0307' is an empirical factor for the spot size at the peak as a function of the beam range; and R_0 is the beam range.

The largest spots in air are found for the lowest energy beams at the greatest distance from the nozzle. In water, however, the highest energies penetrate to deeper ranges, undergo more scattering and thus have larger spots. Equation 2.2 takes account of this effect. It is therefore necessary to assess all of the input lateral profiles in the x- and y-directions to determine the R_{min} required; the result of which is shown in figure 2.6 (for the x-direction only, a similar result was found for the y-direction). It can be seen that the radius of the chamber is not large enough to acquire spots with the highest energies $E \geq 218$ MeV, as it only has a radius of 40.8 mm. For these higher energies, parts of the spot tails will not be acquired. The radius is up to 6% too small, resulting in dose not being collected on the order of $\sim 1\%$. The systems assume a Gaussian profile that extends to infinity, which will affect the fitting (discussed in section 2.5.1.2). The required radius of the detector is > 43.4 mm and > 42.8 mm in the x- and y-directions respectively, for the highest energy (226.7 MeV).

2.3.1.1.1 Pinnacle³

The depth dose fitting process in Pinnacle³ was fully automated, with little input required by the user. After loading in the data in OmniPro-Accept RFA-300 ASCII

Figure 2.5: PTW 34070 Bragg peak chamber used for IDD measurements. PTW [2013]

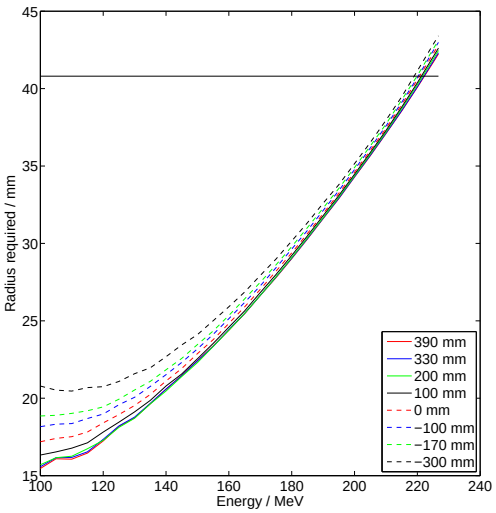


Figure 2.6: Minimum radii required in the x-direction to ensure whole spot is captured in depth dose measurements at various distances from the nozzle. The solid black line at 40.8 mm illustrates the radius of the detector used.

format and entering the possible range bands, Pinnacle[3] fitted all profile parameters to the Bortfeld function that considers both the initial energy spread and range straggling (section 2.1.6.1.1). Fitted values were generated for R_0, σ, σ_1 (a correction term to better fit the proximal tails of the peaks, especially at higher energies) and ϵ. In the automated beam modelling process it was possible to reach a minimum whereby the result is not accurate, but this was remedied by altering the initial value.

In addition to the automated fitting process it was necessary to enter a number of parameters manually. These include: (i) the z offset [g/cm^2]. This is the distance that the IDD is shifted due to measurement conditions, such as the chamber wall thickness. In our case z = 4.2 g/cm^2. (ii) R90 [g/cm^2] (depth of the distal 90% of the Bragg peak) for each energy in water. This is the depth in water that the curve nominally reaches without the measurement device wall thickness. (iii) The max Dose/MU [cGy/MU]. This is dose rate at the peak of the Bragg curve, which is used to scale the IDD curve into absolute units.

2.3.1.1.2 Eclipse

The modelling process in Eclipse was also completely automated. The only free parameters for the user are the initial energy, E_0, the straggling caused by the initial energy spread, $\tau_{incident}$ and (optionally) a coefficient called I_{Lan_1}. The first two of these were included in the input data. It was found, however, that another term I_{Lan_2} had the greatest effect on the entrance dose, but this could not be controlled by the user (in the version evaluated, v10.0.39). This proved problematic, as discussed

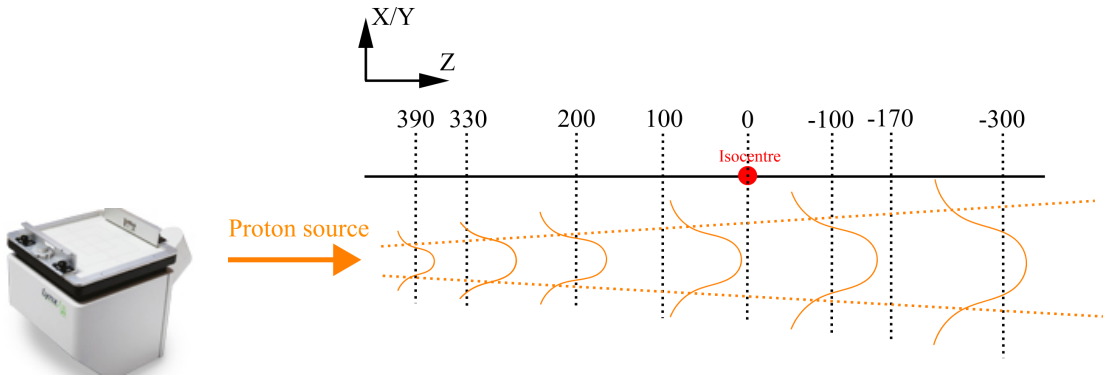
in section 2.5.1.1.1.

2.3.1.1.3 Evaluation

Typically the quality of a model's fit to data would be assessed using the chi-squared test, but such a metric would be inappropriate for assessing the fit of proton depth doses as it places too much weighting on the fitting to the Bragg peak. It could be argued that such weighting is a good thing, because the large dose deposition and sharp distal fall off of the Bragg peak makes it the most important part of the depth dose curve. However, a more complete assessment of the fitting was performed using gamma analysis. More details of this process can be found in appendix A. In quality assurance tests a 3%/3 mm stipulation is typically applied as the pass/fail criterion (Low et al. [1998]). In this analysis a tighter 2%/2 mm criterion was used in an attempt to reveal differences between the fitting.

2.3.1.2 In-Air Spot Profiles

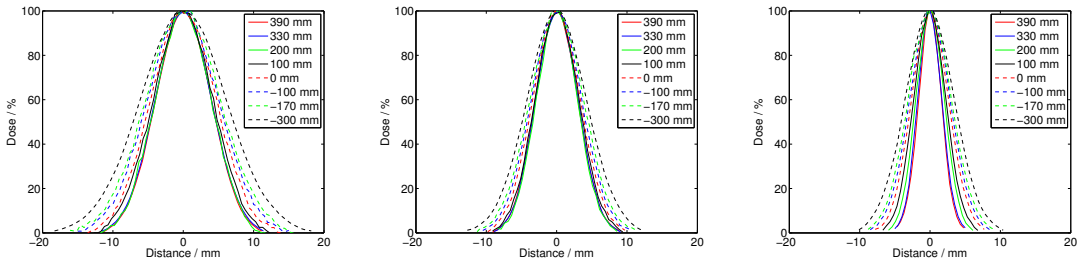
Proton spot profile measurements were acquired in air using IBA's scintillator/CCD camera system, 'Lynx', shown in figure 2.7a. Two-dimensional fluence arrays were acquired in air for planes perpendicular to the beam direction, as shown schematically in figure 2.7b.



(a) Lynx system (b) In-air spot profile measurement set up

Figure 2.7: (a) Proton spot profile measurement device. (b) Measurement positions in air. The proton beam is directed horizontally and measurements of the spot profile (orange Gaussian profiles) are made at a range of positions in air using ionisation chambers.

Measurements were made at positions from the nozzle of 390, 330, 200, 100, 0 (isocentre), -100, -170 and -300 (all mm) (negative numbers indicating distances further from the nozzle). Example spot profiles (in the x-direction) are shown in figure 2.8, demonstrating the expected decrease in profile size closer to the nozzle (higher positive numbers) and with increasing energy. Based on the 8 input measurements in air, the beam divergence and thus the initial source size was modelled, for the 27 energies available and in both the x- and y-directions.



(a) (b) (c)

Figure 2.8: Spot profiles in the x-direction acquired at a range of distances from the nozzle for (a) 100 MeV, (b) 150 MeV and (c) 220 MeV.

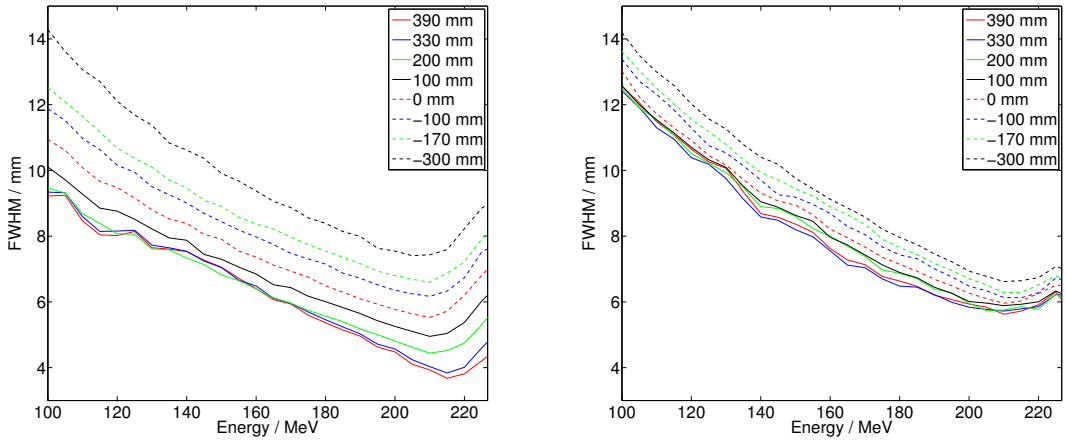


Figure 2.9: Spot sizes (FWHM) in the x- and y-directions (left and right respectively), for a range of energies and measurement positions.

Assuming a Gaussian shape, the spot width (full-width at half-maximum, FWHM) was determined. The variations of x- and y-FWHMs with energy and measurement position for the 512 measured profiles (8 measurement positions x 27 energies x 2 directions (x/y)) can be seen in figure 2.9. The spot size is dependent on the collimator slit settings at the energy selector system and on the tuning of the quadrupoles in the nozzle, both of which depend on energy. The general decrease with increasing energy is to be expected as higher energy beams undergo less scattering in air than lower energy beams, however the increase in spot size towards the highest energies suggests a different set of collimator/energy selector settings are used above 210 MeV. Additionally it can be seen that the profiles are much more divergent in the x-direction than the y-direction (different measurement positions are more spread out). This is because the horizontal quadrupole (x-direction) is further upstream than the vertical quadrupole (y-direction).

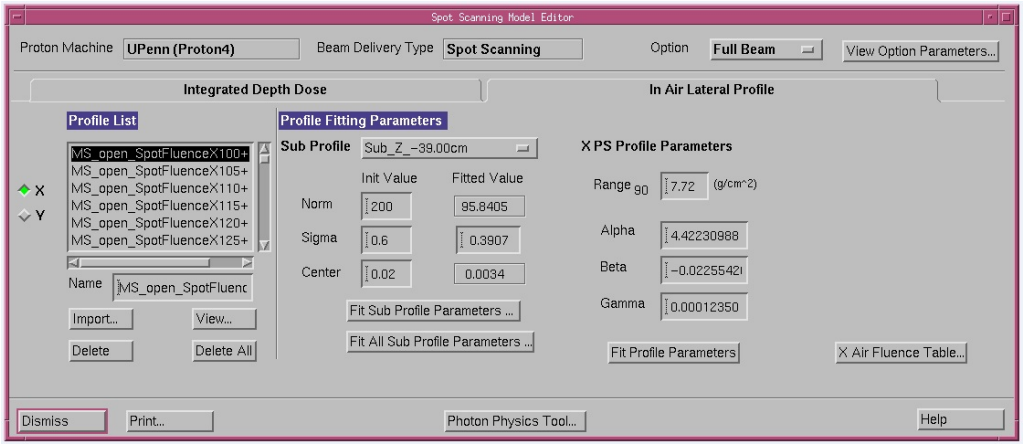


Figure 2.10: Pinnacle^3 user entry screen for fitting lateral profiles.

2.3.1.2.1 Pinnacle^3

Similar to depth doses, the fitting process for lateral profiles was fully automated in Pinnacle^3, with the user screen shown in figure 2.10. Two-dimensional Gaussian functions are fitted to OmniPro-Accept RFA-300 ASCII format profiles (same format as IDD curves). Values are fitted for the centre, σ and the normalisation factor for all measurements at different z positions.

After entering the R90 and with the profile shapes and range known, Pinnacle^3 modelled the source size as a function of distance for each energy using a hyperbolic equation. If the data is too sparse the result will be less accurate and so it was recommended that the user imports sets of profiles in X and Y every 1-2 g/cm^2 ($\approx$ 5-10 MeV), which was satisfied as energy measurements were made every 5 MeV.

The research version analysed (April 2012) modelled the source as a two-dimensional Gaussian. Later versions (February 2013, not assessed in this work) used a double Gaussian model to better account for large non-Gaussian tails at wide angles. This caused a problem with the depth doses, as detailed in section 2.5.1.1, and it would be of interest to determine the accuracy of the improved profile modelling in the later versions.

2.3.1.2.2 Eclipse

The fitting process for lateral profiles was also fully automated in Eclipse, requiring the measurement of at least 3 profiles at different positions per energy. Using these profiles the angular divergence was calculated and the source position was modelled by projecting back to a zero field size.

2.3.1.2.3 Evaluation

The fitting of each system to these profiles was analysed using gamma analysis, in the same approach as for depth doses (section 2.3.1.1.3). The gamma index is a com-

bination of the dose difference (DD) and the distance-to-agreement (DTA) between two points, but the 2%/2 mm criteria used for the IDD analysis was not suitable for spot profiles because of the small dimensions. It was therefore necessary to impose a much tighter restriction on the DTA of 0.05 mm. Prior to any gamma analysis both the measured and fitted profiles were resampled to a very high resolution (every 0.01 mm).

2.3.1.3 General Machine Parameters

The attributes of the machine were also defined during beam commissioning. For instance, the range of gantry rotation determines the angles beams can be directed from. Other information entered included: range of couch angles; nozzle components; source-to-axis distance; magnet-to-axis distance; possible range shifters and the range of snout sizes.

As stated previously, a consistent minimum spot MU limitation (0.021 MU) was enforced on all systems to ensure the plans were deliverable. If this practical limitation were to be removed, the plans would look better but they would have no practical consequence.

2.3.2 Treatment Approach

Fixed horizontal SFUD fields were employed in the study because, at the time of writing, they were probably the most common implementation of spot scanning proton beams due to their relative robustness over IMPT fields to range uncertainties (Trofimov et al. [2007]). Using a fixed horizontal beam line reduces the flexibility of angles available to the planner, but this could be seen as an advantage as there is less choice and therefore a greater possibility of the optimum angle being selected. Additionally, SFUD plans are easier to plan than IMPT plans, which is advantageous due to the planner's inexperience. Also, by removing the additional degrees of freedom provided by IMPT plans, it was hoped differences between systems could be more easily identified.

2.3.3 Patients

Ten meningioma patients, treated at our institution using a RapidArc X-ray radiotherapy treatment, were selected for the comparison. The case load included tumours, with and without overlapping structures, of different sizes (target structures ranged from 45.5 to 248.4 cm^3) and grades. Full details can be found in table 2.3.

Patient	Age [yrs]	Grade†	Details	Beam direction(s)
1	57	2	Parafalcine	LSO, RSO
2	67	2	Frontal	Vertex
3	40	-	Cavernous sinus	LLat, LSO
4	67	1	Cavernous sinus	LLat, LSO
5	40	1	Sphenoid wing	RLat, RSO
6	31	2	Cavernous sinus	LLat, LSO
7	19	2	Sphenoid wing (brain invasion)	LLat, LSO
8	40	-	Cerebellopontine angle	RLat, RSO
9	55	-	Left orbital apex / sphenoid	LLat, LSO
10	62	-	Cavernous sinus	LLat, LSO

Table 2.3: Patient and tumour characteristics. †Grade only measured for patients in which a biopsy was practically possible. Beam direction abbreviations: L=left; R=right; S=superior; O=oblique; Lat=lateral.

2.3.4 Target volumes and OARs

The gross tumour volume (GTV) and clinical target volume (CTV) were outlined according to departmental protocol. Each CTV was expanded by 5 mm to a pencil beam scanning target volume (PBSTV). This expansion, which is 2 mm larger than that employed to construct the departmental PTV used for X-ray radiotherapy planning, is to make the plans sufficiently robust to range uncertainties. Conventionally, margins associated with range uncertainties should only be applied in the proton field direction, but none of the systems allowed for the production of beam-specific PTVs. Additionally, it was much more convenient to work with a single target volume. The use of a consistent margin was appropriate for all patients as the distances to the target distal edge were approximately similar (9.9-13.9 cm) and thus the range uncertainties (a prescription of 3.5% × range + 1 mm is employed at UPenn for PBS treatments) did not vary much (4.5-5.9 mm).

The dose prescription was 50.4 Gy in 28 fractions. The dose-volume constraints for the coverage of the PBSTV, together with the OAR constraints, are detailed in table 2.5. All OAR constraints apply to the planning risk volumes (PRVs), formed by 3 mm expansions from the corresponding structures to account for patient motion. Although no specific constraints were applied, attempts were made to keep the brain dose as low as possible without sacrificing target coverage and the mean dose to the healthy brain was assessed.

2.3.5 Planning strategy

The study objective was to compare clinically-equivalent plans between TPSs, individually optimised for each system. The decision was therefore made to try to give the planners the freedom they would have were they to use each TPS for real

planning (i.e. allow choice of objectives and priorities), but with some artificial constraints imposed to keep the test fair. Similar approaches have been used by other authors conducting TPS comparison studies (Fogliata et al. [2003, 2005]).

Beam geometries, patient isocentres and outlined structures were made consistent between all systems. The beam arrangements used are detailed in table 2.3. The same clinical constraints (table 2.5) were used for all patients, however the choice of numerical objectives that achieves these goals differs between systems and so was not fixed.

2.3.6 Layer spacing, spot spacing and spot positions

It was not possible to define the layer spacing in an identical fashion in each system, although Pinnacle^3 and XiO have similar approaches. For Pinnacle^3, the layer spacing is variable and dependent on the Bragg peak width. The system places layers such that the distal 80% of the shallower layer matches the proximal 80% of the next deepest layer (in later versions the user can alter the distal percentage value to alter the spacing, but this was not possible in the version we tested). In XiO, the spacing is also variable and equal to the Bragg peak width (defined in commissioning, with a definition of the user's choosing) times by a peak width multiplier (an integer value selected by the user during planning). To ensure consistency with Pinnacle^3, we set the Bragg peak widths in commissioning as the 80-80% width of each pristine Bragg peak, and used a peak width multiplier of 1. For Eclipse, the layer spacing cannot be defined in the same way, but it can be configured with the following options during commissioning: fixed distance (in mm) available throughout the energy range; fixed change in energy (in MeV); fixed distance (in mm), determined based on the range sigma of the highest or lowest energy per field; or variable distance, equal to the range sigma of the next highest energy layer multiplied by a user-defined multiplication factor. In our study, the variable distance option was selected in an attempt to maintain consistency with Pinnacle^3 and XiO. The range sigma used in Eclipse only accounts for the energy spread of the initial beam, whereas the Bragg peak width used in Pinnacle^3 and XiO also accounts for the range straggling. To determine what multiplication factor should be used we experimented with beams of different range and modulation, both with and without a range shifter. The further the layers were spaced apart, the larger were the dose ripples along the beam direction. The closer they were together, the more layers there were (hence longer delivery time) and the greater becomes the sensitivity to the minimum spot MU problem (more layers mean fewer MUs per layer and, hence, per spot). This led to spikes and troughs as spots below the minimum MU threshold are either rounded up to the minimum MU or rounded down to zero. The dominating phenomenon depends on the range

and modulation. However, because Eclipse's beam configuration only allowed for the choice of a global value, it was found that using 4 × the range sigma was a good compromise between these competing factors across all the beams studied.

Symmetric lateral spot spacings were defined according to the spot size at the most distal layer of the target, such that there was sufficient overlap between neighbours to avoid appreciable dose ripples. Lateral margins were then set to be equal to this spacing to ensure at least one additional ring of spots was located outside the PBSTV, reducing the possibility of highly weighted spots close to the target edge.

For Eclipse and XiO the available spot positions are defined by a 3D rectangular grid passing through the isocentre, within the target boundaries (and margins). Pinnacle^3 defines 2D square grids for each layer, starting from the left hand side of the target (in the beam's eye view), such that the available positions of successive layers are often offset.

An attempt was made to minimise any variation in planners' ability with each system by asking for feedback on the plan quality from the individual vendors.

2.3.7 Optimisation options and dose calculation parameters

Optimisation involves minimising some variable that quantifies target volume coverage and OAR sparing. During this optimisation process, the stopping tolerance was set to a suitably small value relevant to each system (0.001 in Eclipse, 10^{-5} in Pinnacle^3 and 0.0001% in XiO), so that an optimal dose distribution was ensured but also that the optimisation did not take longer than 30 minutes (the maximum number of iterations was never reached). Although an important factor in optimisation processes, differences between computing powers of the individual workstations made timing comparisons infeasible.

Each system has different optimisers and/or options available, as summarised in table 2.4. In this study the plans were quantified through dose volume histogram (DVH) metrics, so for systems in which there is a choice we selected the optimiser that gave the best dosimetric plan. No robustness options and/or features were tested. Eclipse has two available optimisers: (i) the simultaneous spot optimisation (SSO) algorithm, which is based on a scanning optimisation algorithm (Lomax et al. [1996]); and (ii) the conjugate gradient (CG) algorithm. Both optimisers can produce SFUD and IMPT plans, but only SSO was used in this study as it gives better DVHs (Varian [2011]). Pinnacle^3's optimiser, IMPT, allows for the production of either SFUD or IMPT plans (only the former was tested in this study). A robustness option and the creation of erroneous patient setup scenarios was also available, but was not tested in this study. XiO has three options when optimising: (i) beamwise optimisation of fluence, which produces SFUD plans; (ii) full intensity

System	Optimiser	Comments	Used in study
Eclipse	Simultaneous spot optimisation (SSO)	Better DVH	✓
Eclipse	Conjugate gradient (CG)	Smoother distributions	×
Pinnacle^3	IMPT	Can produce IMPT and SFUD plans	✓
XiO	Beamwise optimisation of fluence	SFUD plans only	✓
XiO	Full intensity modulated proton therapy	IMPT plans only	×
XiO	Sequel beamwise optimisation	Compromise between SFUD robustness and IMPT coverage	×

Table 2.4: Optimisers available in each proton TPS.

modulated proton therapy, which produces IMPT plans; and (iii) sequel beamwise optimisation, which provides a compromise between the robustness of SFUD plans and the coverage of IMPT plans. In this study only the first option was utilised as only SFUD plans were produced. A smoothing option was available during the optimisation, but was not employed.

Dose was calculated on all systems with a grid size of 2.5 mm, using each system's most accurate dose calculation algorithm (for Eclipse this is 'Proton Convolution Superposition', for Pinnacle^3 'Proton PBS', for XiO 'Pencil Beam Algorithm'). All three systems account for heterogeneities and model nuclear interactions. As the accuracy of each dose calculation algorithm was not assessed, full details of the different algorithms can be found in the relevant literature (Philips Medical Systems [2013]; Soukup et al. [2005]; Varian [2011]).

The MUs of each system were normalised to be identical for a uniformly-irradiated 5 cm × 5 cm × 5 cm cube, 5 cm deep within a water phantom. The MUs needed to be comparable between systems because the uniformity of spot weights was to be assessed (see section 2.3.8 below). The 5 cm depth was chosen to ensure that the available 74 mm (water-equivalent) thick range shifter was included, as the targets in all patients extend more proximally than the range of the lowest available energy and thus required the use of this device.

2.3.8 Evaluation

To avoid discrepancies in the final volume building of DVHs, datasets were exported with consistent dose bin widths (0.1 Gy) and analysed independently using CERR, the Computational Environment for Radiotherapy Research (Deasy et al. [2003]). For each patient a set of parameters was computed from the DVHs, as detailed in

table 2.5.

The uniformity and distribution of spot weights from each field was also analysed using 3D spot maps. The spot MUs were determined by multiplying the spot weight by the calibrated MUs for the given field. A parameter, C, was defined to allow quantitative comparisons between systems:

$$C = \frac{\sum_{i=1}^{N} (w(i) \times d(i))}{\sum_{i=1}^{N} w(i)} \tag{2.3}$$

where w is the weight and d the distance from the isocentre for a spot i. The total number of spots is N. The distance is calculated using the x- and y-co-ordinates and energy (converted to a water-equivalent Bragg-Peak depth) for each spot.

2.4 Results

2.4.1 Proton beam commissioning

2.4.1.1 Integrated depth doses

For clarity it is easiest to assess the fitting by splitting the energies up into ranges. At low and mid-range energies (100-160 MeV) Pinnacle^3 fits well on the proximal slope of the Bragg peak, but overestimates the dose on the distal slope. Figure 2.11 shows an example set of fits at 110 MeV, with the gamma analysis statistics for these fits in figure 2.12. In figure 2.12a there are four plots illustrating the breakdown of the gamma analysis: (top left) the distance-to-agreement and (top right) dose difference along the depth of the Bragg peak; (bottom left) the agreement and (bottom right) absolute agreement (whereby the origin represents perfect agreement). Figure 2.12b shows the composite of the distance-to-agreement and dose difference along the Bragg peak, known as the gamma index. Pinnacle^3's overestimation of the Bragg peak position (figure 2.11, at 9-9.5 cm) is reflected by an increase in the gamma index at these depths (figure 2.12b). It is important the system models the position and dose around the Bragg peak well as one of the advantages of proton therapy is the ability to treat tumours very close to critical structures. While underestimating the Bragg peak depth could have very damaging results (a deeper actual depth could irradiate distally-adjacent critical structures), overestimating will have a similarly poor result of underdosing the target. In this energy range (100-160 MeV) the fitting of Eclipse is much better, as demonstrated by the greater density of points at the origin in the lower left plot of figure 2.12a. Its fitting to the distal slope is much better, but sometimes there is a slight overestimation of the plateau dose, as can be seen in figure 2.11 (≈ 5.5 cm), with an increase in the gamma index in figure 2.12b

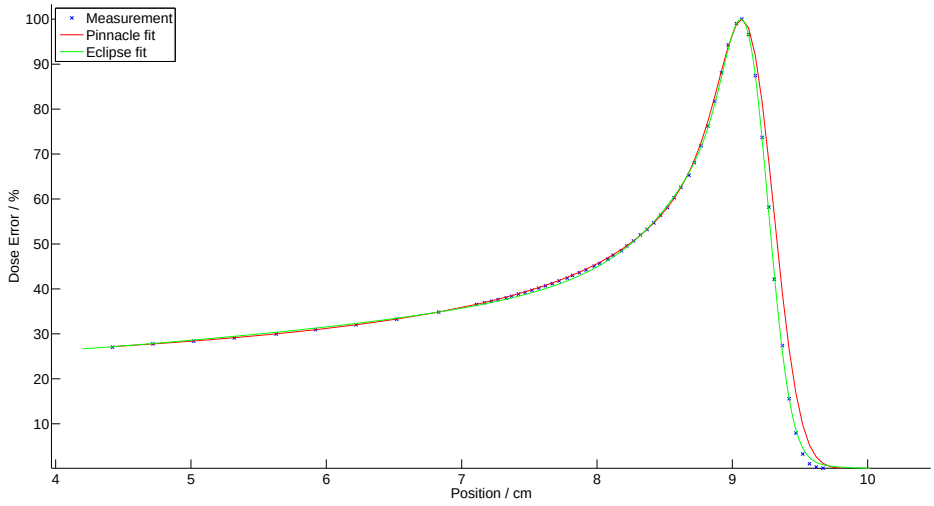


Figure 2.11: Integrated depth dose curves for a 110 MeV proton beam in water.

at this depth.

At higher energies (165-226.7 MeV), Pinnacle^3 still suffers from an overestimation of the distal dose (figure 2.13). Eclipse slightly overestimates the plateau dose, but also overestimates the effect of build-up. As detailed in section 2.1.6.1.2, one of the benefits of using the beamlet model is that build-up is incorporated into the depth dose, unlike the Bortfeld model (section 2.1.6.1.1). However, there is an underestimation of dose at the shallowest depths that gets larger (reflected by an increase in the gamma index at shallow depths in figure 2.14b) until about 200 MeV, where it reaches 2.5%. It remains at this value for the higher energies.

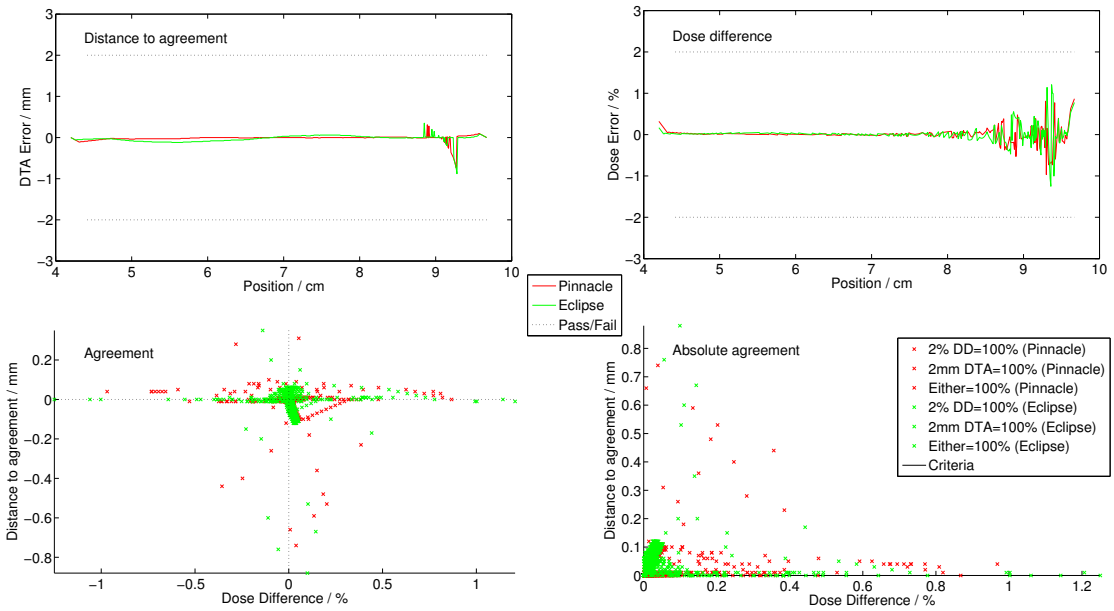
Figure 2.15 shows the variation of gamma index with energy. It can be seen that, for both systems, the fitting worsens with increasing energy. Pinnacle^3 suffers from the biggest single errors, but Eclipse suffers more often at higher energies because of the poorly modelled build-up.

2.4.1.2 In-air spot profiles

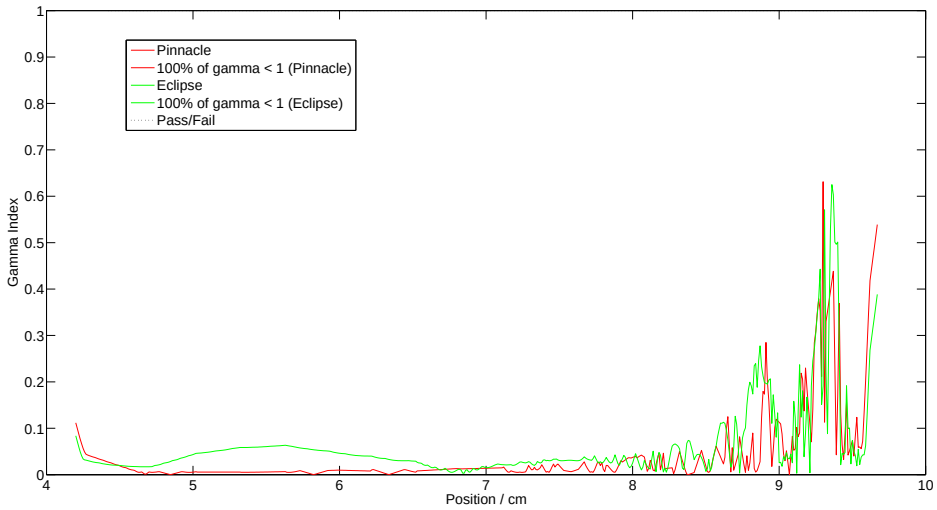
A representative example of the fitting of each system to the measured y-profile data can be seen in figure 2.16a, for 145 MeV. It can be seen that, in general, both systems do not fit well to the tails. In the same format as the depth dose analysis, the statistics and variation of gamma index with position can be seen in figure 2.17.

The fitting of each system to the x-profile is generally better, as shown in figure 2.16b.

The mean/max gamma index for different energies is shown in figure 2.18. The mean gamma for each energy versus the measurement position can be seen in figure 2.19 (where negative values are further from the nozzle). There appears to be little



(a) 110 MeV statistics



(b) 110 MeV gamma index

Figure 2.12: Gamma index analysis for the 110 MeV proton IDD.

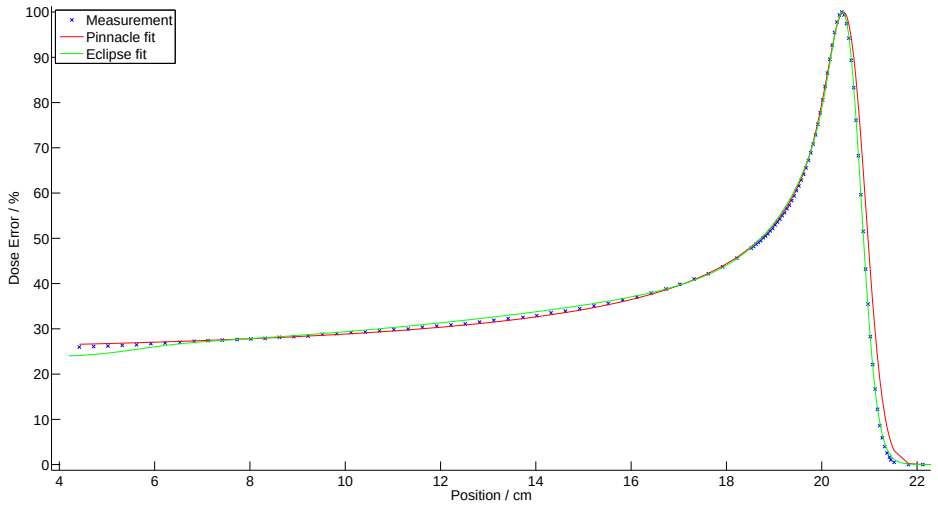


Figure 2.13: Integrated depth dose curves for a 175 MeV proton beam in water.

correlation with position.

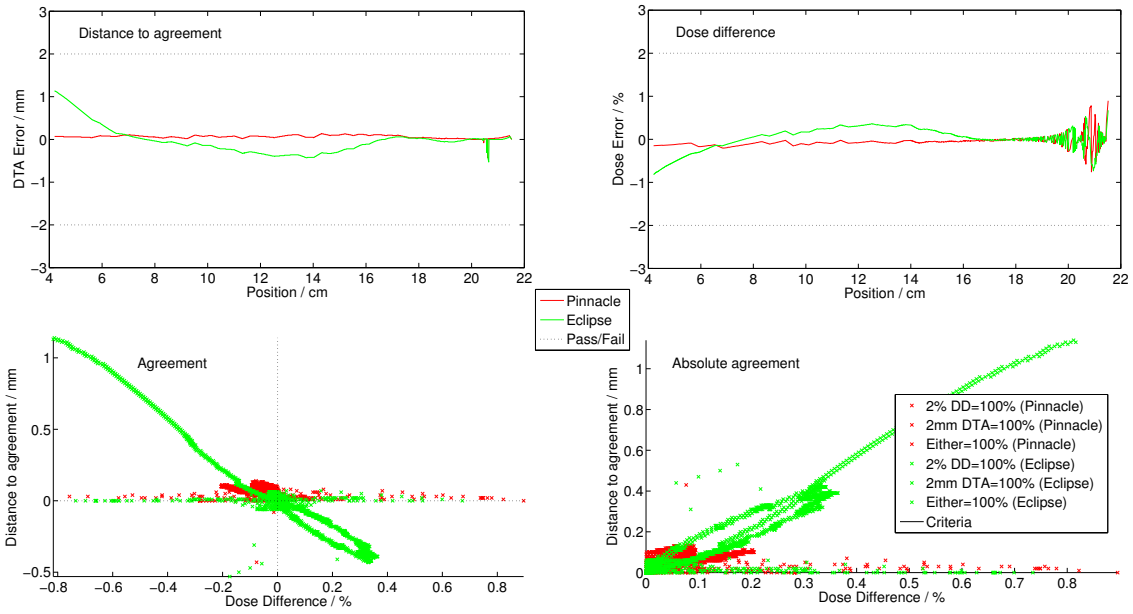
2.4.1.2.1 Summary

In summary, both systems fit to the measured data within clinical tolerances. For the integrated depth dose curves, Pinnacle[3] has a tendency to overestimate the distal Bragg peak depth while Eclipse overestimates the effect of build up at high energies. Both systems generally fit well to the spot profiles, although neither are able to model the tails in the y-direction.

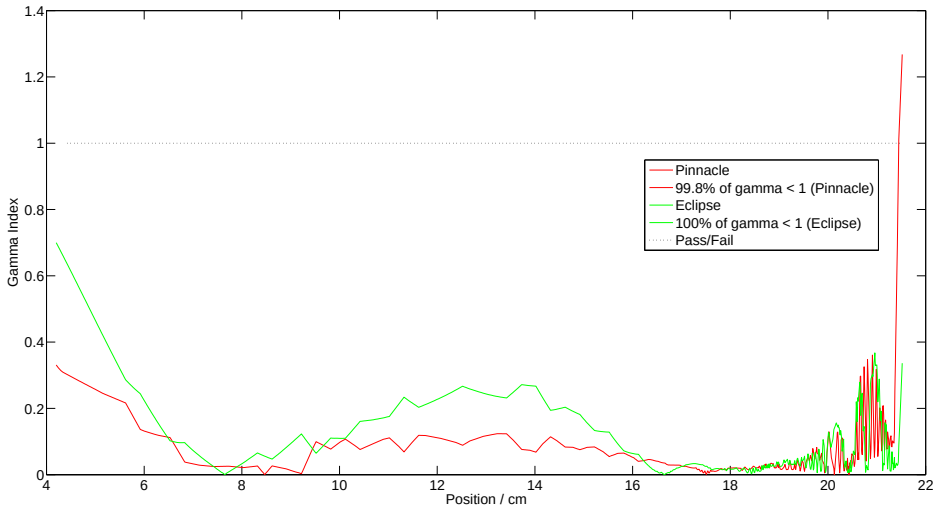
2.4.2 Treatment planning comparison

Table 2.5 details the mean target coverage and OAR sparing for the three systems, for a variety of parameters typically assessed during treatment planning. A mean value for each statistic is given in units of Gy, with an error defined by the standard error on the distribution across the ten patients. Although the errors are larger for the OARs because each case requires different organs to be spared, the values are still useful for comparison between systems because each system had the same range of cases. The bracketed values in the constraint column are the percentages of the prescribed dose (50.4 Gy), while the bracketed values in the other columns are the numbers of patients (out of ten) that fail to meet the higher/lower constraints. One-way analyses of variances (ANOVAs) were calculated between the three systems for each metric, with p-values shown in the far right column.

Figures 2.20 and 2.21 provide a graphical representation of these metrics, with the boxplot showing the distribution of results across the ten patients. The edges of each box are formed by the 75^{th} (q_3) and 25^{th} (q_1) percentiles; the whiskers extend



(a)



(b)

Figure 2.14: Gamma index analysis for the 175 MeV proton IDD.

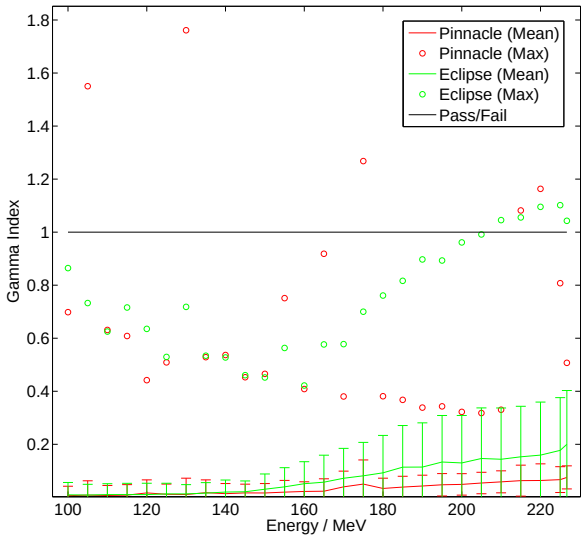
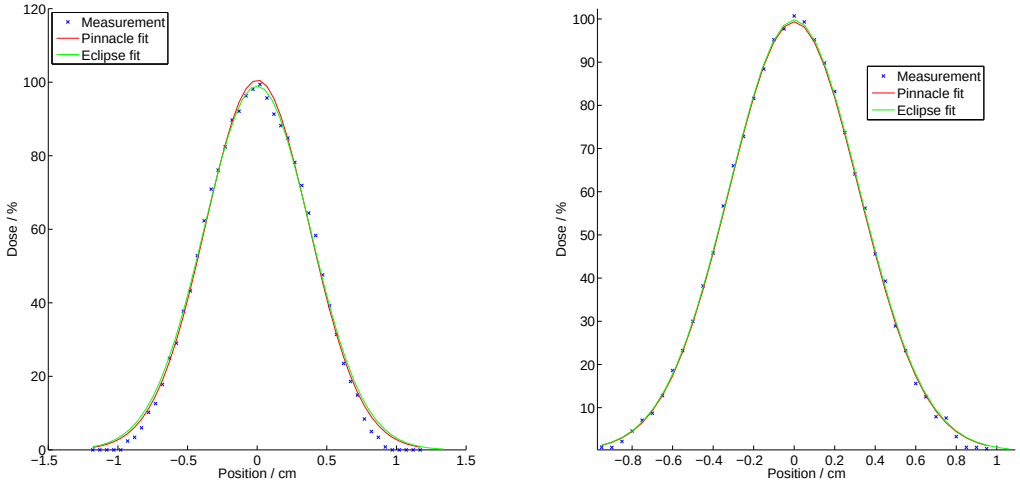
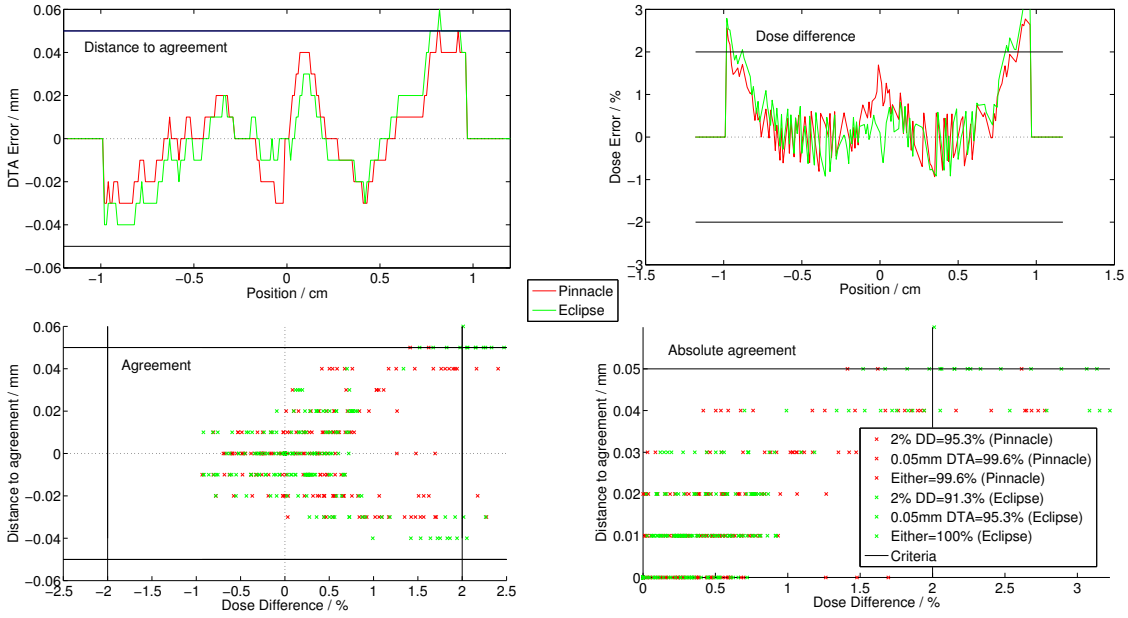


Figure 2.15: Variation of mean (line) and max (open circles) gamma index with energy for depth doses. Error bars are formed by the standard deviations of gamma index values across all measurement positions.

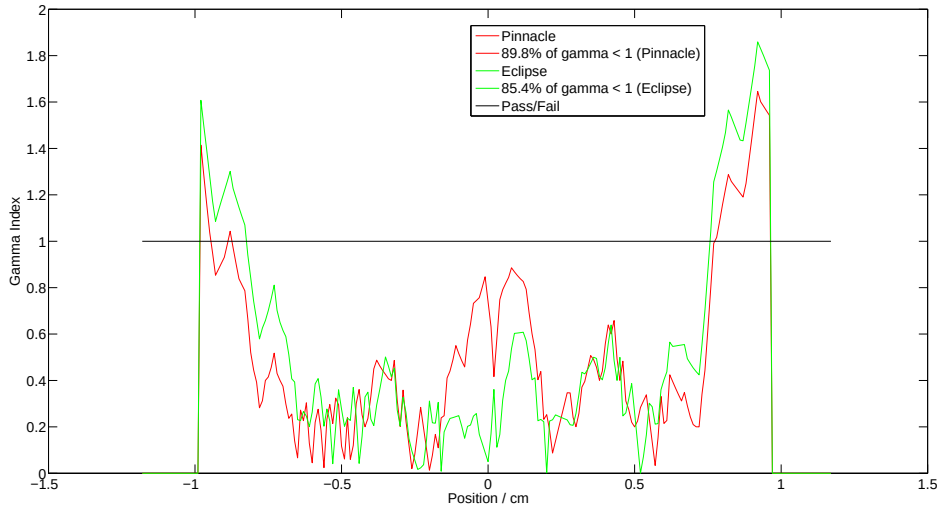


(a) (b)

Figure 2.16: (a) Y profiles at 145 MeV. (b) X profiles at 135 MeV.

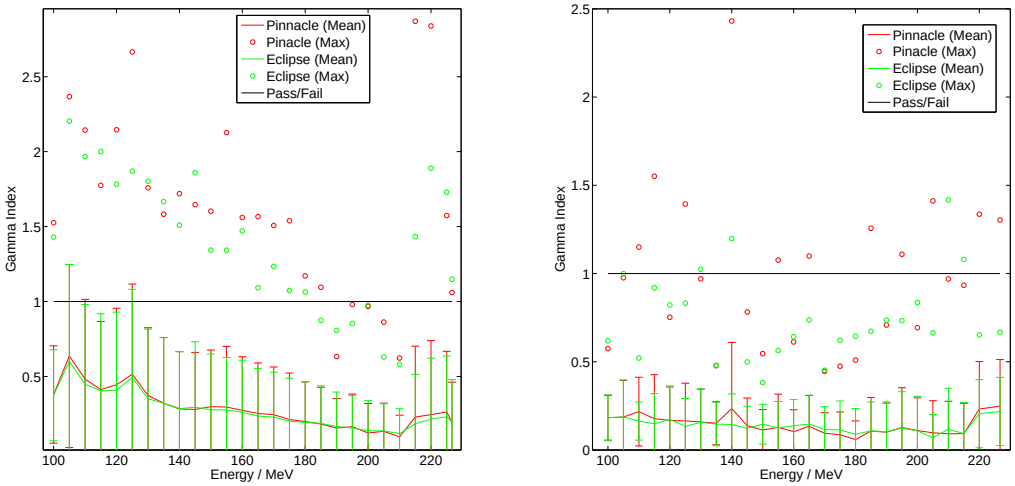


(a)



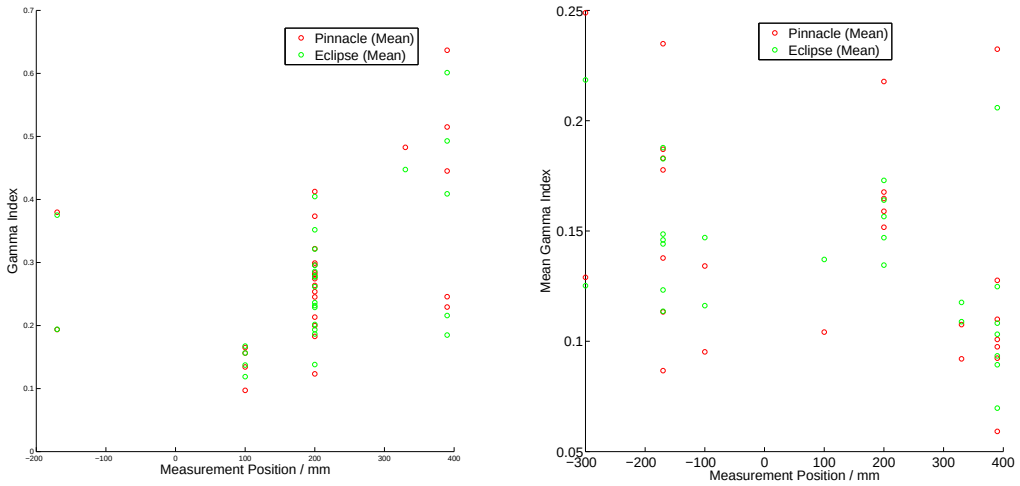
(b)

Figure 2.17: Gamma index analysis for the 145 MeV y-profile. The solid black lines indicate the pass/fail criteria.



(a) (b)

Figure 2.18: Variation of mean (line) and max (open circles) gamma index with energy for spot profiles in the y- (a) and x-direction (b).



(a) (b)

Figure 2.19: Variation of mean gamma index with position for spot profiles in the y- (a) and x-direction (b).

Structure	Metric	Constraint / Gy	Eclipse / Gy	Pinnacle^3 / Gy	XiO / Gy	p
PBSTV	D2 (107%)	< 53.9	52.4±0.1 (0)	51.4±0.3 (0)	51.0±0.1 (0)	$1.7\text{x}10^{-4}$*
	D5 (105%)	< 52.9	52.0±0.1 (0)	50.9±0.2 (0)	50.7±0.1 (0)	$5.6\text{x}10^{-7}$*
	D95 (95%)	> 47.9	46.5±0.8 (7)	48.0±0.6 (2)	48.0±0.6 (1)	0.21
	D99 (90%)	> 45.4	43.2±1.6 (6)	43.1±1.9 (3)	43.7±2.1 (4)	0.97
	V95	-	91.3±1.9%	95.8±1.4%	95.6±1.0%	0.041
BStem	Max	< 55	41.5±6.8 (0)	41.7±6.5 (0)	41.0±6.5 (0)	1.00
	Mean	20.0±4.0	14.7±3.1	21.2±4.0	0.43	
R Glo	Max	< 45	22.0±4.5 (0)	14.9±4.8 (0)	18.6±4.9 (0)	0.58
	Mean	6.1±2.7	3.6±2.5	4.7±3.0	0.81	
L Glo	Max	< 45	33.8±4.8 (1)	31.7±5.5 (1)	31.6±4.9 (0)	0.94
	Mean	12.4±2.9	8.8±2.8	7.0±1.7	0.31	
R ON	Max	< 50	31.8±5.7 (1)	27.4±6.0 (0)	33.1±5.3 (1)	0.76
	Mean	15.5±3.8	10.3±3.3	13.9±3.6	0.58	
L ON	Max	< 50	42.8±4.9 (2)	41.9±5.2 (2)	43.5±4.6 (3)	0.98
	Mean	34.3±5.4	32.8±5.8	32.0±5.1	0.95	
O Chi	Max	< 50	44.2±4.9 (0)	44.6±4.9 (1)	45.9±4.1 (2)	0.97
	Mean	38.1±5.5	37.3±5.6	39.7±4.9	0.95	
R Lens	Max	< 6	6.7±2.3 (5)	1.7±1.0 (1)	1.2±0.4 (1)	0.024*
	Mean	2.7±1.4	0.5±0.3	0.3±0.1	0.094	
L Lens	Max	< 6	11.9±3.3 (6)	6.2±3.2 (4)	2.6±0.6 (5)	0.063
	Mean	6.6±2.2	3.0±2.1	1.0±0.3	0.092	
Brain	Mean		8.2±0.6	6.5±0.6	8.9±0.6	0.022*

Table 2.5: Dose-volume constraints and coverage statistics. All OARs refer to PRVs. The bracketed values in the constraint column are the percentages of the prescribed dose (50.4 Gy). The bracketed values in the Eclipse, Pinnacle^3 and XiO columns are the number of patients that failed to meet the criteria. Abbreviations: PBSTV = pencil beam scanning target volume; BStem = brainstem; R/L Glo = right/left globe; R/L ON = right/left optic nerve; O Chi = optic chiasm; R/L Lens = right/left lens; Brain = healthy brain. *Significant.

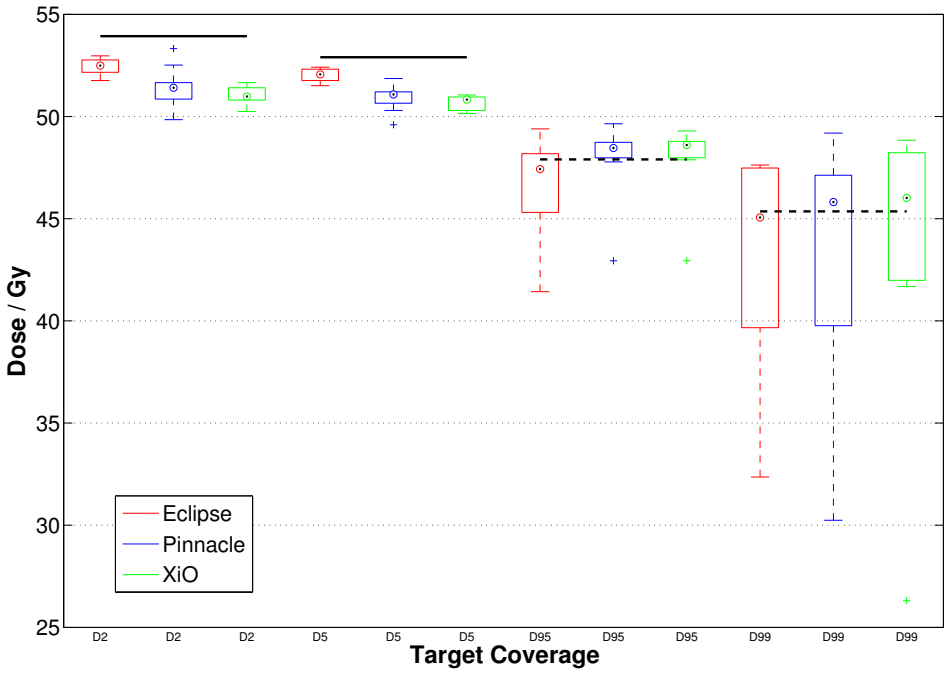


Figure 2.20: Target coverage statistics for the ten patients, for each system. Boxplots are as described in the text, with the upper/lower constraints of each statistic (from table 2.5) shown by the solid/dashed black lines, respectively.

to the most extreme value that is not an outlier; points are considered outliers if their results are greater than $q_3 + w(q_3 - q_1)$ or smaller than $q_1 + w(q_3 - q_1)$ (where w is set to 1.5) and are shown by crosses; and the median is shown by the black circle within the box. As an illustration, figure 2.22 shows dose distributions for patient 8 planned on all three systems.

To assess the uniformity and distribution of spot weights of the plans produced by each system, the weight of each spot (normalised to each system's mean) was plotted against its absolute distance from the isocentre. This is shown in figure 2.23(a), together with the calculated values for C (equation 2.3), for all patients and all fields. The mean weight of all spots at every 1 mm, is shown in figure 2.23(b) for all systems.

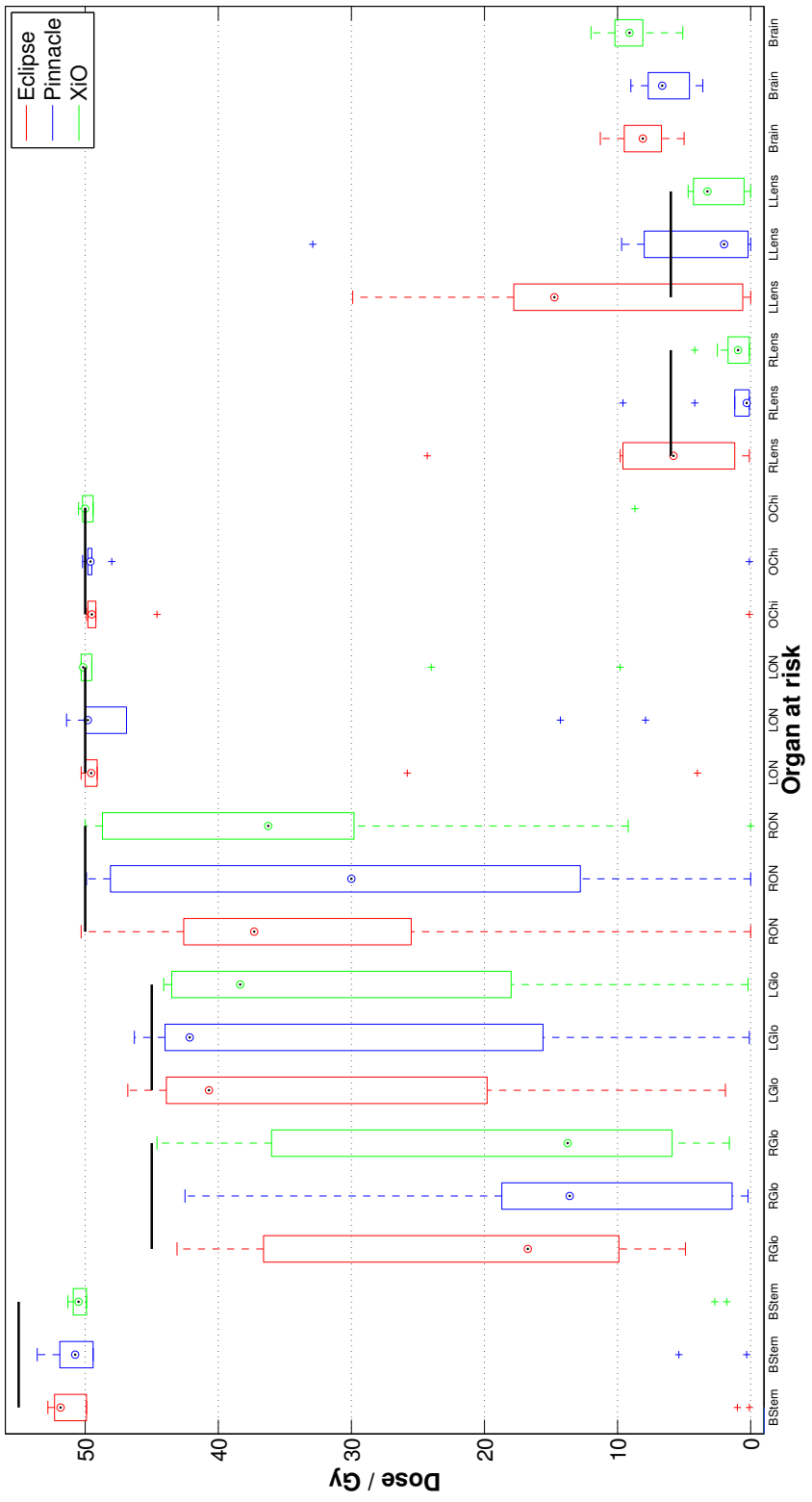


Figure 2.21: OAR doses for the ten patients, for each system. Maximum doses are shown for the PRVs of (left to right) brainstem, right globe, left globe, right optic nerve, left optic nerve, optic chiasm, right lens and left lens. Mean doses are shown for the healthy brain. Boxplots are as described in the text, with the constraints of each OAR (from table 2.5) shown by the corresponding solid black line.

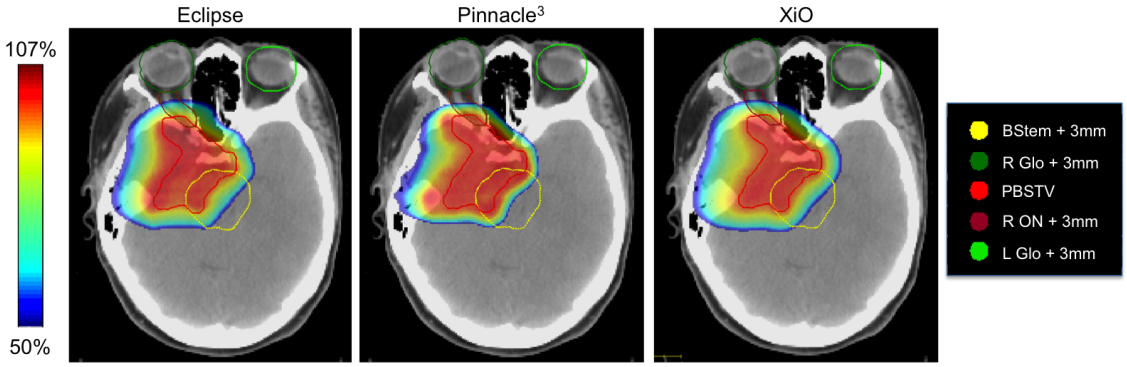


Figure 2.22: Dose distributions for patient 8. Structure abbreviations can be found in table 2.5.

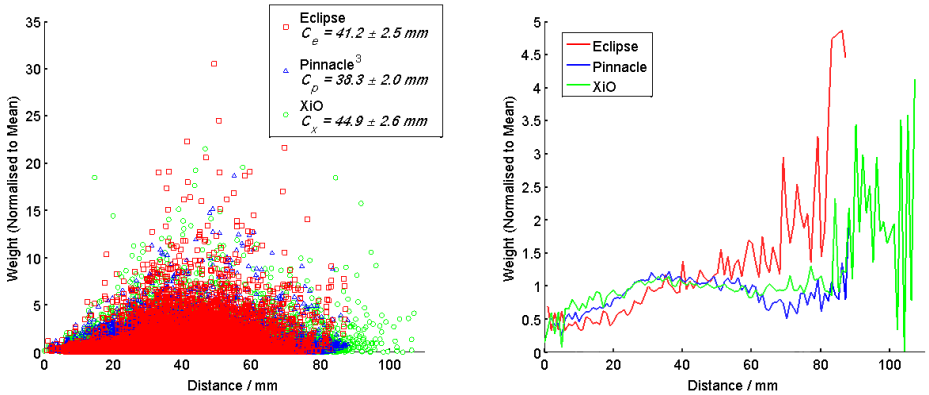


Figure 2.23: (a) Relationship between the normalised spot MUs and distance from the isocentre, for all patients, for Eclipse (red squares), Pinnacle^3 (blue triangles) and XiO (green circles). The calculated values for C (equation 2.3) are displayed on the figure for Eclipse (C_e), Pinnacle^3 (C_p) and XiO (C_x). (b) The mean weight of all spots at every 1 mm, for each system (same colour scheme).

2.5 Discussion

2.5.1 Proton beam commissioning

2.5.1.1 Integrated depth doses

In general the fitting of both systems was good, with both models within accepted clinical tolerances. There were a number of areas of concern, however.

2.5.1.1.1 Build-up

If the Bragg peak is measured with little or no solid material before the water (no tank wall and very thin dosimeter wall) the depth dose curve should demonstrate electron and nuclear build-up in the entrance region (Gottschalk [2004]). Electrons generated by electromagnetic interactions build-up to equilibrium very quickly in proton beams ($\approx 1\,\text{mm}$), which is one of the disadvantages of protons over photon beams (little or no skin sparing). Nuclear build-up occurs over a larger distance ($\approx 1-2\,\text{cm}$ in water), but it is smaller because nuclear secondaries account for only a small fraction of the absorbed dose. To date there have been very few measurements of nuclear build-up. Carlsson and Carlsson [1977] used a 185 MeV proton beam incident on a stack of $(C_2F_4)_n$ sheets interleaved with thermoluminescent detectors. They observed a nuclear build-up of 7% over a characteristic length of $2\,\text{g/cm}^2$ (i.e. 2 cm in water).

The measured data acquired at UPenn does not demonstrate any build-up because the measurements do not go below 4.2 cm (water tank wall thickness + surface offset + chamber entrance window thickness), at which point the build-up has reached equilibrium. The Bortfeld model does not account for any build-up of dose, so the model (somewhat fortuitously) fits well. One of the major failings in the Eclipse modelling was an overestimation of the build-up, as detailed in section 2.4.1.1. For energies $E \geq 175\,\text{MeV}$ the Eclipse model introduces a build-up of 2-3% up to a depth of $\approx 7\,\text{cm}$, as shown in figure 2.24.

A summary of the contributions of the six main components to the primary proton depth dose model of Eclipse is shown in figure 2.25. It can be seen that the I_{Lan_2} term takes account of the build-up at higher energies, but as stated in section 2.3.1.1.2, this term could not be controlled by the user. From our results it is clear that the Eclipse model has not been benchmarked against datasets for which you do not expect to see build-up. This issue was relayed to Varian and the I_{Lan_2} term became controllable in later versions of the software.

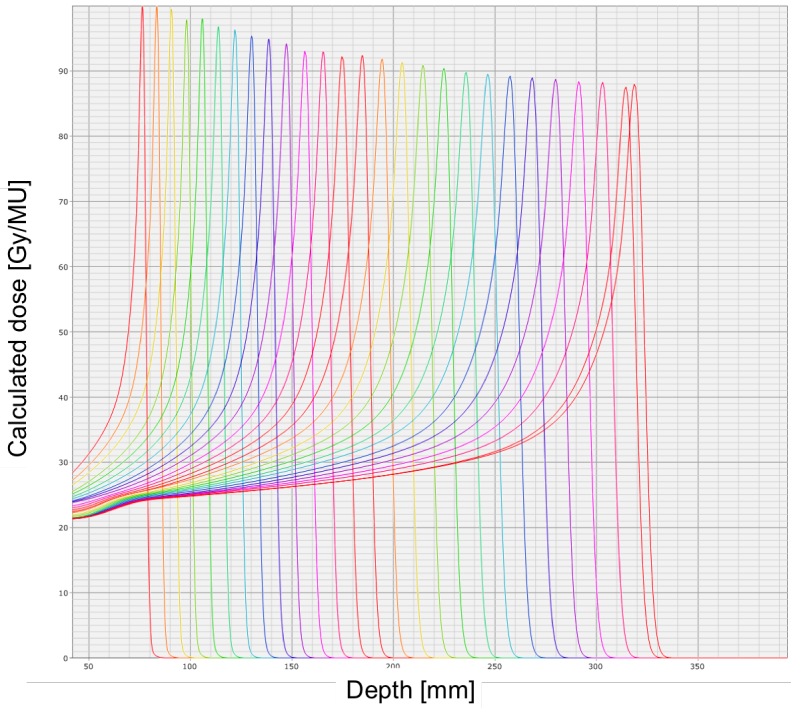


Figure 2.24: Eclipse calculated depth dose curves for UPenn measured data (100-226.7 MeV), demonstrating an erroneous build-up region up to 70 mm for the highest energies.

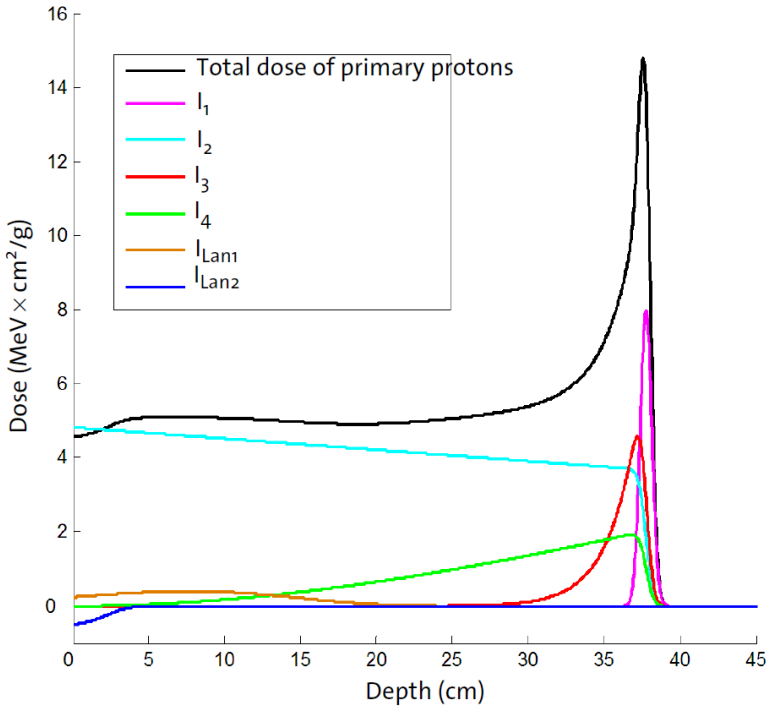


Figure 2.25: Contribution of components to the total dose deposited by primary protons, for a 250 MeV proton beam in water, using the beamlet model. Definitions of the individual components can be found in the Varian [2011] user manual.

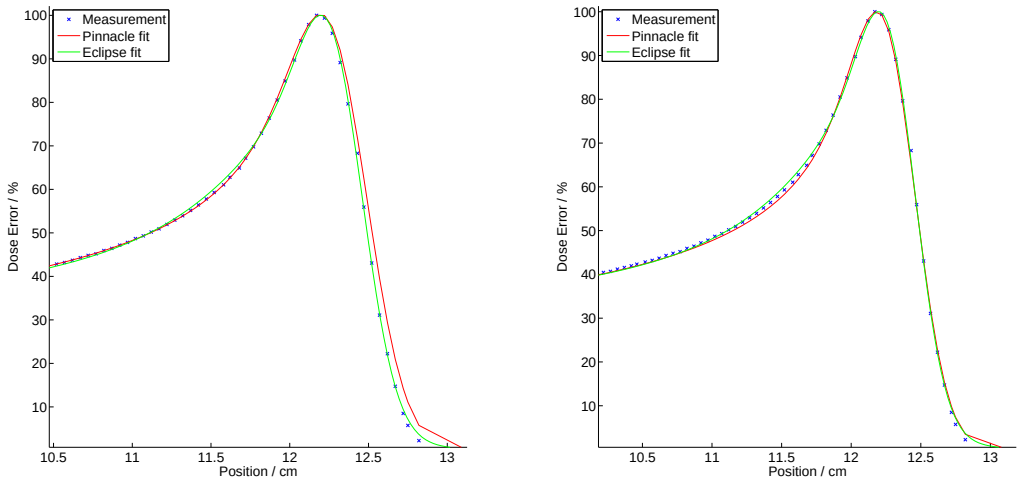


Figure 2.26: Pinnacle^3 fitting for 130 MeV integrated depth dose using (a) Feb 2012 version and (b) July 2012 version.

2.5.1.1.2 Bragg peak depth

Pinnacle^3 consistently overestimated the distal Bragg peak depth, with a mean/max distal R50 error of 0.3 mm/0.5 mm over the 27 energies. These errors are small but noticeable, an example of which can be seen in figure 2.13. The fitting is otherwise good. It can be seen in figure 2.15 that the fitting fails (gamma index > 1) for 5/27 energies: 110, 130, 175, 215 and 220 MeV.

A possible reason for the overestimation in Bragg peak depth could be due to an underestimation of the spread of initial energies. Extreme cases of this are shown in figure 2.3, where different forms of the Bortfeld equation do not take account of the initial spread or range straggling at all, and it is clear that this affects the Bragg peak depth. The level of range straggling is mathematically defined, but the initial energy spread can be input manually into the model. It seems plausible that this could result in an overestimation of the Bragg peak depth, but it does not explain why it only happens at particular energies. The results have been reported to Philips and are currently under investigation.

2.5.1.1.3 Future versions

In Pinnacle^3's July 2012 update, it was possible to place an increased weighting on the distal edge during the fitting process. This option was tested and an example of the fitting improvement of the distal part of the Bragg peak can be seen in figure 2.26. The compromise is that there is a slight underestimation of proximal dose, between 11.5-12.0 cm. The effect of this on the mean and max gamma with each energy is shown in figure 2.27. When comparing the two fittings (Feb fitting = black, July fitting = magenta) the mean gamma is almost unaffected, but the maxima are

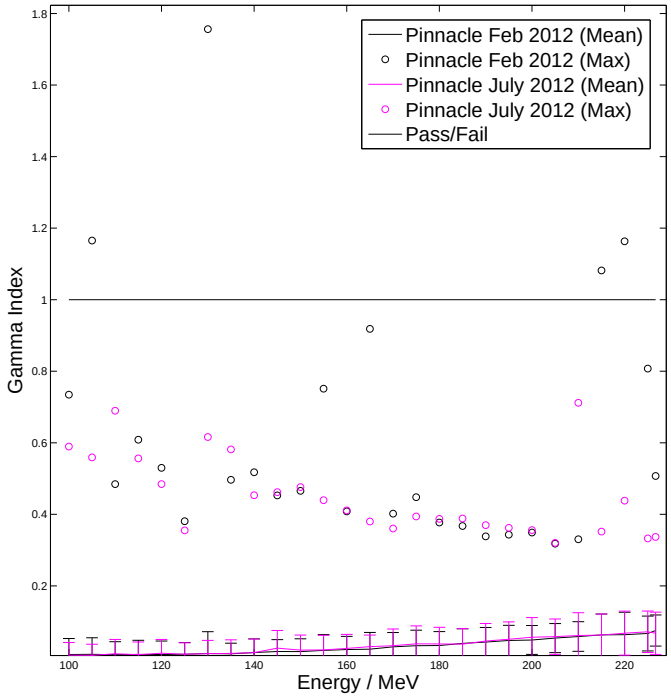


Figure 2.27: Variation of mean (line) and max (open circles) gamma index with energy, for February 2012 (black) and July 2012 (magenta) Pinnacle^3 versions.

significantly reduced for the previously affected energies (110, 130, 175, 215 and 220 MeV). By placing greater weight on the distal edge fitting the mean/max distal R50 errors were reduced from 0.3 mm/0.5 mm to 0.03 mm/0.1 mm.

As detailed in section 2.1.6.1.1 the Bortfeld model does not incorporate build-up. This did not impact on Pinnacle^3's fitting of our data because it did not go to depths low enough where build-up would be seen, but a more recent version of Pinnacle^3 (Feb 2013) includes a correction factor to model build-up, whilst still using the Bortfeld model. It would be of interest to see how this performs compared to the Eclipse beamlet model for data that extends to the shallowest 1-2 g/cm^2 depths.

As discussed in section 2.5.1.1, Eclipse's main failing was an underestimation of the entrance dose. This was attributed to the term I_{Lan_2}, which could not be controlled by the user. In version 11 (the version being tested in our work was 10.0.39), an additional correction curve can be used to control possible build-up artefacts. An example is shown in figure 2.28.

2.5.1.1.4 Impact

The impact of Eclipse incorrectly modelling the build-up is likely to have a small effect in proton therapy planning. As discussed in section 2.5.1.1.1, the nuclear build-up is estimated to be around 7% over the first 2 cm in water, which is a small

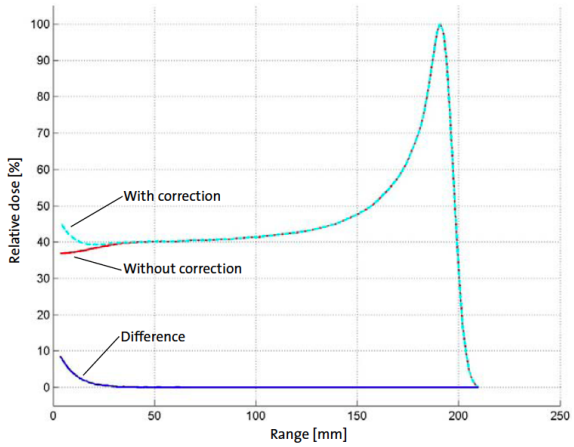


Figure 2.28: An additional curve is available in Eclipse version 11 to correct for build-up artefacts. The entrance dose can thus be correctly modelled by the user.

contribution to the total dose. Although small, this build-up is measurable and it is likely to be detected during patient quality assurance (QA). It is also possible that incorrect modelling could lead to unexpected skin side effects.

The marginally-incorrect modelling of the Bragg peak depth by Pinnacle^3 is likely to have a bigger impact on patient dose. An underestimation of the depth would lead to under dosing at the distal edge of the target, while an overestimation would lead to irradiation of normal tissues distal to the target.

2.5.1.2 In-air spot profiles

2.5.1.2.1 General

As detailed in section 2.4.1.2, the spot profile fitting of each system is good, but not perfect. For the y-profiles both systems have problems fitting to the tails, but the fitting for the x-profiles is good. This is suggestive that the spots do not demonstrate a two-dimensional Gaussian shape. Upon closer inspection, it appears the measurements demonstrate a deviation from the Gaussian shape at the highest energies ($E \geq 215\,\mathrm{MeV}$) in the y-direction, although the reason for this is unclear. At other energies however the profile shape remains Gaussian.

From figure 2.19 there also appears to be little to no dependence on the profile measurement position, which is surprising. Further from the nozzle it is expected the spot profile would be more Gaussian-like in shape (following more scattering) and thus the systems would find it easier to fit to.

Both systems modelled the FWHM well. The spot width (defined by the FWHM) is used in dose calculations and it is possible that the system places a greater emphasis on ensuring this is correct (possibly at the sacrifice of the tail modelling).

2.5.1.2.2 Input data

An issue affecting both systems is the radius of the chamber, which was smaller than the spot size for the highest energies and deepest depths, as detailed at the start of section 2.3.1.1. If this was the only problem, however, the small spot fraction not being collected (of the order of $\sim 1\%$), would translate into insignificant clinical consequences. A bigger problem is that the spots have since been found by UPenn to have long, non-Gaussian tails, caused by large angle scattering and/or nuclear interactions in the monitor chambers. These events affect all energies and all depths and could result in a few per cent of the signal being lost even for low-energy beams at shallow depths. In the versions analysed (Pinnacle^3, Feb 2012; Eclipse, 10.0.39) fitting of the in-air spot profiles was restricted to a two-dimensional Gaussian fit, so these non-Gaussian tails would not be modelled.

2.5.1.2.3 Impact

A good model of the tails is very important if one wants to deliver IMPT (or even widely-varying field sizes) accurately. At the time of the work the fixed beam line at UPenn was mostly used for treating prostate cancer patients, whose field sizes are fairly standard and whose spot intensities did not vary much. Provided this is the case, the models commissioned on the assumption that the spots do not have long tails is acceptable. However, upon the treatment of more complicated sites it was found that discrepancies between the calculated and measured doses appeared, the origin of which was attributed to the low-level tails not being modelled. UPenn are currently acquiring additional data to refine their Eclipse model for such treatments in the future.

2.5.2 Treatment planning comparison

2.5.2.1 Overall performance

The aim of the study was to compare the dose distributions produced by three proton TPSs, Eclipse, Pinnacle^3 and XiO, with a common set of planning guidelines, specifically for meningioma patients. The use of spot scanning protons for meningiomas has been shown to be beneficial (Weber et al. [2004b]), and the cases selected challenged each system as there were often many overlapping structures. Plan differences could be attributed to system differences, which would both inform new proton centers deciding which TPS to purchase and encourage further development of proton TPSs.

With a consistent planning strategy, all systems showed a good capacity to produce satisfactory plans that sufficiently respected the constraints for both the target

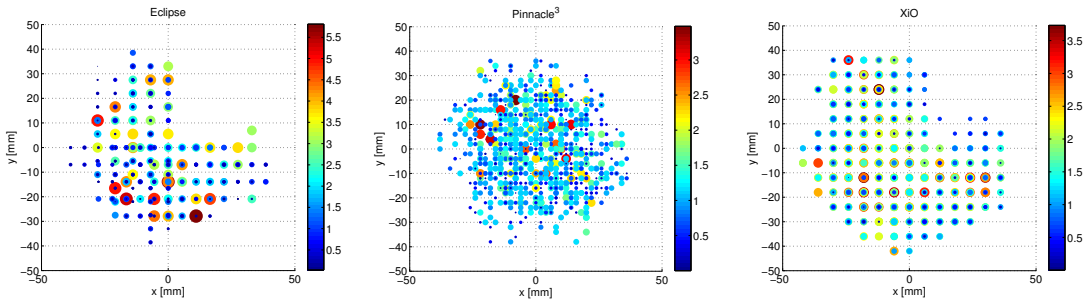


Figure 2.29: Beam's eye views of the spot distributions, with the spots in all layers overlaid for (a) Eclipse, (b) Pinnacle^3 and (c) XiO, for a particular patient and field. Each spot map has been normalised to its own mean, with spot colour and size proportional to the spot weight. The colour scale range differs between figures.

and OARs, despite difficult and conflicting objectives, with large overlapping regions. The operation of these different systems and the options available to the planner do differ, but the final results were similar.

Statistically significant differences were found for the high target doses, D2 (p=1.7x10^{-4}) and D5 (p=5.6x10^{-7}), the maximum dose for one of the lenses (p=0.024) and the mean brain dose (p=0.022). Although not significant, there was a general tendency for Pinnacle^3 to deliver lower OAR doses (as can be seen in figure 2.21). Mean integral doses outside the PBSTV, across all patients, were found to be 4.4 ± 1.5 Gy in Pinnacle^3 (mean $\pm$ standard deviation), compared to 6.0 ± 2.2 Gy in Eclipse and 6.3 ± 2.3 Gy in XiO. This can also be seen in figure 2.22, with Pinnacle^3's dose distribution showing marginally better conformality to the tumour than those of Eclipse and XiO. A possible reason for this is the flexibility of available spot positions, which are more staggered than the fixed 3D grids available in Eclipse and XiO, as explained in section 2.3.6. An illustration of the effect of this is shown in figure 2.29. Other possible reasons, such as different energy layer spacings and other system differences, are detailed below.

2.5.2.2 Energy layers

As stated in section 2.3.6, the layer spacing could not be defined in a consistent manner for all systems. Attempts were made to make the resultant layer spacing of each system consistent, however this proved difficult. It was found (by analysing mean $\pm$ standard error across all patients) that Pinnacle^3 (21 ± 2) used fewer energy layers than XiO (31 ± 2). A variable spacing of 4 times the range sigma was used by Eclipse, but this was perhaps not high enough as the average number of layers (17 ± 1) was lower than both Pinnacle^3 and XiO. This was a potential source of disadvantage to Eclipse. In the plans produced, the number of spots per layer was

similar: Eclipse (68 ± 9), Pinnacle^3 (86 ± 12) and XiO (79 ± 11) (mean $\pm$ standard error across all patients).

2.5.2.3 System differences

Eclipse and Pinnacle^3 allow multiple fields to be simultaneously optimised as separate SFUD fields, whereas XiO requires separate optimisation of each field, with half the prescription dose and half the tolerance doses to the OARs, followed by summing of the different field doses at the end. This effectively doubles the optimisation computational time.

Pinnacle^3 is the only system that has an option to allow the relative weighting of the two fields to be adjusted by the optimiser. As sometimes one field can better spare an OAR while delivering more dose to the target than another, this option could be useful and is a potential explanation for the generally lower OAR doses discussed in section 2.5.2.1.

It is known that, for this version of Eclipse, the optimiser has a tendency to form one or two highly weighted spots. Ordinarily, the planner would assess the effect of removing these, but no post-processing of spot weights was completed in this study in order to specifically test the algorithms. This was less common for Pinnacle^3 and XiO.

2.5.2.4 Spot uniformity and distribution

An assessment was made of the uniformity and distribution of spot weights between TPSs. A parameter, C, was defined to quantify this difference, which involves analysing the spot weight as a function of the distance from the isocentre (equation 2.3). In figure 2.23(b) it can be seen that Pinnacle^3 has a more uniform distribution of spot weights than Eclipse and XiO. This is backed up by the values of C, in which Pinnacle^3 has a lower value (38.3 ± 2.0 mm) than for both Eclipse (41.2 ± 2.5 mm) and XiO (44.9 ± 2.6 mm). Although not verified in this work, it is hypothesised that such a metric could be used as a heuristic measure of field robustness. For a field to be robust, the highly weighted spots should generally be located further from the edge of the target (and thus closer to the isocentre), so that changes in patient position and range uncertainty within a given field are then less critical to the target coverage and dose to the surrounding OARs. This would lead to a lower value of C. Verification of this hypothesis, however, and the determination of a threshold value for C, were beyond the scope of our work.

It should be added that this measure of robustness is only really applicable to SFUD plans and is in contrast with the desire for an optimal plan. The larger number of degrees of freedom in IMPT plans allow for generally better coverage

and OAR sparing (Yeung et al. [2014]), but it has also been shown that IMPT techniques such as distal edge tracking (in which the intention is to deliver highly weighted spots to the distal edge of the target) are less robust to uncertainties (Liu et al. [2012a]; Lomax [2008]). A variety of methods to handle such uncertainties have been suggested (Chen et al. [2012]; Fredriksson et al. [2011]; Meyer et al. [2010]; Pflugfelder et al. [2008]; Unkelbach et al. [2009]; Zhang [2012]), including the reduction of the intensity of spots close to tissue heterogeneities (Pflugfelder et al. [2007]), which is along similar lines to our hypothesis.

2.5.2.5 Study limitations and future work

As stated previously, comparing different TPSs is a very difficult task due to the many variables. Any study trying to perform such a task will have limitations and care should be taken to rank systems based on these preliminary results.

In an ideal study, the dose calculation would be performed in a single or independent engine, to allow testing of the optimiser alone. This is difficult to achieve in practice, however, as dose calculation is a necessary part of the optimisation process and the two cannot be easily disentangled. As mentioned in the methods, the dose calculation algorithm differs between systems and it is inevitable this will impact on the plans. For instance, it may be possible to attribute the higher OAR doses in Eclipse and XiO to deficiencies in their dose calculation algorithms leading to an overestimation of the lateral penumbra, as has been reported to be the case for uniform scanning proton therapy (Rana et al. [2013]). This would lead to OARs lateral to the beam direction receiving a higher dose in the plan. Without verification of the dose calculation algorithm, it is possible Pinnacle^3 may suffer from the opposite problem (i.e. the dose calculation algorithm may underestimate the lateral penumbra, leading to reduced OAR doses). The aim and scope of our work was to compare the dose distributions produced by three proton TPSs using equivalent beam data and (as best as possible) matched configuration settings. A thorough, detailed assessment of the dose calculation algorithms of each system is thus necessary to validate our findings.

As stated in the methods, time is an important factor in the optimisation process, but this could not be assessed due to the hardware differences between systems. Also, the layer spacing could not be controlled in a consistent way for each system, and the resulting number of layers available to each optimiser differed.

To calculate the distance for the quantitative metric C (equation 2.3) it was necessary to convert the spot energy to a water-equivalent range, which does not necessarily correspond to the physical co-ordinate of the pristine peak in the patient relative to the isocentre. It should also be noted that each system has different options available during optimisation, stated in the methods, which may improve

the result of C. Eclipse offers the use of the CG optimisation algorithm, Pinnacle^3 has a 'robustness' option and XiO has a 'smoothing' function, but none of these were tested. It was not possible in this work to test the validity of this metric C as a measure of robustness, but it could be incorporated into future work by introducing shifts in the patient position.

The plans were made robust to range uncertainties using a uniform 5 mm expansion from the CTV, however a full assessment of robustness requires shifts in the patient position, and systematic and statistical variations in the patient density and chemical composition. Such assessments could not be carried out within all TPSs tested, and it is a procedure that we would like to carefully control in an independent scheme. This is an area of future work.

The study only looked at meningioma brain treatments using fixed horizontal SFUD fields. It is anticipated there would be bigger differences in the performance of each system when producing full-gantry IMPT plans, and this is also suggested as an area of future work. How systems cope with different treatment sites will be of interest because of the different OARs and heterogeneity issues that must be considered. Also, as with any TPS comparison study, comparison results age quickly because of the continual evolution of each system.

2.5.2.6 Impact on the proton physics community

In this dose comparison study it has been shown that there are clear differences between the three TPSs and a new proton centre should take these into consideration during the procurement process. For instance, each system has different options to control the layer spacing and it may be the preference of the centre that this can be varied during planning, particularly if the caseload is widely varying. This is not possible in the version of Eclipse tested. Another clear difference was the staggering of spot positions layer-by-layer. In Eclipse and XiO each layer exactly overlaid the layer below, while in Pinnacle^3 the layers could be offset. Although the lower OAR doses in Pinnacle^3 were attributed to this offsetting feature, the implications of this on performing patient QA and (potentially) on delivery times should also be considered by new proton centres. It is hoped that highlighting these differences will generate interest and further work in proton TPS production. It is possible that these system differences could impact on multi-institutional studies if different TPSs are used. Care should therefore be taken to ensure that any TPS differences are documented at the outset of the study, as these may account for different results (as was seen in this work) between centres.

2.6 Conclusions

2.6.1 Proton beam commissioning

The two systems involved in the commissioning (Eclipse and Pinnacle^3) demonstrated a good capability to produce beam models within clinically-accepted tolerances. However, the differences in algorithms lead to minor fitting differences. Perhaps the most important differences are Eclipse's overestimation of the build-up at high energies and Pinnacle^3's consistent overestimation of the Bragg peak position (mean/max 0.3/0.5 mm), for the depth dose curves. The spot profiles were generally modelled well, although additional measurements must be acquired so that the long, non-Gaussian tails can be modelled. Both systems modelled the FWHM well.

2.6.2 Treatment planning comparison

The study compared the plans produced by three proton TPSs, Eclipse, Pinnacle^3 and XiO, for the treatment of meningiomas with an SFUD horizontal fixed beam arrangement. Few statistically significant differences were found, but Pinnacle^3 generally gave lower OAR doses, with an integral dose outside the target 27% lower than Eclipse and 30% lower than XiO on average, across all patients. Possible reasons for this are the flexibility of available spot positions and the option that the optimiser can adjust the relative weighting of the two fields, however the dose calculation algorithms of each system must be assessed in future works to validate our findings. Pinnacle^3 was found to distribute its spots more uniformly than Eclipse and XiO, although the effect of this on any possible robustness has not been explored. In highlighting the differences between the systems we believe the study will prove to be useful both to new proton centers and to the improvement of the TPSs themselves.

Chapter 3

Uncertainties in computing the relative stopping power

Work in this chapter can be found in the following journal article and was presented at the following meetings:

- (Under review) **Doolan P**, Collins-Fekete C-A, Dias M, Royle G, Seco J **2014** Range effects in proton therapy caused by systematic errors in the stoichiometric calibration procedure *Physics in Medicine and Biology*
- **Doolan P**, Dias M, Collins-Fekete C-A, Seco J **2014** Range effects in proton therapy caused by systematic errors in the stoichiometric calibration *American Association of Physics in Medicine 56th Meeting*, Austin, TX, USA (Poster)
- Collins-Fekete C, Dias M, **Doolan P**, Hansen D, Beaulieu L, Seco J **2014** Combining proton radiography and X-ray CT information to better estimate relative proton stopping power in a clinical environment *American Association of Physics in Medicine 56th Meeting*, Austin, TX, USA (Poster)
- Dias M, Collins-Fekete C, Riboldi M, **Doolan P**, Hansen D, Baroni G, Seco J **2014** Ion imaging to better estimate in-vivo relative stopping powers using X-ray CT prior-knowledge information *American Association of Physics in Medicine 56th Meeting*, Austin, TX, USA (Poster)

C-A Collins-Fekete, M Dias and J Seco developed and validated the code that could compute the stopping power using the Bichsel, Janni and ICRU formulae. J Seco conducted the Bragg peak shift measurements to measure the RSP of the Gammex inserts. Computation of the stopping power using the Schneider method and all investigations into the impact on the range were computed by myself.

3.1 Background

As stated in the introduction, the principal benefit of proton therapy derives from one of the physical characteristics of the depth dose curve: the dose peak at a well-defined depth in tissue. The position of this 'Bragg peak' and the range that protons reach in tissue is controlled by their initial energy and by the integrated stopping power of the tissues along the beam path. Provided the initial proton energy is known, to predict the range of protons in tissue it is necessary to measure the tissue stopping power. An ideal procedure to do this would consist of a direct measurement using a proton beam. Currently this can only be conducted in 2D, using proton radiographs, as proton CT is not currently available. However, this is not useful for treatment planning as 3D datasets are required. A more detailed review of proton radiography can be found in section 4.1.

As such, the current approach is to acquire a conventional X-ray CT image, which is a dataset of photon attenuation. To perform any proton dose calculations this dataset (in HU) has to be converted into a dataset of proton energy loss, which is described by the RSP. The single calibration curve used to do this will henceforth be referred to as the 'HU-RSP calibration curve'. The quality of the calibration curve directly affects the accuracy of determining the RSP values in the patient's body, thus affecting the accuracy of dose calculations and the accuracy of the proton range prediction. This area of research is therefore of great importance.

3.1.1 Producing the calibration curve

The ambiguity in transferring from CT numbers to RSPs arises because the CT number is related to the X-ray absorption coefficients, while the RSP is dependent primarily on the relative electron density (Shinoda et al. [2006]). This conversion results in a uncertainty of where the protons will stop in the patient, known as 'range uncertainty'. As discussed in the introduction, in clinical practice a 3.5% margin is added to account for the proton beam range uncertainty, with this HU-RSP conversion process contributing $\approx 0.5\%$ to the uncertainty (Paganetti [2012]), or up to 1.8% if the uncertainty in the mean ionisation energy of the tissue (henceforth referred to as 'I-value') is also considered (Schaffner and Pedroni [1998]).

There are a number of different methods to produce the HU-RSP calibration curve, with different levels of complexity and effort required by the user. An overview of these techniques is described here.

3.1.1.1 Stoichiometric calibration

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 (Yang et al. [2012]). The process consists of four main steps:

1. *Image tissue-substitute materials* with known chemical composition.

2. *Parameterise the response of the CT as a function of chemical composition*, using the equation,

$$\mathrm{HU} = \rho_e^{\mathrm{rel}} \left(AZ^{3.62} + BZ^{1.86} + C \right) \tag{3.1}$$

   where ρ_e^{rel} is the volumetric electron density relative to water; Z is the effective atomic number (calculated using the fraction by weight of the individual elements for compounds); and A, B and C are constants that characterise the contributions of photo-electric, Rayleigh scattering and Compton scattering to the total attenuation, respectively.

   These constants, that parameterise the response of the CT system, depend on the X-ray energy spectrum and are determined by fitting equation 3.1 to a plot of measured HU values of the tissue-substitute samples (Schaffner and Pedroni [1998]). Example values, from Schneider et al. [1996], for these constants are $A = 1.227 \times 10^{-5}$, $B = 4.285 \times 10^{-4}$ and $C = 0.5$. These approximately agree with the relative contributions of photo-electric, coherent scattering and Compton scattering to the total attenuation shown in figure 3.1.

3. *Calculate the CT numbers of human biological tissues.* In an ideal scenario these would be measured directly in the CT scanner, but this is rarely practical. Even if real tissues could be readily handled, it would still require the separate scanning of individual tissues (to avoid problems such as beam hardening). Therefore, the CT numbers for real biological tissues are theoretically predicted using equation 3.1, together with the fitted constants. Chemical compositions and effective densities are taken from literature such as Woodard and White [1986], White et al. [1987] and ICRU [1989].

4. *Calculate the RSPs of human biological tissues.* As with step 3, ideally these would be measured directly with a proton beam, but again this is not practical. Therefore, these values have to be theoretically calculated using the same chemical compositions from literature (Woodard and White [1986], White et al. [1987] and ICRU [1989]). The absolute stopping power depends on

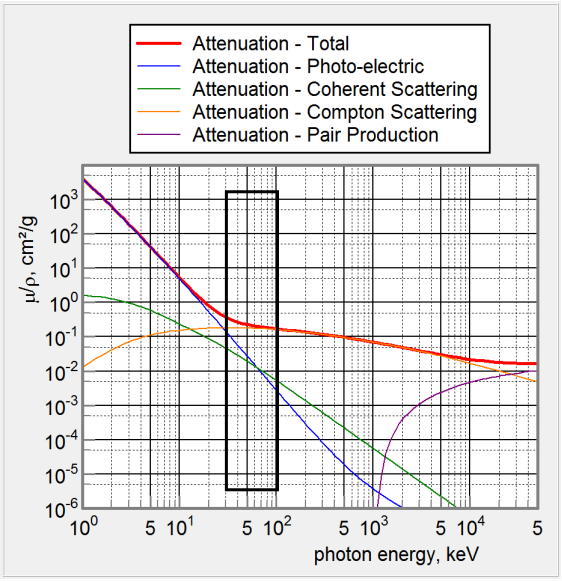


Figure 3.1: Contributions to the mass attenuation coefficient of water, produced using Nowotny [1998] with data from Boone and Chavez [1997]. The diagnostic energy range is highlighted by the black box.

the energy of the particle, but it has been shown that the RSP is almost independent of β (particle velocity as a fraction of the velocity of light) for the range of particle energies relevant to radiation therapy, as shown in figure 3.2.

As such, the RSP is a much more useful quantity for proton therapy treatment planning, where the beam energy varies with depth in the patient. A few different formulae currently exist in literature for this calculation.

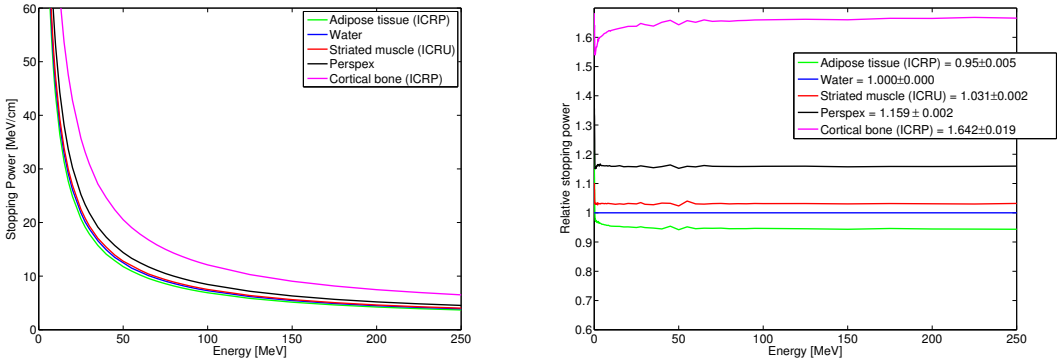
(a) The first proposal was given by Bichsel [1972], with the absolute stopping power S_B given by:

$$S_B = \rho \frac{4\pi e^4}{m_e c^2 u \beta^2} z^2 \frac{Z}{A} \left\{ \ln \left[\frac{2 m_e c^2 \beta^2}{(1-\beta^2)} \right] - \beta^2 - \ln I^t - \frac{C}{Z} - \frac{\delta}{2} \right\} \tag{3.2}$$

where ρ is the mass density; e is the electron charge; $m_e c^2$ is the rest mass energy of the electron; u is the atomic mass unit; z is the charge of the projectile (+1 for protons); Z and A are the atomic number and relative atomic mass of the target atom; I^t is the I-value of the tissue; C/Z is the shell correction; and $\delta/2$ is the density correction. To allow for comparison with other equations, we rewrite the formula of Bichsel (equation 3.2) can be written as,

$$S_B = K \times B \tag{3.3}$$

where K is the term outside the brackets and B represents the terms inside the brackets (the letter B used to represent Bichsel).



(a) (b)

Figure 3.2: Variation of (a) absolute stopping power and (b) relative stopping for a range of tissue substitutes. In (b) the mean and error (calculated as the standard deviation) are calculated over the range 5 MeV to 250 MeV. Graphs produced with data from Berger and Zucker [2004].

This formulae, as well as the Janni and ICRU formulae in described in sections (b) and (c) to follow, all compute the absolute stopping power. The RSP can then be calculated by dividing by the stopping power of water over the same energy range.

(b) Using these definitions, it is possible to write the stopping power as defined by Janni [1982] S_J as,

$$S_J = K\left\{B - \frac{1}{2}\ln\left[1 + \frac{2m_e}{M\sqrt{(1-\beta^2)}} + \left(\frac{m}{M}\right)^2\right] + \frac{\pi\alpha z\beta}{2} + \frac{zZ\alpha^3 F(\beta, Z)}{\beta^3}\right\} \tag{3.4}$$

where M is the proton rest mass and the second term inside the brackets forms part of the factor that accounts for the maximum kinetic energy that can be transferred to an unbound electron at rest; α is the fine structure constant, equal to 1/137.036, and the factor it is part of is important relativistically; $F(\beta, Z)$ is a function to account for the use of the second Born approximation and is important only at low energies (it is usually set to zero); and all other parameters are as defined previously. To allow comparison with the ICRU formula (see below), the formula of Janni (equation 3.4) can be rewritten as,

$$S_J = K\left\{B + J_1 + J_2 + J_3\right\} \tag{3.5}$$

where $J_{1,2,3}$ are the second, third and fourth terms inside the brackets of equation 3.4 (the letter J used to represent Janni).

(c) Similarly, the stopping power as defined by ICRU [1993] Report 49 can

be written as,

$$S_I = K\left\{B + J_1 + \gamma J_3 - \left(\frac{z\alpha}{\beta}\right)^2 \sum_{n=1}^{\infty}\left[n\left(n^2 + \left(\frac{z\alpha}{\beta}\right)^2\right)\right]^{-1}\right\} \quad (3.6)$$

where γ comes from the use of the free-electron model and is approximately equal to $\sqrt{2}$; and all other parameters are as defined previously. In the ICRU Report, there are I-values for elements (defined here as 'condensed') and for atomic constituents of compounds in the liquid and solid phase (defined here as 'compound'). Both can be used to compute the stopping power S_I.

(d) Alternatively, and more conveniently, the RSP can be approximated directly. As far as we are aware, this was first proposed by Schneider et al. [1996],

$$RSP = \rho_e^{\text{rel}} \frac{\{\ln\left[2m_e c^2\beta^2/I^t(1-\beta^2)\right] - \beta^2\}}{\{\ln\left[2m_e c^2\beta^2/I^w(1-\beta^2)\right] - \beta^2\}} \quad (3.7)$$

where ρ_e^{rel} is the volumetric electron density relative to water; I^t and I^w are the I-values of tissue and water respectively; and all other symbols are as defined previously.

To account for the different formulae described in (a), (b) and (c), and because of their date of publication, each source uses different elemental I-values. These values, together with the computed values for a range of Gammex tissue substitutes and, importantly, water (I^w), are detailed in table 3.1.

One of the disadvantages of the stoichiometric method is the complexity of the process and there is quite a lot of manual work required on the part of the user. As such, many centres still use a tissue substitute calibration, described in section 3.1.1.2. Alternative methods have therefore been developed that try to incorporate the accuracy of the stoichiometric method, but are easier to use, such as the poly-binary model and dual energy imaging. These have not been readily taken up in clinical practice and so are only briefly described in section 3.1.1.3.

3.1.1.2 Tissue-substitute calibration

The oldest, and most convenient, method of calibration is to use tissue-substitutes. This process consists of three main steps:

1. Image tissue-substitute materials with known relative electron density ρ_e^{rel} and density ρ. The composition is rarely known because it is considered confidential information by the vendor, but estimates have been made by Watanabe [1999].

Element (Z) or Material	I-value / eV			
	Bichsel [1972]	Janni [1982]	ICRU [1993]	
			Condensed	Compound
H (1)	19.2	20.48	19.2	19.2
C (6)	78	73.8	81	81
N (7)	80	97.8	82	82
O (8)	95	115.7	106	106
F (9)	119	124.8	112	112
Na (11)	148	143	149	168
Mg (12)	156	151.1	156	176
Si (14)	174.5	174.5	173	195
P (15)	172	179.1	173	195
S (16)	180	183.6	180	203
Cl (17)	187	182.6	180	180
K (19)	193	186.8	190	214.7
Ca (20)	196	191.9	191	216
Fe (26)	280	278.2	286	323
I (53)	510	515.2	491	491
LN-300 Lung	70.3	71.2	73.1	74.1
LN-450 Lung	70.3	71.2	73.1	74.1
Adipose	65.7	66.4	68.4	68.4
CT Solid Water	67.1	68.1	70.1	70.2
Inner bone	72.4	72.4	74.5	74.5
CB2 30% $CaCO_3$	76.3	78.5	79.6	80.7
CB2 50% $CaCO_3$	87.1	90.9	91.0	93.2
Cortical Bone	98.2	103.8	102.7	106.1
Water	69.0	81.8	75.3	75.3

Table 3.1: I-values of elements, tissue substitutes and water, calculated using the different sources.

2. Calculate the RSP, for each tissue substitute using any of the formulae described in section 3.1.1.1 - equations 3.2, 3.4, 3.6 or 3.7.

3. Plot the points (HU versus RSP) for the tissue-substitute materials and interpolate between the points to get a calibration curve.

Many authors have adopted this approach, with different tissue-substitute materials. Chen et al. [1979] used solutions of: water and potassium phosphate for bone; ethanol and water mixtures for adipose tissue; and cork for lung. Constantinou et al. [1992] used a commercial phantom with various solid water materials of known atomic composition to deduce a relationship relating the CT number to electron density for radiotherapy. Jakel et al. [2001] used solid tissue substitutes to determine the relationship between CT number and RSP for heavy ion radiotherapy.

Schneider et al. [1996] and Matsufuji et al. [1998] estimated the inaccuracy in using these tissue-substitutes to be a few percent in effective density, with a strong dependence on the choice of materials used. The materials are selected as substitutes for given tissues based on their electron density. However, photon attenuation is not only caused by Compton scattering, which is proportional to electron density, but also by photoelectric effect and coherent scattering, which depend on the effective atomic number.

To address this problem, efforts have recently been made to manufacture tissue-equivalent materials with greater consideration of their radiological properties and less consideration of the chemical composition (Yohannes et al. [2012]). It was found these materials allowed for the production of calibration curves that could match the stopping powers of ICRU real tissues with a mean difference of 0.8%. The Radiological Physics Center (RPC) in the United States has also approached the problem for proton therapy, by measuring the RSP of 18 common tissue substitute materials using a Bragg peak shift measurement (Grant et al. [2014]). It was found that only 9 had a difference of less than 5% from the RSP assigned from their stoichiometric calibration curve, and thus only these 9 are similar to human tissues. These 9 tissues have been since used in phantoms designed to assess the HU-RSP calibration curve at proton therapy facilities across the United States.

3.1.1.3 Other methods

The polybinary model, developed by Kanematsu et al. [2003], approximates all body tissues by their mixtures with respect to muscle, because muscle occupies the largest content of the body (ICRP [1975]) and is in the middle of the HU curve. Lung tissues, soft tissues and bones are approximated to mixtures of muscle with air, fat and bone mineral (40% K_2HPO_4) respectively. The authors suggest it offers

the accuracy of the stoichiometric method (accuracy found to be 1.6% by Shinoda et al. [2006]) and the convenience of the tissue-substitute method.

The accuracy of dual-energy CT (DECT) imaging in determining the RSP of materials was tested by Yang et al. [2011]. Using a kV-MV pair, it was shown that the calculation of the RSP was less sensitive to tissue composition variation than the tissue-substitute method. It is, however, sensitive to CT number variations due to image noise and other uncertainties in the imaging and image reconstruction processes. Although still in its infancy, recent works investigating the use of kV-kV DECT to improve RSP prediction have showed promising results (Bourque et al. [2014]; Hünemohr et al. [2014]). This may not be the optimal solution however, as unpublished work by Hansen et al. [2014] shows that proton CT offers an even better accuracy.

3.1.2 Errors in the calibration curve

There are many potential errors when producing calibration curves, which can be conveniently divided into uncertainties in the imaging and uncertainties in calculating the RSP. Summaries of these uncertainties from literature, often with quantitative values, are listed in the following sections.

3.1.2.1 Uncertainty in the measured CT number

The distinctive Bragg peak shape means that proton beam dose calculations are very sensitive to small changes in the CT number. The uncertainty in the number from a CT scanner is a well-researched field, with many sources of error summarised below.

1. Variations in CT numbers, for the same materials, has been shown in different diagnostic scanners (Faulkner and Moores [1985]; Levi et al. [1982]). X-ray attenuation depends not only on the electron density but also on the atomic composition of the material (Kanematsu et al. [2003]; Schaffner and Pedroni [1998]). As a result, different energy spectra (dependent on the tube voltage and filter) from different scanners will produce varying HU values. Qi et al. [2006] found that by changing the scanning voltage from 120 kVp to 140 kVp, the HU of cortical bone changed by 12.7% (differences were lower for other materials). Provided the scanning voltage is kept consistent, however, the X-ray spectrum has been shown to have a smaller effect on the HU than the chemical compositions of the tissue substitute materials. Cheng et al. [2012] demonstrated this by separately measuring the HU of the same RMI 467 (Gammex Inc., Middleton, WI) phantom on different scanners, with variations $< 5\%$ for lung substitutes, $< 2\%$ for soft tissues and $< 1\%$ for boney tissues.

2. There are known to be CT scanner variations with time. Yang et al. [2012] measured the same cylindrical inserts during monthly QA measurements and found variations in the CT number of 1.0% for lung, 0.3% for soft tissue and 0.6% for bone.

3. The CT number of an object is shown to vary with its position in the scan and the patient size, due to the beam hardening effect (Schaffner and Pedroni [1998]).

   Yang et al. [2012] estimated the effect of the patient size by scanning the same insert at isocentre inside head and body phantoms (diameter 16 cm and 32 cm respectively). They found relatively large variations in the CT number of 2.6%/0.3%/1.9% for lung, soft tissue and bone, respectively. A potential solution to try to compensate for this, used at PSI, is to have different HU-RSP calibration curves depending on the size of the object being imaged. Dose distributions are then produced for three different calibration curves (standard, maximal range error and minimal range error) to get an estimate of the effect of uncertainty in the proton range (Schaffner and Pedroni [1998]).

   Yang et al. [2012] also estimated the effect of the position within the scan by scanning the same insert at the centre and periphery of a body phantom (32 cm diameter). They found small variations in the CT number of 1.3%/0.1%/1.0% for lung/soft tissue/bone.

4. The support table top also has an effect on the HU. CT scanners are typically calibrated by the manufacturer using phantoms suspended on a support (essentially in air), but scans are conducted with patients on couches or, for radiation treatment planning purposes, flat table tops (the latter may enhance the effect). Qi et al. [2006] showed that this can result in HU changes of 26% for water (although percentage errors are misleading for water as the HU value is so close to zero). Yang et al. [2012] scanned the RMI 467 phantom with and without the couch in the field of view, and found small variations in the CT number of 1.1%/0.3%/0.5% for lung, soft tissue and bone, respectively.

5. Noise is present in most CT images, and Chvetsov and Paige [2010] investigated its effect on the range of proton beams. They reported a standard deviation of the range between 0.3% and 0.7%, assuming a CT grid of 3 mm and a 2.5% noise level.

6. CT artefacts, such as streaking, can be caused by the presence of metallic implants (Verburg and Seco [2012]).

7. The detector sensitivity and (possibly) reconstruction algorithm can affect the measured HU value (Kanematsu et al. [2003]).

3.1.2.2 Uncertainty in calculating the RSP

Potential sources of uncertainty in the calculation of the RSP are listed below.

1. There is uncertainty in the real tissue composition, which is used to find the theoretical CT values for the stoichiometric calibration method. The ICRU tables (ICRU [1989], ICRU [1992]), Woodard and White [1986, 1982], White et al. [1987] and White and Widdowson [1991] are often taken as accepted compositions, however they contain uncertainties. In Report 44, ICRU [1989] states explicitly that 'the elemental compositions of most body tissues are known to vary considerably between individuals of the same age' and 'the composition of a given tissue within one individual may vary from one body site to another'.

   In their thorough analysis of the uncertainties affecting the RSP of tissues, Yang et al. [2012] concluded that, for soft tissues, the dominant factor was the uncertainties induced by tissue composition variation. 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 (Yang et al. [2010]). This is an important result for our work as we intend to produce patient-specific HU-RSP calibration curves that remove the reliance on the average tissue compositions. The production of patient-specific calibration curves at the time of treatment planning imaging, first suggested by Schneider et al. [2005], is the subject of our work in chapter 5.

2. In calculating the RSP, the I-values of the element/tissue and water are required (equation 3.7), which are both difficult to obtain experimentally and are known to vary between sources.

   As detailed in section 3.1.1.1 and table 3.1, elemental I-values differ between sources such as Bichsel [1972]; ICRU [1993]; Janni [1982]. Andreo [2009] suggests that uncertainties in the I-value for tissues, estimated using Bragg's additivity rule (ICRU [1992]; Seltzer and Berger [1982]), are potentially 10-15%, which translates to uncertainties in the RSP of 1.4-1.5% (Paganetti [2012]; Yang et al. [2012]).

   As explained previously, the stopping power is often quoted relative to water, so the I-value of water is therefore of critical importance. A range of values have been suggested for water: 75 ± 3 eV (ICRU [1993]), 77.8 eV (Emfietzoglou et al. [2009]), 78.4 ± 1.0 eV (Kumazaki et al. [2007]), 80 ± 2 eV (Bichsel and Hiraoka [1992]), 80.8 eV (Paul et al. [2007]) and 81.8 eV (Janni [1982]); which can have a large impact on proton treatments. Andreo [2009] demonstrates

that variation of I-values between 67.2 and 80 eV can have a substantial impact on the proton beam range in water; while Paganetti [2012] shows that a variation between 75 and 80 eV will result in a 0.8-1.2% difference in the predicted stopping power. Previous works, such as that by Schaffner and Pedroni [1998], have taken the average of literature data.

As can be seen, there are many sources of uncertainty in the calculation of the RSP. Yang et al. [2012] combined uncertainty in the real tissue composition and uncertainty in the I-value to give combined uncertainties (1σ) of 5.0/1.6/2.4% for lung/soft tissue/bone respectively (these uncertainties also include the uncertainty from CT imaging). These uncertainties could be significantly reduced if the RSP is measured directly. A potential method for this is proton radiography, which is explored in chapters 4 and 5.

3.1.2.3 Other uncertainties

1. All of the methods detailed previously require the imaging of tissue-substitute materials. There is a large difference between the energies used in the CT image calibration, which is acquired at low kV, and the treatment, which uses protons at MeV, and these materials respond differently in these different energy ranges. It is difficult to produce tissue-equivalent materials that respond like real tissues over such a wide energy range.

2. A single calibration curve for the whole body is not really applicable, because there is not a perfect one-to-one correspondence between CT numbers and RSPs of human tissues (Yang et al. [2012]). The CT number and RSP describe two different physical properties and, even though both values are dominated by the relative electron density, the elemental composition also matters. As such, different organs can have the same HU value but different RSPs. A possible solution to this is to produce different calibration curves for different organs in the body. Schneider and Pedroni [1995] demonstrated that this offered improvements over the stoichiometric method, however it requires segmentation of each organ on the planning CT by the clinician, which takes time.

3.2 Aim

As detailed in section 3.1.2, the HU-RSP calibration curve is subject to many uncertainties. These lead to an error in the calibration curve of between 0.5-1.8% (dependent on tissue type), which yields itself as a range uncertainty. Currently,

clinical centres account for this uncertainty with increased planning margins, irradiating more tissue than necessary.

In this chapter an investigation is made into the error introduced by making theoretical estimations of RSP, compared to the ideal scenario of measuring directly with a proton beam. Most previous works have looked at the errors introduced by the complete stoichiometric calibration procedure, but this work looks specifically at the error introduced by step four of the stoichiometric procedure: the theoretical calculation of the RSP.

3.2.1 Objectives

This work is separated into three components:

1. We look at the error introduced if the RSP is calculated using the separate formulae of Bichsel, Janni and ICRU, but with their individual I-values. We make comparisons with the Schneider approximation of the RSP, equation 3.7, but with the I-values of the different sources. We compute these errors for tissue substitutes, by measuring the RSP directly in a proton beam.

2. Such systematic errors translate into proton range errors within the patient, which could be critical to the target coverage and thus patient outcome. The impact of these errors on the proton range is then investigated within the Gammex phantom and on two patient sites, by implementing different calibration curves within the proton treatment planning software, XiO. We look only at the most-often used approach, the Schneider approximation. In the patient cases the I-value of water is calculated: (i) using the elemental I-values of the different sources; or (ii) set to the most recently recommended value, in ICRU [2009] Report 73, of 78 eV.

3. Based on the previous set of results, it was possible to define a gold standard that best matched our measurements. We compare this gold standard to the conventional Schneider approximation, both with an I-value of 75.3 eV (calculated using the elemental I-values of ICRU [1993] Report 49) and 78 eV (the most recently recommended value), for a range of human tissues. In this way we demonstrate the impact that systematic errors have in step four of the stoichiometric calibration procedure, the theoretical calculation of the RSP of human tissues.

Substitute material	ρ_e	ρ [gcm^{-3}]	Measured CT number [Hounsfield units]
LN-300 Lung	0.276	0.280	-673
LN-450 Lung	0.462	0.470	-489
Adipose	0.926	0.943	-73
CT Solid Water	0.989	1.018	11
CB2 30% $CaCO_3$	1.277	1.332	404
CB2 50% $CaCO_3$	1.470	1.560	761
Cortical Bone	1.695	1.823	1036

Table 3.2: Gammex RMI 467 phantom tissue substitute materials.

3.3 Materials and methods

3.3.1 Computation of the RSP

A selection of the tissue substitute materials of the Gammex RMI 467 phantom (Gammex Inc., Middleton, WI) were imaged using the standard arrangement suggested by the manual. B200 bone mineral was excluded because it has a significant proportion of fluorine (16.7%) and none of the human body tissues contain fluorine. Measured CT values at 140 kVp, together with other information about our particular batch of tissue substitutes (inserts differ from batch to batch), are shown in table 3.2.

Using estimates of the chemical composition from Watanabe [1999], the stopping powers for these tissue inserts were calculated using the different formulae and their listed mean excitation energies: (i) Bichsel [1972], equation 3.2; (ii) Janni [1982], equation 3.4; and (iii) ICRU [1993], equation 3.6, with condensed and compound I-values. The RSP for each tissue was then determined by dividing by the stopping power of water (calculated using the same source) and taking the average over the energy range relevant to proton therapy and proton imaging (80-330 MeV), as we intend to investigate the applicability of proton radiography in chapters 4 and 5. The calibration curves were formed by joining the individual points together. These calibration curves will be referred to as the 'Bichsel', 'Janni', 'ICRU condensed' and 'ICRU compound' calibration curves for the remainder of this chapter. The RSP of each tissue was also calculated directly, using equation 3.7, but with the I-values from each of the four sources (Bichsel [1972], Janni [1982], ICRU [1993] (condensed and compound)). These calibration curves will be referred to collectively as the 'Schneider calibration curves' for the remainder of this work. These eight calibration curves are shown in figure 3.3.

To investigate the systematic error introduced by this process, RSPs were measured directly using a Bragg peak shift measurement. The Bragg peak from a pencil proton beam, smaller than the diameter of the Gammex insert, was recorded

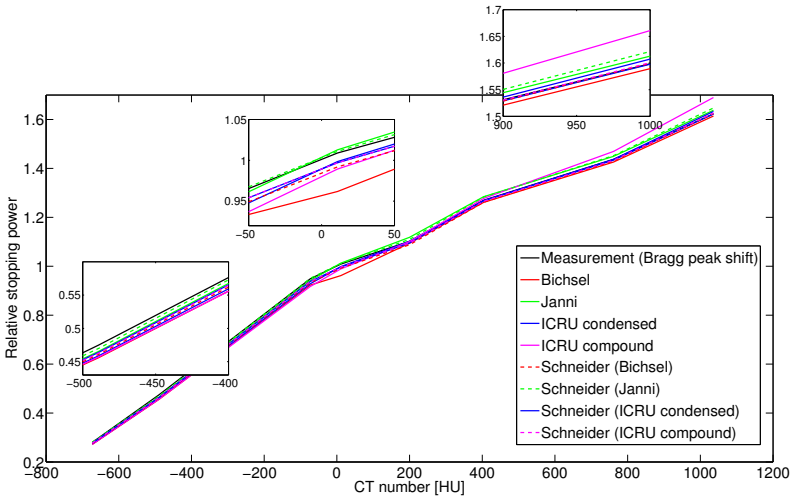


Figure 3.3: Calibration curves with RSP values calculated using the various formulae. The inserts show zooms of particular regions for clarity (axes the same as main figure).

using the commercial stack of parallel plate ionisation chambers, the Zebra (IBA Dosimetry, Belgium). The inserts were positioned end-on, so that the protons traversed through the long side of the insert. The shift in the Bragg peak position was recorded with and without the insert in front of the Zebra. This water equivalent shift, together with the known length of the insert, allowed for computation of the insert RSP.

3.3.2 Range errors

To investigate how differences in RSP translate into range errors, double scattering plans were then made in XiO using single beams treating an octagonal target at the centre of the Gammex phantom, as shown in figure 3.4. Separate plans were created with beams at gantry angles 45° apart, so that the single beam passes through one insert in each plan. An octagonal target was chosen so that the target appears identical from each beam direction. All inserts were in the inner set of holes (so that the effect of beam hardening was approximately consistent between plans). Apertures and compensators were created using the measured calibration curve and kept consistent for the other calibration curves. Apertures were generated with a small 1 mm lateral margin around the target to ensure the beam only passed through the insert of interest. Compensators were generated to conform to the distal edge without any range uncertainty margins, with distances between 17.86 g/cm^2 (LN-300 Lung) to 20.38 g/cm^2 (Cortical Bone). The dose was computed using a fast pencil algorithm (Hong et al. [1996]). Dose profiles were determined through the centre of the beam in each plan for each of the calibration curves and the distal R80 was assessed (the range at which 50% of the protons are stopped). This value was used as it is invariant of the initial energy spread of the protons (Berger [1993];

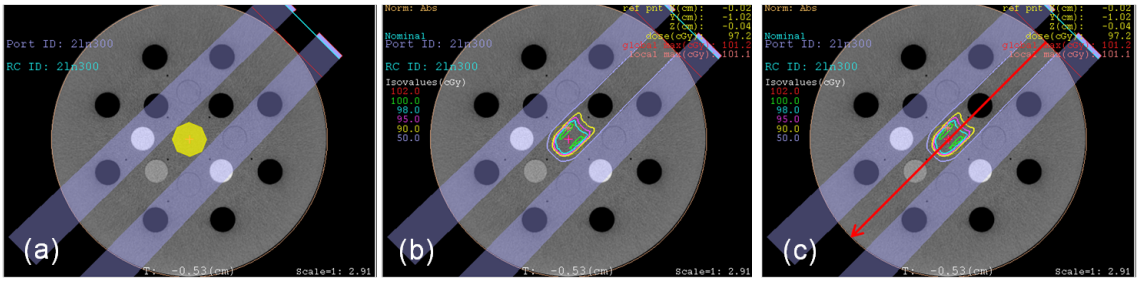


Figure 3.4: Example plan (in this case through a LN-300 lung insert) to investigate the impact on the range within the Gammex phantom. Apertures are shown by the blue rectangles, with (a) the octagonal target shown in yellow. (b) Dose distribution. (c) Dose profile along the red arrow.

Bortfeld [1997]; Paganetti [2012]).

Two patients were tested to investigate how such range errors would translate in the patient. We used the same approach as for the tests in the Gammex phantom: (i) direct a single double scattering beam at the target and compute the apertures and compensators required to treat to the distal edge, with no range uncertainty margin, when using the measured calibration curve; (ii) compute the dose profile through the centre of the beam for this measured calibration curve; (iii) alter the calibration curve to one of Schneider calibration curves; (iv) recompute the dose; and (v) compute the new dose profile. For both patients tested, apertures were created with a clinically-realistic 1 cm lateral margin, but (as with the Gammex plans) compensators were given no margin to account for range uncertainty. Additionally, no smearing was applied to the compensators as this would only apply to the measured calibration curve and we wanted to probe the range effects specifically.

The lung patient plans can be seen in figure 3.5. For this case, two realistic beam arrangements were chosen: posterior-anterior (PA) and a right lateral (RLat) beam. In the PA set up the radiological distance to the distal edge was 8.9 g/cm^2 (set as our range) and the modulation to cover the tumour was 3.7 g/cm^2 (both for the measured calibration curve). For the RLat beam the range/modulation was 13.2/2.8 g/cm^2.

The prostate patient plans can be seen in figure 3.6. For this case, two beam arrangements were chosen. One plan was based on the typical opposing-pair arrangement, but using just a single left lateral (LLat) beam so that the range effect through thick boney areas, such as the femoral heads, could be determined (it would be impossible to investigate the impact on the range for overlapping beams). In this set up the radiological distance to the distal edge was 21.4 g/cm^2 and the modulation to cover the tumour was 6.7 g/cm^2. Another plan was created using a single anterior-posterior (AP) beam, which has been suggested by Tang et al. [2011] as an alternative to the opposing lateral beam pair arrangement commonly used in proton therapy prostate treatments. For this case the range/modulation was

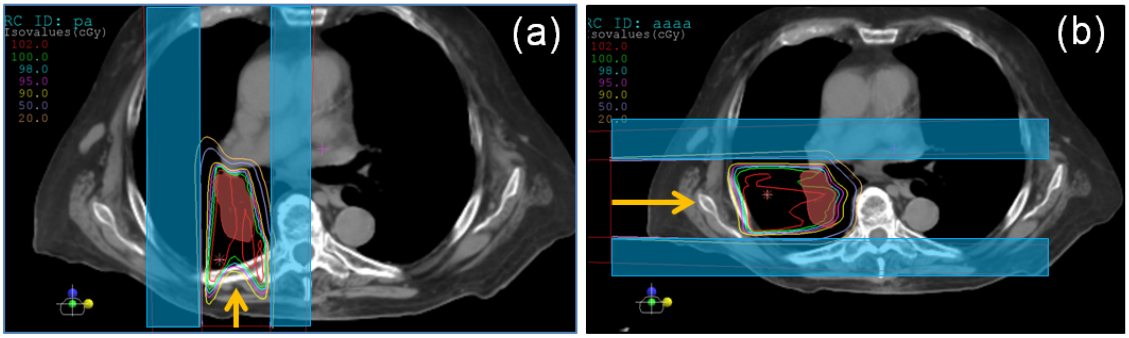


Figure 3.5: Example dose distributions for the (a) PA and (b) RLat double scattering plans for the lung case. The target to treat is shown by the red shaded region, apertures by the blue rectangles and the beam direction by the orange arrows.

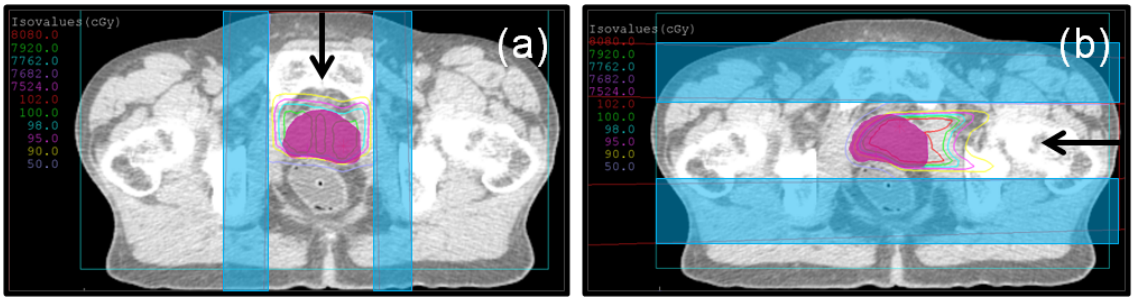


Figure 3.6: Example dose distributions for the (a) AP and (b) LLat double scattering plans for the prostate case. The target to treat is shown by the pink shaded region, the apertures by the blue rectangles and the beam direction by the black arrows.

$11.5/3.6\,\mathrm{g/cm^2}$.

3.3.3 Systematic errors in the current stoichiometric calibration procedure

To quantify the error in step four of the stoichiometric calibration procedure, we required a gold standard that best matched our measurements. In assessing the impact on the range, it was found that the errors were lowest when using the elemental Janni I-values, but it was also shown that the I-value of water plays a significant role (these results can be found later in the report, in section 3.4.2). Given that the I-value of water is currently disputed (section 3.1.2.2), we varied it between 78 eV and 82 eV (in steps of 1 eV) to create 5 calibration curves. Plans were created in the Gammex phantom in the same manner as previously described in section 3.3.2 and the distal R80 was assessed. It was found that an I-value of 80 eV, together with the elemental Janni I-values, produced RSP values that best matched our measurements of tissue inserts. We therefore used this as the gold standard.

As described in detail in the introduction, the final step of the stoichiometric calibration procedure is the theoretical calculation of the RSP of real tissues. As discussed, the most convenient (and likely most frequently used in clinical practice)

formula for performing such calculations is the Schneider approximation (equation 3.7). It is recommended that elemental I-values are taken from ICRU Report 49 and the tissue I-values should be determined based on theoretical tissue compositions from literature (ICRU [1993]; White et al. [1987]; Woodard and White [1986]). In this manner the I-value of water is calculated to be 75.3 eV, (table 3.1), however in a more recent publication, ICRU [2009] Report 73, a water I-value of 78 eV was recommended. We therefore calculated the RSPs of 73 human tissues using Bragg's additivity rule (ICRU [1992]; Seltzer and Berger [1982]), using the ICRU compound I-values together with these two water I-values. The accuracy of these estimates was assessed by comparison with the RSPs of real tissues theoretically calculated using the gold standard described above.

3.4 Results

3.4.1 Computation of the RSP

The theoretical values of RSP for the tissue substitutes, calculated using the Bichsel, Janni and ICRU stopping power formulae (equations 3.2, 3.4 and 3.6) and averaged over the energy range 80-330 MeV, are detailed in table 3.3. Using the I-values from these sources, but with the Schneider approximation for the RSP directly, equation 3.7, gives the theoretical values for RSP as detailed in table 3.4. It can be seen that for both cases this theoretical calculation of RSP, step four of the stoichiometric procedure, generally (not including Janni) results in a systematic underestimation that varies across the CT numbers.

3.4.2 Range errors

A zoom of the fall-offs for the different Bichsel, Janni and ICRU calibration curves, through the different inserts, can be seen in figure 3.7. The same set of results for the Schneider calibration curves are shown in figure 3.8. In these two figures the fall-offs for only three insert plans are shown, but the trend was similar for all other inserts. The errors of each R80 with respect to the plans created using the measured calibration curve is shown in figure 3.9.

The fall-offs for the different relative calibration curves, for the four patient plans, can be seen in figure 3.10. The errors of these values with respect to the plans created using the measured calibration curve is shown in figure 3.11.

Material	Meas.	Bichsel [1972]		Janni [1982]		ICRU [1993] Condensed		ICRU [1993] Compound	
	RSP	RSP	Err	RSP	Err	RSP	Err	RSP	Err
LN-300 lung	0.280	0.272	-2.97	0.277	-0.96	0.273	-2.36	0.272	-2.73
LN-450 lung	0.475	0.456	-3.99	0.466	-2.00	0.459	-3.39	0.457	-3.75
Adipose	0.949	0.923	-2.73	0.942	-0.72	0.929	-2.14	0.917	-3.33
CT Solid Water	1.009	0.962	-4.67	1.013	0.38	0.999	-1.04	0.989	-1.95
Inner Bone	1.102	1.094	-0.71	1.118	1.49	1.103	0.04	1.101	-0.10
CB2 30% $CaCO_3$	1.270	1.261	-0.69	1.284	1.13	1.268	-0.12	1.278	0.59
CB2 50% $CaCO_3$	1.437	1.426	-0.77	1.449	0.85	1.433	-0.23	1.469	2.25
Cortical Bone	1.623	1.614	-0.55	1.638	0.90	1.623	-0.03	1.690	4.11
		RMSE†	2.13		1.05		1.17		2.35
		Mean	-2.13		0.13		-1.16		-0.61
		Max‡	-4.67		-2.00		-3.39		4.11

Table 3.3: RSP for the Gammex tissue substitutes, calculated using the individual source formulae. The percentage error ('Err') with respect to the measurement is listed (may not match the values in the table precisely because of rounding). †Root-mean-square-error. ‡Maximum error (in either positive or negative direction).

Material	Meas.	Bichsel [1972]		Janni [1982]		ICRU [1993] Condensed		ICRU [1993] Compound	
	RSP	RSP	Err	RSP	Err	RSP	Err	RSP	Err
LN-300 lung	0.280	0.275	-1.79	0.281	0.36	0.277	-1.07	0.277	-1.07
LN-450 lung	0.475	0.461	-2.95	0.470	-1.05	0.464	-2.32	0.463	-2.53
Adipose	0.949	0.933	-1.67	0.951	0.21	0.938	-1.16	0.938	-1.16
CT Solid Water	1.009	0.992	-1.68	1.011	0.20	0.997	-1.19	0.997	-1.19
Inner Bone	1.102	1.088	-1.27	1.110	0.73	1.095	-0.64	1.095	-0.64
CB2 30% $CaCO_3$	1.270	1.262	-0.63	1.283	1.02	1.268	-0.16	1.266	-0.32
CB2 50% $CaCO_3$	1.437	1.430	-0.49	1.452	1.04	1.437	0.00	1.433	-0.28
Cortical Bone	1.623	1.625	0.12	1.647	1.48	1.632	0.55	1.626	0.18
		RMSE†	1.33		0.76		0.89		0.91
		Mean	-1.30		0.50		-0.75		-0.87
		Max‡	-2.95		1.48		-2.32		-2.53

Table 3.4: RSP for the Gammex tissue substitutes, calculated using the Schneider approximation for the RSP, equation 3.7, but with the I-values of the different sources. Same format as table 3.3.

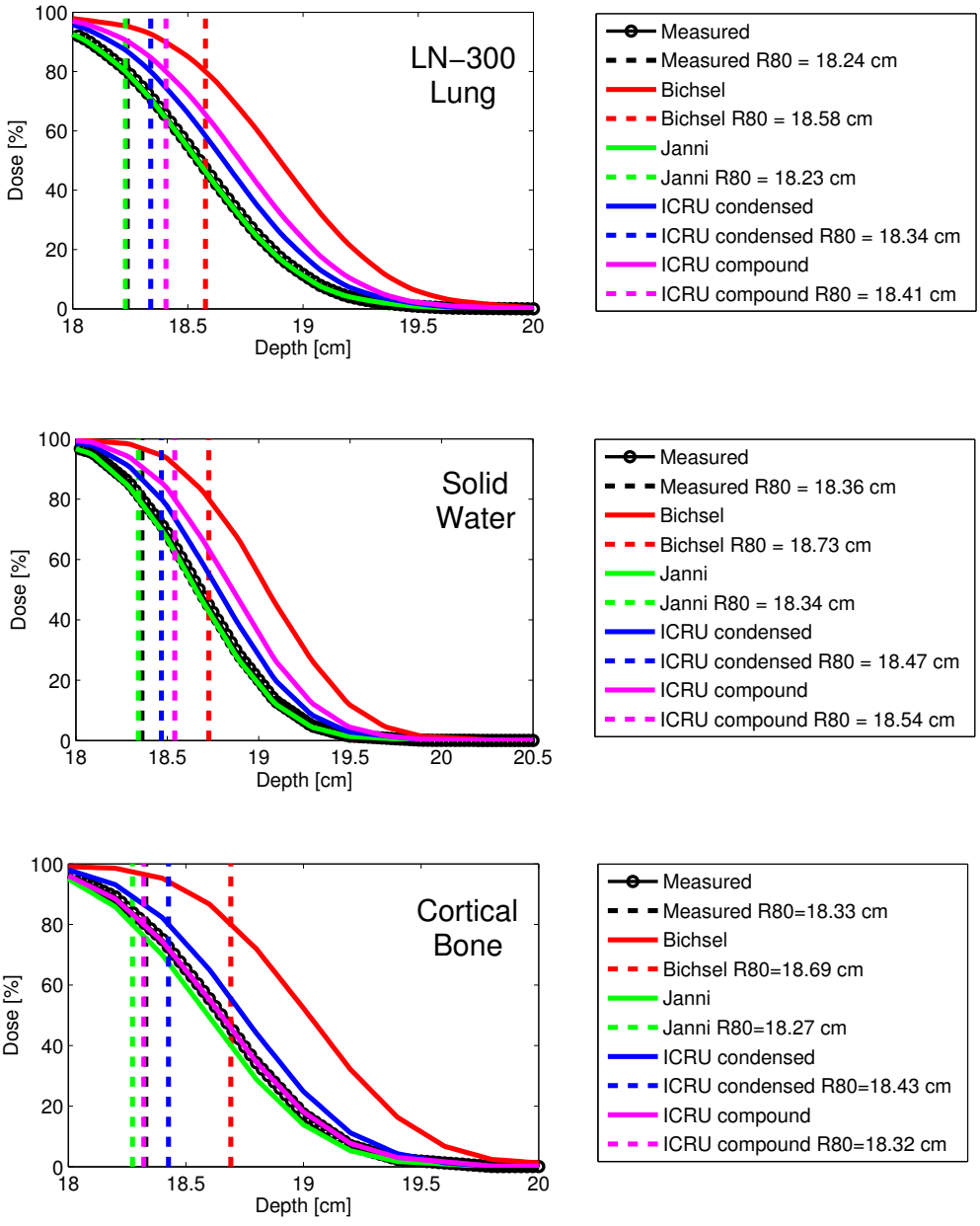


Figure 3.7: Dose profiles showing the distal fall offs for the Bichsel, Janni, ICRU (condensed) and ICRU (compound) calibration curves.

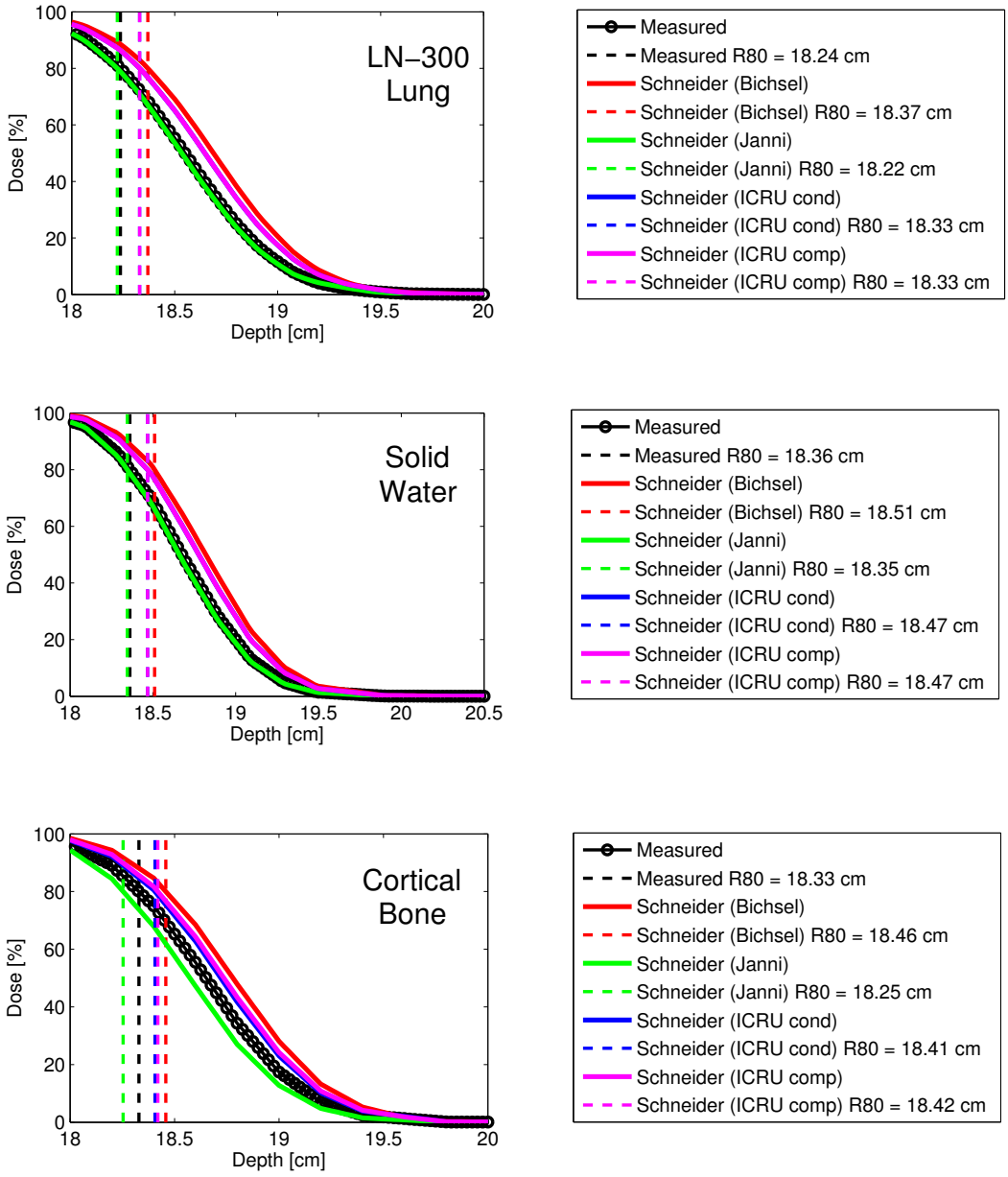


Figure 3.8: Dose profiles showing the distal fall offs for the Schneider calibration curves.

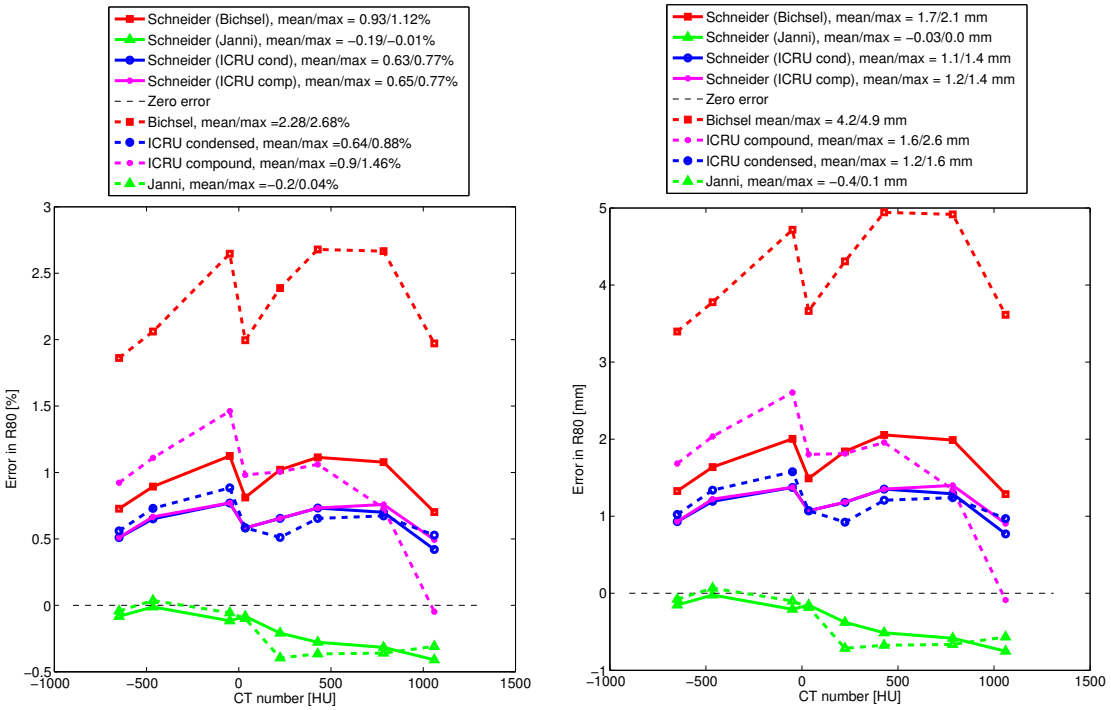


Figure 3.9: Errors in the distal R80 of each plan, compared to the measured values, for all calibration curves. Results are shown in units of [%] (a) and [mm] (b), with mean/max errors listed in the legend.

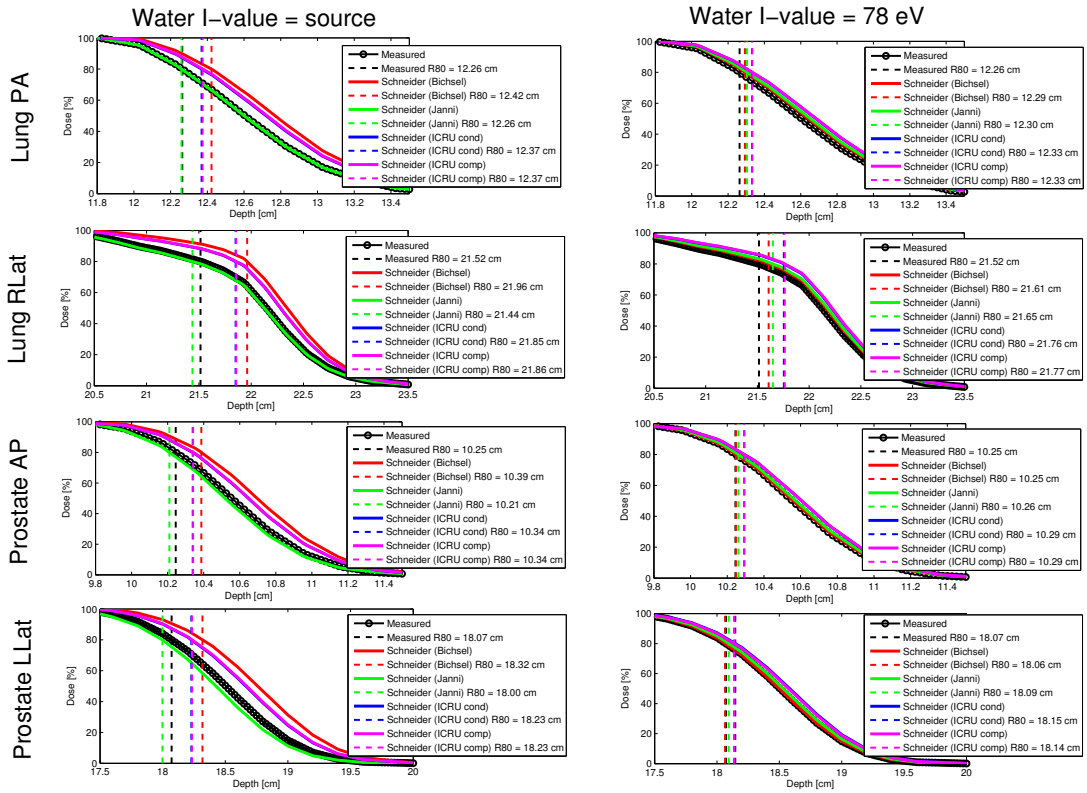


Figure 3.10: Dose profiles showing the different fall offs in the patient cases, for the Schneider calibration curves.

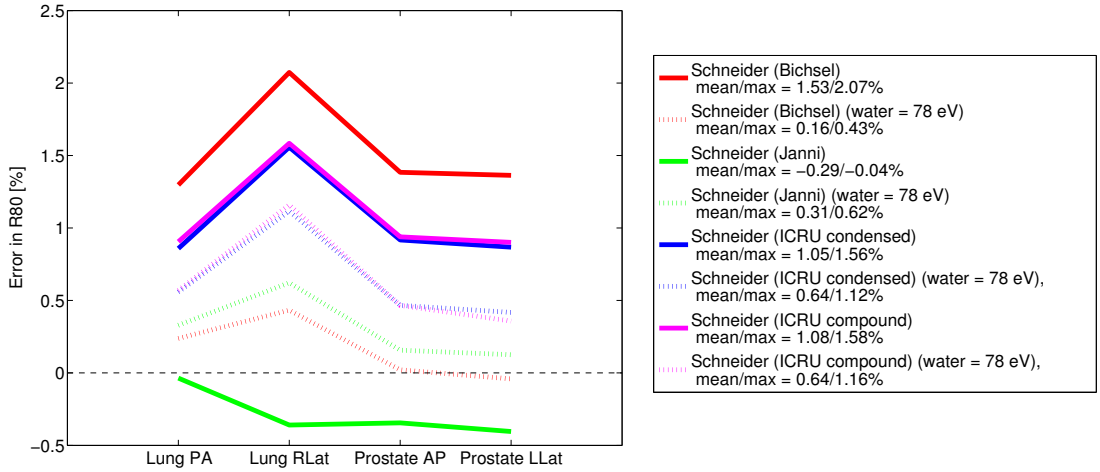


Figure 3.11: Errors in the distal R80 of each patient plan, compared to plans created with the measured calibration curves.

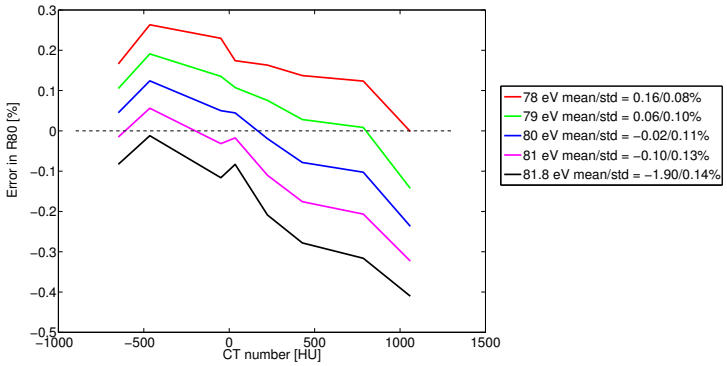


Figure 3.12: Errors in the distal R80 for the 5 calibration curves with different water I-values. The mean and standard deviation of errors across the HU range are listed in the legend.

3.4.3 Systematic errors in the current stoichiometric calibration procedure

To determine the gold standard that best matched our measurements for tissue inserts, we used the elemental Janni I-values with 5 different I-values of water to create 5 calibration curves. The errors in the distal R80 with respect to the measured values for these 5 calibration curves can be seen in figure 3.12.

From figure 3.12 it was determined that the gold standard was using a water I-value of 80 eV with the Janni elemental I-values. Figure 3.13 shows the comparison between the RSPs of real tissues theoretically calculated using the gold standard and those calculated using ICRU elemental I-values with a water I-value of 75.3 eV and 78 eV.

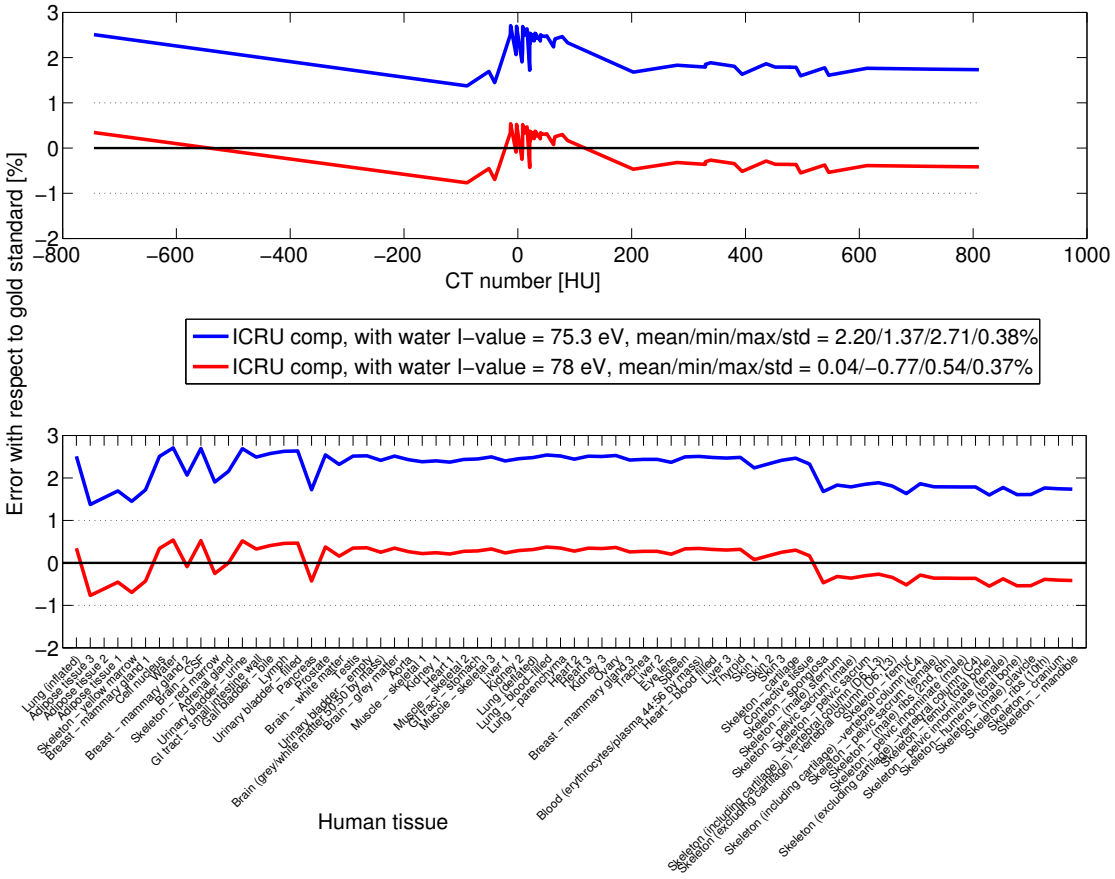


Figure 3.13: Errors in human tissue RSPs, calculated using the Schneider approximation with the ICRU compound I-values and with two water I-values.

3.5 Discussion

3.5.1 Computation of the RSP

Different methods of calculating the stopping power result in different RSP values and, thus, different calibration curves. It can be seen from inspection of tables 3.3 and 3.4 that these differences vary across the CT number range. Comparing to direct measurements in a proton beam, the largest differences were found in the LN-450 Lung insert (2-4% underestimation in the RSP). Such errors could have a big impact on patient outcome, as it has been shown that correct proton beam range prediction is critical in the lung (Seco et al. [2012]).

By first calculating the stopping power, the theoretical calculation of the RSP, one of the fundamental steps of the stoichiometric calibration procedure, has a mean/max error of: -2.1/-4.7% for Bichsel (equation 3.2); +0.1/-2.0% for Janni (equation 3.4); -1.2/-3.4% for ICRU condensed (equation 3.6); and -0.6/+4.1% for ICRU compound (equation 3.6). If calculated using the Schneider approximation (equation 3.7), the mean/max errors are generally smaller: -1.3/-3.0% for Bichsel I-values; +0.5/+1.5% for Janni I-values (overestimation); -0.8/-2.3% for ICRU condensed I-values; and -0.9/-2.5% for ICRU compound I-values.

For both methods, it can be seen that there is a general underestimation of the RSP (except when using the Janni formulae and its I-values). Of course, a possible cause of this systematic underestimation could be due to an error in the measurement procedure, which should not be ruled out. However, what cannot be denied is that there is a difference between the sources and techniques for computing the RSP. Currently, ICRU guidelines suggest the use of their values, but we have shown that use of the Janni formulae or Janni I-values (when computing the RSP directly using equation 3.7) clearly shows the best match with measurements.

The likely approach for most centres will be to use the Schneider formula (for convenience), together with one set of the ICRU I-values (given its later publication date and recommendation for use). However, our results suggest that this results in underestimations of up to -2.3/-2.5% (condensed/compound). Use of the Bichsel set of I-values is not recommended, with an even poorer match to measurements, but the Janni set of I-values proved to be a better match for our measurements.

3.5.2 Range errors - Gammex phantom

The impact on the range was assessed with single-beam double scattering plans on a Gammex phantom, with the beam passing through the Gammex phantom holder and the relevant insert. As expected, the errors in the RSP translated into range

errors. For the majority of cases the RSP was underestimated, which resulted in a systematic overestimation of the beam range (prescribed ranges between 17.86-20.38 g/cm^2). The mean/max errors were: Bichsel = 2.4/3.7%, 4.3/6.7 mm; Janni = -0.1/1.6%, -0.2/3.0 mm; ICRU (condensed) = 0.7/2.3%, 1.3/4.2 mm; ICRU (compound) = 1.0/2.7%, 1.8/4.9 mm. The mean/max errors for the Schneider calibration curves were: Bichsel = 1.0/2.5%, 1.8/4.5 mm; Janni = -0.1/1.6%, -0.02/2.9 mm; ICRU condensed = 0.7/2.3%, 1.3/4.1 mm; ICRU compound = 0.7/2.3%, 1.3/4.1 mm. It can be seen that, as one would expect from the results of the computation of the RSP, both sets of Bichsel results perform the poorest. Also, as expected, the Janni set of values matched our measurements best.

In our plans the set up is such that the errors are dominated by the errors in the RSP of CT Solid Water. The largest geometric distance to the distal edge was 18.5 cm (the target had a curved distal edge), of which only 2.8 cm passed through the specific insert being investigated. This equates to $\approx 15\%$ of the total geometric distance or $\approx 4 - 24\%$ of the radiological distance (range between LN-300 and Cortical Bone). We believe that although there is a bias towards the error in CT Solid Water, this approach is still suitable because these proportions are similar to those of water found in an actual patient.

3.5.3 Range errors - patient cases

It can be seen in figure 3.10 that using the water I-value of the source gives similar results to the Gammex phantom, with mean/max errors across the four plans of: Bichsel = 1.5/2.1%; Janni = -0.3/-0.4%; ICRU condensed = 1.1/1.6%; ICRU compound = 1.1/1.6%. Altering the I-value of water to the most recently recommended value of 78 eV resulted in a big reduction of range errors for most cases. The mean errors were reduced by: Bichsel = 90%; ICRU condensed = 39%; and ICRU compound = 41%. The absolute error for Janni changed from an underestimation to an overestimation (with a minor increase in the absolute error of 7%). These results are suggestive that the I-value of water has a very important role in predicting the RSP of tissues.

3.5.4 Systematic errors in the current stoichiometric calibration procedure

The first task was to determine the gold standard that best matched our measurements. It was clear from the results of section 3.4.1 that, when using the Schneider approximation, use of the elemental I-values of Janni had the closest match to our measured calibration curve. These values, together with the alteration of the water

I-value from 81.8 eV to 80 eV, enabled a very good match to our measurements, with a mean error ± standard deviation of $-0.02 \pm 0.11\%$. Thus our gold standard formulation involved: using the Schneider approximation for the RSP (equation 3.7); elemental Janni I-values; and a water I-value of 80 eV.

Using this gold standard the RSP values of 73 human tissues were theoretically computed. The same values were computed twice more using the ICRU compound I-values, once with a water I-value of 75.3 eV and another with a water I-value of 78 eV. The mean error (± standard deviation) between these and the gold standard was found to be $2.20 \pm 0.38\%$ and $0.04 \pm 0.37\%$ for the 75.3 eV and 78 eV calibration curves, respectively. It is clear that the most recently recommended ICRU water I-value of 78 eV is better than the previous value, although there are still differences of up to 0.8% from our gold standard (the gold standard itself has an uncertainty of up to 0.1%). These differences are tissue dependent, with underestimations for fatty and boney tissues and overestimations for soft tissues.

3.5.5 How to correct the stoichiometric calibration

In the previous sections it can be seen that systematic errors can be introduced by the theoretical RSP calculation of human tissues. However, it is possible to significantly reduce these errors by applying a correction when calculating the RSP of any tissues that form the calibration curve. This requires an additional step in the stoichiometric calibration procedure, between current steps 3 and 4, in which the user conducts measurements to determine the RSP values of the tissue inserts imaged in step 1. A comparison of the theoretical RSP values (calculated by the method of the user's choosing) and the measured RSP values will allow for the systematic scaling and/or offset to be determined. This scaling/offset would then be applied to the theoretically-calculated values for human tissues, step 4 of the stoichiometric calibration procedure.

3.6 Conclusions

There are four clear methods for calculating the RSP of a given tissue. Three involve the computation of the absolute stopping power of the tissue, which must then be divided by the absolute stopping power of water over the same energy range (Bichsel [1972], Janni [1982] and ICRU [1993]). To account for the different formulae and corrections, each source has its own set of I-values (ICRU has two, for condensed and compound materials). Comparing the RSPs of Gammex inserts determined using a simple Bragg peak shift measurement in a proton beam, it was found that all of these techniques, with the exception of the Janni formulae, result in a mean systematic

underestimation of the RSP: Bichsel = -2.1%; Janni = +0.1%; ICRU (condensed) = -1.2%; ICRU (compound) = -0.6%. The fourth method is an approximation in which the RSP is computed directly (Schneider et al. [1996]), using any set of I-values. The errors for this approach followed almost the same pattern but were generally lower in magnitude. For our set of measurements, the Janni I-values showed the best match.

We showed that these RSP errors translate into proton beam range errors of similar magnitudes, for plans within a Gammex phantom and in two patient cases (lung and a prostate, each with two different beam directions). Mean range errors across the four patient plans, for RSPs calculated using the fourth method only, unsurprisingly followed the trend of the errors in the RSP computation.

Using the Schneider approximation, Janni elemental I-values and a water I-value of 80 eV we produced a gold standard method that could match our measured RSP values to within $0.02 \pm 0.11\%$ (mean $\pm$ standard deviation). We used this gold standard to compute the RSPs of real tissues, and compared them with sets of values computed using the ICRU elemental I-values together water values of either 75.3 eV or 78 eV. The latter matched to within $0.04 \pm 0.37\%$ (mean $\pm$ standard deviation), confirming that a value of 78 eV should be used for the water I-value.

It is proposed that an additional step should be added to the current stoichiometric calibration procedure that involves actual measurement of the tissue insert RSP in a proton beam. Any scaling/offsets required to match the theoretical calculation of the tissue insert RSP to the measured RSP should then be also applied to step 4 of the current stoichiometric procedure, the theoretical calculation of the RSP of human tissues.

Chapter 4

Dose ratio proton radiography

Work in this chapter can be found in the following journal article:

- (Under review) **Doolan P**, Royle G, Gibson A, Lu H-M Bentefour E H **2014** Dose ratio proton radiography using two pristine Bragg peaks *Medical Physics*

Practical measurements were made at MGH with E-H Bentefour and H-M Lu. The complete theoretical basis and all analysis was conducted by myself.

4.1 Background

Proton radiography, first proposed by Cormack [1963], is based on measuring the position and energy loss of protons after passing through a medium. This information is then utilised to estimate the integral energy loss along the trajectory, which is converted into a material WET.

Proton radiography offers the potential for higher tissue-to-tissue contrast and better density resolution than X-ray imaging because of the sharpness of the Bragg peak near the end of the proton range (Schulte et al. [2005]; Seco and Depauw [2011]). The shape of the proton depth dose curve also enables the use of less dose (Schneider et al. [2004]; Schulte et al. [2005]). One of the major advantages of proton radiography is the potential to generate RSP maps of the patient directly, rather than the current process of converting the patient's X-ray CT dataset using a calibration curve (Hurley et al. [2012]). As explained previously, this process is thought to contribute between 0.5-1.8% to the proton beam range uncertainty, depending on whether the uncertainties in the material I-values are considered (Paganetti [2012]; Schaffner and Pedroni [1998]). In addition to reducing the proton beam range uncertainty in proton treatment planning, proton radiography has a number of other potential applications, including: convenient patient set-up (Depauw et al. [2014]); passive range verification (Schneider et al. [2005]); and active range verification (Testa et al. [2013]).

4.1.1 Classical proton radiography

One of the major disadvantages of proton radiography is due to the fact that protons undergo MCS in the patient, which limits the spatial resolution compared to X-ray imaging (Schneider et al. [2012]). This physical process has driven the classical approach to proton radiography, in which attempts are made to correct for MCS by tracking protons before and after the patient and by measuring the residual proton energy at the patient exit. The proton trajectory and the energy loss across that path can then be estimated. To keep the patient dose to a minimum this technique requires the tracking of every proton, which places high demands on the speed of device (typical proton dose rates are on the order of 10^9 protons per second). A range of designs have been used for this high speed proton tracking: scintillators (Pemler et al. [1999]; Schneider et al. [2004]; Shinoda et al. [2006]); semiconductors (Poludniowski et al. [2014]; Schulte et al. [2004]; Sipala et al. [2011]; Talamonti et al. [2010]); plasma panels (Ball et al. [2014]; Friedman et al. [2012]); nuclear emulsion films (Braccini et al. [2010]); and gas-based devices (Hollebeek et al. [2012]; Watts et al. [2009]). The energy measurement is either performed directly, using a calorimeter, or inferred from the proton range, in a range telescope design. A variety of direct methods have been investigated: magnetic spectrometry (Akisada et al. [1983]); NaI calorimeters (Schneider and Pedroni [1995]); CsI calorimeters (Talamonti et al. [2010]), (Heimann [2005]; Petterson et al. [2006]; Talamonti et al. [2010]); and scintillating crystals (Shinoda et al. [2006]; Vanzi et al. [2013]). Range telescope designs have included: stacks of scintillators (Bashkirov et al. [2007]; Watts et al. [2009]); and multi-layer ionisation chambers (Rinaldi et al. [2013, 2014]).

4.1.2 Single-detector proton radiography

Despite the promises and the extent of research on the topic, proton radiography has not been routinely introduced into the clinical environment. The classical devices are generally complex, they can often be bulky and the projected cost is high. As such, there has been a recent movement towards using only a single detector positioned beyond the patient. Individual proton paths cannot be reconstructed, but maybe they can be approximated sufficiently well that the proton radiograph is still useful. A variety of technologies have been explored, including: fluorescent screens coupled to CCD cameras (Bentefour et al. [2013]; Muraishi et al. [2009]; Ryu et al. [2008]); commercial flat-panel detectors (Telsemeyer [2012]); and complementary metal oxide seminconductor active pixel pensors (CMOS APS) (Gelover-Reyes et al. [2011b]; Poludniowski et al. [2014]; Seco and Depauw [2011]).

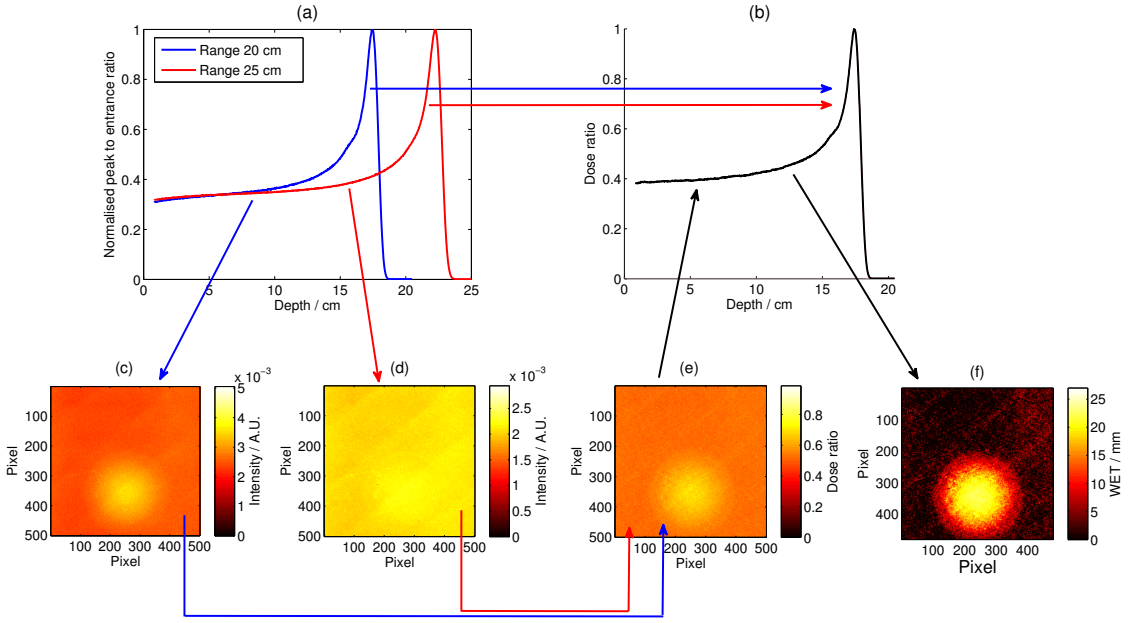


Figure 4.1: Dose ratio method steps to produce the dose ratio image. (a) Measure depth dose curves in water for two pristine Bragg peaks. (b) Determine the dose ratio curve. (c-d) Record the dose map beyond an object for each of the two energies. (e) Form a dose ratio map by dividing (c) by (d). Finally, convert the dose ratio map to a WET map (f) using the dose ratio curve.

4.1.3 Dose measurement for proton radiography

The measurement of dose offers a simple approach to proton radiography. A range of such techniques have been tested. The use of pairs of sloped spread-out Bragg peaks (SOBPs) was suggested by Lu [2008a]; Lu et al. [2010]. Another technique involves measuring the time-resolved dose from a proton beam passively scattered by a range modulator wheel (Gottschalk et al. [2011]; Lu [2008b]; Testa et al. [2013]), which is discussed in much greater detail in section 5.1.2 as it was the technique used to generate patient-specific calibration curves for proton treatment planning. Another technique, suggested by Bentefour et al. [2012b], involves the measurement of the ratio of doses of two pristine Bragg peaks, which is the subject of this work. The technique works as follows: (1) measure depth dose curves in water for two pristine Bragg peaks; (2) determine the 'dose ratio curve' for the energy pair by dividing the lower energy by the higher energy; (3) record the dose map beyond an object of unknown thickness/material for each of the two energies and calculate the dose ratio map; and (4) convert the map to one of water-equivalent thicknesses (WETs) using the dose ratio curve. This approach, known as the dose ratio method, is shown schematically in figure 4.1.

This method has a number of advantages. It requires the use of a single imaging detector, which is more convenient and more likely to fit into the clinical environment than the conventional trackers and energy measurement set up. Also, by taking the dose ratio, the detector does not need to be able to measure absolute dose, rather it must demonstrate high reproducibility. The technique also has the potential to be applied to PBS treatments, as it uses pristine Bragg peaks rather than the spread-out Bragg peaks used in passive scattering.

4.1.4 Aim

It was previously shown by Bentefour et al. [2012b] that this technique can achieve millimetre accuracy in determining the WET, using the distal edge of the lower energy Bragg peak. However, in this approach the dynamic range is limited by the steep fall-off and the approximate material WET must be known to a high accuracy to appropriately select the energy pair. Although on the proximal side of the Bragg peak the fall-off is only from 100% to $\approx$ 30% (compared to the distal fall-off from 100% to 0%), the fall-off is much less steep, leading to a potentially much larger dynamic range. It has not yet been explored how much of the proximal side of the Bragg peak can be reliably used and for which patient treatment sites this imaging technique could be applied. In this study we aim to do the following:

1. Investigate the theoretical limits of the dose ratio method for the proximal side of the Bragg peak. We determine the dynamic range for different energy pairs, different required accuracies in the WET and for a range of uncertainties in the dose ratio map. Comparisons are made with use of the distal side of the Bragg peak.

2. Determine the applicability of this technique. We follow the same theoretical analysis as in step 1, for the energy pair most appropriate to the particular clinical site, and then determine the area that can be imaged under different sets of clinical requirements and proton imager performances.

3. Validate the theoretical predictions with real measurements using a CMOS APS. After measuring the noise in the CMOS APS, we generate theoretical limits and verify that the accuracy in imaging a sapphire sphere matches these predictions.

4.2 Materials and methods

4.2.1 Theoretical limits

The dose ratio method relies on the measurement of pristine Bragg peaks in water. To assess the theoretical limits of the technique, independent of the detector used, pristine Bragg peaks of different energies were modelled using the analytical approximation that considers both range straggling and an initial spread of beam energies (Bortfeld [1997]). The range R_0 was approximated by the range-energy relation $R_0 = \alpha E^p$, for a proton beam with initial energy E, where the constants $\alpha = 0.0022$ and $p = 1.77$ are for a proton beam in water (Bortfeld [1997]). The width of the Gaussian energy spectrum was set to $\sigma_E = 0.01E$. Taking the ratio of normalised depth doses curves of two energies, a theoretical dose ratio curve can be generated.

Our approach to determining the theoretical limits, shown schematically in figure 4.2, is based on introducing uncertainties in dose ratio measurement and inspecting the change in the WET estimate from the theoretical dose ratio curve (in figure 4.2(a) the original dose ratio value is shown in green and the upper/lower uncertainties are shown in blue/red). Then, very simply, right/left horizontal error bars are added to the dose ratio curve for that particular dose ratio value (shown as blue/red arrows in figure 4.2(b)). For a given uncertainty introduced, we used the spread in the proton WET values to define the maximum achievable accuracy of the WET measurement (figures 4.2(c) and 4.2(d)). If the user requires a very high accuracy in the WET (i.e. the horizontal error bars cannot be too large), then the amount of the proximal side of the Bragg peak that can be used for imaging is small. Conversely, if the accuracy required in the WET is relaxed (i.e. the horizontal error bars can be bigger), more of the proximal side of the Bragg peak can be reliably used for imaging. Dependent on the accuracy required in the WET, the amount of the proximal side of the Bragg peak that can be used for imaging (henceforth referred to as 'proximal dynamic range') will vary.

In this analysis, the impact of an uncertainty in the dose ratio map of 1 to 10% (in steps of 1%) is investigated. Uncertainties in the individual energy images will combine in quadrature in the dose ratio map, with typical values of standard deviation (σ) in the dose ratio map found to be 1-3% in our measurements. In real measurements, uncertainties will arise in the raw measured images from both statistical uncertainties (dependent on the number of protons) and detector uncertainties (such as quantum noise, shot noise, read noise, fixed pattern noise, etc.). To remove the effect of statistical uncertainties, which would otherwise vary across the Bragg peak, the error in the dose ratio was divided by the normalised dose at each depth.

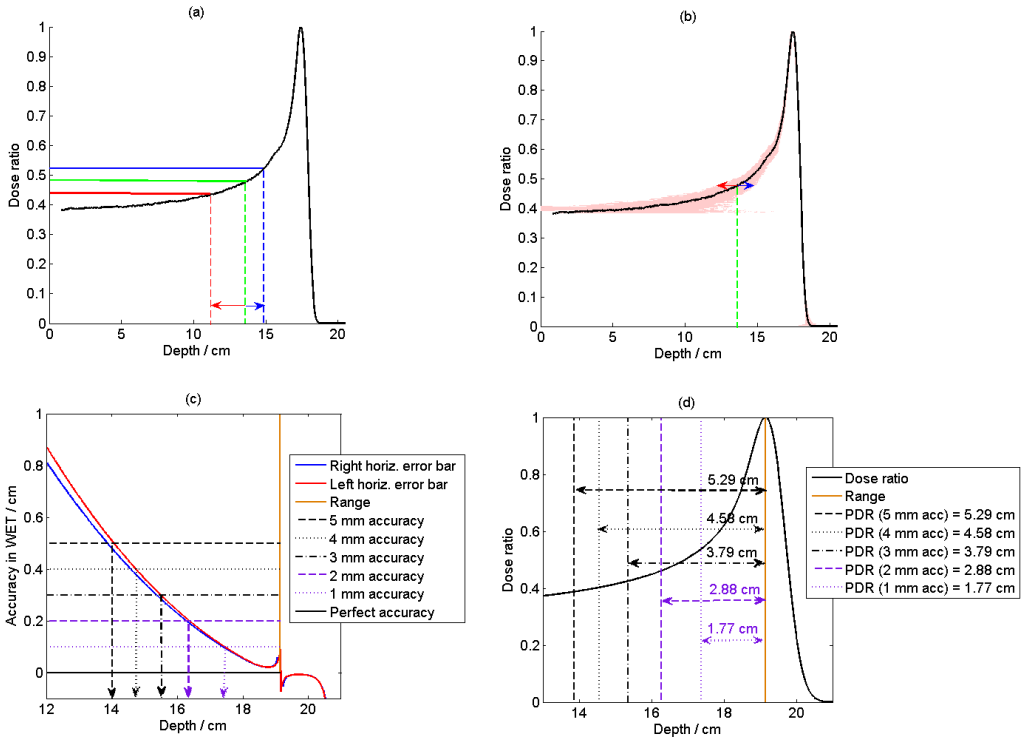


Figure 4.2: Method of determining the proximal dynamic range. (a) Uncertainties in the dose ratio translate into uncertainties in the WET (depth). (b) These are added to the dose ratio curve as horizontal error bars. (c) The error bar magnitude defines the accuracy in the WET. (d) Dependent on the required accuracy, the proximal dynamic range can be determined.

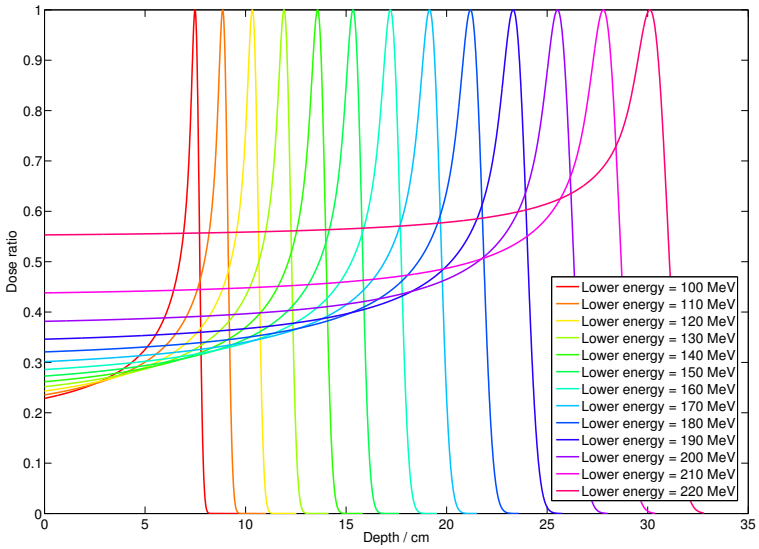


Figure 4.3: Normalised theoretical dose ratio curves (all paired with 230 MeV).

Setting	Values used
Energy (low)	100-220 MeV (in steps of 10 MeV)
Energy (high)	230 MeV (fixed)
Uncertainty in the dose ratio map	1-10% (in steps of 1%)
Required accuracy in WET	1-5 mm (in steps of 1 mm)

Table 4.1: Settings used in the theoretical analysis.

It is desirable to set the higher energy to always be the highest deliverable by the machine, so that a flat proximal dose contributes to the dose ratio. Most proton machines have a maximum deliverable energy of 230 MeV, so in this analysis dose ratio curves were formed using pairs where the maximum energy was fixed at 230 MeV. The lower energy was varied between 100 and 220 MeV, in steps of 10 MeV, producing 13 theoretical dose ratio curves, as shown in figure 4.3. For all these settings, summarised in table 4.1, proximal dynamic ranges were calculated for required accuracies of 1 to 5 mm. For comparison, the analysis was also repeated on the distal edge of the Bragg peak and distal dynamic ranges were calculated.

Most proton centres currently add a margin of between 2.5-3.5% of the proton range to account for range uncertainty (Paganetti [2012]), so a radiography technique that could predict the range to a higher accuracy than this would be an improvement on current practice. For a typical therapeutic proton beam with a range of 20 cm, accuracies we are investigating (1-5 mm) correspond to range uncertainties of 0.5-2.5%, which would offer an improvement in current proton beam range predictions.

4.2.2 Application to patient sites

The results of the theoretical analysis (of section 4.2.1) are two curves for each energy pair, in which the proximal dynamic range is calculated for a given required WET accuracy and for a given percentage error in the dose ratio. Using these curves, it is possible to illustrate the application of the technique for different patient sites.

For any patient site there will be a range of WETs. Due to the shape of the proton dose ratio curve (figure 4.3), the energy pair must be carefully selected so that the lower energy exits the patient and an image can be recorded. This lower energy must therefore have a range higher than that of the thickest WET in the patient site. Once the appropriate energy pair has been selected, the proximal dynamic range can be retrieved from the analysis of section 4.2.1, for different required accuracies in the WET and for different noise levels in the dose ratio image. The steps for determining the region that can be imaged is as follows:

1. Convert the X-ray CT dataset of each patient to RSPs using a calibration curve. For this we used the tissue-substitute calibration curve used clinically at MGH, which is formed of two straight lines (one with a gradient of 0.001 between HU=-1000 and 40, and one with a gradient of 0.0005 from HU=40 to 2995) and a correction above HU=2996 to account for titanium.

2. Parallel line ray tracing is then conducted through the volume using the open-source software, Plastimatch (`plastimatch.org`).

3. Compute the digitally reconstructed radiograph WET ($\mathrm{DRR_{WET}}$) by summing up the RSPs along the ray traces through the patient.

4. Calculate the maximum WET through the patient and convert to a minimum energy to the distal edge using the range-energy relation (we rounded these up to the nearest 10 MeV).

5. Using the theoretical results for the specific energy pair (analysis of section 4.2.1), determine the WETs in the patient site that lie within the proximal dynamic range.

All $\mathrm{DRR_{WET}}$s were produced with parallel ray tracing and with a pixel size of 1 mm, as shown in figure 4.4. The cropped field sizes are: prostate, $61 \times 101\,\mathrm{mm} = 6161\,\mathrm{mm}^2$; lung, white cropped field, $41 \times 46\,\mathrm{mm} = 1886\,\mathrm{mm}^2$; lung, green cropped field, $34 \times 35\,\mathrm{mm} = 1190\,\mathrm{mm}^2$. Details of the maximum WET in the image, the corresponding range and the lower energy selected are provided in table 4.2.

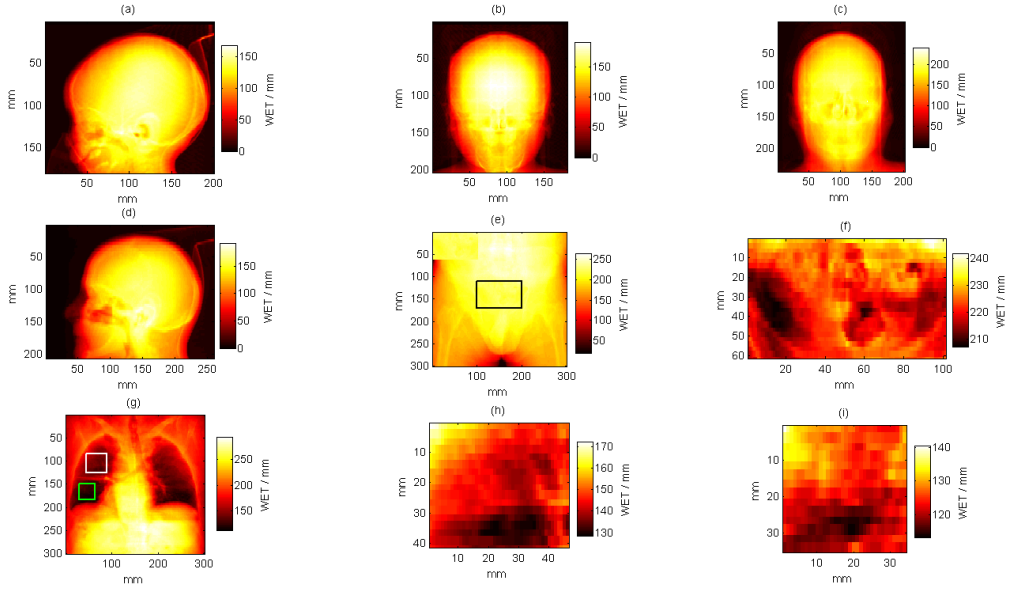


Figure 4.4: DRR_{WET}s of the different patient sites to be analysed: paediatric brain in (a) LR and (b) AP directions; adult brain in (c) AP and (d) LR directions; (e) full prostate field-of-view and (f) cropped region to analyse; (g) full lung field-of-view, with (h) white and (i) green cropped regions of interest.

Case	Field	Maximum WET [cm]	Minimum energy to distal edge [MeV]	Minimum energy rounded up to nearest 10 MeV
Pead. brain	AP	19.2	168.2	170
	LR	16.8	156.1	160
Adult brain	AP	22.0	181.9	190
	LR	19.2	168.5	170
Prostate	AP	24.2	191.9	200
Lung	AP_w	17.2	158.5	160
	AP_g	14.1	141.2	150

Table 4.2: Clinical sites investigated for applicability of dose ratio proton radiography. Abbreviations: AP = anterior-posterior; LR = left-right; the cropped fields for the lung, AP_w and AP_g are the white and green ROIs for the lung patient from figure 4.4.

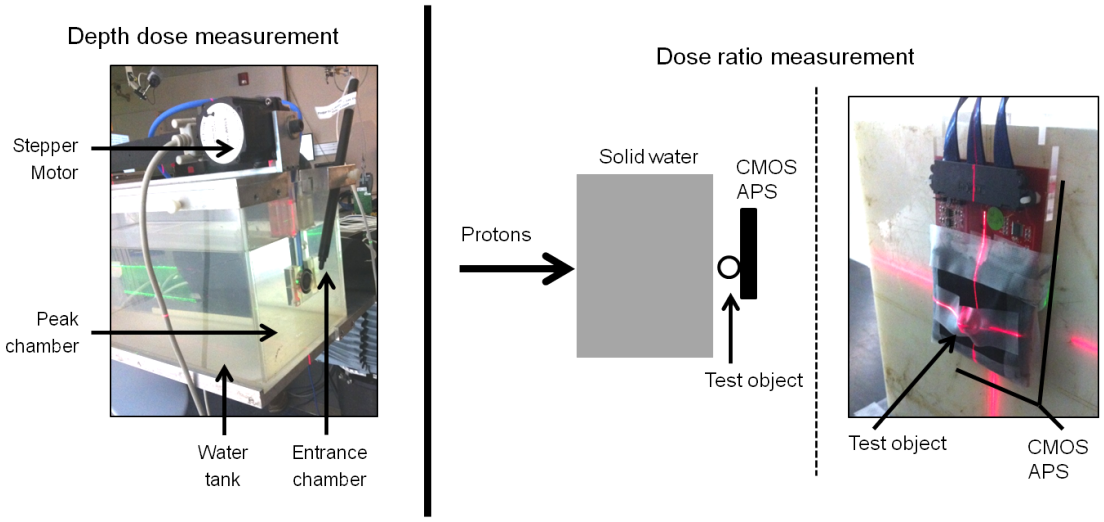


Figure 4.5: (Left) Experimental set up for depth dose measurements. (Right) Schematic and photograph of experimental set up for dose ratio measurements.

4.2.3 Experimental demonstration in a single-detector CMOS APS

The purpose of the experimental measurements was twofold: (1) to demonstrate that it is possible to acquire accurate images using the dose ratio proton radiographic method with a single CMOS APS; and (2) to show that the uncertainties in these measurements follow the theoretical predictions.

4.2.3.1 Experimental details

Measurements were made at the Francis H. Burr Proton Therapy Facility, MGH. To ensure the spot size was sufficiently large (to cover the whole sensor) the beam was used in single scattering mode. Depth dose curves for the two energies, with ranges (R90) of 17.7 and 22.5 cm, were acquired using a Markus chamber on a stepper motor in a water tank, together with a reference entrance chamber. This arrangement is shown in figure 4.5.

Dose ratio measurements were made using a CMOS APS, 'Vanilla'. CMOS APSs have shown promise in imaging for proton radiography (Poludniowski et al. [2014]; Seco and Depauw [2011]), with advantages such as low manufacturing costs, efficient charge collection, high speed imaging, high spatial resolution and region of interest (ROI) addressability. Full details of Vanilla and an assessment of its performance can be found in literature (Bohndiek et al. [2007]), but for our application the important features are: a sensitive area of 13 mm × 13 mm, comprised of an array of 520 × 520 pixels, each of 25μm pitch; and repetition rates between 1-1000 frames per second (fps) (dependent on the ROI addressed). In our measurements we imaged the full sensitive area at a frame rate of 44 fps.

4.2.3.2 Dose ratio measurements with a sapphire sphere

Due to our small sensor size we were limited in the test objects that could be used. A small red sapphire (Al_2O_3, doped with Chromium, Cr_2O_3) sphere with diameter 6.35 mm was imaged. As detailed in sections 4.2.1 and 4.2.2, there is a limited dynamic range using the dose ratio radiographic method, so to ensure we used the sensitive part of the Bragg peak we needed to place solid water slabs between the object and Vanilla, as shown in figure 4.5. The level of Cr_2O_3 doping was unknown for our sample so a percentage weight of 0.05% was assumed, in accordance with typical literature values for such synthetically-grown crystals (Basun et al. [2007]; Ganschow et al. [2011]). For such a composition, the effective atomic number Z_{eff} was calculated to be 11.40 (Boone and Chavez [1997]) and the relative electron density ρ_e was calculated to be 3.51 (Berger and Zucker [2004]). The RSP was estimated using the approximation of Schneider et al. [1996], equation 3.7, defined previously. The sapphire RSP was found to be 3.44, giving a maximum sphere WET of 21.8 mm (sphere diameter, 6.35 mm, multiplied by RSP, 3.44). As we wanted to demonstrate the use of the proximal side of the Bragg peak only, solid water was added before the phantom with a thickness of 14.27 cm WET. The total thickness (solid water + sphere diameter) was therefore then close to, but still on the proximal side of, the Bragg peak of the lower energy beam (17.7 cm).

4.2.3.3 Detector sensitivity calibration

The CMOS sensor is an optical sensor, not designed for the direct detection of charged particles, and as such, it is not radiation hard. Decreases in detector sensitivity have been observed with increasing dose in other CMOS sensors (Esposito et al. [2012, 2014]; Meroli et al. [2013]), and the effect was also evident in our detector. The grayscale of the output image of Vanilla is usually distributed over 4096 digital numbers, but this dynamic range decreases as the sensitivity decreases. As our detector could not be triggered to acquire when the beam switches on, however, the dose to the detector was not carefully controlled and it was impossible to apply a sensitivity correction. To resolve this problem the detector was calibrated before each acquisition, using a minimal dose (10 frames only) through a known material thickness.

4.2.3.4 Image processing

After averaging over 190 frames, images were processed in the same manner as the dose ratio curve was produced: (1) normalise the grayscale images for each energy; and (2) divide the lower energy by the higher energy to form a dose ratio image. Before converting the dose ratio image to a WET map we applied two corrections.

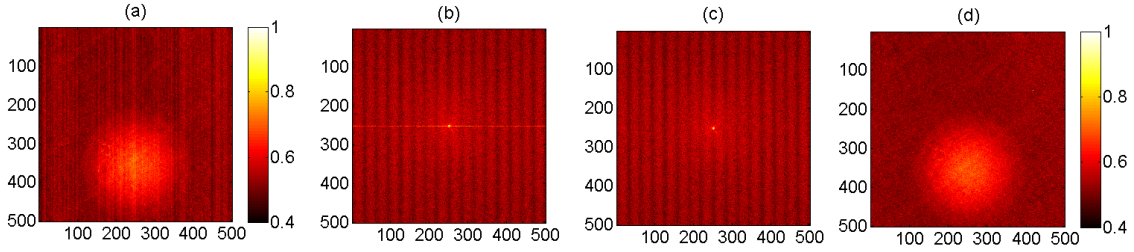


Figure 4.6: Fixed pattern noise removal. (a) Raw dose ratio image. The vertical stripes are caused by fixed pattern noise. (b) Fourier transform of the dose ratio image, before correction, and (c) after applying the interpolation notch reject filter. (d) Corrected, dose ratio image. In all figures the x- and y-axes are pixel numbers.

The first correction was to remove the fixed pattern noise, induced by the manner of the read out of the CMOS sensor. If the flux had been identical for the two energies, the fixed pattern noise would have been identical (Stewart [2013]) and the ratio of the images would have led to it cancelling out. As this was not the case, we needed to use an interpolation notch-reject filter to the affected region on the Fourier transform of the image (Konstantinidis et al. [2010]); the process of which is shown in figure 4.6. In the second correction a 3×3 median filter was passed over the image to remove the salt and pepper noise and the outer 20 pixels (in every direction) were cropped to remove discrepancies at the detector edge. The dose ratio image was then converted to WET using the measured dose ratio curve and finally the image was rebinned into 0.1 mm pixels.

4.3 Results

4.3.1 Theoretical limits

After following the steps in section 4.2.2, the user has the correct energy pair for the site to be imaged. An example dose ratio curve (with a lower energy of 170 MeV), with right and left error bars for 1% and 3% uncertainties in the dose ratio, are shown in figure 4.7. Plotting these error bars across the Bragg peak (also shown in figure 4.7) allows for calculation of the proximal dynamic range.

The proximal dynamic range is a function of the lower energy selected, the noise in the dose ratio map and the required accuracy in the WET. The theoretical relationship between all of these factors is shown in figure 4.8. It should be noted that, for some energy pairs, if the noise level is too high and the required accuracy in the WET is too strict, no part of the Bragg peak can be used for imaging (thus the lines in figure 4.8 are shortened).

For comparison the analysis was repeated on the distal side of the Bragg peak

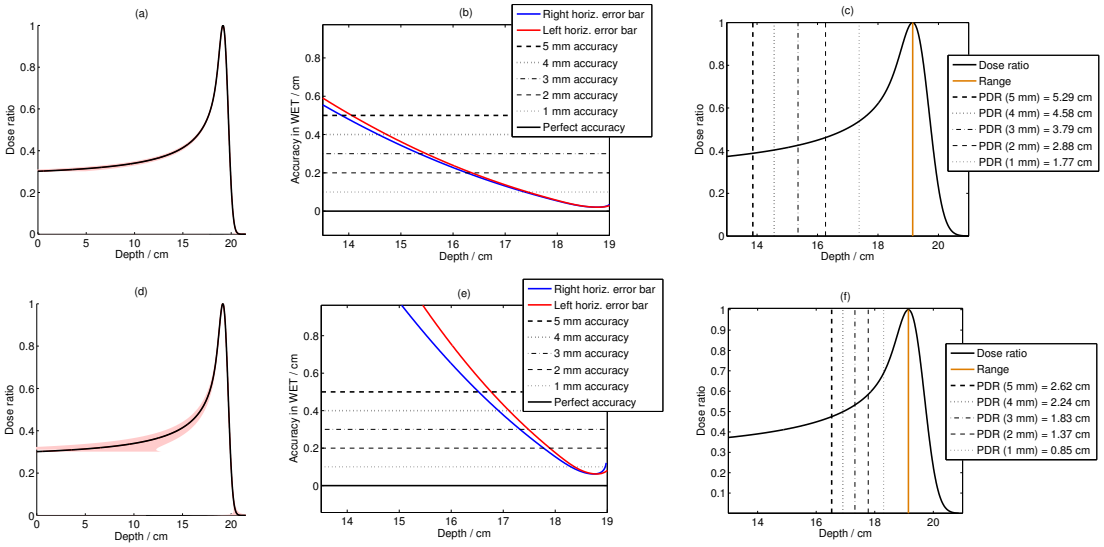


Figure 4.7: (a) Dose ratio curve for the 170/230 MeV energy pair, with horizontal error bars indicating the change in WET for a 1% uncertainty in the dose ratio. (b) Magnitude of error bars plotted as a function of the depth, with different accuracies indicated. (c) Proximal dynamic ranges, for different required accuracies, shown with respect to the original dose ratio curve. (d-f) Same as (a-c), but for a 3% uncertainty in the dose ratio.

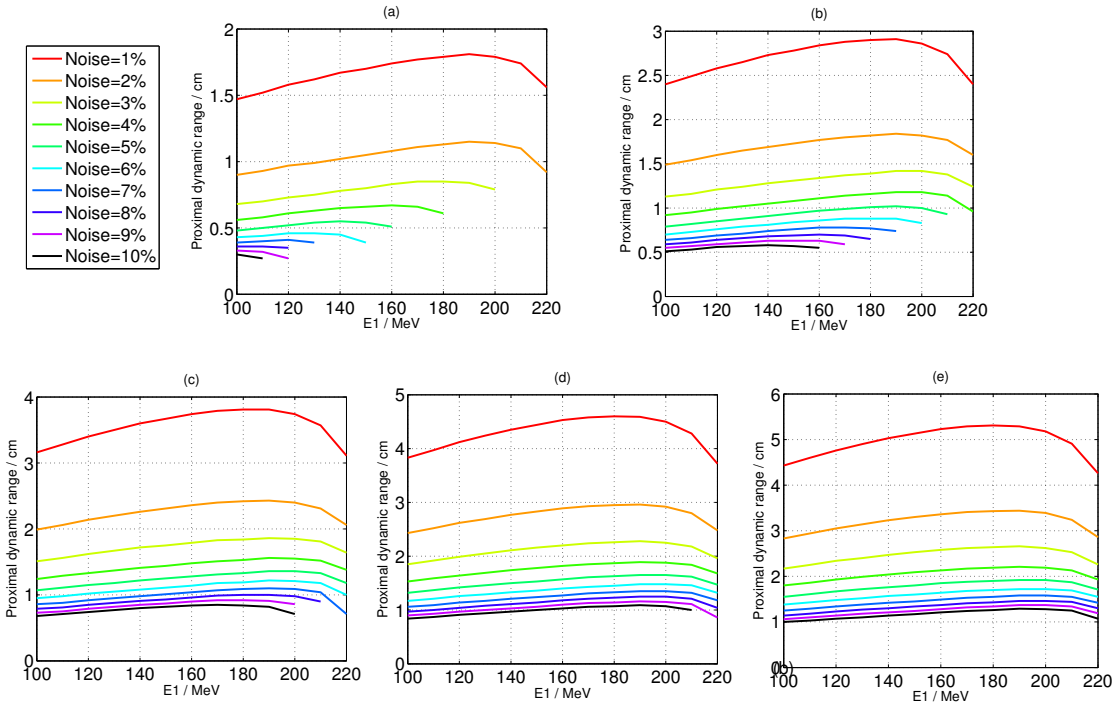


Figure 4.8: Proximal dynamic ranges for required accuracies in the WET of (a) 1 mm, (b) 2 mm, (c) 3 mm, (d) 4 mm and (e) 5 mm.

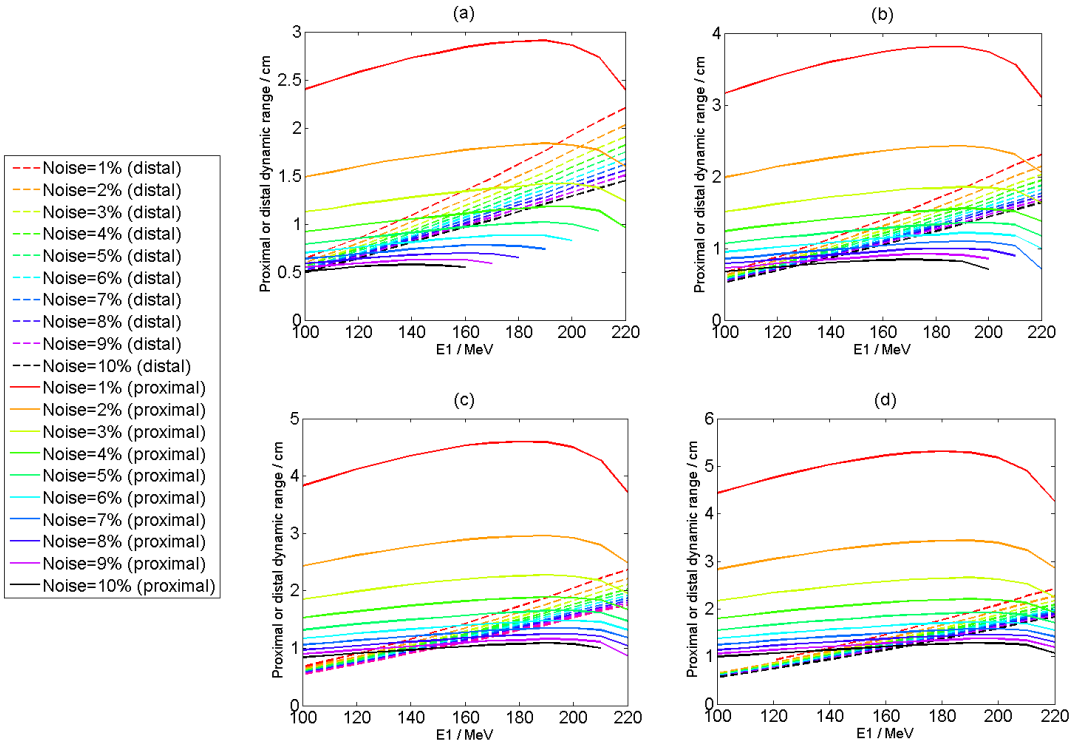


Figure 4.9: Comparison of proximal and distal dynamic ranges for required accuracies in the WET of (a) 2 mm, (b) 3 mm, (c) 4 mm and (d) 5 mm.

and distal dynamic ranges were calculated. These results, together with those on the proximal side of the Bragg peak, are shown in figure 4.9. Unlike the proximal side, the distal dynamic ranges are almost independent of the required accuracy in the WET.

4.3.2 Application to patient sites

The areas that can be imaged for the AP paediatric brain are shown in figure 4.10 as black regions superimposed on the original DRR_{WET}, for different noise levels in the dose ratio and for different required accuracies in the WET. Similar analyses were conducted for all the patient sites tested.

For noise levels 1-3% and for WET accuracies of 1-3 mm, the theoretically maximum size of the area that can be imaged for each of the clinical sites is detailed in table 3. For the lung and prostate cases, the region being analysed is already a cropped version of the full DRR_{WET}, so results are displayed in terms of percentages (of the cropped region size). Areas greater than an equivalent 50 mm × 50 mm field, or covering more than 75% of the cropped region, are shown in bold and underlined typefaces. These cases, considered to be clinically useful, are shown in figure 4.11 (the cropped regions are excluded from this figure).

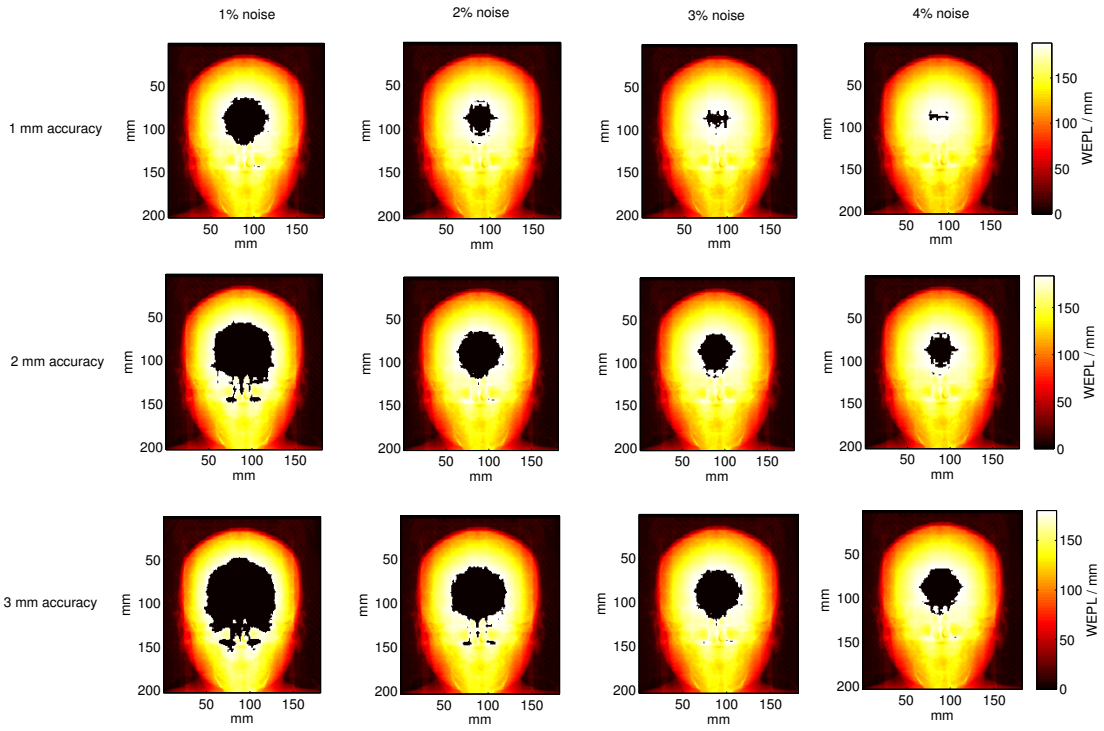


Figure 4.10: Areas that can be imaged (shown by dark region superimposed on brain) for an AP paediatric brain using dose ratio proton radiography.

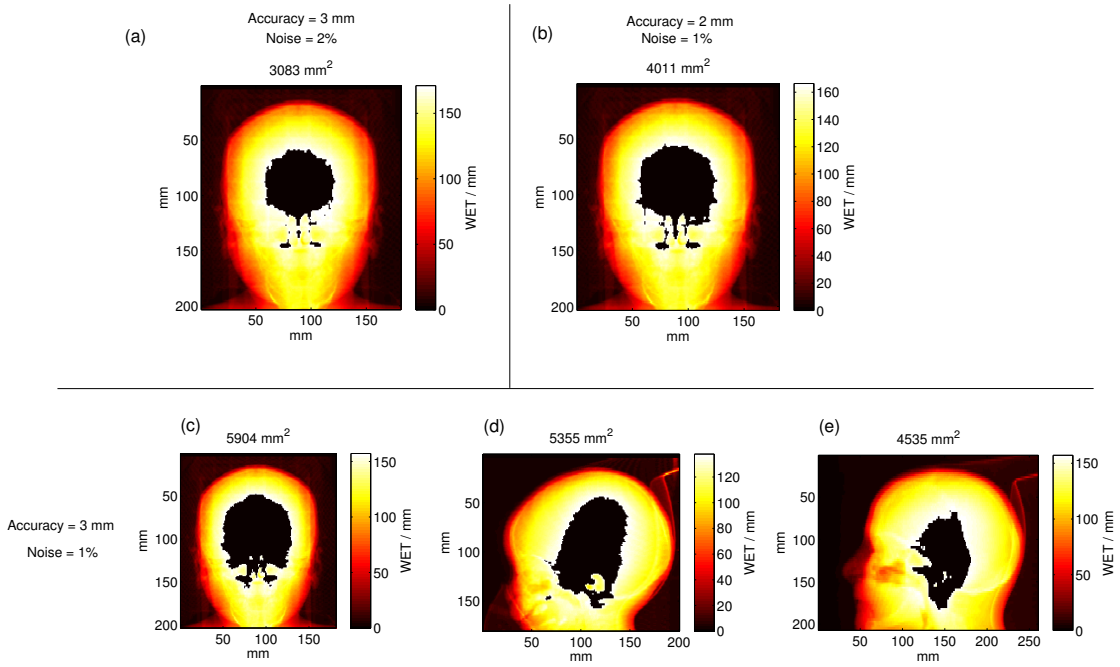


Figure 4.11: Clinical sites in which the usable area is greater than an equivalent $50 \times 50\,\mathrm{mm}^2$ field, with the areas that can be imaged shown by dark region superimposed on the $\mathrm{DRR}_{\mathrm{WET}}$. (a-c) AP paediatric brain; (d) LR paediatric brain; and (e) LR adult brain.

Case	Field	Accuracy in WET [mm]	Area 1% Noise	Area 2% Noise	Area 3% Noise
Paed. brain	AP	1	339 mm^2	820 mm^2	2016 mm^2
		2	**4011 mm^2**	2060 mm^2	1319 mm^2
		3	**5904 mm^2**	**3083 mm^2**	2122 mm^2
	LR	1	66 mm^2	4 mm^2	3 mm^2
		2	2305 mm^2	71 mm^2	9 mm^2
		3	**5355 mm^2**	614 mm^2	73 mm^2
Adult brain	AP	1	3 mm^2	3 mm^2	—
		2	116 mm^2	—	—
		3	1401 mm^2	13 mm^2	—
	LR	1	336 mm^2	63 mm^2	22 mm^2
		2	1683 mm^2	370 mm^2	134 mm^2
		3	**4535 mm^2**	1071 mm^2	396 mm^2
Prostate	AP	1	—	—	—
		2	3.4%	—	—
		3	44.2%	1.1%	0.1%
Lung	AP_w	5.5%	1.7%	0.9%	
		2	29.6%	5.5%	2.7%
		3	**78.8%**	14.5%	5.5%
	AP_g	1	0.2%	—	—
		2	18.2%	0.2%	—
		3	**77.6%**	2.9%	—

Table 4.3: Magnitude of areas, for a variety of clinical sites, that can be imaged using dose ratio proton radiography. Areas greater than an equivalent 50 mm × 50 mm field, or covering more than 75% of the cropped region, are shown in bold and underlined typefaces. Abbreviations: AP = anterior-posterior; LR = left-right; Small fields, AP_w and AP_w, are shown in figure 4.4.

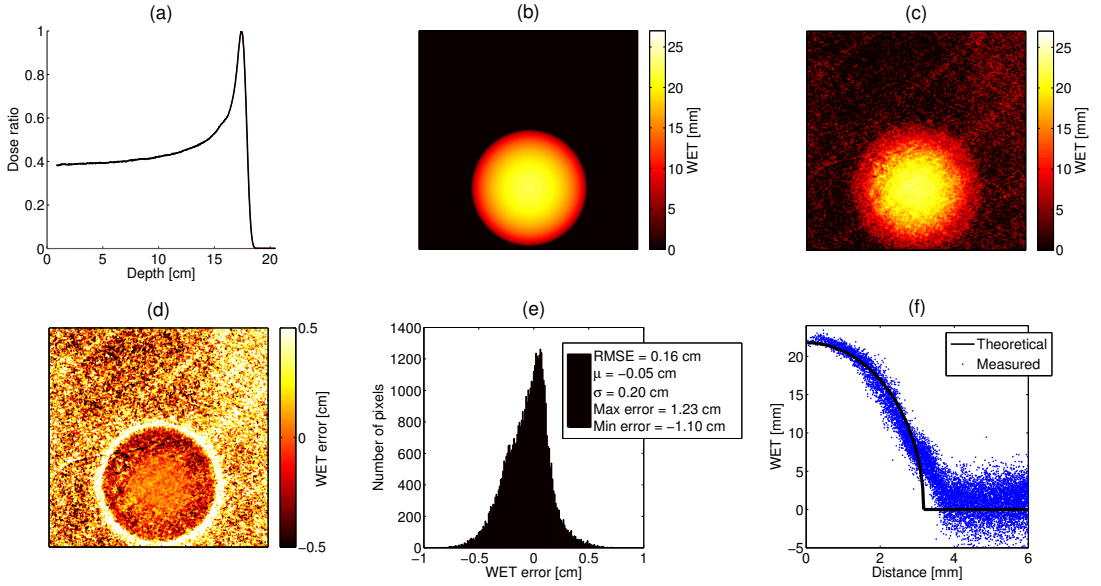


Figure 4.12: (a) Measured dose ratio curve. (b) Theoretical and (c) measured WET maps for the sapphire sphere. (d) Difference (measured-theoretical) WET map. (e) Difference histogram within the sphere radius. (f) Theoretical and measured radial profiles of the 0.1 mm pixel WET maps. The x-axis is the distance from the sphere centre.

4.3.3 Experimental demonstration in a single-detector CMOS APS

4.3.3.1 Accuracy and suitability of CMOS APS

The measured dose ratio curve is shown in figure 4.12(a). The theoretical and measured WET maps are shown in figures 4.12(b) and 4.12(c), respectively. The difference between the theoretical and measured WET maps for the sapphire sphere is shown in figure 4.12(d). The max/min/mean errors were found to be 1.23/-1.10/-0.05 cm. The RMSE and standard deviation were found to be 0.16 cm and 0.20 cm, respectively. In figure 4.12(f), the measured WET values as a function of distance from the centre of the sphere were determined.

4.3.4 Validation of theoretical predictions

The calibration dose ratio image (acquired using a single solid water thickness, with no sphere), after rebinning into 0.1 mm pixels, was found to have a standard deviation, $\sigma = 1.2\%$. It can be seen in figure 4.12(e) that the error in the WET follows an approximately Gaussian shape. If we set the uncertainty in the dose ratio to be 1.96σ, or 2.35%, we would expect 95% of our results to fall within the confidence intervals of our theoretical predictions. To test this, we determined the theoretical limits using the same procedure as in section 4.2.1, for the two energies used in the measurements (160.8 and 184.2 MeV). As a reminder, these steps were:

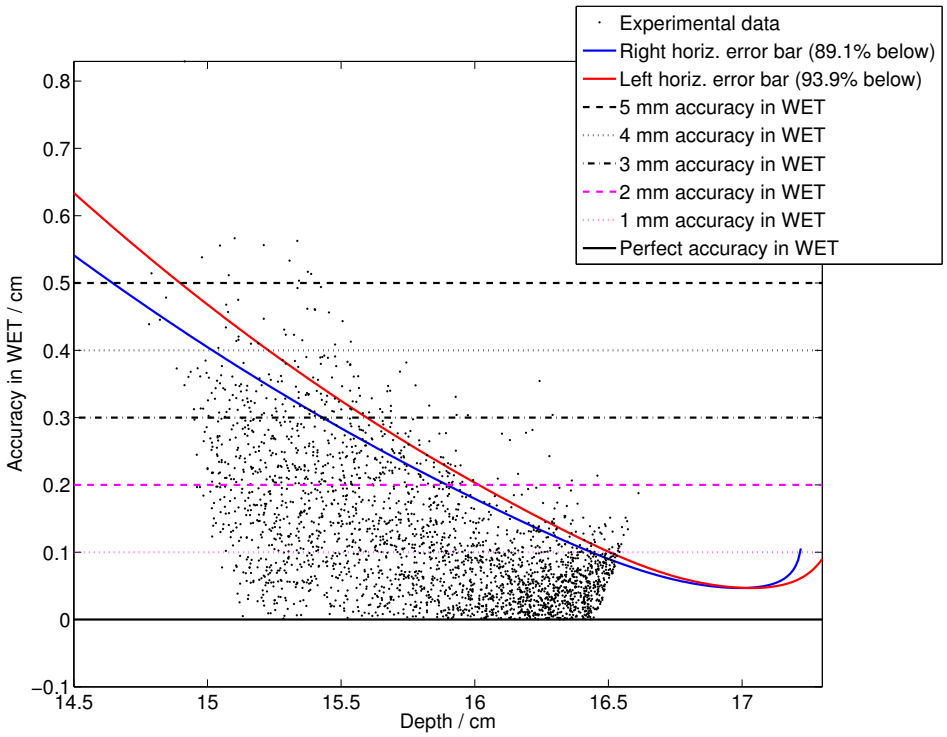


Figure 4.13: Theoretical limits (red and blue lines) for a 2.35% (1.96σ) noise in the dose ratio, for the energies used in measurements. The measured data of each 0.1 mm pixel are shown by the black dots. Given the measured data follows an approximately Gaussian distribution, we would expect 95% of the pixels to fall below the theoretical limits.

(1) determine the spread in WET values for a given change in the dose ratio (here 2.35%); and (2) add right and left horizontal error bars to the dose ratio curve for that particular dose ratio value. The theoretical error bars for a change of the dose ratio of 2.35%, together with the absolute measured WET errors, are shown in figure 4.13. The experimental measurements fall below the theoretical right/left horizontal error bars in 89.1/93.9% of cases, respectively.

4.4 Discussion

4.4.1 Theoretical limits

Using the analytical approximation for the Bragg peak, we were able to define theoretical limits for the dose ratio method on the proximal side of the Bragg peak. Horizontal right and left error bars were added to theoretical dose ratio curves and used to determine the proximal dynamic range for varying energy pairs and different levels of noise in the dose ratio map. A requirement from the user is knowledge of the expected uncertainty in their dose ratio map (defined in our study to be dose independent), which is a quadratic sum of the uncertainties in the images from each

energy. For a given required WET accuracy, the user can then look up the proximal dynamic range for their energy pair (figure 4.8) and set up their measurement so that the object/region to be imaged has the correct total WET. Provided the object has a WET falling within this proximal dynamic range, as defined in figure 4.2, the accuracy remains within the required level.

It was found that the maximum proximal dynamic range for a given noise level was not obtained by the largest separation between the energy pair; rather there is an optimum value (typically $\approx 190\,\text{MeV}$). The reason for this can be explained with reference to figure 4.3. Although the proximal value of the dose ratio decreases with a greater separation in energies, at lower energies there is less range straggling and so the Bragg peak is narrower, shortening the proximal dynamic range. The optimum value arises because, although it is beneficial to have a lower value for the dose ratio in the plateau region, some Bragg peak broadening is also desirable.

For comparison, the analysis was repeated on the distal side of the Bragg peak and distal dynamic ranges were calculated. Across all settings, the largest distal dynamic range was 2.42 cm (lower energy of 220 MeV, 5 mm required accuracy in the WET and 1% noise level). Comparatively, with the same required accuracy and noise level, a proximal dynamic range of 5.31 cm could be utilised. It can be seen from figure 4.9 that the use of the proximal side of the Bragg peak allows for the use of a larger dynamic range, but only for low noise levels (1-2%). When the noise levels become too high, the use of the distal side of the Bragg peak becomes more reliable and larger dynamic ranges can generally be utilised.

4.4.2 Application to patient sites

As mentioned in section 4.2.1, most proton centres currently employ a margin of between 2.5-3.5% of the proton range to account for range uncertainty (Paganetti [2012]). Therefore, proton radiography would offer an improvement on current clinical practice if it could predict the proton range with an error of less than 2.5%, or 5 mm for a 20 cm range proton beam. Given the investment (both time and money) required to start using proton radiography routinely, the technique should offer a significant improvement on current clinical practice. We consider that a 1% accuracy, or 2 mm of a 20 cm range beam, would be a sufficiently high improvement to enable its introduction.

For such an accuracy, for a noise level in the dose ratio map of 1%, dose ratio radiography on the proximal side of Bragg peak has a proximal dynamic range of up to 2.91 cm (dependent on the lower energy). This is quite a small range and it was found the technique can only feasibly be applied to a specific set of patient sites and within each site only a specific region can be imaged. The following sites

were tested for feasibility: brain (paediatric and adult, AP and LR); prostate (AP only); and two small regions in the lung (AP only). The brain cases (in particular the paediatric brain) demonstrated the best possible application of the technique. Provided the noise in the dose ratio map is no more than 1%, equivalent square field sizes in the paediatric brain of $6.3/7.6\,\mathrm{cm}^2$ could be imaged under required accuracies of $2/3\,\mathrm{mm}$, respectively. It should be commented, however, that the shapes of these areas were not always connected (figure 4.11).

Although not explored in this work, it would be possible to increase the region that can be imaged by using additional energy pairs, although this has the drawback of increasing the patient dose. Another possibility is the use of a proton beam with a higher maximum energy. All of our analysis was based on using a fixed higher energy of 230 MeV, but some proton systems can generate energies up to 330 MeV. This would reduce the contribution of the higher beam to the dose ratio, which may allow for a larger dynamic range for imaging.

One disadvantage of the technique is the requirement of prior knowledge of the approximate WET of the patient site, so that the appropriate energy pair for imaging can be selected. The suggested approach for this is to convert the patient X-ray CT dataset into a dataset of RSPs, and generate a $\mathrm{DRR}_{\mathrm{WET}}$ that can then be used as an indication to select the best energy pair.

4.4.3 Experimental demonstration in a single-detector CMOS APS

Our experimental work demonstrated the use of the dose ratio method on the proximal side of the Bragg peak. We imaged a small sapphire sphere in a high energy proton beam line using a small ($13\,\mathrm{mm} \times 13\,\mathrm{mm}$) CMOS APS sensor. Using two Bragg peaks, of 160.8 and 184.2 MeV, a dose ratio curve was formed and, after some image processing, a WET map was generated.

4.4.3.1 Accuracy and suitability of CMOS APS

After determining the precise sphere location, radial profiles of the measured data were compared to the theoretical WET radial profile, showing excellent agreement (RMSE = 1.6 mm). This accuracy is not much worse than previous works in which the distal edge of the Bragg peak was used (WET accuracy up to 0.9 mm in Bentefour et al. [2012b]). It can be seen in figures 4.12(d) and 4.12(f) that the measurements overestimate the sphere WET at the wider radii. This is likely to be caused by MCS, which has the effect of scattering the protons outwards from the sphere centre (where the WET is larger). This effect would have less of an impact if larger objects

were imaged (and a larger CMOS sensor could be used to image them). Despite this, the CMOS APS demonstrated good suitability as a detector for dose ratio proton radiographic imaging, with a WET mean error of 0.5 mm.

As discussed above in section 4.4.2, to be used clinically, the noise level in the dose ratio map generated needs to be not higher than 1%. If it exceeds this level, the region that can be imaged (for the clinical sites assessed) is too small to prove useful. In this regard, the CMOS APS, with a 1.96σ noise level of 2.35% (see section 4.4.3.2 below), is probably too noisy.

Another problem of the CMOS APS used was the fact that it is not radiation hard. To be able to make reliable measurements, the sensitivity of the detector had to be characterised with a calibration image prior to each measurement. Although inconvenient, detector calibration would be required to account for other detector factors such as temperature and day-to-day fluctuations anyway. Provided the measurement is quick and automated, we do not consider this to be a big problem. Alternatively, radiation-hard detectors could be used or careful measurements of the sensitivity response of the detector with dose could be made.

4.4.3.2 Validation of theoretical predictions

The noise level in our dose ratio map was estimated from a calibration image, in which a single solid water thickness was imaged, to be $\sigma = 1.2\%$. Following the same procedure as the theoretical analysis of section 4.2.1, theoretical error bars could be determined for the energy pair and for a noise level of 1.96σ. It was found that the measured error in the WET agreed with the theoretical predictions well, falling under the horizontal right/left errors bars 89.1/93.9% of the time (figure 4.13). These values are in good agreement with the prediction that a noise level of 1.96σ should encompass 95% of the measured data points, given that the measured errors follow an approximately Gaussian shape. It can be concluded that the method of analysis to generate the theoretical predictions, in section 4.2.1 and 4.2.2, is appropriate.

4.4.4 Study limitations

One of the limitations of our theoretical application to patient sites is that the WET maps were generated using straight line DRRs, but it is well known that protons take longer paths due to the effect of MCS. This would have the effect of spreading the range of WETs in the patient, thus decreasing the region of the patient we could image. A potential solution to this, as suggested in section 4.4.2, would be to use multiple energy pairs, with each pair designed to image different regions.

Our theoretical analysis was verified using a CMOS APS. However, because of the small detector size, 13 mm × 13 mm square, we were restricted to imaging objects

with a small maximum WET (a sapphire sphere with a maximum WET of 21.8 mm was imaged). It is expected that the verification would still hold at high WET, but this requires confirmation, particularly as the WETs used in the theoretical analysis were much higher (maximum 242 mm).

4.4.5 Outlook of the dose ratio technique

Based on this theoretical analysis, the potential use of this technique appears quite limited. Only in the imaging of sites with a small maximum WET, namely paediatrics, it is possible to image reasonably-sized fields (greater than an equivalent 5 cm × 5 cm square). This also requires the accuracy in the WET to be no stricter than 2 mm, and the user to have a detector with a noise level of no higher than 1%. In addition to the limited dynamic range of this technique, it requires the measurement of dose from two pristine Bragg peaks, and consequently twice the amount of dose with respect to any radiographic techniques that use only one. If, as has been suggested previously, the dynamic range were to be increased using additional energy pairs, the dose would be increased further.

Despite these limitations, the technique does have some potential. Firstly, it only requires the use of a single detector, which, as discussed in the introduction, can be more conveniently introduced into the clinical environment than classical proton radiographic devices. Additionally, it provides a potential solution for PBS, which appears to increasingly become the standard mode of delivery for proton therapy. Dose was not explored in this work, but given that relative images are required, it may be possible to deliver a very small dose for each energy (provided the signal-to-noise ratio is not too low).

4.5 Conclusions

The dose ratio method has been systematically assessed using a set of theoretical pristine Bragg peaks. Look-up graphs that list the range that one can use on the proximal side of the Bragg peak, dependent on the energy pair, noise level in the dose ratio image and the required accuracy in the WET, were generated. These theoretical predictions were validated with experimental measurements using a CMOS APS, which was found to have a noise of $\sigma = 1.2\%$. The CMOS APS proved to be a sufficiently accurate detector for this technique, with a RMSE of 1.6 mm WET in the imaging of a small sapphire sphere.

Provided the noise level is less than 2%, it was shown that the proximal side of the Bragg peak has a larger dynamic range (maximum 5.31 cm) than when using the distal side of the Bragg peak (maximum 2.42 cm). The dose ratio on the proximal

side of the Bragg peak has some potential as a proton radiographic method, although the technique can only be applied to clinical sites with small maximum WETs such as for paediatric brains. For an example paediatric brain it is possible to use the technique to image a region with a square field equivalent size of 7.6 cm^2, for a 3 mm accuracy in the WET and 1% noise level in the dose ratio image.

Chapter 5

Patient-specific calibration using proton radiography

Work in this chapter can be found in the following journal article and was presented at the following meetings:

- (Under review) **Doolan P**, Testa M, Sharp G, Bentefour E H, Royle G, Lu H-M **2014** Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography *Physics in Medicine and Biology*

- **Doolan P**, Sharp G, Testa M, Bentefour E H, Royle G, Lu H-M **2014** Reducing proton beam range uncertainty with patient-specific CT HU to RSP calibrations based on single-detector proton radiography *American Association of Physics in Medicine 56th Meeting*, Austin, TX, USA (Oral presentation)

- **Doolan P**, Sharp G, Testa M, Bentefour E H, Royle G, Lu H-M **2014** An optimization scheme to produce patient-specific calibration curves for proton therapy *Particle Therapy Co-Operative Group (PTCOG) 53rd Meeting*, Shanghai, China (Oral presentation)

- **Doolan P**, Sharp G, Testa M, Bentefour E, Royle G, Lu H-M **2014** A method to produce patient-specific calibration curves for proton therapy using proton radiography *NEAAPM Young Investigators' Symposium*, Boston, MA, USA (Oral presentation)

- Testa M, **Doolan P**, Bentefour E H, Paganetti H, Lu H-M **2014** Experimental investigation of proton radiography based on time-resolved dose measurements *American Association of Physics in Medicine 56th Meeting*, Austin, TX, USA (Poster)

- Testa M, **Doolan P**, Bentefour E H, Paganetti H, Lu H-M **2014** Proton radiography based on time-resolved dose measurements *Particle Therapy Co-Operative Group (PTCOG) 53rd Meeting*, Shanghai, China (Oral presentation)

As can be seen, there were many contributors to the work in this chapter. Practical radiographic measurements were made at MGH with M Testa, E-H Bentefour and H-M Lu, while some CT scans were conducted with G Sharp. M Testa developed the code to analyse the radiographic data. G Sharp edited Plastimatch to allow its use in our project and helped with the deformable registration of the real tissue phantom. All other work and analysis was conducted by myself.

5.1 Background

5.1.1 Improving the calibration curve with proton radiography

As discussed in detail in chapter 3, the final step of the stoichiometric calibration procedure involves calculation of the theoretical RSPs of human biological tissues, using composition data for average, healthy, adults (ICRU [1989]; White et al. [1987]; Woodard and White [1986]). 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.

One solution to remove the reliance on literature tissue compositions, and the uncertainties associated with patient-to-patient variations, is to generate an RSP map directly. Proton CT would be able to generate such a 3D map, but the technology is not currently available to acquire this for the size of a patient. Proton radiography is much more readily available but, due to MCS, it lacks the required spatial resolution for treatment planning.

However, it has been shown in one previous work by Schneider et al. [2005] (we are not aware of any other publications) that the RSP information from the proton radiograph can be combined with the high resolution spatial information from the X-ray CT to optimise the HU-RSP calibration curve. In their work, a digitally reconstructed radiography (DRR) through the X-ray CT dataset (in the same direction as the proton radiograph) was converted to a 'DRR_{WEPL}' map. This involved conversion of the HU values to RSPs using a calibration curve, followed by summation of the RSPs along the ray trace. Starting with a curve produced using the stoichiometric method (section 3.1.1.1), portions of the curve across a

small CT number range (800-1135 HU) were randomly varied 20,000 times. Chi-squared differences between the DRR and proton radiograph were determined until a minimum solution was found. The resultant calibration curve is then an optimised one, specific to that patient.

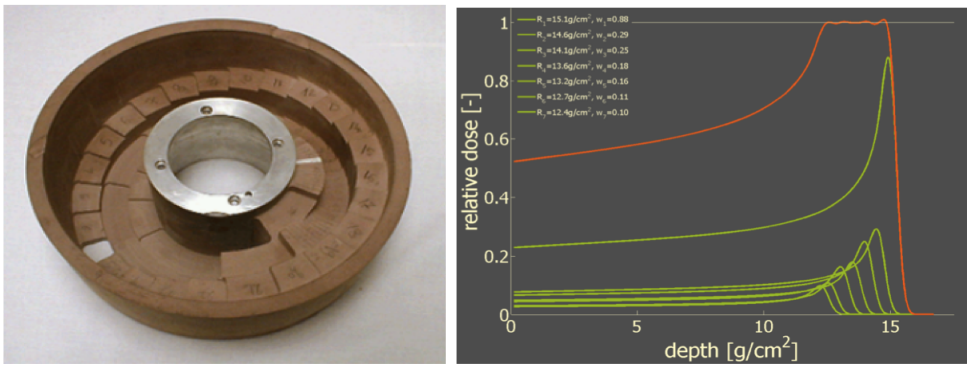
The technique showed promise, with a reduction in the mean difference between the DRR_{WEPL} and the proton radiograph from 3.6 mm to 0.4 mm following optimisation. However, upon speaking to the corresponding author of the work, the procedure was not taken up clinically for a number of reasons. (i) The speed of the electronic read-out was limited because each proton had to be detected in coincidence. (ii) It was difficult to include an imaging device into a control system designed for radiotherapy, because the dose rates used were much lower. (iii) Also, the radiographic system was too bulky. In our work we propose the use of a detector that addresses all three of these issues, by (i) measuring a bulk dose over 10 cycles of the range modulator wheel, so that the total acquisition time is 1 second (ii) at a clinical dose rate (iii) using a single detector behind the patient.

5.1.2 The 'time-resolved' principle

It was first suggested by Lu [2008b] that range verification measurements could be made using the time dependence of the dose distribution delivered by range-modulated passively scattered fields. As detailed in section 2.1.3, passively scattered beams consist of a beam of fixed energy intercepted by various beam modifiers. A range modulator (RM) wheel, an example of which is shown in figure 5.1(a), is usually used for this task. It is composed of a series of stepwise segments of absorber material with increasing thickness. When a mono-energetic proton beam passes a particular step, its range is reduced, or pulled back, by the WET of the absorber. Rotation of the RM wheel, at a typical 600 revolutions per minute, results in the beam passing through the wheel one segment at a time along the track, delivering a series of Bragg peaks spread out in depth. The RM wheel rotation frequency, step sizes and thicknesses and the beam current are all carefully controlled so that the beams of different energies combine to form a spread-out Bragg peak (SOBP), as shown in figure 5.1(b).

At different depths in the target volume, the dose received is dependent on the pattern of the range modulator wheel and the incoming energy. Lu [2008b] suggested that by measuring the time dependence of the dose at any point in the target volume, the residual range of the beam at that point could be determined. Provided the initial energy is known, the integrated energy loss, and therefore the WEPL of materials traversed and the approximate depth in tissue, could be estimated.

The dose rate pattern as a function of time, shown in figure 5.2, varies consid-



(a) (b)

Figure 5.1: (a) Range modulator wheel (courtesy of Ion Beam Applications, Louvain-La-Neuve, Belgium). (b) Demonstration of how individual pristine Bragg peaks of differing energy and weight (green) can be combined to form a spread-out Bragg peak (orange). Reproduced from Slopsema [2009].

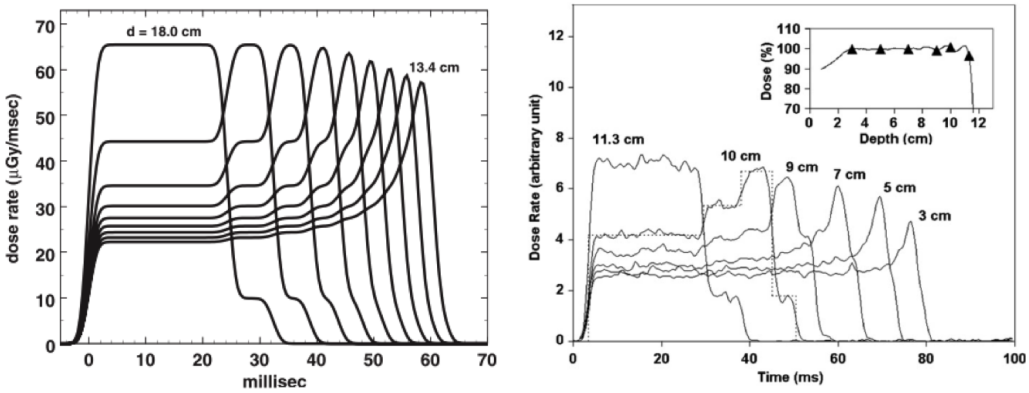
erably and uniquely with depth. Differences between theoretical (figure 5.2(a)) and measured patterns (figure 5.2(b)) arise because of range straggling, range mixing caused by the beam spot size being larger than some of the step sizes, and also beam current modulation in some systems.

To determine the water-equivalent depth it is necessary to first acquire a library of patterns in water. The measured dose rate pattern can then be compared to the library using a least squares minimisation. It has been shown that such an approach can achieve an accuracy in the WEPL (water-equivalent path length) of $< 1\,\text{mm}$ using ionisation chambers, provided the depths being measured are within the SOBP region (Lu [2008b]).

Alternatively, a calibration curve can be formed based on statistical measures of the dose rate patterns in the library. Gottschalk et al. [2011] showed that using semiconductor diodes it was possible to achieve a 0.3 mm accuracy using the inverse of the time-width of each pulse, over 10 rotations of the RM wheel. Further details of this technique, used in our study, are provided in section 5.2.3.

5.1.2.1 Applications

So far there have been numerous suggested uses of the method. One use is as an *in vivo* pre-treatment proton range check for prostate treatments, the most common malignancy in elderly males. Treatments are generally formed of two laterally opposed proton beams, but a higher conformality (the lateral penumbra are not as sharp as the distal Bragg peak) and sparing of the concavely-shaped anterior rectal wall could be achieved using anterior beams (either directly anterior or slightly oblique, $\pm 30°$) (Tang et al. [2010, 2011]). Such a technique is risky because of the high dose deposition at the end of Bragg peak. Additionally, these proton beams



(a) Theoretical (b) Experimental

Figure 5.2: Dose rate patterns at a series of depths in a SOBP. Images reproduced from Gottschalk et al. [2011] (a) and Lu [2008b] (b).

must pass through tissue whose WEPL can vary considerably (e.g. variable bladder filling, Andreo [2009]). It has been suggested that detectors attached to endo-rectal balloons could provide a measure of the proton range of anterior beams below the prostate based on this time-resolved technique (Bentefour et al. [2010, 2009]; Samuel et al. [2013]). Once the proton range has been verified pre-treatment, the range required for treatment on the particular day could be tuned to avoid giving dose to the rectum. Additionally, these detectors could provide a backup dose measurement during treatment.

It has also been suggested that the technique could be used as a crude imaging method (Lu [2008b]). Recently this has been demonstrated on simple geometric objects and a human skull by Testa et al. [2013]. Spatial resolution was limited by the detector used, a 2D diode array (Sun Nuclear Corporation - Melbourne, Florida, USA) with 249 semiconductor diodes, to the smallest linear pitch between the diodes (7.07 mm), but the results were promising. The same equipment is used in this work.

5.1.2.2 Range mixing

As detailed in section 4.1, using only a single detector positioned beyond the patient is a more convenient way to introduce proton radiography into the clinical environment than the classical approach, which involves trackers before and after patient and some energy measurement device. However, while convenient, not tracking protons means the individual paths cannot be reconstructed. 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 monoenergetic) and it is possible to define the WEPL. If, however, protons can get to

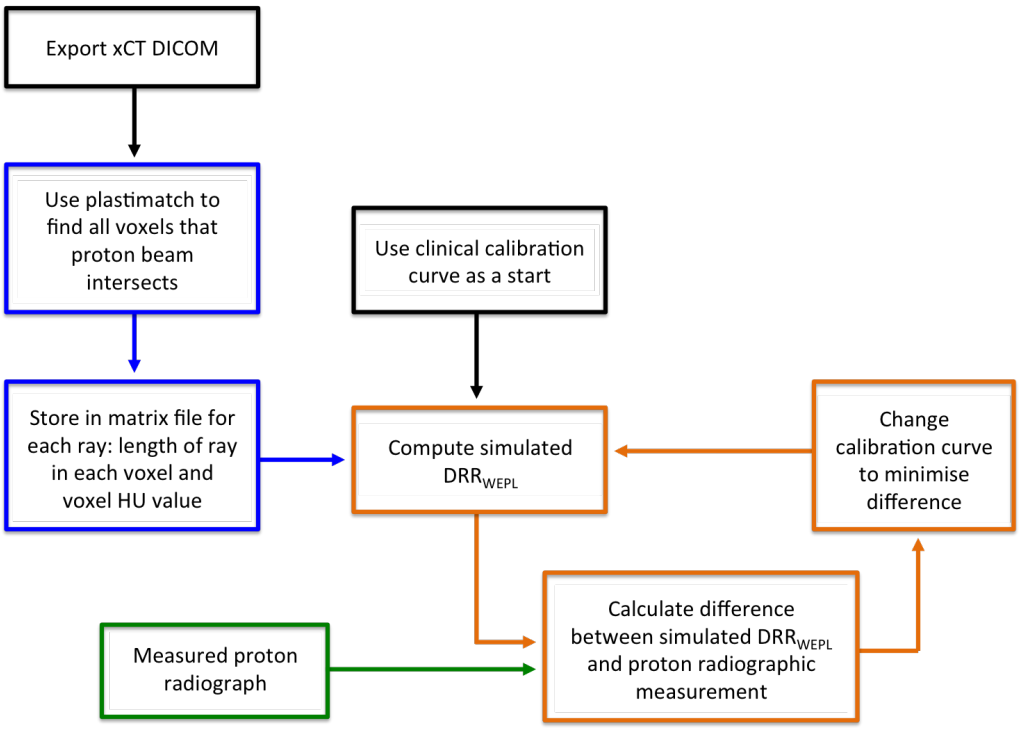


Figure 5.3: Schematic of the DRR optimisation procedure. Plastimatch operations are shown in blue, measurements in green and optimisation processes in orange. xCT = X-ray CT.

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. Methods therefore have to be introduced to detect and handle this effect, known as 'range mixing' (Bentefour et al. [2012a]). In our work, parts of the image that are subject to range mixing needed to be identified as the WEPL values are not reliable and should not be used as part of the optimisation.

5.2 Materials and methods

5.2.1 Optimisation scheme

The modular structure of the DRR optimisation procedure is shown schematically in figure 5.3. DRRs are created using the Siddon [1985] transform implemented in Plastimatch (`plastimatch.org`) and the length of intersection of each ray trace with every voxel, together with the voxel HU value, is stored in a matrix file for future computations. The HU values are then converted into RSPs using the generic 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:

$$w_c^i = \sum_j l_j F(H_j) \tag{5.1}$$

where l_j is the length of intersection and H_j the HU value of voxel j; and $F(H_j)$ is a function that converts the HU values to RSPs. As described previously, we call the WEPL map produced in this manner the $\mathrm{DRR_{WEPL}}$. The number of ray traces matches the resolution required in the $\mathrm{DRR_{WEPL}}$ map. In equation 5.1, F is a function that describes the calibration curve and is typically a single piecewise linear function F,

$$F = \begin{cases} 0 & \text{if } H \leq -1000; \\ a + bH & \text{if } -1000 < H \leq 0; \\ a' + b'H & \text{if } 0 < H \leq 40; \\ \vdots & \vdots \end{cases} \tag{5.2}$$

The division of the linear sections is a choice of the user. In this work all results are compared to a 'generic' calibration curve, 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 HU for titanium). For optimisation, this curve was split into 25 linear sections or 'material bins', based on the material binning suggested for Monte Carlo simulations by Schneider et al. [2000], with an additional bin edge at the point of water. The 26 bin edges 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 optimisation procedure as this would ordinarily be overridden as part of treatment planning.

For the optimisation, the parameters $a, b, a', b' \ldots$ in equation 5.2 are varied until the difference between the $\mathrm{DRR_{WEPL}}$ map and the proton radiography WEPL image reaches a minimum. The cost function describing this difference Δ, which must be minimised, was thus defined:

$$\Delta_i = \frac{(w_i^m - w_i^c)^2}{(w_i^m)^2}, \tag{5.3}$$

where w_i^m is the proton radiography (measured) value of the $\mathrm{i^{th}}$ pixel and w_i^c is the $\mathrm{DRR_{WEPL}}$ (calculated) value of the $\mathrm{i^{th}}$ pixel.

The optimisation was performed using Matlab's built-in Nelder-Mead optimisation function, 'fminsearch'. In addition to the convenience of a Matlab built-in function, Nelder-Mead optimisations are more robust to local minima and do not require an equation to be provided for the derivative of the cost function (Lagarias

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

Table 5.1: Input variables for the optimisation. n is the number of parameters being optimised (27 in our case).

et al. [1998]). Details of the input variables used in optimisation scheme can be found in table 5.1. When optimising, 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 through the points of air (HU = -1000, RSP = 0.001) and water (HU = 0, RSP = 1); by introducing a high cost penalty in the optimisation 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 $100 * n$ iterations.

5.2.1.1 Plastimatch validation

Plastimatch was utilised for computation of the DRRs, but it was important to ensure that using this does not introduce any additional errors. Thus a multiple material, uniform shape, phantom (figure 5.4(b)) was used for this purpose. To validate Plastimatch a known WEPL map was produced using the theoretical RSPs for the 28 tissues (based on the literature tissue compositions from ICRU [1989]; White et al. [1987]; Woodard and White [1986]) multiplied by the geometric thickness of the phantom (5 cm). This was compared with a $\mathrm{DRR}_{\mathrm{WEPL}}$ map, created by converting the CT values in each voxel to their corresponding RSP values (based on literature), summing up the (length of intersection) $\times$ (value of voxel) of each ray and summing up across all rays. Two $\mathrm{DRR}_{\mathrm{WEPL}}$ maps were created, with an SAD consistent with that of the proton gantry (2.27 m) and with an SAD of 1000 m. The latter was tested to remove divergence in the ray tracing, which may impact on the final result.

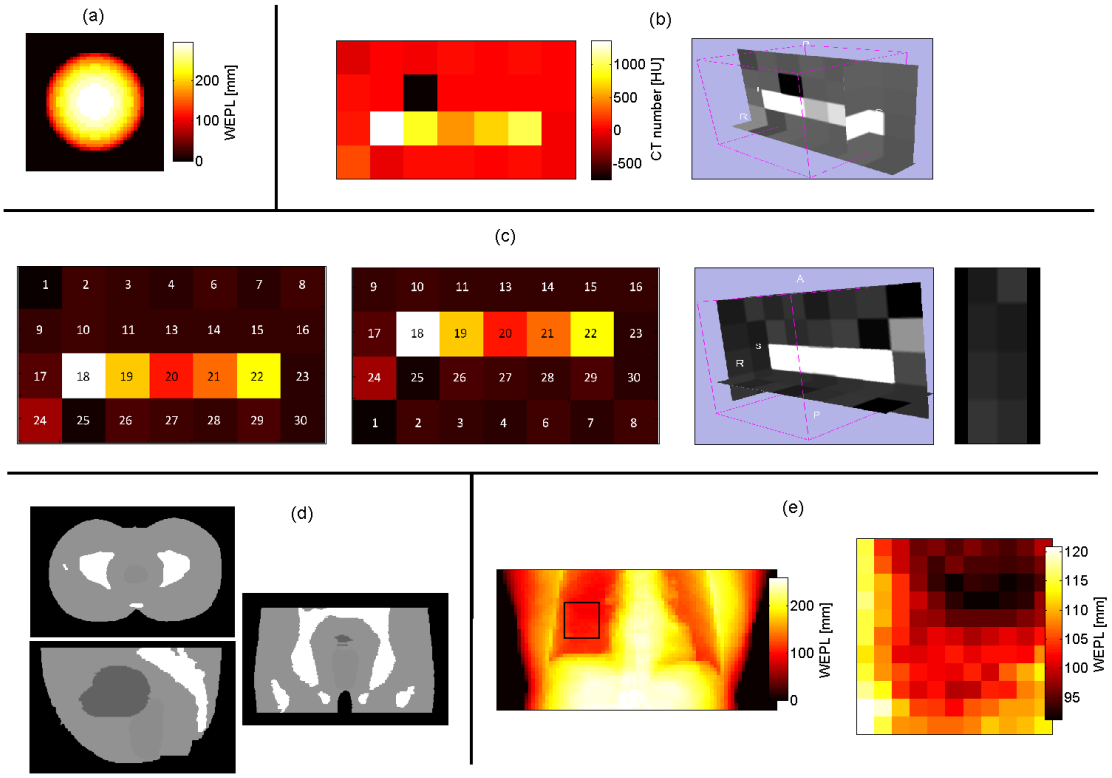


Figure 5.4: Synthetic phantoms used to validate the optimiser. (a) Single, homogeneous material (example shown here is a sphere of breast material) in a 15 cm radius spherical shape. (b) Multiple material phantom in (left) 2D view and (right) 3D view. There are 28 materials arranged in 4 rows and each block is 5 cm thick. (c) Heterogeneous phantom, composed of two layers. (Left) first layer, (middle) second layer, (right) 3D and sagittal views. Each layer is 5 cm thick. (d) Anthropomorphic pelvic phantom with each organ overridden to a single CT number. (e) Lung patient, with each organ overridden to a single CT number. (Right) Cropped region on which optimisation was performed.

5.2.2 Validation of the optimiser against synthetic datasets

To validate the function of the optimiser we worked with synthetic data. Real CT datasets contain noise and real proton radiographs are subject to MCS, so it would be otherwise difficult to discern whether the optimiser had reduced the errors to the theoretical minimum. Using theoretical CT numbers and RSPs for tissue (based on the literature tissue compositions from ICRU [1989]; White et al. [1987]; Woodard and White [1986]) simulated phantoms were constructed and perfect, known, WEPL maps were generated. The calibration curves were optimised against these known WEPL maps, for a range of synthetic datasets shown in figure 5.4.

5.2.2.1 Single materials, multiple materials and heterogeneous phantom

The simplest task for the optimiser 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. An example for the breast tissue is shown in figure 5.4(a). The known WEPL maps were created using the theoretical RSP values and the known geometry.

The next test for the optimiser was to simultaneously reduce the error for multiple materials at once. A synthetic phantom was created, shown in figure 5.4(b), of 5 cm × 5 cm × 5 cm blocks of 28 different materials across the whole HU range. Voxels were 5 mm × 5 mm × 5 mm in size, so each material block contained 10 × 10 × 10 pixels. As before, the known WEPL map was created using the theoretical RSP values, multiplied by the geometric thickness (5 cm).

To verify that the optimiser works through heterogeneous materials, the synthetic phantom shown in figure 5.4(c) was tested. The phantom was composed of two of the multiple material phantoms (described in the paragraph above) in succession, except that the rows of the second phantom were reordered. The known WEPL map was determined using the theoretical RSP values of the first half multiplied by 5 cm, added to the theoretical RSP values of the second half multiplied by 5 cm.

5.2.2.2 Anthropomorphic phantom and real patient

To avoid discrepancies due to CT artefacts (e.g. beam hardening between the femoral heads) and to allow a known WEPL map to be constructed, the CT numbers of each organ in an anthropomorphic pelvic phantom were overridden to a single value and assigned a single RSP: bladder HU=0, RSP=1; prostate HU=58, RSP=1.03; muscle HU=66, RSP=1.05; rectum HU=66, RSP=1.05; cortical bone HU=1194, RSP=1.60). The resultant phantom is shown in figure 5.4(d). As mentioned previously, 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, optimisation was applied over an anterior field.

As with the pelvic phantom, the CT numbers of each organ in a lung patient were overridden to a single value. For real patients we only need to know how the curve should be optimised for the field to be treated, so for speed the optimisation was performed on a smaller cropped region, as shown in figure 5.4(e).

5.2.2.3 Impact of CT noise and multiple Coulomb scattering

In all previous validation tests, the optimisation has involved minimising the difference between two perfect images: (i) a DRR formed through a CT with no noise; and (ii) a simulated known WEPL map. In real patient cases the CT will contain noise and the WEPL map generated by proton radiography will include MCS. The robustness of the optimiser was therefore assessed for these cases.

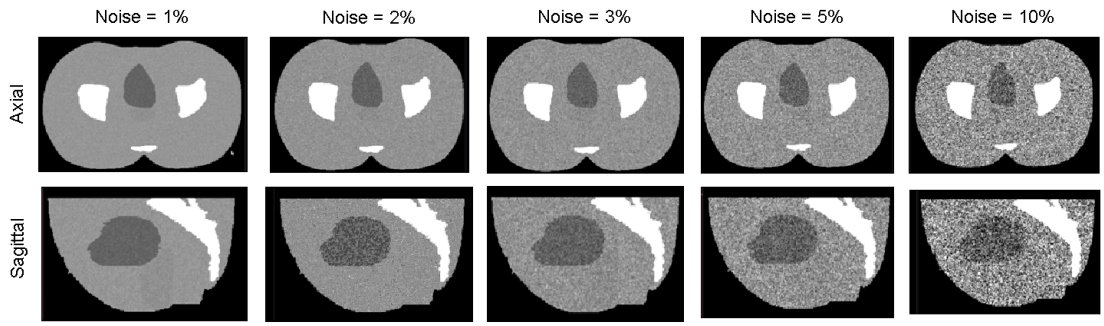
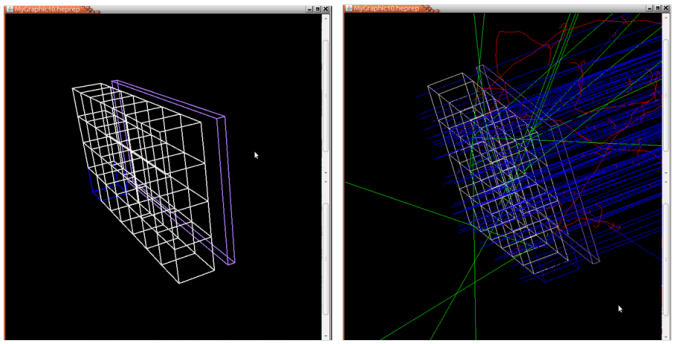


Figure 5.5: Anthropomorphic pelvic phantom with Gaussian noise (standard deviation in percentage) added to the HU values of each organ.

Noise in CT images is a stochastic phenomenon, usually assumed to have a Gaussian distribution (Chvetsov and Paige [2010]). To investigate the impact on the optimisation in a controlled manner, Gaussian noise was introduced into perfect CT datasets by modification of the overriding function of Plastimatch. The CT numbers of each organ in the pelvic phantom (figure 5.4(d)) were overridden to a Gaussian distribution of values. Optimisation was performed on datasets in which a range of noise levels (standard deviations of 1%, 2%, 3%, 5% and 10%) were applied to the scaled CT numbers (+1000 HU compared to usual values) of each organ. These phantoms are shown in figure 5.5.

To investigate the impact of MCS, an ideal proton radiograph was simulated by Monte Carlo using the platform TOPAS (Perl et al. [2012]). The multiple material phantom (figure 5.4) was 'imaged' with a flat beam of 300 MeV protons (an energy sufficient to traverse the object) of no energy spread and with a width sufficient to fully cover the phantom. Monte Carlo simulations require the composition and density of each material, rather than the CT number, and these were acquired from the usual sources (ICRU [1989]; White et al. [1987]; Woodard and White [1986]). The final energy was scored in a plane behind the phantom and 10 million histories were acquired. Figure 5.6 shows the geometry used in the simulation. The 28 material blocks are shown by the white squares, the scoring plane behind the phantom is shown by the purple rectangle and the source begins at the face of the phantom. Different phantom thicknesses were tested: 5 cm, 10 cm, 15 cm and 20 cm.

No detector-specific factors were simulated in an attempt to isolate the effect MCS alone. Subtracting the range in water of the final energy protons from the range in the water of the initial energy allowed the WEPL of the object to be determined. Range mixing will result in a distribution of WEPLs per pixel (Bentefour et al. [2012a]), the mean of which was used as the known WEPL map against which to optimise. This WEPL map is then essentially the best map that could be produced



(a) (b)

Figure 5.6: (a) Geometry of the Monte Carlo simulation of the multiple material phantom (white) and scoring volume (purple). Illustrative events are shown in (b).

by a single-detector proton radiographic system (i.e. in which reconstruction of the individual proton tracks is not possible).

5.2.3 Proton radiographic measurements

All radiographic measurements were made in the proton gantry at MGH, 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 as detailed in section 5.1.2. A diode array was used for imaging, consisting of 249 semiconductor diodes arranged on an octagonal matrix (see figure 5.16, top middle), with a diagonal pitch of 7.07 mm and a linear pitch of 10 mm separating diodes on the same row. The diodes have a sampling time of 2 ms, which is sufficiently small to capture the 100 ms cycle of the RM wheel.

To calibrate each diode in the device, dose rate functions were recorded through different thicknesses of solid water, spanning the range of modulation to be imaged. Dose rate functions were averaged over ten successive RM wheel cycles to reduce the impact of fluctuations in the detector or the proton beam current. The integral acquisition time was thus 1 sec. It has been shown previously that a number of moments can be computed from these dose rate functions (Gottschalk et al. [2011]). In our calibration, we use the third moment, the root-mean-square (RMS) deviation of each modulator cycle, given by,

$$\sigma_t = \left(\frac{1}{S} \sum_{i=1}^{N} m_i (t_i - \mu)^2 \right)^{1/2} \tag{5.4}$$

where S is the sum of all the measurements, m, between the index at the start and end of the RM cycle; m_i is the measurement sampled at time t_i; and μ is the mean

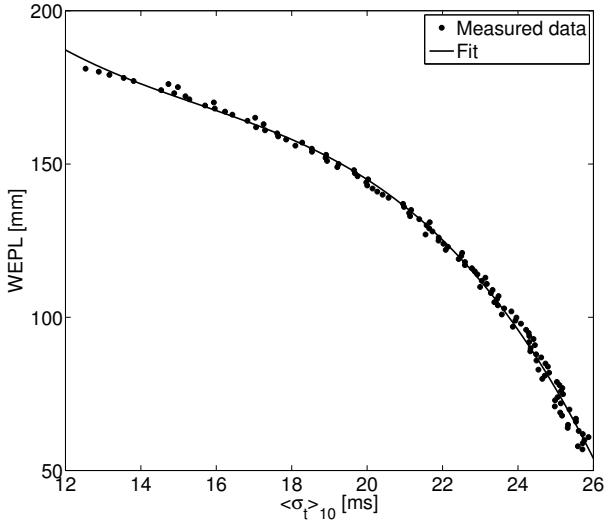


Figure 5.7: Average calibration across all diodes in the detector.

time. Separate calibrations were applied to each diode in the imaging device but the average across all diodes is shown in figure 5.7 as an illustration.

5.2.3.1 Phantoms

To test the optimisation procedure we imaged the phantoms shown in figure 5.8. They are both composed of two halves. The reason for this is that radiographs are made through the entire patient, but we really want to know the WEPL to a therapeutic depth. If we assume 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) optimise the calibration curve by minimising the difference between the $\mathrm{DRR}_{\mathrm{WEPL}}$ and the proton radiographic WEPL; and (iii) predict the WEPL to the distal edge of the top half using the optimised calibration curve. To validate that the optimisation improves the WEPL estimate, we needed an independent measure of the WEPL to the distal edge of the top half. We therefore split apart the phantoms and conducted dose extinction measurements (section 5.2.3.2) on the top half only.

In the top half of the plastic phantom, there are eight inserts from the Gammex 467 phantom (Gammex Inc., Middleton, WI), arranged in a 12 cm lucite box (1 cm thick walls) with a central divider. Water was used to fill in the gaps, up to the insert height (70 mm). The insert materials ranged from RSPs of 0.94 (Adipose) to 1.60 (Cortical bone), which together with the thickness of the lucite box, gave a range of WEPLs from 6.6 to 11.2 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 lucite box as for the

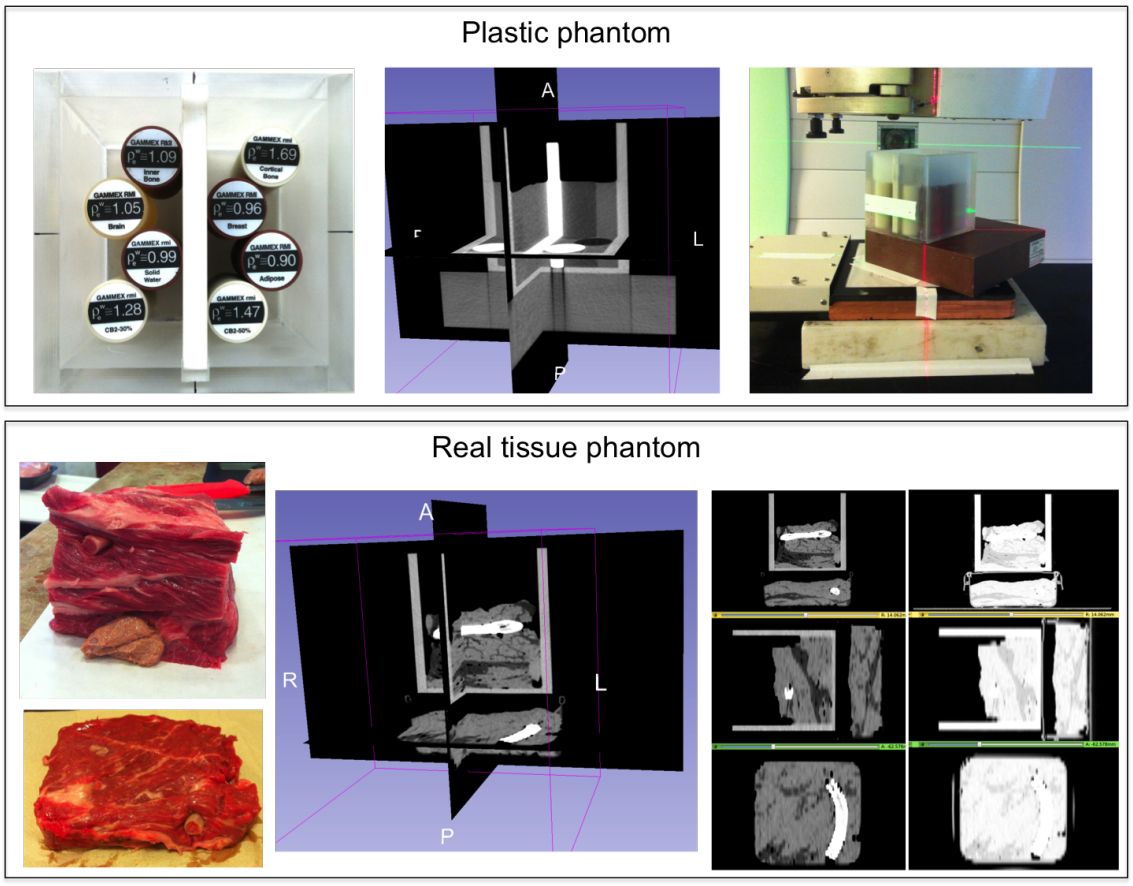


Figure 5.8: Phantoms imaged with proton radiography. (Top) Plastic phantom. Left-right: top-down view showing insert locations; oblique view of CT scan showing water-filling between the inserts; phantom positioned for proton radiography. (Bottom) Real tissue phantom, (left-right) top and bottom halves of real tissue; oblique view of CT scan; transverse/ sagittal/coronal views at two different window levels to show the bone locations and plastic boxes.

plastic phantom) was filled with a single piece of boneless beef chuck (fat left on), a non-smoothed pork bone and a strip of veal liver along the base. 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 hr before and 30 mins after the proton radiographic measurements (that lasted for 4 hrs) 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 in Plastimatch and tissue changes were computed along each direction.

5.2.3.2 Dose extinction measurements

To determine if the optimisation procedure results in a more accurate WEPL map at the therapeutic depth, it was necessary to compare with an independent measurement. For this we used 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}).

5.2.3.3 Handling range mixing

As stated in section 5.1.2.2, range mixing is a problem for single detector proton radiography as individual proton paths cannot be reconstructed. To ensure it did not impact on our optimisation, it was imperative that any diodes subject to range mixing were masked when calculating the cost function. 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_{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 considered to be close to a boundary and thus could be subject to range mixing. These were then masked out during optimisation and any further analysis. This is shown schematically in figure 5.9.

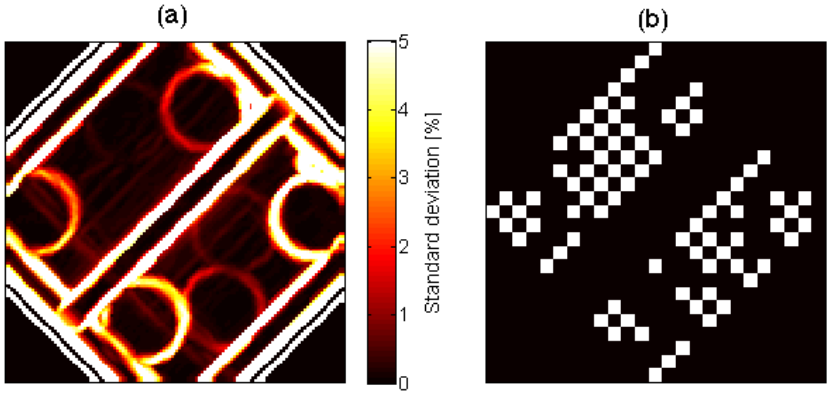


Figure 5.9: (a) Local standard deviation map for plastic phantom, showing edges of objects. (b) Mask based on standard deviations > 1% and diode locations (black = masked).

5.3 Results

5.3.1 Validation of the optimiser

5.3.1.1 Plastimatch validation

The results using a source-to-axis-distance (SAD) of 2.27 m (matches the MGH proton gantry geometry) and 1000 m (to remove divergence), is shown in figure 5.10. At the top is the known WEPL map. In the middle row are the set of results using an SAD of 2.27 m. The difference between the DRR_{WEPL} map and the known WEPL map shows the divergence of the beam, and the histogram of these differences shows the overestimation of the WEPL. On the bottom row is the same set of results for an SAD of 1000 m. The difference image and histogram show that the beam enters parallel and the negligible difference confirms that Plastimatch does not introduce any additional errors. The errors remaining are caused by truncation of floating point numbers and other rounding errors.

5.3.2 Single material, multiple material and heterogeneous phantoms

Optimisation results for the synthetic phantoms are shown in figure 5.11. At the top of the figure are the known WEPL maps, created using the known RSPs and the known geometry. The middle row shows the difference between the DRR_{WEPL} maps created using the generic HU-RSP calibration curve, both in image and histogram format. The bottom row shows the same set of results as the middle row, but for the HU-RSP curves following optimisation. For single materials, it was a simple

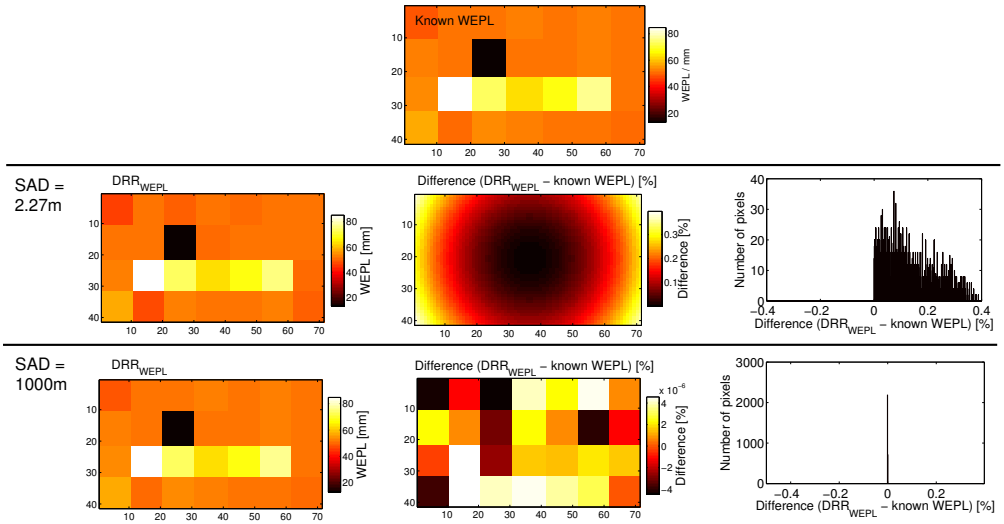


Figure 5.10: Comparison between DRR and physical WEPL maps using an SAD of 2.27 m or 1000 m.

task to converge to the optimal solution. An example result, for breast tissue, is shown on the left of figure 5.11. The root-mean-square-error (RMSE) for each individual material, and its reduction following optimisation, can be seen in figure 5.12. The optimiser can clearly eliminate the errors for all tissues. The result for the multiple material phantom can be seen in the centre of figure 5.11, with the optimised curve reducing the RMSE significantly, from 0.69 mm to 0.09 mm. The optimisation result for the heterogeneous phantom can be seen on the right of figure 5.11, with the optimised curve reducing the RMSE significantly, from 0.94 mm to 0.19 mm.

5.3.2.1 Anthropomorphic phantom and real patient

With an anterior field for the pelvic phantom, it can be seen (figure 5.13, left) that using the generic calibration curve results in an overestimation of the WEPL of RMSE = 4.1 mm, which is reduced to 0.1 mm following optimisation.

The optimisation over a small, field-sized region in a real patient geometry allowed for complete eradication of the RMSE, as shown in figure 5.13 (right). Compared to the case where the entire field-of-view was used for optimisation, the time significantly reduced. The increase in speed is dependent on the relative decrease in image sizes, which in this case was a factor of 7-8.

5.3.2.2 Summary

A summary of the results of the previous validation tests is detailed in table 5.2.

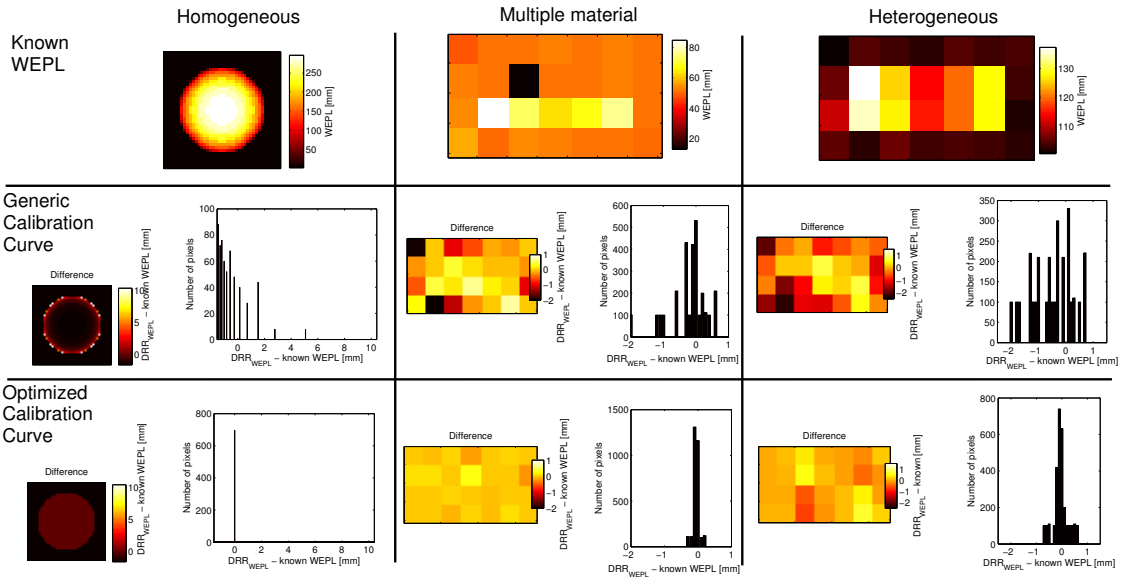


Figure 5.11: Optimisation results for synthetic phantoms: (left) individual homogeneous phantoms; (middle) multiple material phantom; (right) heterogeneous phantom. Top row: known WEPL maps, created using the known RSP values assigned to each CT number. Middle row: difference between the DRR_{WEPL} produced by the generic calibration curve and the known WEPL map, and the histogram of these differences. Bottom row: same as middle row, but for the optimised calibration curves.

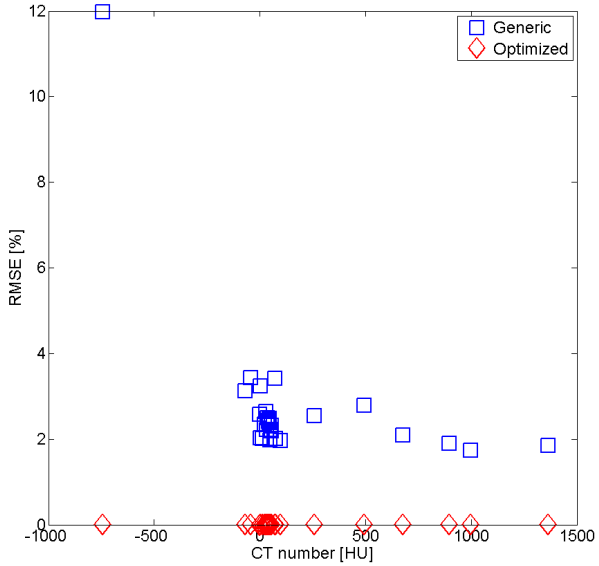


Figure 5.12: RMSE in the difference image for individual uniform spheres of homogeneous human tissues, before (blue squares) and after (red diamonds) optimisation.

Test	Difference [mm] (Generic)		Difference [mm] (Optimised)		Reduction in RMSE	Architecture†	Time
	RMSE	Max/Min	RMSE	Max/Min			
Single materials‡	2.72	12.49/0.05	2.6×10^{-6}	6.1×10^{-5} /6.1×10^{-7}	100.0%	32-bit, 4GB	18.7 s*
Multiple material	0.69	0.68/-2.18	0.09	0.25/-0.25	87.0%	32-bit, 4GB	11 min∩
Heterogeneous	0.94	0.72/-2.30	0.19	0.64/-0.66	79.8%	64-bit, 8GB	13 min∩
Anthropomorphic	4.13	5.03/3.39	0.10	0.63/-0.13	97.6%	64-bit, 8GB	56 min∩
Real patient	3.92	4.23/3.31	0.01	0.03/-0.02	99.7%	64-bit, 8GB	4.1 hrs

Table 5.2: Differences between the DRR_{WEPL} and known WEPL map before and after optimisation, for the range of validation tests. †Architecture is defined by the operating system and the available RAM on which the optimisation was performed. ‡Separate optimisations were run for this test for each material; the values in this row are the average across the 30 different runs. *Range = 5.39/33.4 s. ∩Rounded to the nearest minute.

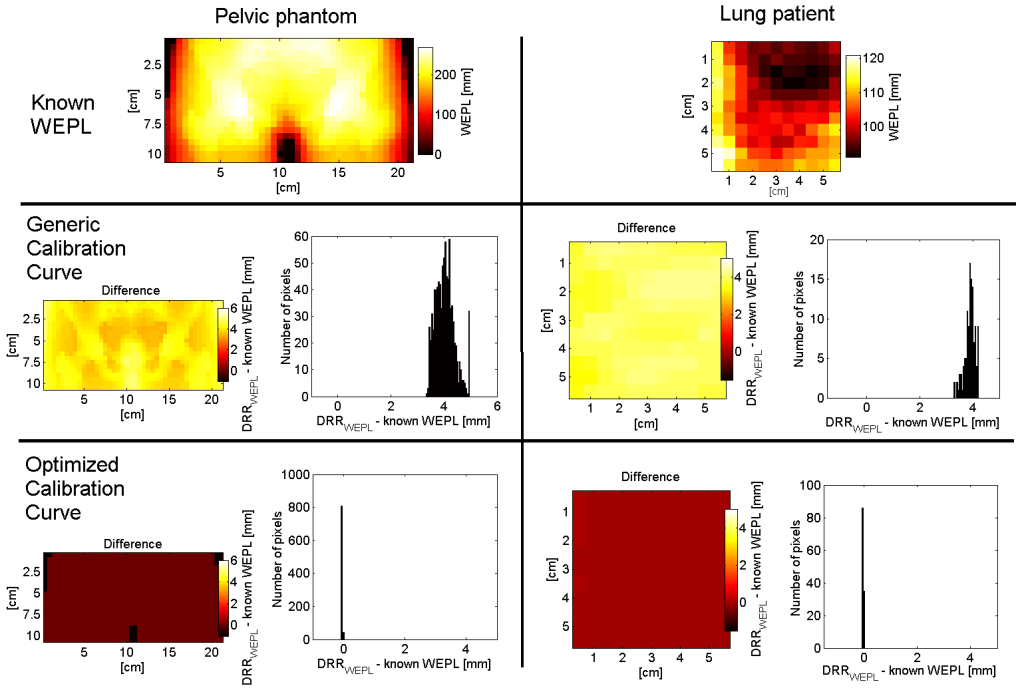


Figure 5.13: Optimisation results for the pelvic phantom (left) and for the lung patient (right).

5.3.2.3 Impact of CT noise and multiple Coulomb scattering

The optimiser was shown to handle noise in the CT number well, with results before and after optimisation detailed in table 5.3. To investigate the impact of MCS on the optimiser, optimisation against a WEPL map produced by Monte Carlo was tested, the results of which can also be found in table 5.3. An example result is shown in figure 5.14 for the 5 cm thick phantom. Before optimisation there was a mean underestimation of -3.8%, which reduced to -0.1% following optimisation. The largest differences were found at the borders of two materials of very different density, where range mixing becomes an issue.

5.3.3 Optimisation against time-resolved proton radiographs

Calibration curves were produced using the optimisation procedure, using real proton radiographs acquired with the time-resolved technique. Figure 5.15 shows the results of the optimisation for the plastic phantom after 120 mins. On the top row are the proton radiographic images, the middle/bottom rows show the results using the generic/optimised calibration curves, and the histogram of differences is shown in the bottom right corner. Further details are given in the figure caption.

Figure 5.16 shows the set of results for the real tissue phantom after an optimisation time of 125 mins (in the same format). From the registration it was found that the tissue remained quite static during the course of the 3 hr experiment. With ref-

Test	Generic		Optimised	
	RMSE	μ/σ	RMSE	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%
Monte Carlo (5 cm)	4.24%	-3.76 / 1.95%	1.95%	-0.13 / 1.95%
Monte Carlo (10 cm)	3.61%	-3.06 / 1.90%	1.69%	-0.11 / 1.69%
Monte Carlo (15 cm)	3.13%	-2.35 / 2.06%	1.81%	-0.15 / 1.80%
Monte Carlo (20 cm)	3.92%	-1.24 / 3.72%	2.75%	-0.38 / 2.72%

Table 5.3: Differences between DRR_{WEPL} and known WEPL maps, before (generic) and after (optimised) optimisation, for the different robustness tests.

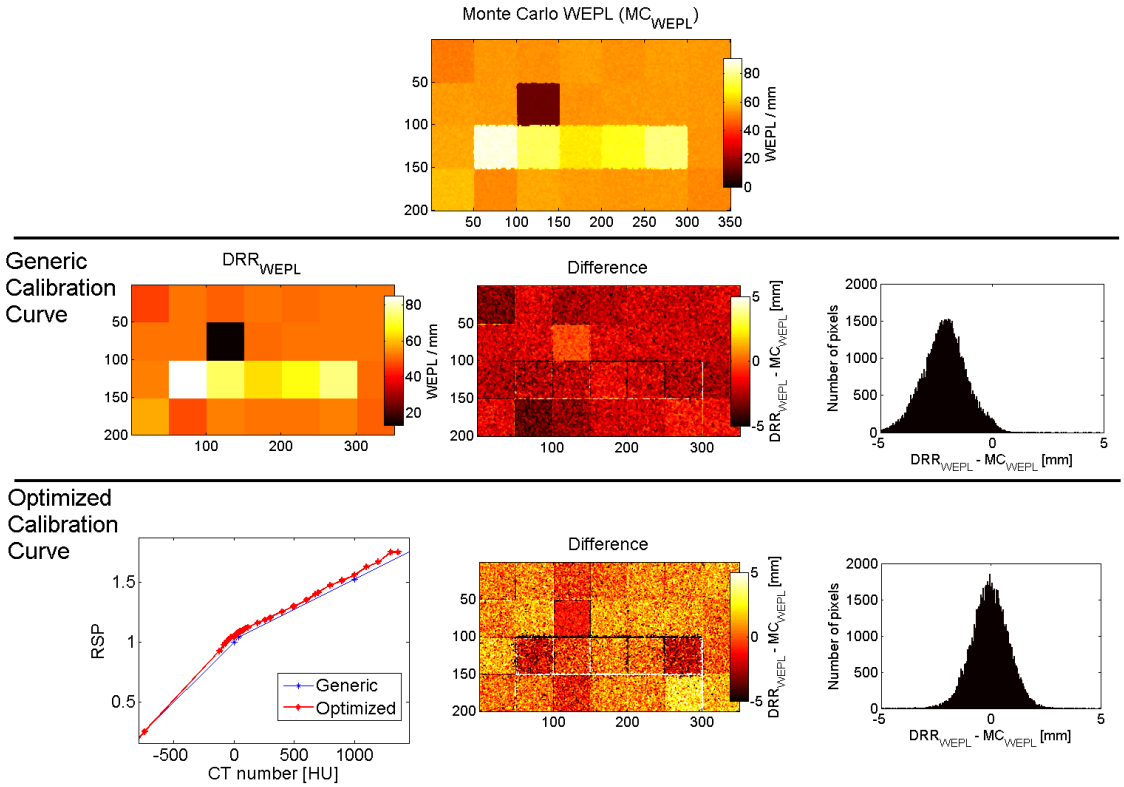


Figure 5.14: Results before and after optimisation for the multiple material phantom, with 5 cm thickness, using WEPL maps produced by Monte Carlo to simulate proton scattering.

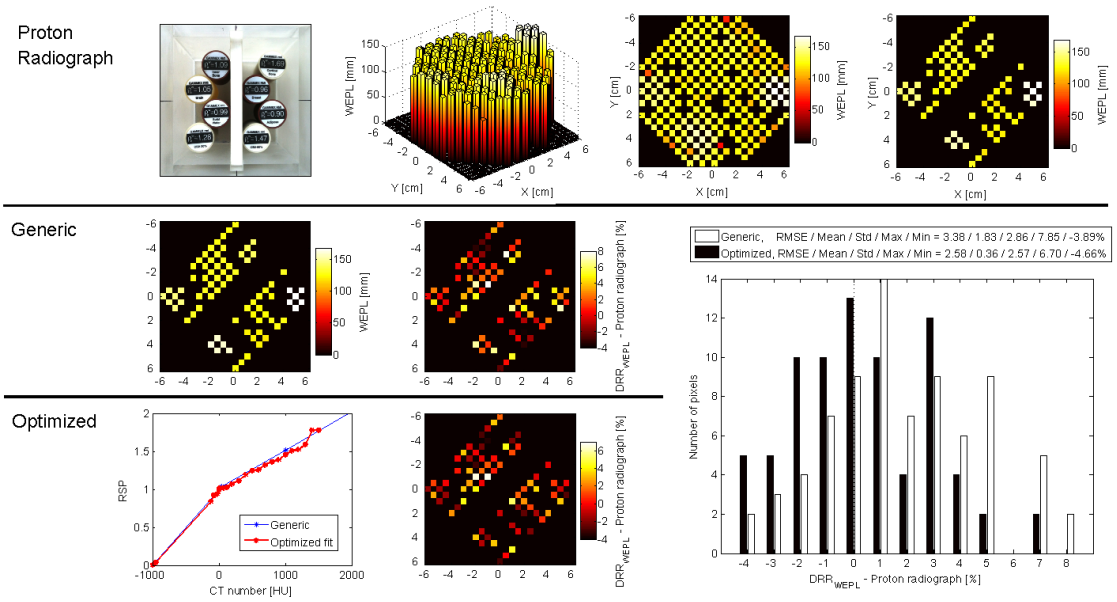


Figure 5.15: Optimisation results for the plastic phantom. (Top, left-right): insert arrangement (rotated 45° counterclockwise with respect to the proton radiographic images); 3D proton radiographic WEPL map; 2D proton radiographic WEPL map; and 2D proton radiographic WEPL map with range mixing mask applied. (Middle, left-centre) DRR_{WEPL} map produced with the generic calibration; and difference between generic DRR_{WEPL} and proton radiographic WEPL. (Bottom, left-centre) Optimised calibration curve; and difference between optimised DRR_{WEPL} and proton radiographic WEPL. (Bottom right) Histogram of differences for the generic and optimised calibration curves.

erence to the directions in figure 5.8, it was found the mean/max deformations were: -0.5/-1.5 mm in the LR direction; -0.5/-2.0 mm in the AP direction; and 0.1/-1.5 mm in the SI direction.

5.3.4 Estimation of the WEPL at the therapeutic depth

To validate that the optimised 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 on top. These dose maps, for the plastic phantom, are shown in figure 5.17. It can be seen that with increasing thicknesses of solid water on top the dose readings in the diodes steadily decreases. The insert with the highest WEPL (Cortical bone) is the first to disappear, while the one with the lowest WEPL (Adipose) is the last. Depth dose curves were created for each diode. The R80 of each fall off, together with the known R80 of the beam, allowed for the calculation of the WEPL to each diode. The WEPL maps for the two phantoms are shown in figure 5.18.

A comparison of the DRR_{WEPL} maps at the therapeutic depth (constructed using the generic and optimised calibration curves) with the WEPL maps measured by

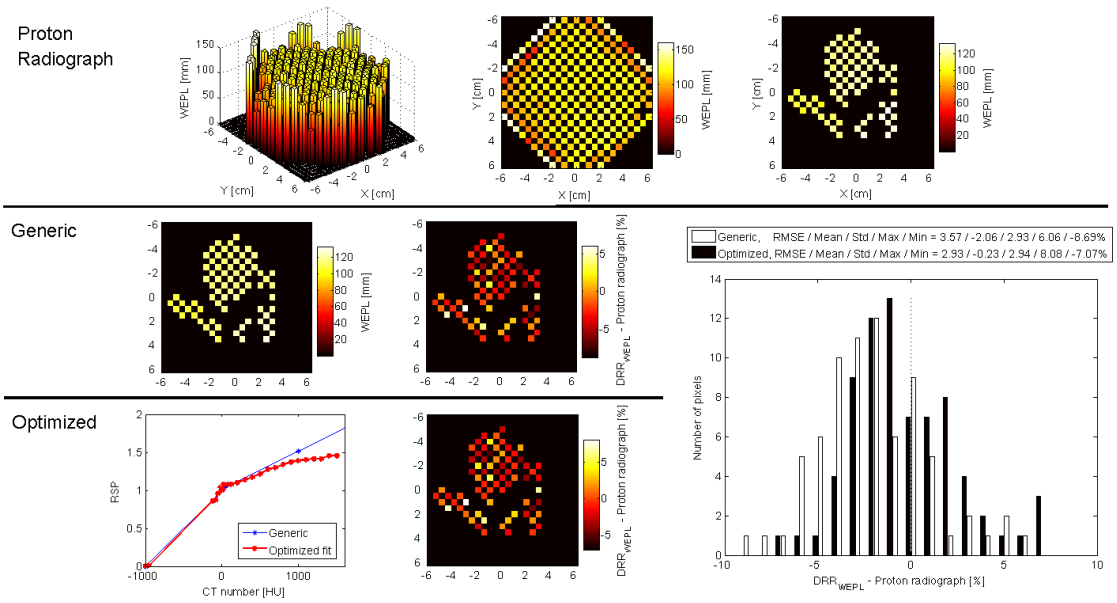


Figure 5.16: Real tissue phantom results. Same format as figure 5.15.

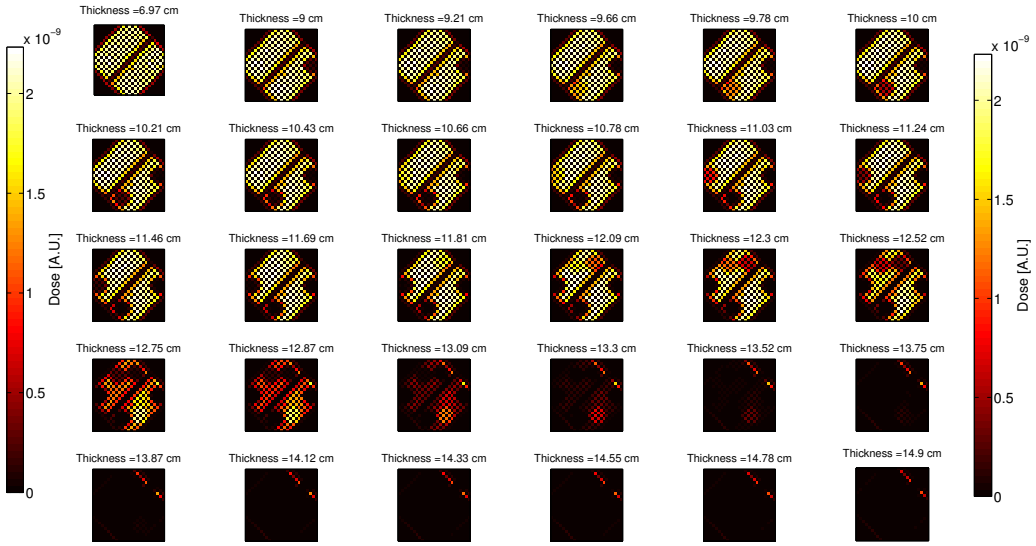


Figure 5.17: Integrated dose maps for the plastic phantom as part of the dose extinction measurement.

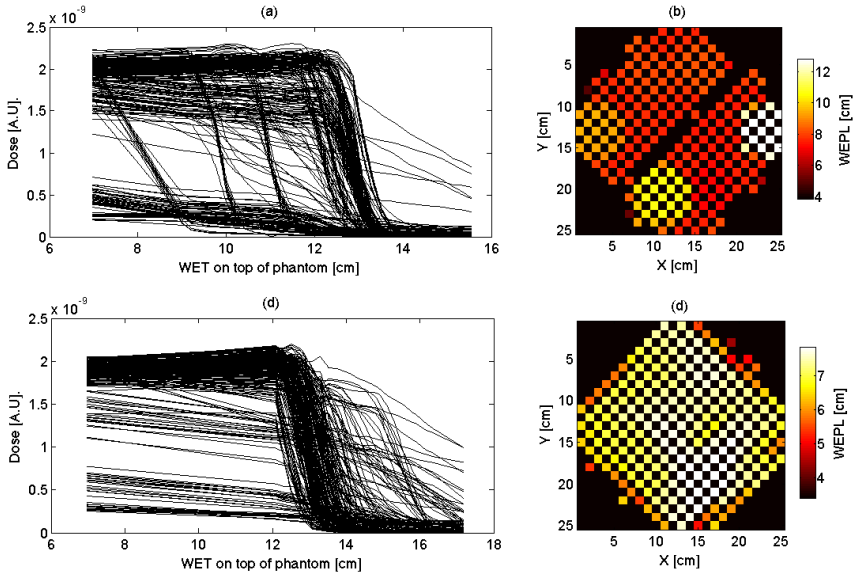


Figure 5.18: (a) Depth dose curves for all diodes behind the top half of the plastic phantom. (b) DE_{WEPL} map. (c,d) Same as (a,b), for the top half of the real tissue phantom.

dose extinction is shown in figure 5.19.

5.4 Discussion

5.4.1 Validation of the optimiser

5.4.1.1 Plastimatch validation

It can be seen (figure 5.10) that an SAD of 2.27 m introduces noticeable divergence in the ray tracing. By increasing the SAD to 1000 m, the rays effectively enter parallel and the error is significantly reduced (max error from 0.4% to 4.6×10^{-6}% in the WEPL). The remaining errors are caused by precision errors in Plastimatch (i.e. truncation of floating point numbers).

5.4.1.2 Validation against simulated datasets

The use of synthetic datasets made it possible to validate the optimiser. Based on the assigned RSPs for the HU values in the synthetic dataset, perfect WEPL maps could be created against which to optimise. In all cases the error defining the cost function was either completely eradicated or significantly reduced. The only cases where the optimiser failed to completely minimise the error were for the multiple material and heterogeneous phantoms. In these phantoms there were 28 different materials, but the optimised calibration curve was only split up into 25

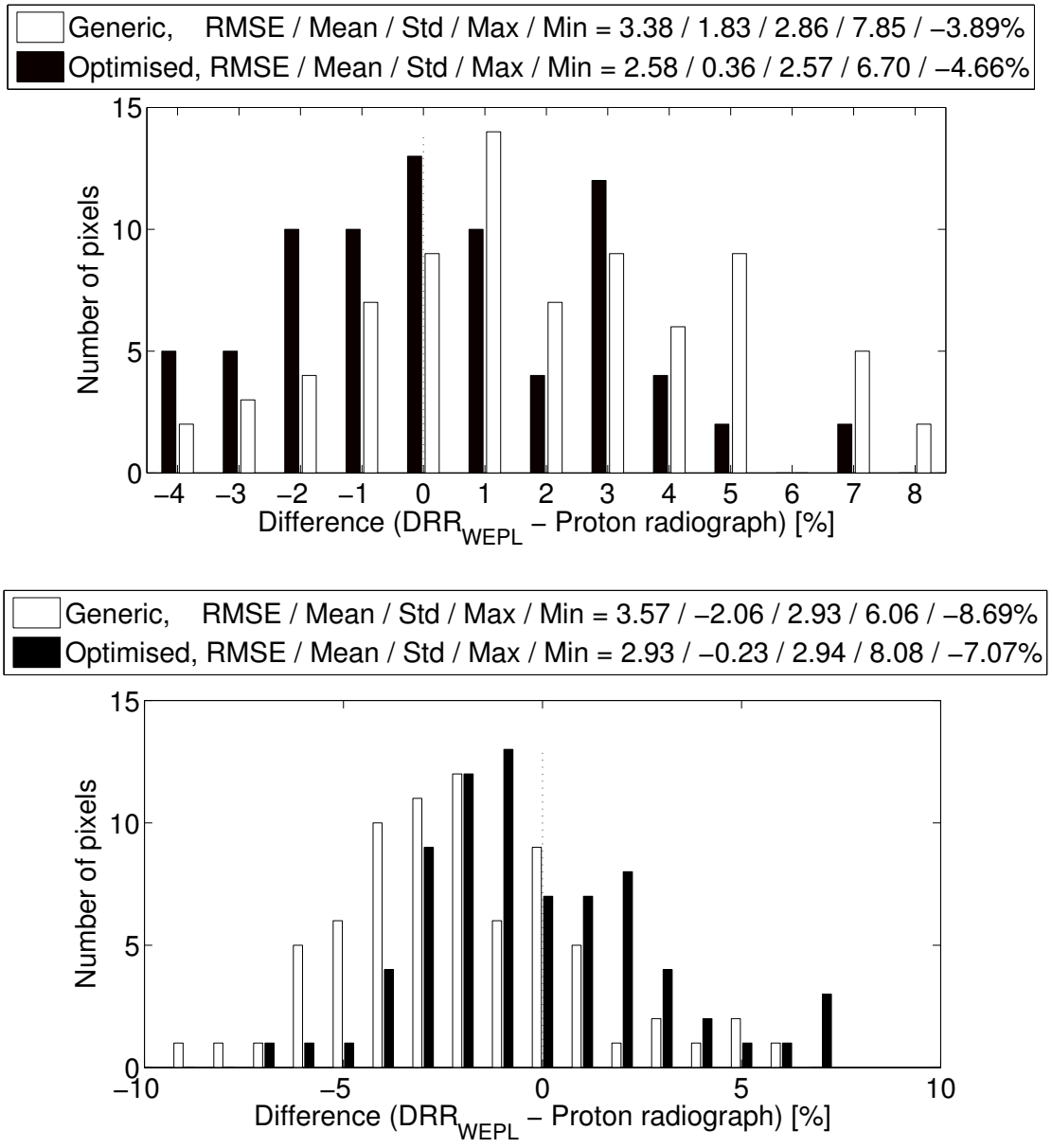


Figure 5.19: Differences between the $\mathrm{DRR_{WEPL}}$ and the $\mathrm{DE_{WEPL}}$, before and after optimisation, for (top) the plastic phantom and (bottom) the real tissue phantom.

different materials. As such, it was not very likely for the optimiser to completely eradicate the errors. The fact that there was a significant reduction demonstrates the optimiser's robustness to potentially conflicting requirements. The results also showed that the complexity of the dataset does not impact on the results, with significant reduction and complete eradication of the error for the anthropomorphic phantom and real patient geometries, respectively.

5.4.1.3 Robustness of the optimiser against CT noise and multiple Coulomb scattering

The perfect datasets were critical in constructing the optimiser, but it was also important to validate that the optimiser was robust in the more realistic case of CT noise. It can be seen from table 5.3 that the optimiser was robust to varying noise levels, with an almost identical final error following optimisation. As to be expected, there is an intrinsic spread in the errors that the optimiser cannot remove due to the noise. Despite this, the optimiser was still able to find a minimal solution and functioned as well as could be expected.

In the validation tests described previously the calibration curve had been optimised against a perfect WEPL map, formed using RSP values assigned to specific CT numbers. However, in the real situation the WEPL map will be generated from a proton radiograph and so will include MCS and uncertainties from the detection method. The optimiser showed robustness against this, as shown by the statistics in table 5.3. It can be seen in figure 5.14 that the WEPL generated from the DRR is, prior to optimisation, underestimated compared to that generated from the Monte Carlo simulation. This is expected as the DRR is a straight line approach and in the simulation the protons traverse a longer path due to MCS. Additionally, similar to when CT noise was introduced, MCS leads to a spreading of the errors. As one would expect, with increasing phantom thickness, there is an increase in the error and spread in errors. Following optimisation, the offset due to the underestimation was removed and the spread in errors was also reduced. Reductions in the RMSE following optimisation were found to be 64/72/77/30% for the 5/10/15/20 cm thicknesses, respectively.

5.4.2 Optimisation against time-resolved proton radiographs

The results in figures 5.15 and 5.16 demonstrate that the procedure works when optimising against real proton radiographs. For both the plastic and real tissue phantoms the optimisation improved the results. Following optimisation on the plastic phantom: the RMSE difference was reduced from 3.4% to 2.6%; the mean was reduced from 1.8% to 0.4%; and the standard deviation was reduced slightly

from 2.9% to 2.6%. For the real tissue phantom: the RMSE was reduced from 3.6% to 2.9%; the mean error was reduced from -2.1% to -0.2%; and the standard deviation remained constant at 2.9%. As expected the improvements are more modest than for the synthetic datasets that had a known solution, but they are sufficiently large to warrant use of the technique.

For the real tissue phantom the tissue shape changes were small during the course of the experiment. Maximum changes of -1.5/-2.0/-1.5 mm in the LR/AP/SI directions were found between the pre- and post-radiographic CT images, which were acquired 4 hours apart in time. The time difference between the pre-radiographic CT image, upon which the optimisation was performed, and the proton radiographic measurement was around 1 hour. 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.

5.4.3 Estimation of the WEPL at a therapeutic depth

After 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 optimised curve matched the DE_{WEPL} better than the generic curve. The RMSEs for plastic phantom were 3.4/2.5% for generic/optimised calibration curves. The same statistics for the real tissue phantom were 2.0/1.4%. These improvements are similar to those of the previous section, suggesting that optimisation of the calibration curve using proton radiography across the whole patient allows for a more accurate estimate of the WEPL at a therapeutic depth. Additionally, this independent measurement validates the accuracy of the original proton radiograph.

5.4.4 Technique limitations

The technique proposed here relies on the ability to produce an accurate proton radiographic image. Optimising 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 utilised in this study was able to produce accurate WEPL values. Previous studies have shown the accuracy in the WEPL to be < 1 mm using ionisation chambers (Lu [2008b]) or 0.3 mm when using diodes (Gottschalk et al. [2011]). Additionally, the agreement with the independently measured DE_{WEPL} was better following optimisation.

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 optimised calibration curve as input for a Monte Carlo simulation so that the effects of MCS are accounted for. The optimised calibration curve could then be fine-tuned, based on 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,

$$\sigma_s = S_2 \frac{\sqrt{X/X_0}}{p\beta c}[1 + \varepsilon log(X/X_0)] \tag{5.5}$$

where $S_2 = 14.1\,\text{MeV}$ and $\varepsilon = 1/9$ are empirical factors; X is the thickness of the scatterer; 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. At the energy used (approximately 175 MeV) and at largest thicknesses of our phantoms, $\sigma_s = 0.0292\,\text{rad}$ for the plastic phantom and $\sigma_s = 0.0253\,\text{cm}$ for the real tissue phantom. An angular deflection leads to a projected lateral deflection D [cm] as described by Leroy and Rancoita [2009]:

$$D = \frac{\sigma_s}{\sqrt{3}} X \tag{5.6}$$

In our experiments the average lateral deflections were thus $D = 0.29\,\text{cm}$ for the plastic phantom and $D = 0.19\,\text{cm}$ 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 optimisation 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 minimised, affects the time to calculate the cost function. (iii) The stopping tolerance of the optimisation directly affects how long the optimisation takes to converge. The first of these, computation of the $\mathrm{DRR}_{\mathrm{WEPL}}$ map using the ray tracing details, was the major factor for slow optimisation 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 optimised. It is also possible that an alternative optimisation scheme, such as a gradient descent, could speed up the minimisation.

Another limitation of our specific set of measurements is the limited 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 CCD cameras (Bentefour et al. [2013]; Muraishi et al. [2009]; Ryu et al. [2008]), commercial flat-panel detectors (Telsemeyer [2012]) or CMOS APS (Gelover-Reyes et al. [2011a]; Poludniowski et al. [2014]; Seco and Depauw [2011]). Additionally, production of the $\mathrm{DRR}_{\mathrm{WEPL}}$ map is reliant on the accuracy of the CT. As such, removal or reduction (with improved reconstruction algorithms or dual-energy CT) of artefacts such as beam hardening or streaking must be considered when producing the DRR.

5.4.5 Application to other proton radiographic techniques

The optimisation 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 optimisation scheme could, in theory, be used for classical proton radiographic devices. Provided the ray tracing of the DRR can be effectively modelled (i.e. the source-to-isocentre and source-to-imager distances are known), patient-specific curves could be generated.

5.4.6 Introduction into the proton treatment planning workflow

While in its infancy, this proof of concept work demonstrates how proton radiography could be utilised and implemented into the proton treatment planning workflow. With a single radiograph imaged using a single detector, it is possible to generate a calibration curve that is specific to the patient being treated. The treatment planning workflow could be updated as follows: (i) acquire 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 optimisation; and (iv) create a treatment plan using the patient-specific calibration curve.

It is clear that adding an additional procedure into the patient treatment workflow will impact on patient throughput. Logistically it may be difficult for a centre to apportion beam time specifically to imaging. We believe that in order for this to be introduced routinely into the clinic it would be necessary to have a dedicated beam line for imaging. If, as has been suggested by this preliminary work, the technique can offer a large improvement in WEPL prediction, companies may develop and eventually offer dedicated single-room imaging beam lines.

5.4.7 Future work

In addition to an improvement on the limitations outlined in section 5.4.4, there are a number of areas of future work. A comparison with DECT is required, as this has shown a great deal of promise recently (Bourque et al. [2014]; Hünemohr et al. [2014]) and offers a more convenient solution than attempting to introduce a new step into the proton treatment planning workflow.

It seems logical that taking radiographs from additional angles would offer an improvement, particularly in the estimation of the WEPL at a therapeutic depth (which may be affected by tissues distal to the target). An investigation into the optimal number of angles and the potential improvement offered is warranted. This would lead to a comparison with proton CT, which is argued by Hansen et al. [2014] to offer better accuracy than DECT. Experimental validation of this claim is required.

Range mixing is a hindrance to single-detector proton radiography and a thorough study into how it can be effectively detected and removed is needed. A variety of methods have been suggested, one of which is the assessment of the 4^{th} and 5^{th} moments of the dose rate pattern. Gottschalk [2014] shows that deviations from the natural (measured in a homogeneous water phantom) skewness and kurtosis can be used to quantitatively determine if a diode is subject to range mixing.

5.5 Conclusions

It has been shown that it is possible to construct a robust optimiser that can produce patient-specific calibration curves, using the information from simple single-detector proton radiographs. Following validation with a range of perfect (containing no noise) synthetic datasets, the robustness of the optimiser was confirmed using

CT datasets containing noise and proton radiographs produced with Monte Carlo simulations.

The technique was demonstrated experimentally 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 optimisation 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 optimised calibration curve. Using this optimised 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 optimised calibration curve.

Based on these promising results further work is justified, with the view to introducing an additional step in the current proton treatment planning procedure. A single proton radiography in the intended treatment direction could be used after the treatment planning X-ray CT, allowing for the production of a patient-specific calibration curve that is subsequently used in treatment planning.

Chapter 6

Overall conclusions and final remarks

In this thesis we looked at two themes. The first involved the assessment of three proton TPSs, with the aim of informing the procurement process for the new proton centre at UCLH. The second theme was centred around the range uncertainty arising from the use of a single HU-RSP calibration curve in proton treatment planning.

6.1 Treatment planning system comparison

In this work we make a timely comparison of the optimisation algorithms of proton TPSs of three major vendors, Varian (Eclipse), Philips (Pinnacle^3) and Elekta (XiO), based on the treatment of meningiomas with an SFUD horizontal fixed beam arrangement. The comparison was built upwards from the beam data and it was found that both Eclipse and Pinnacle^3 modelled the measured data to within clinical tolerances. Pinnacle^3 consistently overestimated the Bragg peak position (max 0.5 mm) for the depth dose curves, while Eclipse overestimated the effect of build up at higher energies (< 175 MeV). It was also concluded that additional measurements are needed to model the long, non-Gaussian tails from large angle scattering and nuclear interactions.

In the treatment plans Pinnacle^3 generally gave lower OAR doses and a lower integral dose outside the target. This was attributed to the flexibility of available spot positions and the ability of the optimiser to be able to adjust the relative weighting of the two fields. It was also found that Pinnacle^3 distributed its spots more uniformly than the other systems. All conclusions are subject to validation of the accuracy of the dose calculation algorithms of each system.

6.2 Errors in the HU-RSP calibration curve

It has been suggested that the use of a single calibration curve, to convert the patient's X-ray CT into RSPs for proton treatment planning, contributes up to 1.8% uncertainty to the proton beam range. Being able to reduce this range uncertainty will allow for tighter planning margins and a reduction in the irradiation of healthy tissue.

6.2.1 Range effects

We began the investigation by first assessing the errors and range effects caused by the stoichiometric procedure, which is used to produce the HU-RSP calibration curve. We specifically look at the theoretical calculation of the RSP.

Four methods of calculating the RSP were assessed (Bichsel [1972], Janni [1982], ICRU [1993] and Schneider et al. [1996]). Comparing to measured values of Gammex inserts, it was found that there were errors in all techniques, with a mean underestimation of up to 2.1% for the Bichsel approach. Using the Schneider approximation with the Janni I-values showed the best match to our measurements. These RSP errors were shown to translate directly into proton beam range errors of similar magnitudes, for plans within a Gammex phantom and in two patient cases.

To assess the error in the theoretical calculation of human tissues, it was necessary to define a gold standard that matched the measurements best. Given that the I-value of water is still disputed, the Schneider approximation, Janni elemental I-values and a water I-value of 80 eV was used to form a gold standard that matched our measurements excellently. Comparing this to the method used by most centres (Schneider approximation, ICRU elemental I-values and a water I-value of 78 eV), there were errors of up to 0.8% across 73 human tissues.

An RSP measurement of the inserts is therefore suggested as a solution to minimise this error, although this assumes the scaling in plastics also applies to human tissues. Even if this may not be the case, patient-to-patient variations from literature tissue compositions point to proton radiography as a more viable solution, leading to the next two chapters of the thesis.

6.2.2 Dose ratio proton radiography

A number of different approaches of proton radiography have been attempted. However the fact that it has not yet been introduced into the clinic suggests that none have so far been successful. In this thesis an investigation was made into a novel

single-detector method termed 'dose ratio' proton radiography that is both simple and has the potential to be applied to PBS treatments.

The application of the technique to a range of patient sites was conducted using theoretical pristine Bragg peaks. The proportion of the site that could be reliably imaged was determined based on the energy pair, noise level in the dose ratio image and the required accuracy in the WEPL. These theoretical predictions were validated with experimental measurements using a CMOS APS. It was showed that use of the proximal side of the Bragg allowed for a wider range of WEPLs than the distal side, but in general the choice of patient sites is limited. A more established proton radiographic method was therefore used in the final chapter of the thesis.

6.2.3 Reducing range uncertainty with proton radiography

In the final chapter of the thesis an approach was developed that uses proton radiographic images to generate patient-specific HU-RSP calibration curves. The function of the optimiser was first validated on synthetic datasets and its robustness to noise was demonstrated. Using a simple diode array, time-resolved proton radiographic images were used to generate specific calibration curves for a plastic and a real tissue phantom. The optimised curve enabled the mean error between a DRR_{WEPL} map and the proton radiograph to be reduced from -2.1% to -0.2% for the real tissue phantom. It was also shown that with this curve a more accurate prediction of the WEPL at a therapeutic depth could be obtained.

If these promising results could be built upon, this approach could lead to an update to the current treatment planning workflow for proton therapy patients. Proton radiographs would be acquired at the same time as the treatment planning X-ray CT and patient-specific HU-RSP calibration curves would be determined. These would then be utilised when creating the treatment plan and the reduced range uncertainty would allow for tighter target margins and hopefully improved patient outcomes.

6.3 Areas for future work

One area of future work for the TPS comparison project is a thorough investigation into the accuracy of the dose calculation algorithm, which was beyond the scope of this work. Throughout the work, the reader was reminded that all conclusions were made on the provision that the dose calculation algorithms were validated. The most straightforward approach to this would be to make identical plans in water on each system, with the only difference being the dose calculation algorithm, and to

then deliver the plans and measure the dose experimentally. Following this, more complicated geometries could be tested to investigate the impact of heterogeneities.

For prospective buyers of the software it would also be of interest to look at a number of other areas: the ability to utilise multiple imaging sets (from MRI or positron emission tomography, PET) and the registration of these to the original X-ray CT; the performance and accuracy of photon treatment planning, as many centres will deliver combined treatments; the integration of photon and proton treatment plans; adaptive planning, taking into account additional daily or weekly imaging; and biological modelling (among others). Comments on the user interface and the planning experience would also be of benefit, although these are subjective and were thus avoided in our work.

In the application of proton radiography project there are many areas of future work. There have been many recent publications investigating the potential of DECT as a tool to estimate the tissue stopping powers, with some promising results. There is also unpublished work comparing the performance of a Monte Carlo-generated proton CT to DECT and single-energy CT, but as yet no experimental comparison. There is certainly a need for an experimental comparison of proton radiography (we believe that the most developed and promising single-detector technique is the time-resolved approach) and DECT.

If proton radiography fares favourably to DECT, there is justification to take the procedure developed here one step closer to the clinical environment. In order for this to happen, there would need to be a number of improvements to the imaging system. The spatial resolution of the detector would need to be increased. A minimum/maximum pitch between diodes of 7/10 mm is too large for a clinical detector and could lead to important physical features being missed. Spatial resolution on the order of 1-2 mm, comparable to X-ray CT, would be more appropriate. Additionally, the detector would need to have a larger imaging area. It is possible that treatment fields could be delivered larger than the area of the current proton imager (12 cm $\times$ 12 cm), so an imaging area of at least 20 cm $\times$ 20 cm would be more appropriate. Time constraints made it difficult in this work (we instead relied on the accuracy determined in previous works), but the accuracy and density resolution of the proton imager should be fully determined. Additionally, a comparison of the dose delivered by both techniques would be essential.

In the optimisation procedure one of the biggest sources of uncertainty is the use of straight-line ray tracing when producing the DRR. Using only one detector, however, does not allow us to reconstruct proton paths so estimates must be made as to the shape of the path. Based on the effective source size and source-to-axis distance, the beam divergence could be estimated and a Gaussian spreading from this path could be assumed. It would then be necessary to write software (or

make modifications to Plastimatch) that could compute the length of intersection of these cubic spline paths with the CT dataset. Improvements in the speed of optimisation should also be introduced. This could involve either introducing field-specific optimisations or by using a more efficient optimiser such as a gradient descent approach.

Appendix A: Gamma Analysis

Gamma analysis, a standard in the radiotherapy field, assesses the quality of the fitting in terms of dose difference (DD) and distance-to-agreement (DTA), throughout two dose maps. Here we use a one-dimensional implementation, by comparing two curves. The DD measures the dose difference between the measured data and the fit, effectively assessing the fit in the y-direction. The DTA measures the spatial distance between a measured data point and a fitted point, and thereby effectively assesses the fitting in the x-direction. Depending on the situation the DD or DTA alone may be insufficient and they are actually complementary to each other. The dose difference gives suitable results in low dose gradient regions but not in high dose gradient regions, with DTA the opposite. The gamma evaluation method is a combination between the dose difference and the DTA evaluation methods.

Perfect fitting at a given point of the curve is defined if both DD and DTA are zero. The gamma index, first proposed by Low et al. [1998], is a function that performs composite analysis of the DD and DTA by collapsing them both into one parameter. The gamma index γ between a given measurement position $\mathbf{r_m}$ and fitted position $\mathbf{r_c}$ is defined as,

$$\gamma(\mathbf{r_m}) = \min\{\Gamma(\mathbf{r_m}, \mathbf{r_c})\} \,\forall\, \{\mathbf{r_c}\} \tag{1}$$

where

$$\Gamma(\mathbf{r_m}, \mathbf{r_c}) = \sqrt{\left(\frac{r(\mathbf{r_m}, \mathbf{r_c})}{\Delta d_M}\right)^2 + \left(\frac{\delta(\mathbf{r_m}, \mathbf{r_c})}{\Delta D_M}\right)^2} \tag{2}$$

and where $r(\mathbf{r_m}, \mathbf{r_c}) = |\mathbf{r_c} - \mathbf{r_m}|$ is the DTA and $\delta(\mathbf{r_m}, \mathbf{r_c}) = D_c(\mathbf{r_c}) - D_m(\mathbf{r_m})$ is the DD, both at the position $\mathbf{r_m}$; Δd_M is the DTA acceptance criteria; and ΔD_M is the DD acceptance criteria. A visual representation of this is shown and explained in figure A.1.

In radiotherapy a two-dimensional gamma analysis is usually used to compare two dose distributions. For example, before an intensity modulated radiotherapy (IMRT) treatment, patient-specific quality assurance involves delivering the planned treatment to a water phantom loaded with an array of diodes. The delivered and planned fluence from a given beam direction are compared using gamma analy-

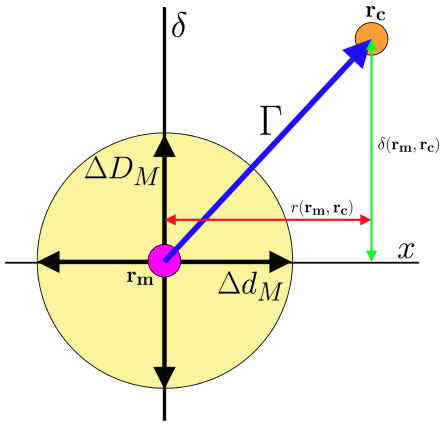


Figure A.1: Geometric representation of the gamma index in one-dimension. All parameters are as defined in equation 2. $\mathbf{r_m}$ (measured point, purple) and $\mathbf{r_c}$ (fitted point, orange) are the points being compared. The red line given by $r(\mathbf{r_m}, \mathbf{r_c})$ is the distance-to-agreement, the green line given by $\delta(\mathbf{r_m}, \mathbf{r_c})$ is the dose difference and Γ is the vector combining these two differences. ΔD_M and Δd_M are the criteria for passing in the dose difference and distance-to-agreement directions respectively. The circle representing the accepted region is shaded in ochre.

sis, typically with a DD/DTA criteria of 2%/2mm or 3%/3mm (depending on the precision required).

To assess the quality of the fitting, a one-dimensional gamma analysis code was written in Matlab. The code compared each measured point with all points in the fits by determining the Γ value. The minimum of these Γ values was selected as the gamma index γ for that measurement position. This process of comparison was repeated for all measured points. To ensure there were no artefacts in the analysis and to make the comparison fair for all systems, all datasets were resampled to a high resolution (points equally spaced, every 0.1 mm). A criteria of 2%/2 mm was used for proton integrated depth doses, but a much tighter 2%/0.05 mm was used for the proton spot profiles.

References

Akisada, M., Ohashi, J., and Kondo, K. (1983). Conceptual design of proton computed tomography with magnetic spectrometer. *Japanese Journal of Applied Physics*, 22(4):752–8. 89

Andreo, P. (2009). On the clinical spatial resolution achievable with protons and heavier charged particle radiotherapy beams. *Physics in Medicine and Biology*, 54(11):N205–15. 70, 115

Archambeau, J., Slater, J., Slater, J., and Tangeman, R. (2000). Role for proton beam irradiation in treatment. *International Journal of Radiation Oncology Biology Physics*, 22(2):287–94. 10

Arjomandy, B., Schultz, T., Park, S., and Gayar, H. (2012). A Comparative Study of Single Verses 2-field Daily Fraction for Treatment of Prostate Cancer Using IMPT, Double-Scattered, and SFUD Delivery Technique. *International Journal of Radiation Oncology Biology Physics*, 84(3):S843. 16

Arvold, N. D., Niemierko, A., Broussard, G. P., Adams, J., Fullerton, B., Loeffler, J. S., and Shih, H. A. (2012). Projected second tumor risk and dose to neurocognitive structures after proton versus photon radiotherapy for benign meningioma. *International Journal of Radiation Oncology Biology Physics*, 83(4):e495–500. 16

Ayyangar, K. M., Fung, A. Y. C., Li, S., Pillai, S., Yoe-Sein, M. M., Zhen, W., and Enke, C. A. (2005). Dose volume histogram comparison between ADAC Pinnacle and Nomos Corvus systems for IMRT. *Australasian Physical & Engineering Sciences in Medicine*, 28(1):1–7. 16

Ball, R., Beene, J., Ben-Moshe, M., Benhammou, Y., Bensimon, R., Chapman, J., Etzion, E., Ferretti, C., Friedman, P. S., Levin, D. S., Silver, Y., Varner, R. L., Weaverdyck, C., Wetzel, R., Zhou, B., Anderson, T., McKinny, K., and Bentefour, E. H. (2014). Development of a plasma panel radiation detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 89

Bashkirov, V. A., Schulte, R. W., Penfold, S. N., Member, S., and Rosenfeld, A. B. (2007). Proton Computed Tomography : Update on Current Status. *IEEE Nuclear Science Symposium Conference Record*, HT6-1:4685–4688. 89

Basun, S., Meltzer, R., and Imbusch, G. (2007). Exchange-coupled chromium ion pairs in ruby revisited. *Journal of Luminescence*, 125(1-2):31–9. 98

Baumert, B., Lomax, A., Miltchev, V., and Davis, J. (2001). A comparison of dose distributions of proton and photon beams in stereotactic conformal radiotherapy of brain lesions. *International Journal of Radiation Oncology Biology Physics*, 49(5):1439–1449. 20

Bentefour, E. H., Samuel, D., Testa, M., and Lu, H.-M. (2013). Methods and device for dose based proton radiography. In *American Association of Physicists in Medicine 55th Annual Meeting & Exhibition*, page 308. 89, 139

Bentefour, E. H., Tang, S., Prieels, D., and Lu, H.-M. (2010). Towards range guided prostate treatment using an AP field. In *American Association of Physicists in Medicine 51st Annual Meeting & Exhibition*, page 3295. 115

Bentefour, E. H., Tang, S., Prieels, D., and Lu, H.-M. (2012a). Effect of tissue heterogeneity on an in vivo range verification technique for proton therapy. *Physics in Medicine and Biology*, 57(17):5473–84. 116, 121

Bentefour, E. H., Testa, M., and Lu, H.-M. (2012b). In-vivo WEPL verification based on point dose measurements in proton treatment by beam scanning. In *PTCOG 51*, page 3855. 90, 91, 107

Bentefour, H., Lu, H.-M., and Prieels, D. (2009). In vivo proton range verification/correction for safe anterior-posterior field treatment of prostate cancer. In *American Association of Physicists in Medicine 50th Annual Meeting & Exhibition*, page 2571. 115

Berger, J. and Zucker, M. (2004). ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (version 1.2.2). Available: http://physics.nist.gov/Star. 64, 98

Berger, M. J. (1993). Penetration of proton beams through water, I. Depth-dose distribution, spectra and LET distribution. *National Institute of Standards and Technology Report*, (NSTIR 5226 Gaithersburg: NIST). 74

Bethe, H. and Ashkin, J. (1953). The Passage of Heavy Particles Through Matter. In Segre, E., editor, *Experimental Nuclear Physics*. Wiley, New York. 24

Bichsel, H. (1972). Passage of charged particles through matter. In Gray, E. D., editor, *American Institute of Physics Handbook*. McGraw-Hill, New York, 3rd edition. 63, 66, 70, 73, 78, 86, 143

Bichsel, H. and Hiraoka, T. (1992). Energy loss of 70 MeV protons in elements. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 66(3):345–351. 70

Bohndiek, S., Arvanitis, C., Royle, G., Speller, R., Clark, A., Crooks, J., Prydderch, M., Turchetta, R., Blue, A., and O'Shea, V. (2007). Characterization studies of two novel active pixel sensors. *Optical Engineering*, 46(12):124003. 97

Boone, J. and Chavez, A. (1997). Comparison of x-ray cross sections for diagnostic and therapeutic medical physics. *Medical Physics*, 23(12). 63, 98

Bortfeld, T. (1997). An analytical approximation of the Bragg curve for therapeutic proton beams. *Medical Physics*, 24(12):2024–33. 16, 21, 22, 75, 92

Bourque, A. E., Carrier, J.-F., and Bouchard, H. (2014). A stoichiometric calibration method for dual energy computed tomography. *Physics in Medicine and Biology*, 59(8):2059–88. 68, 140

Bouyon-Monteau, A., Habrand, J.-L., Datchary, J., Alapetite, C., Bolle, S., Dendale, R., Feuvret, L., Helfre, S., Calugaru, V., Cosset, J.-M., and Bey, P. (2010). Is proton beam therapy the future of radiotherapy? Part I: Clinical aspects. *Cancer Radiothérapie*, 14:727–38. 10

Braccini, S., Ereditato, a., Kreslo, I., Moser, U., Pistillo, C., Studer, S., Scampoli, P., Coray, a., and Pedroni, E. (2010). First results on proton radiography with nuclear emulsion detectors. *Journal of Instrumentation*, 5(09):P09001–P09001. 89

Carlsson, C. and Carlsson, G. (1977). Proton dosimetry with 185MeV protons. Dose buildup from secondary protons and recoil electrons. *Health Physics*, 33:481–484. 49

Chen, G. T., Singh, R. P., Castro, J. R., Lyman, J. T., and Quivey, J. M. (1979). Treatment planning for heavy ion radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 5(10):1809–19. 67

Chen, W., Unkelbach, J., Trofimov, A., Madden, T., Kooy, H., Bortfeld, T., and Craft, D. (2012). Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Physics in Medicine and Biology*, 57(3):591–608. 57

Cheng, C., Zhao, L., and Wolanski, M. (2012). Comparison of tissue characterization curves for different CT scanners: implication in proton therapy treatment planning. *Translational Cancer Research*, 1(4):236–46. 68

Chung, C., Keating, N., and Yock, T. (2008). Comparative analysis of second malignancy risk in patients treated with proton therapy versus conventional photon therapy. *International Journal of Radiation Oncology Biology Physics*, 72(1):S8. 10

Chvetsov, A. V. and Paige, S. L. (2010). The influence of CT image noise on proton range calculation in radiotherapy planning. *Physics in Medicine and Biology*, 55(6):N141–9. 69, 121

Combs, S. E., Ganswindt, U., Foote, R. L., Kondziolka, D., and Tonn, J.-C. (2012). State-of-the-art treatment alternatives for base of skull meningiomas: complementing and controversial indications for neurosurgery, stereotactic and robotic based radiosurgery or modern fractionated radiation techniques. *Radiation Oncology*, 7(1):226. 16

Constantinou, C., Harrington, J., and DeWerd, L. (1992). An electron density calibration phantom for CT-based treatment planning computers. *Medical Physics*, 19(2):325–27. 67

Cormack, A. (1963). Representation of a Function by Its Line Integrals, with Some Radiological Applications. *Journal of Applied Physics*, 34(9):2722–7. 88

Cozzi, L., Fogliata, A., Lomax, A., and Bolsi, A. (2001). A treatment planning comparison of 3D conformal therapy, intensity modulated photon therapy and proton therapy for treatment of advanced head and neck tumours. *Radiotherapy and Oncology*, 61(3):287–97. 16

Damato, B., Kacperek, A., Chopra, M., Campbell, I. R., and Errington, R. D. (2005). Proton beam radiotherapy of choroidal melanoma: the Liverpool-Clatterbridge experience. *International Journal of Radiation Oncology Biology Physics*, 62(5):1405–11. 10

Deasy, J. O., Blanco, A. I., and Clark, V. H. (2003). CERR: A computational environment for radiotherapy research. *Medical Physics*, 30(5):979–85. 36

Depauw, N., Dias, M. F., Rosenfeld, A., and Seco, J. C. (2014). Ion radiography as a tool for patient set-up & image guided particle therapy: a Monte Carlo study. *Technology in Cancer Research & Treatment*, 13:69–76. 88

Doolan, P., Rosenberg, I., Ainsley, C., Gibson, A., and Royle, G. (2013). A comparison of the beam configuration modules of two proton treatment planning systems. *ESTRO 2nd Forum*, pages PD–0570. 24, 26

Eldesoky, I., Attalla, E., Elshemey, W., and Zaghloul, M. (2012). A comparison of three commercial IMRT treatment planning systems for selected paediatric cases. *Journal of Applied Clinical Medical Physics*, 13(2):124–35. 16

Emfietzoglou, D., Garcia-Molina, R., Kyriakou, I., Abril, I., and Nikjoo, H. (2009). A dielectric response study of the electronic stopping power of liquid water for energetic protons and a new I-value for water. *Physics in Medicine and Biology*, 54(11):3451–72. 70

Esposito, M., Anaxagoras, T., and Diaz, O. (2012). Radiation hardness of a large area CMOS Active Pixel Sensor for bio-medical applications. *IEEE Nuclear Science and Medical Imaging Conference Record (NSS/MIC)*, N14(183):1300–4. 98

Esposito, M., Anaxagoras, T., Price, T., Manolopoulos, S., Evans, P. M., Wells, K., and Allinson, N. M. (2014). Ionizing and non ionizing radiation damage in a large area CMOS active pixel sensor for medical applications. *IEEE Transactions on Nuclear Science*, pages 1–8. 98

Faulkner, K. and Moores, B. (1985). Variations in measured computed tomography number values. *Radiography*, 51(597):163–7. 68

Fogliata, A., Bolsi, A., and Cozza, L. (2003). Comparative analysis of intensity modulation inverse planning modules of three commercial treatment planning systems applied to head and neck tumour model. *Radiotherapy and Oncology*, 66:29–40. 16, 34

Fogliata, A., Nicolini, G., Alber, M., Asell, M., Dobler, B., El-Haddad, M., Hå rdemark, B., Jelen, U., Kania, A., Larsson, M., Lohr, F., Munger, T., Negri, E., Rodrigues, C., and Cozzi, L. (2005). IMRT for breast. a planning study. *Radiotherapy and Oncology*, 76(3):300–10. 16, 34

Fogliata, A., Vanetti, E., Albers, D., Brink, C., Clivio, A., Knöös, T., Nicolini, G., and Cozzi, L. (2007). On the dosimetric behaviour of photon dose calculation algorithms in the presence of simple geometric heterogeneities: comparison with Monte Carlo calculations. *Physics in Medicine and Biology*, 52(5):1363–85. 16

Fredriksson, A., Forsgren, A., and Hardemark, B. (2011). Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical Physics*, 38(3):1672–84. 57

Friedman, P. S., Ball, R., Beene, J. R., Benhammou, Y., Bentefour, E. H., Chapman, J. W., Etzion, E., Ferretti, C., Guttman, N., Levin, D. S., Moshe, M. B., Silver, Y., Varner, R. L., Weaverdyck, C., and Zhou, B. (2012). Plasma panel sensors for particle and beam detection. *IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, d:1775–1780. 89

Ganschow, S., Klimm, D., and Bertram, R. (2011). On the effect of oxygen partial pressure on the chromium distribution coefficient in melt-grown ruby crystals. *Journal of Crystal Growth*, 325(1):81–4. 98

Gelover-Reyes, E., Jimenez-Spang, F., Rosenberg, I., and Royle, G. (2011a). Proton radiography with CMOS APS. *Radiotherapy and Oncology*, 99:S531. 139

Gelover-Reyes, E., Jimenez-Sprang, F., Rosenberg, I., D'Souza, D., and Royle, G. (2011b). Proton radiography with silicon active sensors. In *ESTRO Proceedings*, page S531, London. 89

Gottschalk, B. (2004). *Passive Beam Spreading in Proton Radiation Therapy*. Unpublished. 49

Gottschalk, B. (2014). Skewness and kurtosis as measures of range mixing in time resolved diode dosimetry. *arXiv:1407.2931*, pages 1–39. 140

Gottschalk, B., Koehler, A., Schneider, R., Sisterson, J., and Wagner, M. (1993). Multiple Coulomb scattering of 160 MeV protons. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, B74:467–90. 16

Gottschalk, B., Tang, S., Bentefour, E. H., Cascio, E., Prieels, D., and Lu, H.-M. (2011). Water equivalent path length measurement in proton radiotherapy using time resolved diode dosimetry. *Medical Physics*, 38(4):2282–8. 12, 90, 114, 115, 122, 137

Grant, R., Summers, P., and Neihart, J. (2014). Relative stopping power measurements to aid in the design of anthropomorphic phantoms for proton radiotherapy. *Journal of Applied Clinical Medical Physics*, 15(2):121–6. 67

Grassberger, C., Dowdell, S., Lomax, A., Sharp, G., Shackleford, J., Choi, N., Willers, H., and Paganetti, H. (2013). Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. *International Journal of Radiation Oncology Biology Physics*, 86(2):380–6. 17

Hall, E. (2006). Intensity-modulated radiation therapy, protons, and the risk of second cancers. *International Journal of Radiation Oncology Biology Physics*, 65(1):1–7. 10

Hansen, D., Sorensen, T., Seco, J., Verhaegen, F., and Landry, G. (2014). Beyond the stoichiometric method: a comparison of dual energy CT and proton CT for clinical stopping power estimation. *Physics in Medicine and Biology*, (Submitted. 68, 140

Heimann, J. (2005). *Developing an FPGA-based readout for the pCT detector system*. PhD thesis, University of California. 89

Highland, V. (1975). Some practical remarks on multiple scattering. *Nuclear Instruments and Methods*, 29:497–499. 138

Hollebeek, R., Newcomer, M., Mayers, G., Delgado, B., Shukla, G., Maughan, R., and Dolney, D. (2012). A new technology for fast two-dimensional detection of proton therapy beams. *Physics Research International*, 2012(i):1–11. 89

Hong, L., Goitein, M., Bucciolini, M., Comiskey, R., Gottschalk, B., Rosenthal, S., Serago, C., and Urie, M. (1996). A pencil beam algorithm for proton dose calculations. *Physics in Medicine and Biology*, 41(8):1305–30. 74

Hünemohr, N., Krauss, B., Tremmel, C., Ackermann, B., Jäkel, O., and Greilich, S. (2014). Experimental verification of ion stopping power prediction from dual energy CT data in tissue surrogates. *Physics in Medicine and Biology*, 59(1):83–96. 68, 140

Hurley, R. F., Schulte, R. W., Bashkirov, V. A., Wroe, A. J., Ghebremedhin, A., Sadrozinski, H. F.-W., Rykalin, V., Coutrakon, G., Koss, P., and Patyal, B. (2012). Water-equivalent path length calibration of a prototype proton CT scanner. *Medical Physics*, 39(5):2438–46. 88

ICRP (1975). International Commission on Radiation Protection Report 23: Reference man: anatomical, physiological and metabolic characteristics. 23:273–334. 67

ICRU (1989). International Commission on Radiation Units and Measurements Report 44: Tissue substitutes on radiation units and measurement. 62, 70, 112, 118, 119, 121

ICRU (1992). International Commission on Radiation Units and Measurements Report 46: Photon, electron, proton and neutron interaction data for body tissues. 70, 77

ICRU (1993). International Commission on Radiation Units and Measurements Report 49: Stopping powers and ranges for protons and alpha particles. 64, 66, 70, 72, 73, 77, 78, 86, 143

ICRU (2009). Errata and Addenda for ICRU Report 73, Stopping of ions heavier than helium. 72, 77

Isacsson, U., Montelius, A., Jung, B., and Glimelius, B. (1996). Comparative treatment planning between proton and X-ray therapy in locally advanced rectal cancer. *Radiotherapy and Oncology*, 41(3):263–72. 10

Jakel, O., Jacob, C., Schardt, D., Karger, C. P., and Hartmann, G. H. (2001). Relation between carbon ion ranges and x-ray CT numbers. *Medical Physics*, 28(4):701. 67

Janni, F. (1982). Part 1 - Proton range-energy tables, 1 keV-10 GeV. Energy loss, range, path length, time-of-flight, straggling, multiple scattering and nuclear interaction probability. *Atomic Data and Nuclear Data Tables*, 27:147–339. 64, 66, 70, 73, 78, 86, 143

Kacperek, A. (2009). Proton therapy of eye tumours in the UK: a review of treatment at Clatterbridge. *Applied radiation and isotopes*, 67(3):378–86. 10

Kanematsu, N., Matsufuji, N., Kohno, R., Minohara, S., and Kanai, T. (2003). A CT calibration method based on the polybinary tissue model for radiotherapy treatment planning. *Physics in Medicine and Biology*, 48:1053–64. 67, 68, 69

Kleihues, P., Louis, D., Scheithauer, B., Rorke, L., Reifenberger, G., Burger, P., and Cavenee, W. (2002). The WHO classification of tumors of the nervous system. *Journal of Neuropathology and Experimental Neurology*, 61(3):215–25. 19

Knopf, A.-C. and Lomax, A. (2013). In vivo proton range verification: a review. *Physics in Medicine and Biology*, 58(15):R131–R160. 12

Konstantinidis, A. C., Olivo, A., Munro, P. R., Bohndiek, S. E., and Speller, R. D. (2010). Optical characterisation of a CMOS active pixel sensor using periodic noise reduction techniques. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 620:549–56. 99

Kooy, H. M., Clasie, B. M., Lu, H.-M., Madden, T. M., Bentefour, H., Depauw, N., Adams, J. A., Trofimov, A. V., Demaret, D., Delaney, T. F., and Flanz, J. B. (2010). A case study in proton pencil-beam scanning delivery. *International Journal of Radiation Oncology Biology Physics*, 76(2):624–30. 16

Kumazaki, Y., Akagi, T., Yanou, T., Suga, D., Hishikawa, Y., and Teshima, T. (2007). Determination of the mean excitation energy of water from proton beam ranges. *Radiation Measurements*, 42(10):1683–91. 70

Lagarias, J., Reeds, J., Wright, M., and Wright, P. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–47. 117

Lee, M., Wynne, C., Webb, S., Nahum, A., and Dearnaley, D. (1994). A comparison of proton and megavoltage X-ray treatment planning for prostate cancer. *Radiotherapy and Oncology*, 33(3):239–53. 10

Leroy, C. and Rancoita, P. G. (2009). *Principles of Radiation Interaction in Matter and Detection*. World Scientific. 138

Levi, C., Gray, J. E., McCullough, E. C., and Hattery, R. R. (1982). The unreliability of CT numbers as absolute values. *AJR. American Journal of Roentgenology*, 139(3):443–7. 68

Li, T., Liang, Z., Singanallur, J., Satogata, T., Williams, D., and Schulte, R. (2006). Reconstruction for proton computed tomography by tracing proton trajectories: A Monte Carlo study. *Medical Physics*, 33(3):699–706. 138

Lin, R., Hug, E., Schaefer, R., Miller, D., and Slater, J. (2000). Conformal proton radiation therapy of the posterior fossa: a study comparing protons with three-dimensional planned photons in limiting dose to auditory structures. *International Journal of Radiation Oncology Biology Physics*, 48(4):1219–26. 20

Liu, P. Z. Y., Suchowerska, N., Lambert, J., Abolfathi, P., and McKenzie, D. R. (2012a). Reply to the comment on: 'Plastic scintillation dosimetry: comparison of three solutions for the Cerenkov challenge'. *Physics in Medicine and Biology*, 57(11):3667–73. 57

Liu, W., Li, Y., and Li, X. (2012b). Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques. *Medical Physics*, 39(6):3089–101. 18

Lomax, A. (1999). Intensity modulation methods for proton radiotherapy. *Physics in Medicine and Biology*, 44:185–205. 18

Lomax, A., Pedroni, E., Schaffner, B., Scheib, S., Schneider, U., and Tourovsky, A. (1996). 3D treatment planning for conformal proton therapy by spot scanning. In Faulkner, K., Carey, B., Crellin, A., and Harrison, R., editors, *Quantitative Imaging in Oncology*, pages 67–71. BIR Publishing, London. 35

Lomax, A. J. (2008). Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Physics in Medicine and Biology*, 53(4):1043–56. 26, 57

Low, D., Harms, W., Mutic, S., and Purdy, J. (1998). A technique for the quantitative evaluation of dose distributions. *Medical Physics*, 25(5):656–61. 29, 147

Lu, H.-M. (2008a). A point dose method for in vivo range verification in proton therapy. *Physics in Medicine and Biology*, 53(23):N415–22. 90

Lu, H.-M. (2008b). A potential method for in vivo range verification in proton therapy treatment. *Physics in Medicine and Biology*, 53(5):1413–24. 12, 90, 113, 114, 115, 137

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. *Medical Physics*, 37(11):5858. 90

Lynch, G., Dahl, O., and Berkeley, L. (1991). Approximations to multiple Coulomb scattering. *Nuclear Instruments and Methods in Physics Research*, B58:6–10. 16

Matsufuji, N., Tomura, H., Futami, Y., Yamashita, H., Higashi, A., Minohara, S., Endo, M., and Kanai, T. (1998). Relationship between CT number and electron density, scatter angle and nuclear reaction for hadron-therapy treatment planning. *Physics in Medicine and Biology*, 43:3216–3275. 67

McGowan, S. E., Burnet, N. G., and Lomax, A. J. (2013). Treatment planning optimisation in proton therapy. *The British Journal of Radiology*, 86(1021):20120288. 18

Merchant, T. E. (2009). Proton beam therapy in pediatric oncology. *The Cancer Journal*, 15(4):298–305. 10

Meroli, S., Passeri, D., Servoli, L., and Angelucci, A. (2013). Analysis of the performance of CMOS APS imagers after proton damage. *Journal of Instrumentation*, 8(02):1–5. 98

Meyer, J., Bluett, J., Amos, R., Levy, L., Choi, S., Nguyen, Q.-N., Zhu, X. R., Gillin, M., and Lee, A. (2010). Spot scanning proton beam therapy for prostate cancer: treatment planning technique and analysis of consequences of rotational and translational alignment errors. *International Journal of Radiation Oncology Biology Physics*, 78(2):428–34. 57

Min, C.-H., Kim, C. H., Youn, M.-Y., and Kim, J.-W. (2006). Prompt gamma measurements for locating the dose falloff region in the proton therapy. *Applied Physics Letters*, 89(18):183517. 12

Miralbell, R., Lomax, A., and Bortfeld, T. (1997). Potential role of proton therapy in the treatment of pediatric medulloblastoma/primitive neuro-ectodermal tumors:

reduction of the supratentorial target volume. *International Journal of Radiation Oncology Biology Physics*, 38(3):477–84. 10

Molière, G. (1948). Theorie der Streuung schneller geladener Teilchen II. Mehrfach und Vielfachstreuung. *Z. Naturforschung*, 3a:78. 24

Moyers, M., Miller, D., Bush, D., and Slater, J. (2001). Methodologies and tools for proton beam design for lung tumors. *International Journal of Radiation Oncology Biology Physics*, 49(5):1429–38. 10

Muraishi, H., Nishimura, K., Abe, S., Satoh, H., Hara, S., Hara, H., Takahashi, Y., Mogaki, T., Kawai, R., Yokoyama, K., Yasuda, N., Tomida, T., Ohno, Y., and Kanai, T. (2009). Evaluation of spatial resolution for heavy ion CT system based on the measurement of residual range distribution with HIMAC. *IEEE Transactions on Nuclear Science*, 56(5):2714–21. 89, 139

Noa, K., Dolney, D., Baumann, B., Bekelman, J., Deville, C., Vapiwala, N., Both, S., Kassaee, A., and Christodouleas, J. (2012). Adjuvant radiation for bladder cancer: a dosimetry study. *International Journal of Radiation Oncology Biology Physics*, 84(3):S420. 16

Nowotny, R. (1998). XMuDat: Photon attenuation data on PC. In *IAEA-NDS-195*, Vienna, Austria. Available on diskette or at http://www-mds.iaea. 63

Oelfke, U. and Bortfeld, T. (2000). Intensity modulated radiotherapy with charged particle beams: studies of inverse treatment planning for rotation therapy. *Medical Physics*, 27(6):1246–57. 18

Paganetti, H. (2012). Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in Medicine and Biology*, 57(11):R99–R117. 10, 11, 12, 61, 70, 71, 75, 88, 94, 106

Parodi, K., Ponisch, F., and Enghardt, W. (2005). Experimental Study on the Feasibility of In-Beam PET for Accurate Monitoring of Proton Therapy. *IEEE Transactions on Nuclear Science*, 52(3):778–86. 12

Paul, H., Geithner, O., and Jakel, O. (2007). The influence of stopping powers upon dosimetry for radiation therapy with energetic ions. In *Advances in Quantum Chemistry*, volume 52, pages 289–306. Elsevier Inc. 70

Peggs, S. (2002). Proton therapy accelerators. *IEEE Nuclear Science Symposium Conference Record*, 2:654. 10

Pemler, P., Besserer, J., Boer, J., Dellert, M., Gahn, C., Moosburger, M., Schneider, U., Pedroni, E., and Stauble, H. (1999). A detector system for proton radiography

on the gantry of the Paul-Scherrer-Institute. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 432(2-3):483–95. 89

Perl, J., Shin, J., Schumann, J., Faddegon, B., and Paganetti, H. (2012). TOPAS: an innovative proton Monte Carlo platform for research and clinical applications. *Medical physics*, 39(11):6818–37. 121

Petterson, M., Blumenkrantz, N., Feldt, J., Heimann, J., Lucia, D., Seiden, A., Williams, D., Sadrozinski, H., Bashkirov, V., Schulte, R., Bruzzi, M., Menichelli, D., Scaringella, M., Talamonti, C., Cirrone, G., Cuttone, G., Lo Presti, D., Randazzo, N., and Sipala, V. (2006). Proton radiography studies for proton CT. In *IEEE Nuclear Science Symposium Conference Record*, pages 2276–80. 89

Pflugfelder, D., Wilkens, J. J., and Oelfke, U. (2008). Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. *Physics in Medicine and Biology*, 53(6):1689–700. 57

Pflugfelder, D., Wilkens, J. J., Szymanowski, H., and Oelfke, U. (2007). Quantifying lateral tissue heterogeneities in hadron therapy. *Medical Physics*, 34(4):1506. 57

Philips Medical Systems (2013). Pinnacle3 IMPT/Spot Scanning Proton Treatment Planning Prototype User Manual. 36

Poludniowski, G., Allinson, N. M., Anaxagoras, T., Esposito, M., Green, S., Manolopoulos, S., Nieto-Camero, J., Parker, D. J., Price, T., and Evans, P. M. (2014). Proton-counting radiography for proton therapy: a proof of principle using CMOS APS technology. *Physics in Medicine and Biology*, 59(11):2569–81. 89, 97, 139

PTCOG (2013). Proton therapy patient statistics (as of Dec 2013). 10

PTW (2013). Ionising Radiation Detectors. 27

Qi, Z.-Y., Huang, S.-M., and Deng, X.-W. (2006). Calibration of CT values used for radiation treatment planning and its impact factors. *Chinese Journal of Cancer*, 25(1):110–114. 68, 69

Rana, S., Zeidan, O., Ramirez, E., Rains, M., Gao, J., and Zheng, Y. (2013). Measurements of lateral penumbra for uniform scanning proton beams under various beam delivery conditions and comparison to the XiO treatment planning system. *Medical Physics*, 40(9):091708. 16, 57

Rietzel, E., Liu, A. K., Doppke, K. P., Wolfgang, J. A., Chen, A. B., Chen, G. T. Y., and Choi, N. C. (2006). Design of 4D treatment planning target volumes. *International Journal of Radiation Oncology Biology Physics*, 66(1):287–95. 12

Rinaldi, I., Brons, S., Gordon, J., Panse, R., Voss, B., Jäkel, O., and Parodi, K. (2013). Experimental characterization of a prototype detector system for carbon ion radiography and tomography. *Physics in Medicine and Biology*, 58(3):413–27. 89

Rinaldi, I., Brons, S., Jäkel, O., Voss, B., and Parodi, K. (2014). Experimental investigations on carbon ion scanning radiography using a range telescope. *Physics in Medicine and Biology*, 59(12):3041–57. 89

Ryu, H., Song, E., Lee, J., and Kim, J. (2008). Density and spatial resolutions of proton radiography using a range modulation technique. *Physics in Medicine and Biology*, 53(19):5461–8. 89, 139

Samuel, D., Testa, M., Lee, J., Cascio, E., Orban de Xivry, J., Gottschalk, B., Bentefour, E. H., and Lu, H.-M. (2013). Workflow of the in vivo range verification in proton therapy. In *American Association of Physicists in Medicine 55th Annual Meeting & Exhibition*. 115

Schaffner, B. (2008). Proton dose calculation based on in-air fluence measurements. *Physics in Medicine and Biology*, 53(6):1545–62. 16

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. *Physics in Medicine and Biology*, 43:1579–92. 61, 62, 68, 69, 71, 88

Schaffner, B., Pedroni, E., and Lomax, A. (1999). Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation. *Physics in Medicine and Biology*, 44:27–41. 16

Schneider, U., Besserer, J., and Hartmann, M. (2012). Technical note: spatial resolution of proton tomography: impact of air gap between patient and detector. *Medical Physics*, 39(2):798–800. 89

Schneider, U., Besserer, J., Pemler, P., Dellert, M., Moosburger, M., Pedroni, E., and Kaser-Hotz, B. (2004). First proton radiography of an animal patient. *Medical Physics*, 31(5):1046–51. 88, 89

Schneider, U. and Pedroni, E. (1995). Proton radiography as a tool for quality control in proton therapy. *Medical Physics*, 22(4):353–63. 71, 89

Schneider, U., Pedroni, E., and Lomax, A. (1996). The calibration of CT Hounsfield units for radiotherapy treatment planning. *Physics in Medicine and Biology*, 41(1):111–124. 26, 62, 65, 67, 87, 98, 143

Schneider, U., Pemler, P., Besserer, J., Pedroni, E., Lomax, A., and Kaser-Hotz, B. (2005). Patient specific optimization of the relation between CT-Hounsfield units and proton stopping power with proton radiography. *Medical Physics*, 32(1):195–9. 70, 88, 112

Schneider, W., Bortfeld, T., and Schlegel, W. (2000). Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions. *Physics in Medicine and Biology*, 45:459–78. 117

Schulte, R., Bashkirov, V., Loss Klock, M., Li, T., Wroe, A., Evseev, I., Williams, D., and Satogata, T. (2005). Density resolution of proton computed tomography. *Medical Physics*, 32(4):1035–46. 88

Schulte, R., Bashkirov, V., Mueller, K., Heimann, J., Johnson, L., Keeney, B., Sadrozinski, H., Seiden, A., Williams, D., Peggs, S., Satogata, T., and Woody, C. (2004). Conceptual design of a proton computed tomography system for applications in proton radiation therapy. *IEEE Transactions on Nuclear Science*, 51(3):866–72. 89

Schwarz, M. (2011). Treatment planning in proton therapy. *The European Physical Journal Plus*, 126:67–76. 19

Seco, J. and Depauw, N. (2011). Proof of principle study of the use of a CMOS active pixel sensor for proton radiography. *Medical Physics*, 38(2):622–3. 88, 89, 97, 139

Seco, J., Panahandeh, H. R., Westover, K., Adams, J., and Willers, H. (2012). Treatment of non-small cell lung cancer patients with proton beam-based stereotactic body radiotherapy: dosimetric comparison with photon plans highlights importance of range uncertainty. *International Journal of Radiation Oncology Biology Physics*, 83(1):354–61. 84

Seltzer, S. and Berger, M. (1982). Evaluation of the collision stopping power of elements and compounds for electrons and positrons. *The International Journal of Applied Radiation and Isotopes*, 33(11):1189–1218. 70, 77

Shinoda, H., Kanai, T., and Kohno, T. (2006). Application of heavy-ion CT. *Physics in Medicine and Biology*, 51(16):4073–81. 61, 68, 89

Siddon, R. L. (1985). Prism representation: a 3D ray-tracing algorithm for radiotherapy applications. *Physics in Medicine and Biology*, 30(8):817–24. 116

Sipala, V., Brianzi, M., Bruzzi, M., Bucciolini, M., Cirrone, G., Civinini, C., Cuttone, G., Lo Presti, D., Pallotta, S., Randazzo, N., Romano, F., Stancampiano, C., Scaringella, M., Talamonti, C., and Tesi, M. (2011). PRIMA: An apparatus for medical application. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 658:73–7. 89

Slopsema, R. (2009). Basic Physics of Proton Therapy. In *American Association of Physicists in Medicine 50th Annual Meeting & Exhibition*. 114

Soukup, M., Fippel, M., and Alber, M. (2005). A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations. *Physics in medicine and biology*, 50(21):5089–104. 16, 36

St Clair, W. H., Adams, J., Bues, M., Fullerton, B., La Shell, S., Kooy, H., Loeffler, J., and Tarbell, N. (2004). Advantage of protons compared to conventional X-ray or IMRT in the treatment of a pediatric patient with medulloblastoma. *International Journal of Radiation Oncology Biology Physics*, 58(3):727–34. 16

Stewart, G. D. (2013). *Silicon Pixel Detectors for Synchrotron Applications*. PhD thesis, University of Glasgow. 99

Stuschke, M., Kaiser, A., Pöttgen, C., Lübcke, W., and Farr, J. (2012). Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to intensity modulated photon plans. *Radiotherapy and Oncology*, 104(1):45–51. 16

Talamonti, C., Reggioli, V., Bruzzi, M., Bucciolini, M., Civinini, C., Marrazzo, L., Menichelli, D., Pallotta, S., Randazzo, N., and Sipala, V. (2010). Proton radiography for clinical applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 612(3):571–5. 89

Tang, S., Both, S., Bentefour, E. H., and Tochner, Z. (2010). Anterior fields improve rectal sparing in prostate treatment by proton therapy. In *American Association of Physicists in Medicine 52nd Annual Meeting & Exhibition*, volume 37, page 3297. 114

Tang, S., Both, S., Bentefour, H., Paly, J., Tochner, Z., Efstathiou, J., and Lu, H.-M. (2011). Improvement of prostate treatment by anterior proton fields. *International Journal of Radiation Oncology Biology Physics*, 83(1):408–18. 75, 114, 120

Tang, S., Deville, C., McDonough, J., Tochner, Z., Wang, K. K.-H., Vapiwala, N., and Both, S. (2013). Effect of intrafraction prostate motion on proton pencil beam

scanning delivery: a quantitative assessment. *International Journal of Radiation Oncology Biology Physics*, 87(2):375–82. 16

Telsemeyer, J. (2012). *Investigation of an Amorphous Silicon Flat-Panel Detector for Ion Radiography*. PhD thesis, University of Heidelberg. 89, 139

Testa, M., Verburg, J. M., Rose, M., Min, C. H., Tang, S., Bentefour, E. H., Paganetti, H., and Lu, H.-M. (2013). Proton radiography and proton computed tomography based on time-resolved dose measurements. *Physics in Medicine and Biology*, 58(22):8215–8233. 88, 90, 115

Thomas, S. (2006). Margins for treatment planning of proton therapy. *Physics in Medicine and Biology*, 51:1491–1501. 12

Trofimov, A., Nguyen, P. L., Coen, J. J., Doppke, K. P., Schneider, R. J., Adams, J. A., Bortfeld, T. R., Zietman, A. L., Delaney, T. F., and Shipley, W. U. (2007). Radiotherapy treatment of early-stage prostate cancer with IMRT and protons: a treatment planning comparison. *International Journal of Radiation Oncology Biology Physics*, 69(2):444–53. 16, 32

Ulmer, W. and Matsinos, E. (2011). Theoretical methods for the calculation of Bragg curves and 3D distributions of proton beams. *The European Physical Journal Special Topics*, 190(1):1–81. 16

Ulmer, W. and Schaffner, B. (2011). Foundation of an analytical proton beamlet model for inclusion in a general proton dose calculation system. *Radiation Physics and Chemistry*, 80(3):378–89. 16, 23

Unkelbach, J., Bortfeld, T., Martin, B. C., and Soukup, M. (2009). Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. *Medical Physics*, 36(1):149–63. 57

Urie, M., Goitein, M., Holley, W. R., and Chen, G. T. Y. (1986). Degradation of the Bragg peak due to inhomogeneities. *Physics in medicine and . . .*, 31(1):1–15. 11

Vanzi, E., Bruzzi, M., Bucciolini, M., Cirrone, G. P., Civinini, C., Cuttone, G., Lo Presti, D., Pallotta, S., Pugliatti, C., Randazzo, N., Romano, F., Scaringella, M., Sipala, V., Stancampiano, C., Talamonti, C., and Zani, M. (2013). The PRIMA collaboration: Preliminary results in FBP reconstruction of pCT data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 730:184–90. 89

Varian (2011). Proton Algorithm Reference Guide: Eclipse. 23, 27, 35, 36, 50

Verburg, J. M., Riley, K., Bortfeld, T., and Seco, J. (2013). Energy- and time-resolved detection of prompt gamma-rays for proton range verification. *Physics in Medicine and Biology*, 58(20):L37–49. 12

Verburg, J. M. and Seco, J. (2012). CT metal artifact reduction method correcting for beam hardening and missing projections. *Physics in Medicine and Biology*, 57(9):2803–18. 69

Watanabe, Y. (1999). Derivation of linear attenuation coefficients from CT numbers for low-energy photons. *Physics in Medicine and Biology*, 44:2201–11. 65, 73

Watts, D., Amaldi, U., Go, A., Chang, Y.-H., Hajdas, W., Iliescu, S., Malakhov, N., Samarati, J., and Sauli, F. (2009). A proton range telescope for quality assurance in hadrontherapy. In *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 4163–6. 89

Weber, D., Lomax, A., Rutz, H., Stadelmann, O., Egger, E., Timmermann, B., Pedroni, E., Verwey, J., Miralbell, R., and Goitein, G. (2004a). Spot-scanning proton radiation therapy for recurrent, residual or untreated intracranial meningiomas. *Radiotherapy and Oncology*, 71:251–8. 20

Weber, D., Trofimov, A., Delaney, T., and Bortfeld, T. (2004b). A treatment planning comparison of intensity modulated photon and proton therapy for paraspinal sarcomas. *International Journal of Radiation Oncology Biology Physics*, 58(5):1596–606. 26, 54

White, D. and Widdowson, E. (1991). The composition of body tissues, (II) Fetus to young adult. *British Journal of Radiology*, 64:149–59. 70

White, D., Woodard, H., and Hammond, S. (1987). Average soft-tissue and bone models for use in radiation dosimetry. *British Journal of Radiology*, 60:907–13. 62, 70, 77, 112, 118, 119, 121

Williams, D. (2004). The most likely path of an energetic charged particle through a uniform medium. *Physics in Medicine and Biology*, 49:2899–911. 138

Wilson, R. (1946). Radiological use of fast protons. *Radiology*, 47(5):487–91. 10

Woodard, H. and White, D. (1986). The composition of body tissues. *British Journal of Radiology*, 59:1209–18. 62, 70, 77, 112, 118, 119, 121

Woodard, H. Q. and White, D. R. (1982). Bone models for use in radiotherapy dosimetry. *British Journal of Radiology*, 55(652):277–82. 70

Yang, M., Virshup, G., Clayton, J., Zhu, X. R., Mohan, R., and Dong, L. (2010). Theoretical variance analysis of single-and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues. *Physics in Medicine and Biology*, 55:1343–62. 70, 112

Yang, M., Virshup, G., Clayton, J., Zhu, X. R., Mohan, R., and Dong, L. (2011). Does kV-MV dual-energy computed tomography have an advantage in determining proton stopping power ratios in patients? *Physics in Medicine and Biology*, 56(14):4499–515. 68

Yang, M., Zhu, X. R., Park, P. C., Titt, U., Mohan, R., Virshup, G., Clayton, J. E., and Dong, L. (2012). Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration. *Physics in Medicine and Biology*, 57(13):4095–115. 62, 69, 70, 71

Yeung, D. (2009). New developments in proton treatment planning systems. In *AAPM Symposium - The Promises and Perils of Proton Radiotherapy*. 19

Yeung, D., Mckenzie, C., and Indelicato, D. J. (2014). A dosimetric comparison of intensity-modulated proton therapy optimization techniques for pediatric craniopharyngiomas: a clinical case study. *Paediatric Blood Cancer*, 61:89–94. 16, 57

Yohannes, I., Kolditz, D., Langner, O., and Kalender, W. a. (2012). A formulation of tissue- and water-equivalent materials using the stoichiometric analysis method for CT-number calibration in radiotherapy treatment planning. *Physics in Medicine and Biology*, 57(5):1173–90. 67

Yoon, M., Shin, D. H., Kim, J., Kim, J. W., Kim, D. W., Park, S. Y., Lee, S. B., Kim, J. Y., Park, H.-J., Park, B. K., and Shin, S. H. (2011). Craniospinal irradiation techniques: a dosimetric comparison of proton beams with standard and advanced photon radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 81(3):637–46. 16

Zhang, X. (2012). Automatic IMRT, VMAT, IMPT treatment planning. In *American Association of Physicists in Medicine 54th Annual Meeting & Exhibition*. 57

Zhu, X., Palmer, M. R., Makrigiorgos, G. M., and Kassis, A. I. (2010). Solid-tumor radionuclide therapy dosimetry: New paradigms in view of tumor microenvironment and angiogenesis. *Medical Physics*, 37(6):2974–84. 26

Zhu, X. R., Poenisch, F., Lii, M., Sawakuchi, G. O., Titt, U., Bues, M., Song, X., Zhang, X., Li, Y., Ciangaru, G., Li, H., Taylor, M. B., Suzuki, K., Mohan, R.,

Gillin, M. T., and Sahoo, N. (2013). Commissioning dose computation models for spot scanning proton beams in water for a commercially available treatment planning system. *Medical Physics*, 40(4):041723. 16