IOP Publishing Phys. Med. Biol. 65 (2020) 03NT02 (11pp) https://doi.org/10.1088/1361-6560/ab66e1

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

NOTE

Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing

A Meijers[1,3,4], J Free[1], D Wagenaar[1], S Deffet[2], A C Knopf[1], J A Langendijk[1] and S Both

[1] Department of Radiation Oncology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
[2] Université catholique de Louvain, Icteam, Louvain-la-Neuve, Belgium
[3] Meijers Arturs, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
[4] Author to whom any correspondence should be addressed.

E-mail: a.meijers@umcg.nl

Keywords: proton therapy, range uncertainty, CT calibration

Supplementary material for this article is available online

RECEIVED
17 June 2019

REVISED
18 December 2019

ACCEPTED FOR PUBLICATION
2 January 2020

PUBLISHED
4 February 2020

Abstract

Proton therapy is affected by range uncertainty, which is partly caused by an ambiguous conversion from x-ray attenuation to proton stopping power. CT calibration curves, or Hounsfield look-up tables (HLUTs), are institution-specific and may be a source of systematic errors in treatment planning. A range probing method to verify, optimize and validate HLUTs for proton treatment is proposed.

An initial HLUT was determined according to the stoichiometric approach. For HLUT validation, three types of animal tissue phantoms were prepared: a pig's head, 'thorax' and femur. CT scans of the phantoms were taken and a structure, simulating a water slab, was added on the scan distal to the phantoms to mimic the detector used for integral depth-dose measurements. The CT scans were imported into the TPS to calculate individual pencil beams directed through the phantoms. The phantoms were positioned at the therapy system isocenter using x-ray imaging. Shoot-through pencil beams were delivered, and depth-dose profiles were measured using a multi-layer ionization chamber. Measured depth-dose curves were compared to the calculated curves and the range error per spot was determined. Based on the water equivalent path length (WEPL) of individual spot, a range error margin was defined. Ratios between measured error and theoretical margin were calculated per spot. The HLUT optimization was performed by identifying systematic shifts of the mean range error per phantom and minimizing the ratios between range errors and uncertainty margins.

After optimization, the ratios of the actual range error and the uncertainty margin over the complete data set did not exceed 0.75 (1.5 SD), indicating that the actual errors are covered by the theoretical uncertainty recipe.

The feasibility of using range probing to assess range errors was demonstrated. The theoretical uncertainty margins in the institution-specific setting potentially may be reduced by ~25%.

Introduction

Over the last decades, proton therapy is becoming a more widely available treatment modality. However, range uncertainty is commonly regarded as a significant concern. Major contributors to the range uncertainty are linked to the CT calibration, or conversion from CT number to SPR (direct or via mass density), and handling of lateral and longitudinal heterogeneities by the dose calculation engine of the treatment planning system (TPS). By quantifying various contributors, range uncertainty recipes have been proposed and these typically consist of a relative component (relative to the range of the beam) and an absolute component, which is largely influenced by proton beam delivery equipment (Paganetti 2012). The institution-specific range uncertainty

margin will depend on the utilized treatment modalities, the performance of the equipment and the choice of dose calculation engine (differentiating between analytical and Monte Carlo engines).

CT numbers are imperfect input for dose calculations due to images being affected by the noise and lack of an unambiguous assignment of tissue properties based on CT numbers (Unkelbach and Paganetti 2018). As CT calibration is one of the contributors to range uncertainty, typically an effort is invested in this task during the implementation phase of new proton treatment delivery equipment. There are two methodologies frequently used for CT calibration: the tissue-substitute method and the stoichiometric method (Schneider *et al* 1996). The tissue-substitute method relies on establishing a calibration curve based on scans of known density materials, which are meant to substitute specific human-like tissues and correlating those densities to measured CT number. The stoichiometric method also requires scans of tissue-equivalent materials. However, the measured data is used to characterize the CT via three fitting parameters. The obtained parameters are used to pre-calculate expected CT numbers for human-like tissues, considering their elemental composition.

A critical look at the stoichiometric method has recently been taken by Gomà *et al* (2018). It was pointed out that, in order to accurately perform fitting of the measured data and thereby characterize the CT acquisition properly, the fitting procedure itself must be mathematically constrained. However, the approach to set constraints is not well defined. Furthermore, it was shown that the outcome of the calibration may be affected by the selection of phantom for the initial scan. Gomà *et al* applied the stoichiometric method for two commercially available CT calibration phantoms of CIRS (Norfolk, VA, USA) and Gammex (Middleton, WI, USA). It was observed that in case of the CIRS phantom, the tissue-substitute method and the stoichiometric method resulted in two different HLUTs, especially in bone-like tissue section. While in case of the Gammex phantom, the tissue-substitute method and stoichiometric method resulted in nearly the same calibration curve, situated between the two calibration curves determined for the CIRS phantom. Due to the absence of a ground truth, it remains unclear, which of the obtained curves would be the most appropriate for actual clinical use.

The purpose of the current study was to propose and apply a methodology that would allow to verify, optimize and validate a HLUT and its performance using range estimations from the TPS and range probing (proton radiography) measurements.

Furthermore, the intent is to confirm that the range uncertainty recipe derived from literature is applicable in an institution-specific setting for the use of robust optimization during the treatment planning.

Material and methods

The following methodology was developed and applied to create, verify, optimize and validate a site-specific HLUT prior to implementation in the clinic:

1. Creation of an initial HLUT based on the stoichiometric method and its implementation in the TPS
2. Validation of the HLUT
   a. Acquisition of CT scans of animal tissue samples
   b. Execution of range probing calculations in TPS
   c. Performance of range probing measurements for the tissue samples
3. Optimization of the HLUT
   a. Comparison of range probing measurements with TPS calculations and definition of residual range errors
   b. Minimization of range errors and implementation of the optimized HLUT into the TPS
4. Validation of the optimized HLUT
   a. Re-calculation of range probing data in TPS
   b. Re-evaluation of the residual range errors (including independent sample)
   c. Release of the HLUT for clinical use

Initial HLUT

For proton dose calculation during the treatment planning phase, imaging data (most commonly, patient CT) are generally converted to relative proton stopping power ratio (SPR) maps. However, depending on the TPS and dose calculation engine, the user input for calculation of SPR maps may differ. For example, the conversion of CT number may be defined either directly to proton stopping power or to physical density. RayStation 7 (RaySearch Laboratories, Sweden) requires defined CT numbers versus predetermined physical densities as an input for its CT calibration. Afterwards, during the dose calculation itself, the TPS is using built-in material look-

Table 1. Summary of utilized CIRS inserts for definition of initial HLUT.

Insert label	Relative electron density	HLUT section	Small configuration, HU	Large configuration, HU
Lung (inhale)	0.19	Organ-like	−789	−780
Lung (exhale)	0.489	Organ-like	−511	−505
Adipose	0.949	Fat-like	−68	−66
Breast (50% gland/50% adipose)	0.976	Fat-like	−32	−31
Muscle	1.043	Organ-like	40	38
Liver	1.052	Organ-like	52	51
Trabecular bone	1.117	Bone-like	230	204
Dense bone	1.456	Bone-like	914	845

up tables (RaySearch Laboratories 2017) to convert mass density to SPR. According to RayStation Reference manual, built-in material look-up tables have been based on ICRU49, ICRU44 and ICRP23 report.

To create an initial HLUT, the stoichiometric approach was partially used (Schneider *et al* 1996). Since user-provided input only specifies a CT number to mass density curve, but mass density to SPR conversion is performed by TPS, the stoichiometric method was applied only to calculate the CT number to mass density curve. Furthermore, to confirm accurate SPR values calculated by the TPS, the conversion methodology (RayStation 7B) was replicated outside of the TPS and theoretical SPR values were calculated and compared to the values provided by the TPS. No significant variations (<0.2%) were observed.

For the initial CT number measurements, the tissue substitute phantom by CIRS model 062 M was used. The phantom consists of inner and outer cylinders, which can be filled with a set of inserts, made of materials that are substitutes for human-like tissues in terms of density, designed to establish the HLUT.

For every clinically used scan protocol (and reconstruction kernel) the CIRS phantom was scanned in two configurations: (1) only the middle cylinder (representing a small object) and (2) the full phantom, consisting of both cylinders (representing a large object). It was observed that variations between the HU versus mass density curves for large and small objects were most pronounced in calcium rich regions. Overall, the variations due to the choice of reconstruction kernel and size of the scan object were considered acceptable to create an averaged-out curve per scan energy. This also allowed to minimize the number of clinically used curves, simplifying the treatment planning process.

Scans of the tissue phantoms were performed using Somatom Definition AS scanner (Siemens Healthineers, Erlangen, Germany) with following scan settings: 120 kV, reconstruction slice thickness of 2 mm, I40 f or I40 s reconstruction kernels (according with the applicable clinical scan protocols), with enabled iMAR artefact correction setting. Reconstruction kernels include iterative reconstruction algorithm Safire (used strength: (3) and have intrinsic beam hardening correction for water only.

Although, the stoichiometric method was used to establish the initial HLUT, an additional HLUT (figure 2) based on tissue-substitute method was calculated for comparison purpose.

The HLUT was split in three sections: (a) organ-like tissue, (b) fat-like tissue and (c) bone-like tissue. These three sections were linearly fitted; i.e. the HLUT consists of 3 fitted lines and 2 transition areas (Ainsley and Yeager 2014). A list of used CIRS phantom inserts and their assignment to corresponding section of the HLUT is provided in table 1.

Validation data set

The measurements were performed using three fresh vacuum-sealed animal tissue phantoms: (a) a pig's head, (b) a 'thorax', which consisted of ribs, liver, muscle (with cartilage), fat and (c) a femoral bone with soft tissue. CT scans and measurements were performed over a period of two days. Samples were stored in fridge between the activities.

The used tissue phantoms had approximately following diameters at their thickest slices: head phantom 19 cm, thorax phantom 26 cm, femoral bone 16 cm. These dimensions are comparable with anthropomorphic phantoms and representative for cases, such as, intracranial, head and neck or pediatric indications. Larger dimensions of the phantoms would make range probing (or proton radiography) unfeasible, due to limitations imposed by the maximum available energy of the proton beam in the clinical facility (230 MeV, depth of 32 cm). Additionally, some of the scans for femoral bone were performed by placing the phantom on the solid water plates, in order to introduce additional scattering material in the field of view. Average water equivalent thickness (WET) values per phantom along the range probe path were as follows: Head phantom 143.2 mm (SD 50.8 mm), Thorax phantom 43.0 mm (SD 20.5 mm), Femoral bone 113.9 mm (SD 40.9 mm).

All phantom samples were scanned with a CT scanner using clinical scanning pre-sets (120 kV) established already during the definition of the initial HLUT. Afterwards scans were imported into the TPS for (a) the placement of the isocenters for the range probing measurements and (b) for the calculation of the individual pencil beams for comparison with the subsequently measured data.

Range probing measurements (Schneider and Pedroni 1995, Mumot *et al* 2010) were performed following the methodology as described by Farace *et al* (2016a, 2016b). The approach is based on mapping the samples of interest with individual proton pencil beams of an energy high enough to pass through the sampled area and measuring the exit residual range of individual pencil beams. Measurements were performed with the Giraffe multi-layer ionization chamber (IBA dosimetry, Germany). Due to the size of detector (electrode diameter of 12 cm), a maximum area of 4.5 by 4.5 cm^2 can be covered in a single measurement frame. Therefore, the tissue samples were covered by multiple measurement frames, where each frame contained of 81 individual pencil beams (spot spacing 0.5 cm). Measurements were performed in a movie measurement mode with a sampling time of 10 ms. The delivery system (Proteus Plus, IBA, Belgium) was intentionally slowed down to ensure enough delay between the two consecutive spots within the same frame (i.e. field). All spots were delivered with 210 MeV energy.

As stated by Farace *et al* (2016a), nominal range accuracy for this measurement technique is $\pm$0.5 mm. Due to the detector size and selection of the frame size (4.5 $\times$ 4.5 cm^2), there is no need for high positioning accuracy of the MLIC perpendicular to the beam axis. As verified experimentally, displacements in the order of 5 mm, will not have relevant impact on the shape of the measured integral depth dose curve. Furthermore, positioning accuracy of the MLIC along the beam axis also will not have significant impact on the measurement accuracy, as the distance from the isocenter mainly affect the amount of air between the detector and the tissue sample. Appropriate positioning accuracy of the MLIC under such circumstances can be achieved by employing on-board laser positioning systems, which typically are an integral part of proton treatment rooms. Accuracy of the experiment will depend on the alignment of the phantom to the isocenter of the proton treatment room. In our case, according to the commissioning date for rigid anthropomorphic phantoms positioning accuracy of less than 0.5 mm can be achieved. Tissue phantoms, if handled properly and over a short time frame, can be considered nearly rigid.

The tissue material was positioned at the isocenter by using the on-board x-ray imaging system: either CBCT or kV–kV imaging. The samples were repositioned between the frames by applying predefined offsets by the robotic patient positioning system. The patient position system has a positioning accuracy $<$0.5 mm.

To simulate the range probing measurements in the TPS in each animal tissue CT scan, a slab of homogeneous water-equivalent material (40 $\times$ 40 $\times$ 50 cm^3) was added representing the multi-layer ionization chamber (MLIC) measurement device, which is calibrated to output measurements in water-equivalent depth. The used MLIC had an electrode diameter of 12 cm. Every single pencil beam (FWHM 8.2 mm) was calculated individually using a Monte Carlo dose calculation algorithm (version 4.1) with an uncertainty of 0.5% and on a 1 mm isotropic dose grid. Afterwards, dose distributions were integrated on the plane along the beam axis to create integral depth dose curves to compare with the measured integral depth dose distributions. Experimental setup is shown in figure 1 (Suppl. 1).

Raw measurement data and exported data from TPS were processed and analyzed by a dedicated in-house Matlab tool (Farace *et al* 2016b, Deffet *et al* 2017). The measurements and TPS calculations were transformed into proton radiograms and reconstructed radiograms respectively. Per pencil beam, a shift (*shift*) along the beam axis to reach the best alignment between measured ($MLIC_i$) and calculated (TPS_i) depth dose curve was calculated by solving least square cost function (1). This shift was defined as a residual range error.

$$S = \sum_{i=1}^{n} \left(\mathrm{MLIC}_i(\mathrm{depth}_i) - \mathrm{TPS}_i\left(\mathrm{depth}_i + \mathrm{shift}\right)\right)^2. \quad (1)$$

Based on literature (Paganetti 2012) in case of Monte Carlo calculations, the range uncertainty is estimated as 2.4% of beam range +1.0 mm. For every measured pencil beam, a range error margin was calculated as 2.4% of water-equivalent path length (WEPL) +1.0 mm. The WEPL of the individual pencil beam was used instead of the beam range because the definition of range in water is provided as an input to the TPS for beam modelling. Therefore, it is more appropriate to exclude the contribution of range error in water (or residual range itself as measured by the MLIC) from the uncertainty margin calculation. It was considered that the energy reproducibility of the proton delivery system for the measurements is within the absolute (+1.0 mm) component of the range uncertainty margin.

Eventually the range errors, defined as the discrepancy between measurement and calculation, were compared to corresponding range error margins, and for each pencil beam the ratio between error and margin was calculated. A schematic representation of the approach is shown in figure 2.

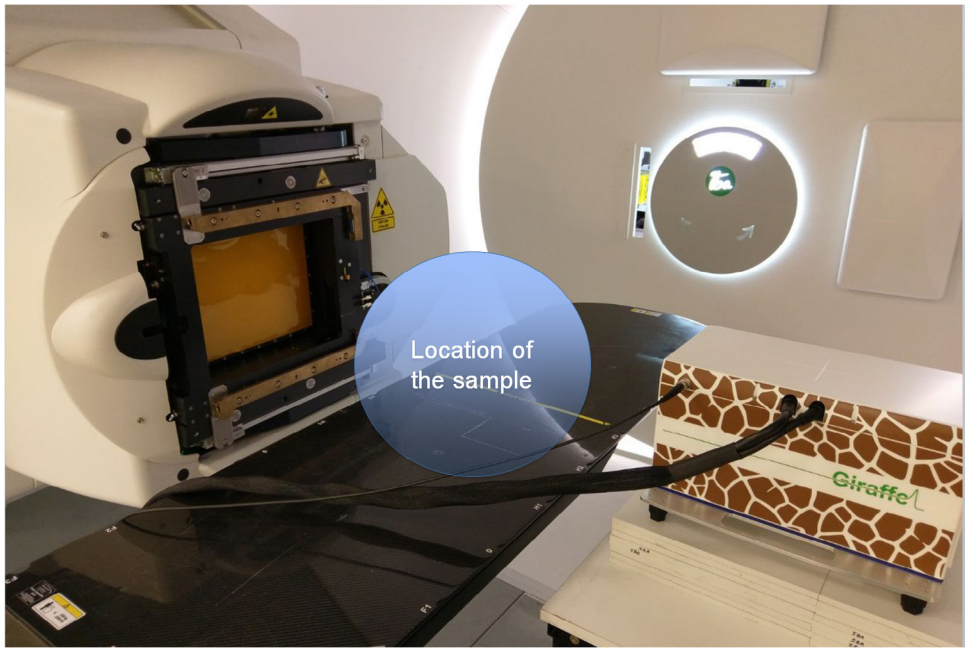

Figure 1. Experimental setup.

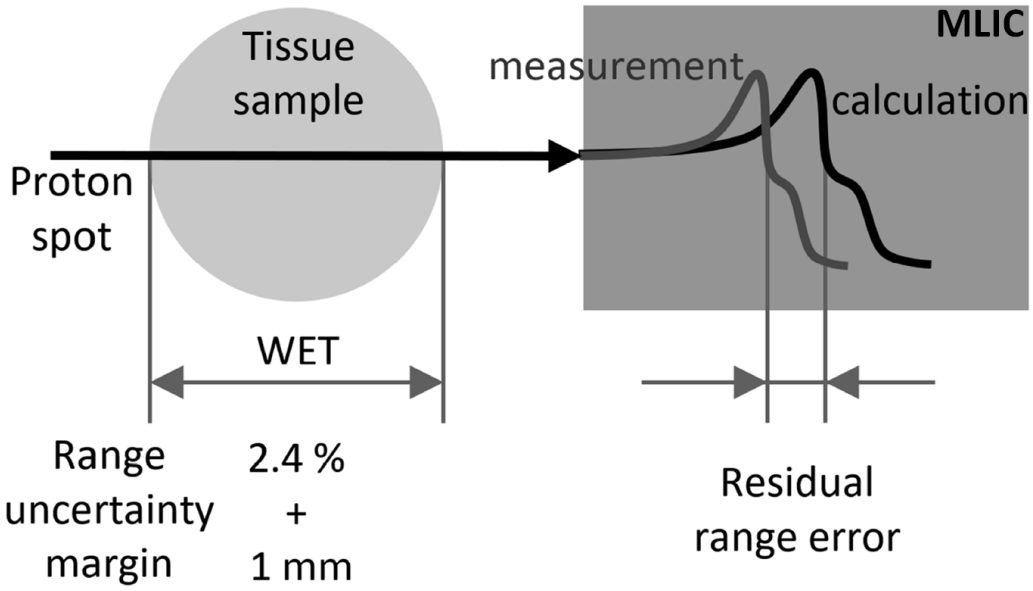

Figure 2. Schematic representation of the performed measurement and analysis. Range uncertainty margin is calculated per every pencil beam as 2.4% of water-equivalent thickness (WET) of the tissue in the beam path plus 1 mm. MLIC was used to acquire integral depth dose distally from the tissue sample. MLIC was simulated as a water slab for the purpose of calculations in the TPS.

If the ratio was larger than 1, the range error exceeds the uncertainty margin. If it is lower, the range error is within the uncertainty margin. In total, over all three samples, approximately 1600 individual pencil beams were measured and evaluated.

Optimization of the HLUT

By reviewing range error and uncertainty margin ratio histograms per tissue sample it was possible to identify inaccurate CT curve sections. Since the HLUT consisted of three fitted line segments (organ-like, fat-like and bone-like tissues), each specific segment of the HLUT could be corrected by adjusting the slope and intercept to achieve an optimal agreement between measurements and calculations. Iterations were performed by determining approximate overshoots or undershoots for a specific HLUT region based on the measurement set and afterwards adjusting the corresponding line segment to compensate previously identified overshoots or undershoots.

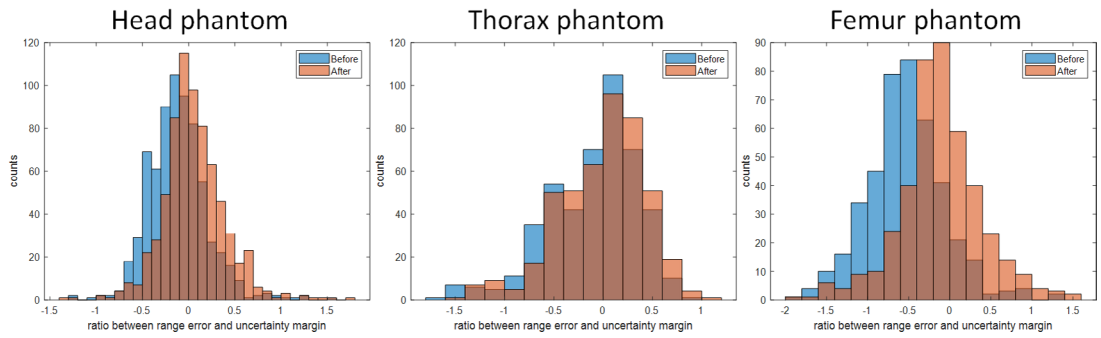

Figure 3. Histograms of ratios between range error and uncertainty margin before and after optimization of HLUT per phantom type.

Afterwards, the optimized HLUT was introduced in the TPS and the entire data set was recalculated using the new HLUT. The analysis as described previously was repeated to validate the modifications. Figure 3 shows histograms of ratios between range error and uncertainty margin per phantom type. It can be observed that the distribution is skewed for the thorax phantom case, which is likely linked to the composition of tissues (some more predominant than other) in the phantom. Furthermore, an independent measurement set, specifically focusing on the tissue-type (bone-like tissues in our case) in the adjusted area of the HLUT, was included in the analysis.

Results

Using the initial HLUT, created with the stoichiometric method, residual range error maps and histograms were determined for all three tissue samples separately.

In case of (a) the pig's head, a mean range error of −0.54 mm with a standard deviation (SD) of 1.5 mm was observed; for (b) the 'thorax' phantom, the mean range error was −0.17 mm (SD of 1.0 mm) and for (c) the femoral bone, the mean range error was −2.37 mm (SD 2.0 mm). Based on these observations, it was concluded that bone-like tissues in TPS are seen denser than they are. Therefore, the slope of bone-like tissue section of the HLUT was adjusted to compensate for this effect. The optimized HLUT is shown in figure 4. The intercept and slope of the segment representing bone-like tissues was adjusted from 1.009 and 0.0007 to 1.029 and 0.0006 respectively.

The optimized HLUT lies in between the initial stoichiometric HLUT and HLUT as calculated by the tissue substitute method. Range error maps were recalculated using the optimized HLUT. After recalculation in case of (a) the pig's head, the mean range error changed to 0.33 mm (SD 1.4 mm); for (b) the thorax phantom, the mean range error dropped to 0.03 mm (SD 0.9 mm) and for (c) the femoral bone, the average range error reduced to −0.43 mm (SD 1.9 mm). Mean range errors as observed prior to and after HLUT optimization are provided in table 2. The results presented in this section and table 2 are based on the initial set of tissue phantoms.

The map of the ratios between range errors and uncertainty margin (defined as 2.4% of WEPL + 1.0 mm), as calculated using the optimized HLUT for head sample, is shown in figure 5. Two areas highlighted in figure 5 (Area A and B) have not been covered in the analysis. In case of Area A, few of the spots acquired in this area could not have been clearly separated timewise while post-processing the measurements. Therefore, to avoid ambiguous sampling, all spots in Area A were excluded from the analysis. Area B was not included in the analysis and measurements in this area were not performed, because most of the spots in this area would have travelled only through the air. Range error and uncertainty margin ratios for a combined data set (all tissue phantoms, including additional femoral-bone phantom, and all measurement points) are shown in a histogram in figure 6.

Discussion

Geometrical localization of range errors

Although the range error distribution seems to be normally distributed, if considering isolated areas of the sample, the magnitude of range errors between these areas varies. For example, increased range errors are well correlated with intersections between materials of clearly different density, such as, high- and low-density bone intersections (see figure 7).

This observation is also consistent with the literature (Knopf *et al* 2008, Farace *et al* 2016a). As pointed out by Farace *et al*, increased range errors based on range probing measurements in a head phantom were observed along the skull contour.

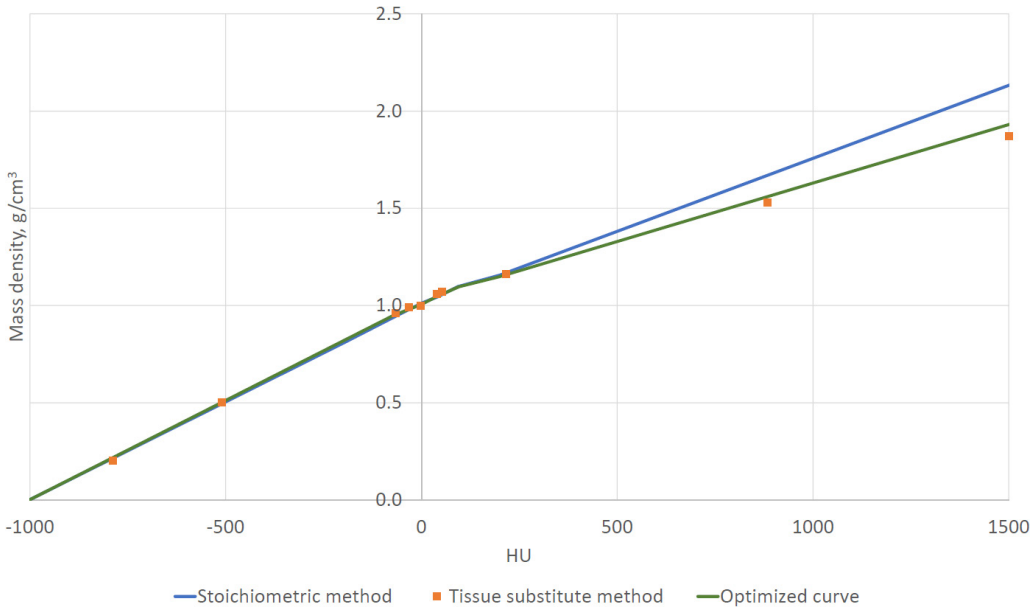

Figure 4. Comparison of obtained HLUTs.

Table 2. Summary of range errors before and after adjustment of HLUT.

	Before HLUT adjustment	After HLUT adjustment
Head phantom	−0.54 (SD 1.5) mm	0.33 (SD 1.4) mm
Thorax phantom	−0.17 (SD 1.0) mm	0.03 (SD 0.9) mm
Femoral bone	−2.37 (SD 2.0) mm	−0.43 (SD 1.9) mm

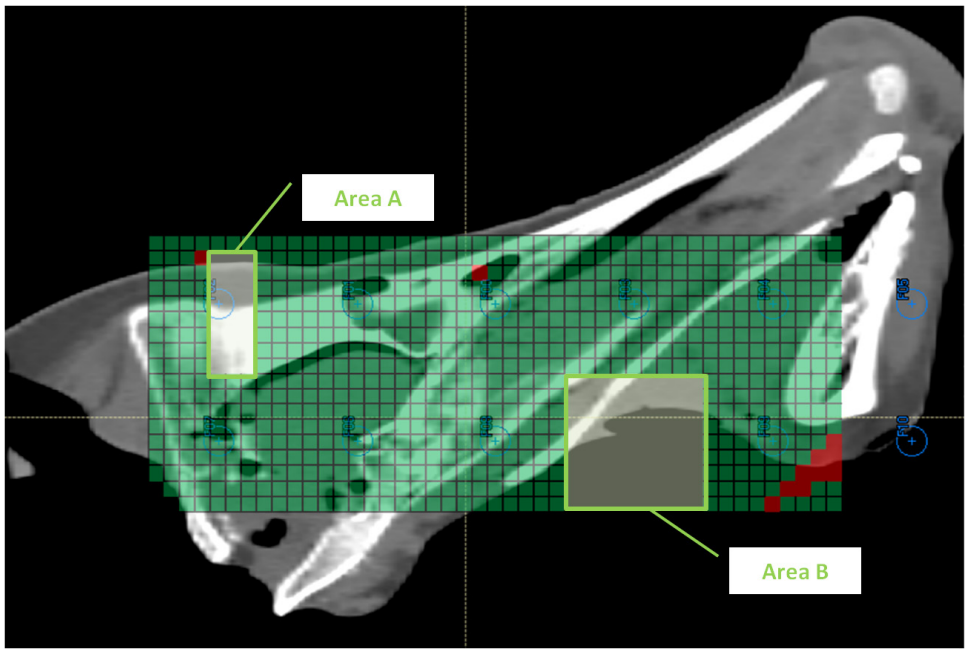

Figure 5. Range error and uncertainty margin ratio map. Red squares indicate spots for which range errors exceeded the uncertainty margin.

This indicates that systematic density scaling, which is broadly used in robust treatment plan optimization, is not the ideal approach to address the range uncertainty problem. Preferably, range uncertainty in the planning process should be applied considering knowledge about the materials in the beam path. Such information as mass density and mass density variation laterally to the beam path should be considered to more realistically account for range uncertainty in the robust optimization process.

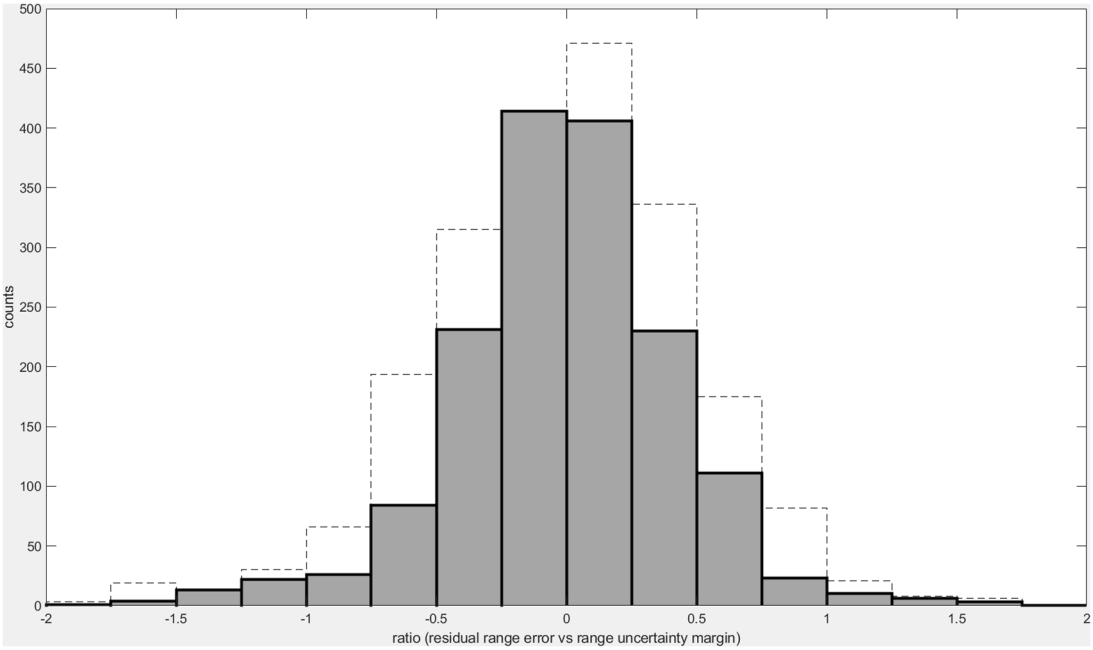

Figure 6. Histogram of ratios between range error and uncertainty margin for a combined data set after adaptation of the HLUT. For 1.5σ of the cases the ratio between the range error and the uncertainty margin is less than 0.75, when assuming an uncertainty margin of 2.4% + 1 mm. The dashed line indicates the upper border of the confidence interval accounting for the uncertainty of the measurement and evaluation.

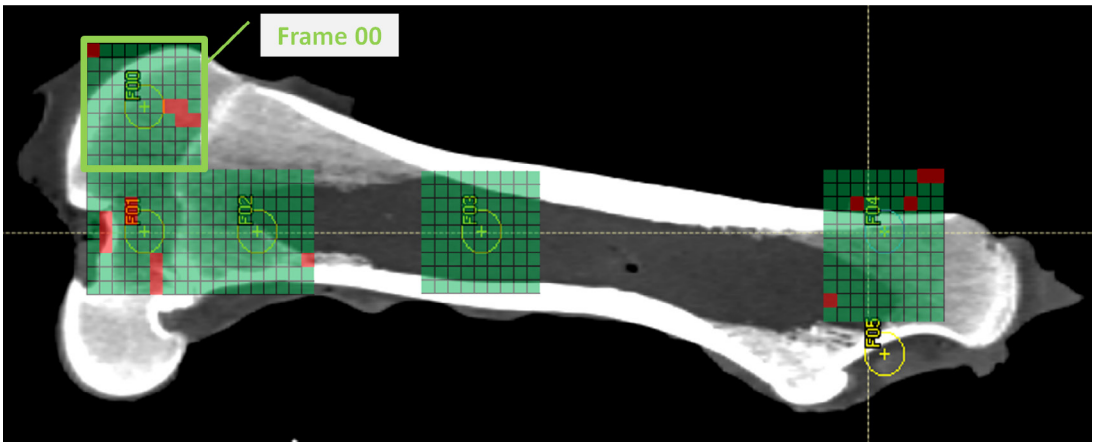

Figure 7. Range error and uncertainty margin ratio map for femoral bone. Red squares indicate spots with range error above uncertainty margin. Four failing spots on the Frame 00 are intersecting a titanium screw, which was attached to the bone.

In the absence of such algorithms in the TPS, an indication specific range uncertainty recipe may be implemented. For example, it may not be necessary to apply the same magnitude of CT number scaling in robust optimization of an intracranial tumour, when the beam path is intersecting the skull perpendicularly and does not intersect cavities, as compared to a base of skull case, where beam may intersect ear canal or to some extent travel parallel to the brain—bone intersection. Therefore, also the beam angle selection may be used to minimize the range uncertainty, which needs to be accounted for by robust optimization.

Figure 8 shows the range error and uncertainty margin ratio map, for a case where the uncertainty margin of 2.4 % + 1.0 mm has been reduced by half. Spots that are intersecting such areas as brain, soft tissue in the snout or are perpendicular to flat and large bony areas would still have range errors within the uncertainty margin. However, beams, which are traveling close to bone—soft tissue intersections or through cavities, have range errors outside the reduced uncertainty margin.

General

Based on the range error and uncertainty margin ratio histogram, it can be observed that 1.5 SD of the spots has a ratio less than 0.75, which indicates that the range uncertainty margin, set as 2.4% + 1.0 mm in our institution-

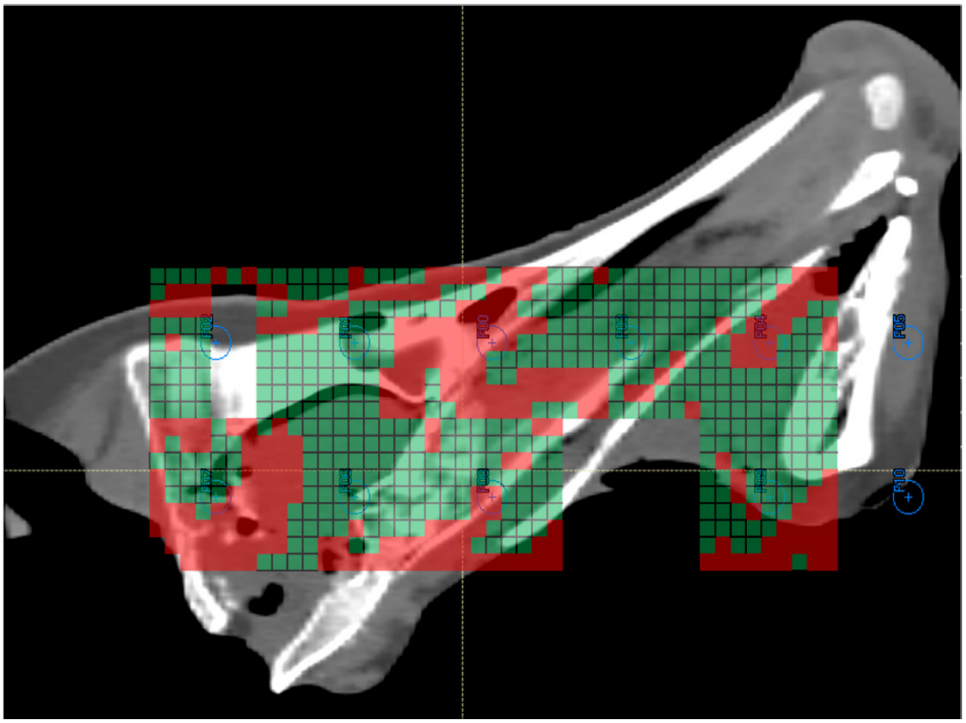

Figure 8. Range error and uncertainty margin ratio map for a case, where uncertainty margin has been reduced by half (1.2% + 0.5 mm).

specific setting, is an overestimation of the actual range errors encountered for the studied tissue phantom set, if the optimized HLUT is used. In case of applying initial, non-optimized HLUT, 1.5 SD of the spots has a range error to uncertainty margin ratio of less than 0.9. The obtain results are representative for the sample dimensions as presented above. Beam hardening effect may introduce additional uncertainty for patients of significantly larger dimensions. Therefore, ultimately range probing checks should be performed for actual patients.

Following this study based on animal tissue, it is worthwhile to consider the application of the methodology using patient-specific data. This would bring more insight in SPR values for human tissue and assess further gains in potentially employing site-specific or patient-specific HLUTs (Schneider *et al* 2005, Doolan *et al* 2015, Collins-Fekete *et al* 2017, Krah *et al* 2019). Using the range probing in patients would require relatively low doses. Without further adjustments of the technique, the dose at the plateau would be less than 1 cGy per frame. Additionally, range probing could be integrated in the workflow to provide information on anatomical variations and assist in the decision-making process for triggering plan adaptation.

Eventually, range probing could enrich CBCT data and improve the quality and reliability of virtual CTs, created based on the CBCT anatomy. Currently, one of the challenges in creating virtual CTs based on CBCT data sets is the ability to validate the accuracy of CT number retrieval, which commonly is done based on image deformation fields (Landry *et al* 2015, Veiga *et al* 2016) or in combination with other methodologies, such as machine learning (Tappeiner *et al* 2019). Similarly, to the use of range probing in the scope of this work for validation of range calculation accuracy based on CT data set, it could be used for the validation of virtual CT data sets.

Currently, there is a substantial interest in the field for the application of dual energy CT imaging for proton treatment planning, to provide more accurate SPR data (Hudobivnik *et al* 2016, Wohlfahrt *et al* 2017a, Wohlfahrt *et al* 2018). It is estimated that the use of dual energy CT would allow to reduce range uncertainties to about 2% (Li *et al* 2017). By using the proposed range probing methodology for optimization and validation of the HLUTs, it was demonstrated that range uncertainty could be reduced to almost 2% as well. This reduction may be further enhanced when site specific or patient specific HLUTs are used. However, combining the range probing approach with the use of dual energy CT imaging could potentially allow to reduce the uncertainty even further. Nevertheless, in order to support comparison between range uncertainty estimation for usage of dual energy CT and results of this study, it would be necessary to also perform the range probing-based evaluation for dual energy CT images.

A limitation of the HLUT optimization is, that within the optimization process contribution of absolute component in range uncertainty, which is affecting the experimental data set, is also 'minimized'. However, in practice adjustment of HLUT cannot reduce the contributions to absolute component of the range uncertainty. By overfitting HLUT to obtain perfect agreement between measurement and simulation, one might incorrectly introduce a compensation for range errors caused by contributors to the absolute component in the HLUT. The

inclusion of independent samples and data sets in the evaluation to some extent provides a possibility to assess the impact of the above-mentioned issue.

The methodology is not intended, nor suitable for identifying tissue-specific ground-truth SPRs. It is rather an end-to-end verification, which looks at tissue compositions in an integral manner. Therefore, the methodology should not be used for extensive optimizations of the HLUT, if the problematic area is not obvious (such as in the current case bone-like tissues). By overfitting the HLUT, solutions might be found that give excellent agreement between measurements and simulations, however, due to integral characteristic of the range probe, still incorrect stopping powers might be assigned to individual tissues. To overcome this limitation, more projections could be acquired, which resample an approach towards proton CT.

Possible deformations of the tissue samples that may happen between CT simulation and treatment delivery can impact the comparison between measurements and simulations. This is a limitation of the proposed methodology. It is difficult to numerically asses the possible impact of phantom deformations, as it can vary significantly depending on type of deformation, extent and localization. For our experiments, possible deformations were investigated by extensively reviewing CBCT image overlaid with the CT image and no significant difference were identified.

Currently one of the drawbacks of the experimental measurement technique is the lack of integration between measurement device and beam delivery system. More reliable sampling during measurement, for example using a trigger mode, would be desirable. This would allow to avoid or limit artefacts in the measurement set and a need to exclude data, as shown in figure 5 area A.

While the range probing based method for assessment of range accuracy in the treatment delivery process was demonstrated in a single energy CT based setting, in principle the technique can also be applied to perform range accuracy evaluations in departments, which use dual energy CT for patient simulation (Wohlfahrt *et al* 2017b).

In conclusion, it has been demonstrated that range probing is an efficient method for institution-specific validation and optimization of HLUTs prior to their use in the clinic, opening possibilities for reducing literature-based range uncertainty margins. Further range probing studies should evaluate the potential range uncertainty reduction for site- or patient-specific HLUT alone and/or in conjunction with DECT.

Disclosures

University of Groningen, University Medical Centre Groningen, Department of Radiation Oncology has active research agreements with RaySearch, Philips, IBA, Mirada, Orfit.

ORCID iDs

J Free https://orcid.org/0000-0002-1573-2613

References

Ainsley C G and Yeager C M 2014 Practical considerations in the calibration of CT scanners for proton therapy *J. Appl. Clin. Med. Phys.* **15** 4721

Collins-Fekete C A, Brousmiche S, Hansen D C, Beaulieu L and Seco J 2017 Pre-treatment patient-specific stopping power by combining list-mode proton radiography and x-ray CT *Phys. Med. Biol.* **62** 6836–52

Deffet S, Macq B, Righetto R, Vander Stappen F and Farace P 2017 Registration of pencil beam proton radiography data with x-ray CT *Med. Phys.* **44** 5393–401

Doolan P J, Testa M, Sharp G, Bentefour E H, Royle G and Lu H M 2015 Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography *Phys. Med. Biol.* **60** 1901–17

Farace P, Righetto R and Meijers A 2016a Pencil beam proton radiography using a multilayer ionization chamber *Phys. Med. Biol.* **61** 4078–87

Farace P, Righetto R, Deffet S, Meijers A and Vander Stappen F 2016b Technical note: a direct ray-tracing method to compute integral depth dose in pencil beam proton radiography with a multilayer ionization chamber *Med. Phys.* **43** 6405

Gomà C, Almeida I P and Verhaegen F 2018 Revisiting the single-energy CT calibration for proton therapy treatment planning: a critical look at the stoichiometric method *Phys. Med. Biol.* **63** 235011

Hudobivnik N *et al* 2016 Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy CT images *Med. Phys.* **43** 495

Knopf A, Parodi K, Paganetti H, Cascio E, Bonab A and Bortfeld T 2008 Quantitative assessment of the physical potential of proton beam range verification with PET/CT *Phys. Med. Biol.* **53** 4137–51

Krah N, Patera V, Rit S, Schiavi A and Rinaldi I 2019 Regularised patient-specific stopping power calibration for proton therapy planning based on proton radiographic images *Phys. Med. Biol.* **64** 065008

Landry G *et al* 2015 Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation *Med. Phys.* **42** 1354–66

Li B, Lee H C, Duan X, Shen C, Zhou L, Jia X and Yang M 2017 Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging *Phys. Med. Biol.* **62** 7056–74

Mumot M, Algranati C, Hartmann M, Schippers J M, Hug E and Lomax A J 2010 Proton range verification using a range probe: definition of concept and initial analysis *Phys. Med. Biol.* **55** 4771–82
Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99–117
RaySearch Laboratories 2017 RayStation 7B Reference manual (Stockholm: RaySearch Laboratories)
Schneider U and Pedroni E 1995 Proton radiography as a tool for quality control in proton therapy *Med. Phys.* **22** 353–63
Schneider U, Pedroni E and Lomax A 1996 The calibration of CT Hounsfield units for radiotherapy treatment planning *Phys. Med. Biol.* **41** 111–24
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 *Med. Phys.* **32** 195–9
Tappeiner E, Pröll S, Hönig M, Raudaschl P F, Zaffino P, Spadea M F, Sharp G C, Schubert R and Fritscher K 2019 Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach *Int. J. Comput. Assist. Radiol. Surg.* **14** 745–54
Unkelbach J and Paganetti H 2018 Robust proton treatment planning: physical and biological optimization *Semin. Radiat. Oncol.* **28** 88–96
Veiga C *et al* 2016 First clinical investigation of cone beam computed tomography and deformable registration for adaptive proton therapy for lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **95** 549–59
Wohlfahrt P, Möhler C, Hietschold V, Menkel S, Greilich S, Krause M, Baumann M, Enghardt W and Richter C 2017a Clinical implementation of dual-energy CT for proton treatment planning on pseudo-monoenergetic CT scans *Int. J. Radiat. Oncol. Biol. Phys.* **97** 427–34
Wohlfahrt P, Möhler C, Stützer K, Greilich S and Richter C 2017b Dual-energy CT based proton range prediction in head and pelvic tumor patients *Radiother. Oncol.* **125** 526–33
Wohlfahrt P, Troost E G C, Hofmann C, Richter C and Jakobi A 2018 Clinical feasibility of single-source dual-spiral 4D dual-energy CT for proton treatment planning within the thoracic region *Int. J. Radiat. Oncol. Biol. Phys.* **102** 830–40