Patient-specific CT calibration based on ion radiography for different detector configurations in $^1$H, $^4$He and $^{12}$C ion pencil beam scanning

Chiara Gianoli$^{1,∗}$, Maximilian Göppel$^1$, Sebastian Meyer$^{1,4}$, Prasannakumar Palaniappan$^1$, Martin Rädler$^1$, Florian Kamp$^2$, Claus Belka$^{2,3}$, Marco Riboldi$^1$ and Katia Parodi$^1$

$^1$ Department of Medical Physics – Experimental Physics, Ludwig-Maximilians-Universität München, Munich, Germany
$^2$ Department of Radiation Oncology, University Hospital, LMU Munich, Marchioninistraße 15, 81377 München, Germany
$^3$ German Cancer Consortium (DKTK) partner site Munich, Germany and German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
$^4$ Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, United States of America

* Author to whom any correspondence should be addressed
E-mail: Chiara.Gianoli@physik.uni-muenchen.de

Keywords: ion radiography, CT calibration, list-mode and integration-mode detector configurations, clinical data

Abstract

The empirical conversion of the treatment planning x-ray computed tomography (CT) image to ion stopping power relative to water causes dose calculation inaccuracies in ion beam therapy. A patient-specific calibration of the CT image is enabled by the combination of an ion radiography (iRad) with the forward-projection of the empirically converted CT image along the estimated ion trajectories. This work investigated the patient-specific CT calibration for list-mode and integration-mode detector configurations, with reference to a ground truth ion CT (iCT) image. Analytical simulations of idealized carbon ion and proton trajectories in a numerical anthropomorphic phantom and realistic Monte Carlo simulations of proton, helium and carbon ion pencil beam scanning in a clinical CT image of a head-and-neck patient were considered. Controlled inaccuracies and noise levels were applied to the calibration curve and to the iRad, respectively. The impact of the selection of slices and angles of the iRads, as well as the choice of the optimization algorithm, were investigated. Accurate and robust CT calibration was obtained in analytical simulations of straight carbon ion trajectories. Analytical simulations of non-straight proton trajectories due to scattering suggested limitations for integration-mode but not for list-mode. To make the most of integration-mode, a dedicated objective function was proposed, demonstrating the desired accuracy for sufficiently high proton statistics in analytical simulations. In clinical data the inconsistencies between the iRad and the forward-projection of the ground truth iCT image were in the same order of magnitude as the applied inaccuracies (up to 5%). The accuracy of the CT calibration were within 2%–5% for integration-mode and 1%–3% for list-mode. The feasibility of successful patient-specific CT calibration depends on detector technologies and is primarily limited by these above mentioned inconsistencies that slightly penalize protons over helium and carbon ions due to larger scattering and beam spot size.

1. Introduction

Although physical and radiobiological properties of ion beams foster an improved targeting in radiation therapy when compared to photon beams, the superior dose conformity offered by scanned ion beams comes at the expense of enhanced sensitivity to inaccuracies (Durante and Loeffler 2010). Among others, inaccuracies are intrinsically related to the current clinical treatment planning, due to uncertainties in the knowledge of the ion stopping power properties of different tissues (Schneider and Pedroni 1995, Schaffner and Pedroni 1998). Treatment planning is typically performed on x-ray computed tomography (CT) images,
expressed as relative photon linear attenuation coefficients (Hounsfield Unit, HU). In order to match the specific physical and radiological properties of the therapeutic radiation, the CT image undergoes an empirical calibration, usually relying on a plastic phantom provided with tissue equivalent inserts. In ion beam therapy, the CT image is converted to ion stopping power relative to water (RSP) (Schneider et al 2000), referred to as rspCT image, as this ratio is needed by analytical dose calculation algorithms to scale the well-known depth dose distribution in water to any arbitrary tissue (Hong et al 1996, Schaffner et al 1999). A stoichiometric calibration is usually adopted, based on a parameterization of the HU in elemental composition and mass density (Schneider et al 1996). Information about different elemental compositions can be also derived from dual-energy CT image (Bourque et al 2014). Although the stoichiometric calibration overcomes the approximation of the elemental composition for biological tissue with tissue equivalent inserts, tissue-specific and patient-specific variations of elemental compositions are responsible for up to 3% of remaining inaccuracies in the calibration curve (Jiang et al 2007, Yang et al 2012), thus translating into inaccuracies in range estimation up to several millimeters (Rietzel et al 2007, Paganetti 2012).

Ion imaging has the potential of directly measuring the tissue stopping power properties to overcome intrinsic inaccuracies in treatment planning (Parodi 2014). The residual energy of energetic ions after traversing the object of interest is measured relying on a total absorption detector and then calibrated to the equivalent ion beam range in water or water equivalent thickness (WET). The WET matches the integral RSP along the traversed object of interest. When the ions are individually tracked, each ion carries a WET. This ‘list-mode’ detector configuration is typically synchronized with a tracking system before and after the object of interest, thus enabling the estimation of the ion trajectories (Schulte et al 2008). Scanned ion beams offer the possibility to simplify the detector configuration by removing the tracking system and synchronizing the total absorption detector with the pencil beam delivery system. With this ‘integration-mode’ detector configuration, the WET is carried by the ensemble of ions in the pencil beam. The ion trajectories are statistically distributed according to the beam spot size (BSS) and the multiple Coulomb scattering (MCS) model. Although ions are not individually tracked, information about WET inhomogeneity can be retrieved by discretizing the total absorption detector (Rinaldi et al 2013), thus detecting range mixing effects in the discrete energy loss signal. For each pencil beam, a WET histogram expressing the weight of each WET component can be obtained by means of a linear decomposition of the signal (Meyer et al 2017). Accounting for the estimation of the ion trajectories provided by different detector configurations, the measurements covering the object of interest at a certain angle are typically organized in an ion radiography, or iRad. The combination of an iRad with the rspCT image enables a patient-specific calibration of the CT image to RSP (Schneider et al 2005). This calibration relies on the forward-projection of the rspCT image along an estimation of the ion trajectory (yielding the water equivalent Digitally Reconstructed Radiography, weDRR). The patient-specific calibration of the CT image to RSP can be implemented as a linear minimization if prior segmentation of the rspCT image is applied. Alternatively, an optimization algorithm embedding forward-projection and segmentation of the iteratively adjusted rspCT image can be adopted. Despite the computational advantage of a linear minimization (Collins-Fekete et al 2017, Zhang et al 2018, Krah et al 2019), the adoption of a non-linear optimization algorithm enables to overcome the limitation of the prior segmentation and the forward-projection of the rspCT image (Schneider et al 2005, Doolan et al 2015). Depending on the detector configuration, the trajectory is either estimated for each ion (list-mode) (Collins-Fekete et al 2017) or assumed to be a straight line for each pencil beam (integration-mode) (Doolan et al 2015, Zhang et al 2018, Krah et al 2019).

In this work, the patient-specific calibration of the CT image based on iRads is extensively investigated in analytical simulations of idealized carbon ion and proton trajectories in a numerical anthropomorphic phantom (Göppel et al 2017). The role of the MCS as well as the BSS in both ideal list-mode and integration-mode detector configurations is assessed. Two different non-linear optimization algorithms are compared to linear minimization. The first one is adopted from a previous work targeted to integration-mode detectors (Doolan et al 2015) and the second one is chosen for its flexibility, as required by CT calibration when no assumption about the inaccuracy model of the calibration curve is made. To make the most of integration-mode detector configuration, a dedicated objective function based on the entire WET histogram is proposed. Relying on the investigation in analytical simulations, the patient-specific calibration of the CT image based on iRads for different detector configurations in proton, helium and carbon ion pencil beam scanning is investigated in Monte Carlo simulations at clinically acceptable imaging dose in a clinical CT image of a head-and-neck patient.
2. Materials and methods

2.1. Analytical simulations in numerical anthropomorphic phantom

The photon linear attenuation coefficients of a numerical anthropomorphic phantom (Segars et al 2002) were linearly transformed to HUs. Subsequently, the generated CT image was converted to the ground truth ion CT (iCT) image relying on a monotonic calibration curve, referred to as true calibration curve. The calibration curve was made of 13 points, one for each linear attenuation coefficient of the numerical anthropomorphic phantom, according to a histogram-driven point selection of the calibration curve. Each point was associated with a point weight equal to the corresponding abundance in the image. The size of the images was $128 \times 128 \times 100$ voxels. The pixel size was $0.3 \times 0.3 \times 0.3$ cm$^3$. An inaccurate calibration curve was created by applying controlled inaccuracies $\Delta$ to the points of the true calibration curve. The inaccurate calibration curve was then used to generate the rspCT image. The inaccuracies were generated under the constraint of monotonically increasing calibration curve, randomly sampled from a uniform distribution over realistic intervals.

The idealized carbon ion and proton trajectories outline the best and the worst cases of MCS, respectively. The idealized carbon ion trajectories were traced as straight lines along the pencil beam direction to simplify the negligible MCS and smaller BSS compared to protons. Therefore, no distinction between list-mode and integration-mode detector configurations was enabled. The idealized proton trajectories were simulated relying on a recently proposed analytical framework for both list-mode and integration-mode detector configurations in proton pencil beam scanning (Gianoli et al 2019). The proton trajectories were traced within the statistical distribution of the BSS and MCS model, neglecting energy straggling and nuclear interactions. The physical parameters of the analytical simulations were summarized in table 1. The iRad was simulated as forward-projection of the ground truth iCT image along the simulated trajectories. The iRad for list-mode corresponded to the WET for each ion of the pencil beam. In contrast, a WET histogram for each pencil beam was obtained for integration-mode, thus referred to as integration-mode hist. The WET histogram enabled the computation of a weighted mean WET (WET components weighted by the occurrences and averaged) and a mode WET (WET component with the maximum occurrence), thus referring to integration-mode mean and mode, respectively.

For the analytical simulations of idealized carbon ions, the weDRR was based on the straight lines coinciding with the simulated carbon ion trajectories. For the analytical simulations of idealized protons, the weDRR for list-mode was based on proton trajectories estimated using the Most Likely Path (MLP) algorithm (Schulte et al 2008). For integration-mode mean and mode, the estimated proton trajectories were traced straight along the pencil beam direction. Conversely, for integration-mode hist the proton trajectories were estimated according to a random sampling of the statistical distribution within BSS and MCS model along the pencil beam direction. The number of these samples, or proton statistics for the weDRR, was set equal to 100 protons per pencil beam. The weDRR was calculated as the forward-projection of the rspCT image along the estimated trajectories. For integration-mode hist, the forward-projection of the rspCT image generated a WET histogram for each pencil beam, similar to the corresponding iRad.

2.2. Monte Carlo simulations in clinical data

A clinical CT image of a head-and-neck patient from the Department of Radiation Oncology of the university hospital of the LMU Munich undergoing intensity modulated photon therapy treatment was considered. The CT image was converted to the ground truth iCT image relying on a calibration curve defined by 11 points, similar to the one used at the Heidelberg Ion Beam Therapy Center for treatment planing of head-and-neck patients. This calibration curve was assumed as true calibration curve. The size of the CT image was $314 \times 314 \times 40$ voxels. The pixel size was $1.074 \times 1.074 \times 3$ mm$^3$. The iRads for proton, helium and carbon ion were simulated in FLUKA (Ferrari et al 2005, Böhlen et al 2014) (public release version 2011.2c.6 using HADROTHE defaults without explicit delta-ray production), relying on a customized simulation framework (Meyer et al 2019) provided with phase space files of the Heidelberg Ion Beam Therapy Center (Tessonnier et al 2016). The weDRRs for proton, helium and carbon ions were based on trajectories estimated according to the cubic spline approximation of the MLP algorithm (Fekete et al 2015). The ion statistics for the weDRR, as

| Table 1. Physical parameters of the analytical simulations of iRads for idealized proton trajectories. |
|------------------------------------------------------|
| Parameter                             | Value                  |
| Idealized protons                      |                        |
| Proton pencil beam energy              | 280 MeV                |
| Number of protons per pencil beam      | 100, 75, 50, 25, 10, 5 |
| Standard deviation of the BSS ($\sigma_{BSS}$) | 0.0 cm (i.e. ideal pencil beam), 0.3 cm |
Table 2. Physical parameters of the Monte Carlo simulations of iRads for proton, helium and carbon ion pencil beam scanning.

|                        | Protons     | Helium ion | Carbon ion |
|------------------------|-------------|------------|------------|
| Nominal pencil beam energy | 199.44 MeV  | 199.03 MeV/u | 385.18 MeV/u |
| Number of ions per pencil beam (ion statistics) | 400 (low), 10 000 (high) | 100 (low), 2500 (high) | 20 (low), 500 (high) |
| Full width at half maximum of the BSS (FWHM<sub>BSS</sub>)<sup>a</sup> | 8.5 mm      | 5.1 mm     | 3.5 mm     |

<sup>a</sup>In air at the isocenter (Meyer et al 2019).

required for the objective function dedicated to integration-mode hist, was set equal to the ion statistics of the iRad. The physical parameters summarized in table 2 resulted in a comparable physical dose for proton, helium and carbon ion iRads (0.01 mGy for low ion statistics and 0.26 mGy for high ion statistics).

### 2.3. Patient-specific CT calibration based on iRads

The patient-specific CT calibration was implemented as an optimization algorithm that estimated the inaccuracies $\Delta$ to adjust the weDRR based on the iRad. Expressing the weDRR as the forward-projection (matrix-vector product) of the rspCT image (column vector, function of $\Delta$) along the estimated trajectory (column vector in the matrix $A$, for all the trajectories), $\text{weDRR} = A \cdot \text{rspCT}(\Delta)$, the optimization algorithm searches for the inaccuracies $\Delta$ as

$$\Delta = \text{argmin} \left( \frac{1}{N} \sum_{n} (\text{weDRR} - \text{iRad})^2 \right),$$

$N$ being the dimension of the weDRR and the iRad.

Accordingly, the optimization algorithm is expected to adjust the inaccuracies of the calibration curve towards the true calibration curves. The forward-projection and the segmentation of the iteratively adjusted rspCT image were embedded in the optimization algorithm. The objective function of the optimization algorithm was the root mean square error (RMSE) between the iRad and the weDRR. This objective function was expressed in voxels for list-mode as well as for integration-mode mean and mode. For integration-mode hist, the dedicated objective function was expressed as normalized WET occurrences (probability) for each pencil beam.

Five different seeds were selected to initialize the random number generator in order to obtain reproducible inaccuracies of the inaccurate calibration curve. The accuracy of the CT calibration was quantified as mean absolute relative error (MRE) between the points of the true and the optimized calibration curves. Median and interquartile range (iqr) MRE among the different random seeds were quantified. Successful CT calibration was defined as MRE below 1%.

### 2.4. Investigation in numerical anthropomorphic phantom

Relying on idealized carbon ion trajectories, two different non-linear optimization algorithms implemented in MATLAB (Matlab, MathWorks, Natick, MA, USA), namely Nelder–Mead (Lagarias et al 1998) and particle swarm (Kennedy 1995), were compared. In addition, linear minimization relying on the least-squares solver implemented in MATLAB was considered. The angles of the iRads was $0^\circ$ (frontal iRad), $45^\circ$, $90^\circ$ (lateral iRad) and $135^\circ$. The impact of the number of iRads at different angles and the number of slices for each iRad were investigated. In addition to the histogram-driven point selection of the calibration curve, an equally-spaced point selection within the minimum and the maximum of the linear attenuation coefficients was considered. The controlled inaccuracies $\Delta$ were applied according to both additive ($\text{RSP} + \Delta$) and multiplicative ($\text{RSP} \times (\Delta + 1)$) inaccuracy models in combination with histogram-driven and equally-spaced point selection of the calibration curve. To investigate the effect of an increase in inaccuracy level, the inaccuracies were applied within an interval spanning from $\pm 0.05$ to $\pm 0.30$ at additive steps of $+0.05$. The inaccuracies were applied to a progressively increasing number of points of the calibration curve (from one to all the 13 points of the calibration curve), according to both increasing and decreasing point weights. The effect of an increase in noise level on the iRads was also assessed. Either the experimental or the statistical nature of the noise was considered. Relying on idealized carbon ion trajectories, nine multiplicative noise levels simulating experimental noise were investigated. The variance spanned from $10^{-4}$ to $10^4$ mm$^2$, logarithmically spaced by $10^1$ mm$^2$. Relying on idealized proton trajectories, the effect of the statistical noise was assessed by decreasing proton statistics. Moreover, list-mode, integration-mode hist and integration-mode mean and mode were compared.

An overview of the investigation in numerical anthropomorphic phantom is reported in figure 1.
2.5. Investigation in clinical data

Inaccuracies were applied according to multiplicative inaccuracy models. The particle swarm was selected as optimization algorithm. The standard deviation of the WET for the ions of the pencil beam was calculated. In list-mode detector configuration, the ion trajectories were also selected based on 1, 2 and 3 times this standard deviation. The pencil beam scanning enabled the implementation of a data-driven, instead of the originally proposed model-driven, sigma cut (Schulte et al 2008). The lateral and the frontal iRads, as well as one slice crossing the nasal air cavity (nasal cavity slice) and one slice crossing the brain (brain slice) were selected (figure 2). For this selection, the normalized root mean square error (NRMSE) between the iRad and
Table 3. Selection of slices and the corresponding number of points of the calibration curve included in the iRad. Slices 16,76, corresponding to abdomen and thorax respectively, contain all the 13 points of the calibration curve.

| Selected slices | Number of points included in the iRad | Median MRE (iqr MRE) (%) |
|-----------------|--------------------------------------|--------------------------|
| 15,75, 16,76, 17, 77 | 12, 13, 12 | $3.1 \times 10^{-1}$ ($1.1 \times 10^{-1}$), $2.8 \times 10^{-9}$ ($4.2 \times 10^{-9}$), $3.8 \times 10^{-1}$ ($8.6 \times 10^{-2}$) |

Figure 3. The abdomen slice 16 and the thorax slice 76 of the numerical anthropomorphic phantom.

Figure 4. Median MRE (red line) and iqr MRE (blue box) for increasing noise (A) and inaccuracy (B) levels.

the weDRR of the ground truth iCT image, normalized by the latter, was quantified. The Wilcoxon signed rank test (5% significance level) was applied to support the statistical interpretation of results.

An overview of the investigation in clinical data is reported in figure 1.

3. Results

3.1. Analytical simulations of idealized carbon ion trajectories in numerical anthropomorphic phantom

With an inaccuracy level within $\pm 0.05$, the linear minimization resulted in median MRE equal to $3.30 \times 10^{-9}$% and iqr MRE equal to 0.70% for the multiplicative inaccuracy model. The corresponding results for the additive inaccuracy model were 1.53% and 0.51%, respectively. With the same inaccuracy level, the Nelder–Mead algorithm demonstrated different results depending on additive or multiplicative inaccuracy models, reaching median and iqr MRE in the order of $10^{-11}$% for the multiplicative inaccuracy model but unstable results among the different random seeds for the additive inaccuracy model, with median MRE equal to $1.53 \times 10^{-2}$% and iqr MRE equal to $8.47 \times 10^{-2}$%. In contrast, the particle swarm algorithm reached instead median and iqr MRE in the order of $10^{-5}$% for both additive and multiplicative inaccuracy models. Relying on the particle swarm algorithm, the number of iRads at different angles and the
number of slices demonstrated to have no impact, given that the iRad included all the points of the calibration curve. A CT calibration based on points that are not included in the iRads generated suboptimal results, as demonstrated by the equally-spaced point selection (median MRE of 2.09% and iqr MRE of 0.43%) or, although successful, the selection of slices that do not contain all the 13 points of the calibration curve (table 3). The effect of applying inaccuracies to a progressively increasing number of points of the calibration curve, according to both increasing and decreasing point weights, was negligible, with median MRE less than $10^{-8}$. Selecting all possible combinations of number of iRads at different angles, the median and iqr MRE remained in the order of $10^{-9}–10^{-10}$. Selecting two slices of the lateral iRad (figure 3), an increase in the inaccuracy level did not affect the CT calibration (figure 4). Instead, the increase in noise level (with an inaccuracy level within ±0.05) severely affected the CT calibration, showing successful results only for small variance, equal to or less than 0.1 mm² (figure 4).

3.2. Analytical simulations of idealized proton trajectories in numerical anthropomorphic phantom

Selecting two slices of the frontal iRad (figure 3), list-mode and integration-mode hist performed differently under reduction of proton statistics with an inaccuracy level within ±0.05. Whereas list-mode yielded stable results (median MRE 0.12%, iqr MRE $1.33 \times 10^{-4}$%) for all considered proton statistics, integration-mode hist revealed generally larger MRE and successful CT calibration for proton statistics down to five protons per pencil beam and (figure 5). Integration-mode mean and mode resulted in unsuccessful CT calibration. For 100 protons per pencil beam, the median MREs were 2.12% and 1.84%, respectively (the iqr MREs were $1.46 \times 10^{-1}$% and $1.59 \times 10^{-1}$%, respectively). The BSS caused unsuccessful CT calibration for
Table 4. NRMSE (expressed in %) between the iRad and the weDRR of the ground truth iCT image, for different detector configurations in proton, helium and carbon ion pencil beam scanning at low and high ion statistics, for selected slices and angles of the iRad.

| Selected slices   | Nasal cavity | Brain |
|-------------------|--------------|-------|
|                   | Frontal iRad | Lateral iRad | Frontal iRad | Lateral iRad |
| Low proton statistics | Integration-mode mode | 4.09 | 4.29 | 2.50 | 2.56 |
|                    | Integration-mode mean | 4.79 | 8.93 | 4.23 | 8.18 |
|                    | Integration-mode hist | 13.16 | 8.91 | 18.74 | 4.89 |
|                    | List-mode | 2.12 | 2.54 | 1.44 | 2.01 |
| High proton statistics | Integration-mode mode | 3.36 | 5.51 | 2.69 | 3.41 |
|                    | Integration-mode mean | 5.38 | 9.50 | 4.24 | 7.96 |
|                    | Integration-mode hist | 4.74 | 13.33 | 11.58 | 30.27 |
|                    | List-mode | 2.13 | 2.64 | 1.48 | 2.06 |
| Low helium statistics | Integration-mode mode | 2.61 | 2.99 | 1.52 | 3.67 |
|                    | Integration-mode mean | 2.30 | 3.60 | 1.55 | 2.81 |
|                    | Integration-mode hist | 24.36 | 17.90 | 35.17 | 24.66 |
|                    | List-mode | 1.24 | 1.39 | 0.87 | 1.06 |
| High helium statistics | Integration-mode mode | 3.46 | 3.95 | 3.89 | 2.48 |
|                    | Integration-mode mean | 2.49 | 4.51 | 1.67 | 3.29 |
|                    | Integration-mode hist | 3.29 | 13.42 | 12.72 | 25.89 |
|                    | List-mode | 1.50 | 1.85 | 1.16 | 1.49 |
| Low carbon statistics | Integration-mode mode | 2.59 | 2.59 | 2.23 | 2.07 |
|                    | Integration-mode mean | 2.36 | 2.77 | 1.69 | 2.57 |
|                    | Integration-mode hist | 53.81 | 64.09 | 50.44 | |
|                    | List-mode | 0.95 | 0.96 | 0.77 | 0.80 |
| High carbon statistics | Integration-mode mode | 2.36 | 3.13 | 1.77 | 2.49 |
|                    | Integration-mode mean | 1.90 | 3.73 | 1.62 | 2.51 |
|                    | Integration-mode hist | 4.59 | 15.37 | 7.18 | 14.40 |
|                    | List-mode | 1.36 | 1.80 | 1.15 | 1.53 |

Integration-mode, whereas the accuracy for list-mode was better than 1% down to proton statistics equal to five protons per pencil beam (figure 5).

3.3. Monte Carlo simulations in clinical data

The optimization algorithm successfully converges in both detector configurations. List-mode enabled an improved accuracy of the CT calibration compared to inaccurate calibration curve (figure 6), but the accuracy was not always improved to be better than 1% as defined as successful calibration. Integration-mode gave worse results than simply relying on the inaccurate calibration curve (figure 7). The inaccuracies remained above 1% with an inaccuracy level within ±0.05, thus determining an unsuccessful CT calibration. The NRMSE between the iRad and the weDRR of the ground truth iCT image expressed inconsistencies that explained the unsuccessful CT calibration (table 4). These inconsistencies reached about 9% for integration-mode mean and mode in proton pencil beam scanning and were hardly reduced below 1% for list-mode detector configuration in carbon ion pencil beam scanning. Larger NRMSE was quantified for integration-mode hist due to the normalization by smaller WET occurrences. These inconsistencies enabled a fluctuation of the optimized calibration curve in the order of the inaccuracies, thus preventing successful CT calibration. In general, integration-mode mean and mode performed worse than integration-mode hist. The latter remained worse than list-mode. Helium and carbon ions demonstrated slightly better accuracy than protons. The high ion statistics did not significantly benefit the CT calibration compared to low ion statistics. The statistical test applied to proton pencil beam scanning showed that there were no significant differences between low and high ion statistics. Statistically meaningful differences were instead observed in high ion statistics between protons and helium ions, as well as between protons and carbon ions.

To improve the accuracy of the CT calibration the constraint of monotonically increasing calibration curve was embedded in the optimization algorithm (appendix-table A1). Significant differences between constrained and unconstrained optimization, although small, were found in low proton statistics. The selection of multiple slices (appendix-table A2) and multiple angles (appendix-table A3) did not decisively improve results. Even the results for the lateral iRad resulted comparable to the results for the frontal iRad. The choice of the lateral iRad over the frontal one was supported by the separation of the treatment couch and the supporting pillow from the patient. Statistically meaningful differences between single angle and multiple angles as well as single slice and multiple slices were detected in low proton statistics when considering the highly inhomogeneous nasal cavity slice. For this slice, the lateral iRad statistically differed from the frontal iRad. The statistical test applied to high proton statistics showed that there were no
Figure 6. Inaccurate and optimized calibration curves for list-mode detector configuration in proton pencil beam scanning at low ion statistics for nasal cavity slice (A) and brain slice (B) and for selected angles of the iRad and for inaccuracy level within ± 0.05.

significant differences between selections based on 1, 2 and 3 times the standard deviation of the ion energy (appendix-table A4).

4. Discussions

In this study, a patient-specific CT calibration based on iRad was investigated in a CT image of a head-and-neck patient, at clinically acceptable imaging doses, for both list-mode and integration-mode detector configurations. Proton, helium and carbon ion pencil beam scanning, with realistic energy compatible to current accelerator and detector technologies, were considered.

The selection of slices and angles of the iRads, as well as the choice of the optimization algorithm, was guided by an extensive investigation in analytical simulations in a numerical anthropomorphic phantom. In particular, the iRad should include all the points of the calibration curve matching the HUs of the CT image. Hence, the calibration curve should be defined as a function of the tissue linear attenuation coefficients or HUs of the CT image. Once this condition is met, a single iRad could be sufficient for the patient-specific CT calibration. To further investigate in this direction, a selection of the points of the calibration curve guided by their relevance in treatment planning (i.e. dose calculation inaccuracies) is currently under consideration.

When comparing different optimization algorithms, better performance was obtained with the particle swarm algorithm independently of the inaccuracy models. Particularly, the multiplicative inaccuracy model can add more inaccuracies to RSP values larger than one and less inaccuracies to RSP values smaller than one. In other words, the multiplicative inaccuracy model shortens the axes of the multidimensional variable space corresponding to the points of the calibration curve from air to water (with smaller inaccuracies) and stretches those from water on (with larger inaccuracies). The update of the simplex of the Nelder–Mead algorithm seems favored by these shortening and stretching of the axes (appendix). This was likely the cause of the better performance with respect to the additive inaccuracy model. With the particle swarm algorithm, the motion of the particles is always consistent to the axes of the multidimensional variable space, either with
additive or relative inaccuracy models. However, since the inaccuracy model in reality is unknown, an optimization algorithm sufficiently insensitive to the inaccuracy model, as demonstrated by the particle swarm, should be chosen.

The robustness against controlled inaccuracy levels of the calibration curve observed in analytical simulations of idealized carbon ion trajectories put forward the CT calibration feasibility based on less accurate in-room cone beam CT systems or even non-ionizing magnetic resonance imaging (Dowling et al 2012). However, in clinical data and in analytical simulations of idealized proton trajectories the CT calibration, although robust against high inaccuracy level (within ± 0.30), did not succeed at realistic inaccuracy level (within ± 0.05) because the inconsistencies between the iRad and the weDRR of the ground truth iCT image were in the same order of magnitude as the applied inaccuracies. In analytical simulation of idealized carbon ion trajectories, these inconsistencies were not present by construction.

These inconsistencies are differently present in list-mode and integration-mode detector configurations. Integration-mode is affected by range mixing, caused by MCS and BSS. The range mixing in integration-mode mean is literally mixed by assigning the most probable straight ion trajectory to the most probable range component. On the other hand, in integration-mode mode the range mixing for relatively high ion statistics is solved as the most probable range component is replaced by the range component with maximum occurrence. It should be noted that these inconsistencies are differently present in proton, helium and carbon ion pencil beam scanning. Particularly, larger MCS and BSS slightly penalize protons over helium and carbon ions, although ion statistics (for the same physical dose) work in favor of them. Energy straggling and non-ideal detectors, the latter not considered in this study, are sources of noise in the iRad that translates into additional inconsistencies in both detector configurations. For the integration-mode detector configuration, the noise increases in presence of low ion statistics, as the linear decomposition of the discrete energy loss signal is required. List-mode is affected by inaccuracies of the estimation of ion trajectory under the assumption of uniform water, with particular reference to the unavoidable treatment couch and the
supporting pillow. These inaccuracies are also present in integration-mode hist, when the proposed dedicated objective function is applied. Finally, anatomical changes between treatment planning and delivery scenarios, not considered in this study, could be an additional source of inconsistencies, demanding appropriate compensation (Gianoli et al 2014, 2016, Palaniappan et al 2019, 2020).

The proposal of the dedicated objective function for integration-mode detector configuration based on the entire WET histogram (hence, integration-mode hist) found the motivation in the poorer results obtained with integration-mode mode and integration-mode mean. Results benefited from the full exploitation of the information (the entire WET histogram). Neglecting some of the WET components increased the inconsistencies due to range mixing between the weDRR and the iRad that affected the CT calibration. Particularly, fluctuations of the estimated inaccuracies have been observed (Krah et al 2019). These fluctuations have been mitigated by regularization in the optimization (Zhang et al 2018), that causes undesired insensitivity to inaccuracies when based on the (actually unknown and not unique in reality) true calibration curve (Krah et al 2019), and/or minimized by removing from the optimization the regions relevant to range mixing (difficult in patient anatomy) (Doolan et al 2015, Zhang et al 2018).

The CT calibration based on proton radiography has been so far experimentally obtained in phantom and numerically observed in anthropomorphic phantoms for list-mode detector configurations at relatively high energy (Collins-Fekete et al 2017) and for integration-mode detector configuration guided by information about range mixing (Doolan et al 2015, Zhang et al 2018) and/or provided with information about the inaccurate calibration curve (Zhang et al 2018) or the true one (Krah et al 2019). Realistic calibration curves have been adopted as inaccurate, without controlling the applied inaccuracies. The CT calibration has been also performed by excluding from the optimizations non-straight ion trajectories in a pioneer prototype of list-mode detector in a dog patient at high dose (Schneider et al 2005) or by excluding regions relevant to range mixing in real tissue samples in an integration-mode detector (Doolan et al 2015). However, in these two latter works the ground truth iCT image was not available, thus making the optimization of the objective function hardly interpretable as a minimization of the inaccuracies in the optimized calibration curve.

5. Conclusions

The feasibility of successful patient-specific CT calibration depends on detector technologies and is primarily limited by the inconsistencies between the iRad and the ground truth, unknown in reality. Only in absence of these inconsistencies, the optimization algorithm, which is desirably insensitive to the inaccuracy model and provided with reasonable constraints, robustly adjusts the calibration curve. The importance of the definition of the calibration curve, as well as the selection of angles and slices, is demonstrated. The integration-mode detector configuration is limited by range mixing due to MCS and the BSS when limited to the weighted mean (integration-mode mean) or the mode (integration-mode mode) of the WET components. Potentials of integration-mode detector configuration are entrusted to the full exploitation of the entire WET histogram of components (integration-mode hist). List-mode detectors represent the configuration of choice. However, their technological complexity does not work in favor of clinical translation. Protons are slightly penalized over helium and carbon ions due to larger MCS and BSS. Ion statistics (for the same physical dose) play a secondary role. Clinically acceptable imaging dose, patient anatomy and setup do not generally prevent the feasibility of successful patient-specific CT calibration.

Acknowledgments

The work is supported by the Deutsche Forschungsgemeinschaft (DFG) project Hybrid ImaGing framework in Hadrontherapy for Adaptive Radiation Therapy (HIGH ART). Authors acknowledge Dr. Andrea Mairani and Dr. Thomas Tessonnier for providing phase space files of the Heidelberg Ion Beam Therapy Center.

Appendix

Optimization algorithm for the patient-specific CT calibration based on iRads

The CT calibration is based on optimization algorithms searching for the minimum of the objective function $f : \mathbb{R}^n \to \mathbb{R}$ in an unconstrained multidimensional variable space $\mathbb{R}^n$, so that $n$ represents the variable degree of freedom in terms of points of the calibration curve (in this work $n$ is equal to 11 or 13). A linear solver embedded in Matlab (lsqlin.m) based on Hessian matrix and provided with the optimization algorithm Interior-Point without inequality constraints was chosen for linear minimization. Two direct (without objective function derivative) non-linear optimization algorithms embedded in Matlab were also considered.
First, the Nelder–Mead algorithm \((\text{fminsearch.m})\) considers a simplex with \(n + 1\) vertices, where each vertex is described in terms of position \(x_n \in \mathbb{R}^n\) in the multidimensional variable space. The algorithm iteratively updates the simplex based on the evaluation of the objective function at each vertex position. In particular, the worst vertex is replaced with the reflected centroid of the remaining \(n\) vertices across the opposite best face of the simplex (reflection). Afterwards, either expansion or contraction and shrinkage are applied. As the volume of the simplex decreases, the minimum of the objective function is found. The convergence is defined when the relative changes in both the objective function and the vertex position are less than \(10^{-10}\). Second, the particle swarm algorithm \((\text{particleswarm.m})\) considers a population (swarm) of candidate solutions (particles) and describes them in terms of position \(x_n \in \mathbb{R}^n\) and velocity \(v_n \in \mathbb{R}^n\) in the multidimensional variable space. The algorithm iteratively searches for the best current position at the minimum of the defined objective function. The particles move toward the best current position according to their current position and velocity. The best current position is updated as a better position is found by other particles. The convergence is defined when the relative change in the objective function is less than \(10^{-10}\).

### Additional results

**Table A1.** Median/iqr MRE (%) of the inaccurate and optimized calibration curves provided with constraints in the optimization algorithm for different detector configurations in proton pencil beam scanning at low ion statistics for selected slices and angles of the iRad and for inaccuracy level within \(\pm 0.05\).

| Selected slices | Nasal cavity | Brain |
|----------------|--------------|-------|
| Selected angle | Frontal iRad | Lateral iRad | Frontal iRad | Lateral iRad |
| Inaccurate calibration curve | | 2.36/0.08 |
| Integration-mode mode | 3.20/0.08 | 2.95/0.46 | 4.10/0.15 | 2.76/0.57 |
| Integration-mode mean | 4.26/1.18 | 3.82/0.87 | 3.33/1.00 | 3.90/1.13 |
| Integration-mode hist | 2.54/0.32 | 2.78/0.31 | 2.23/0.19 | 1.91/0.07 |
| List-mode | 2.08/0.71 | 1.98/0.50 | 2.59/0.37 | 1.92/1.27 |

**Table A2.** Median/iqr MRE (%) of the inaccurate and optimized calibration curves for different detector configurations in proton pencil beam scanning at low ion statistics for selected angles but multiple slices of the iRad and for inaccuracy level within \(\pm 0.05\). Slice 41 corresponds to the nasal cavity slice, slice 61 corresponds to the brain slice.

| Selected angles | Frontal iRad | Lateral iRad | Frontal iRad | Lateral iRad | Frontal iRad | Lateral iRad |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Multiple slices | Frontal iRad 41,61 | Lateral iRad 36,46 | Lateral iRad 56,66 |
| Inaccurate calibration curve | 2.54/0.68 | 2.38/0.24 | 2.24/0.24 |
| Integration-mode mode | 3.49/0.36 | 2.73/0.39 | 3.14/0.78 | 2.40/0.53 | 3.55/0.59 | 3.64/0.47 |
| Integration-mode mean | 4.26/0.75 | 3.86/0.37 | 3.24/1.29 | 2.88/1.18 | 3.70/0.57 | 4.27/0.77 |
| Integration-mode hist | 4.11/0.48 | 3.36/0.68 | 2.73/1.01 | 2.55/2.22 | 2.94/0.81 | 2.56/0.40 |
| List-mode | 1.88/0.44 | 1.48/1.44 | 1.47/0.24 | 1.85/0.69 | 1.43/0.36 | 2.39/1.03 |

**Table A3.** Median/iqr MRE (%) of the inaccurate and optimized calibration curves for different detector configurations in proton pencil beam scanning at low ion statistics for selected slices but multiple angles of the iRad and for inaccuracy level within \(\pm 0.05\).

| Selected slices | Nasal cavity | Brain |
|----------------|--------------|-------|
| Multiple angles | Nasal cavity Brain |
| Frontal and lateral iRads 0°,90° 0°,20°,40°,60°,80°,90°,100°,120°,140°,160° |
| Inaccurate calibration curve | 2.39/0.72 | 2.51/0.33 |
| Integration-mode mode | 3.00/0.67 | 3.52/0.15 | 3.96/1.08 | 2.89/0.86 |
| Integration-mode mean | 4.20/0.55 | 4.27/0.29 | 4.10/0.77 | 3.64/1.15 |
| Integration-mode hist | 4.91/0.74 | 4.62/0.05 | 5.28/0.69 | 5.29/1.10 |
| List-mode | 2.34/0.32 | 2.41/0.42 | 1.72/0.28 | 2.25/0.37 |
Table A4. Median/interquartile range (IQR) MRE (%) of the inaccurate and optimized calibration curves for different detector configurations in proton pencil beam scanning at high ion statistics, selected based on 1, 2 and 3 times the standard deviation of the ion energy, for selected slices and angles of the iRad and for inaccuracy level within ± 0.05.

| Selected slices | Selected angle | Nasal cavity | Brain |
|-----------------|----------------|--------------|-------|
|                 | Frontal iRad   | Lateral iRad | Frontal iRad | Lateral iRad |
| 1× Inaccurate calibration curve | 2.17/0.47 |
| Integration-mode mode | 3.80/1.07 | 2.55/1.17 |
| Integration-mode mean | 4.14/0.83 | 3.89/0.71 |
| Integration-mode hist | 4.48/0.56 | 3.26/0.42 |
| List-mode | 2.72/0.39 | 2.24/0.69 |
| 2× Inaccurate calibration curve | 2.38/0.26 |
| Integration-mode mode | 4.21/0.83 | 3.20/1.14 |
| Integration-mode mean | 4.75/0.87 | 4.25/0.61 |
| Integration-mode hist | 3.37/0.57 | 3.94/0.72 |
| List-mode | 1.93/1.13 | 2.03/1.00 |
| 3× Inaccurate calibration curve | 2.17/0.16 |
| Integration-mode mode | 3.67/0.89 | 3.15/0.09 |
| Integration-mode mean | 4.67/0.71 | 4.20/0.38 |
| Integration-mode hist | 4.41/2.26 | 4.36/0.52 |
| List-mode | 1.89/0.14 | 2.15/1.04 |

**References**

Böhlen T T, Cerutti E, Chin M P W, Fassó A, Ferrari A, Ortega P G and Vlachoudis V 2014 The FLUKA code: developments and challenges for high energy and medical applications Nucl. Data Sheets 120 211–4

Bourque A E, Carrier J F and Bouchard H 2014 A stoichiometric calibration method for dual energy computed tomography Phys. Med. Biol. 59 2059

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

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

Dowling J A, Lamberti J, Parker J, Salvado O, Frippe I, Cap P and Greer P B 2012 An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy Int. J. Radiat. Oncol. Biol. Phys. 83 e5–11

Durante M and Loeffler J S 2010 Charged particles in radiation oncology Nat. Rev. Clin. Oncol. 7 37–43

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

Ferrari A, Sala P R, Fassó A and Runft J 2005 FLUKA: a multi–particle transport code (Program version 2005) (No. INFN-TC-05-11)

Giancoli C, Dedes G, Meyer S, Magallanes L, Landry G, Nijhuis R and Parodi K 2016 PO-0909: merging proton radiographies with magnetic resonance imaging (MRI) alone treatment planning and adaptive MRI-based prostate radiation therapy Int. J. Radiat. Oncol. Biol. Phys. 83 e5–11

Göppel M, Meyer S, Magallanes L, Kopp B, Parodi K and Giancoli C 2017 Session 23: radiation therapy III – particle therapy II Biomedical Engineering/Biomedizinische Technik p 2421

Hong L, Goitein M, Bucciolini M, Comiskey R, Gottschalk B, Rosenthal S and Urie M 1996 A pencil beam algorithm for proton dose calculations Phys. Med. Biol. 41 1305

Jiang H, Seco J and Paganetti H 2007 Effects of Hounsfield number conversion on CT based proton Monte Carlo dose calculations Med. Phys. 34 1439–49

Kennedy J and Eberhart R 1995 Particle swarm optimization Proc. ICNN’95 Int. Conf. on Neural Networks (IEEE) 4 1942–8

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

Lagaris I C, Reeds J A, Wright M H and Wright P E 1998 Convergence properties of the Nelder–Mead simplex method in low dimensions SIAM J. Optim. 9 112–47

Meyer S, Giancoli C, Magallanes L, Kopp B, Tessonnier T, Landry G and Parodi K 2017 Comparative Monte Carlo study on the performance of integration- and list-mode detector configurations for carbon ion computed tomography Phys. Med. Biol. 62 1096

Meyer S, Kamp F, Tessonnier T, Mairani A, Belka C, Carlson D J and Parodi K 2019 Dosimetric accuracy and radiobiological implications of ion computed tomography for proton therapy treatment planning Phys. Med. Biol. 64 125008
Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations Phys. Med. Biol. 57 R99
Palaniappan P, Meyer S, Kamp F, Belka C, Riboldi M, Parodi K and Gianoli C 2019 A 2D-3D deformable image registration framework for proton radiographies in adaptive radiation therapy Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC), 2019 IEEE (IEEE) pp I–4
Palaniappan P, Meyer S, Kamp F, Belka C, Riboldi M, Parodi K and Gianoli C 2020 Deformable image registration of the treatment planning CT with proton radiographies for adaptive radiation therapy Phys. Med. Biol. in press (https://doi.org/10.1088/1361-6560/ab8fc5)
Parodi K 2014 Heavy ion radiography and tomography Phys. Med. 30 539–43
Rietzel E, Schardt D and Haberer T 2007 Range accuracy in carbon ion treatment planning based on CT-calibration with real tissue samples Radiat. Oncol. 2 1
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 Phys. Med. Biol. 58 413
Schaffner B and Pedroni E 1998 The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power Phys. Med. Biol. 43 1579
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 Phys. Med. Biol. 44 27
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
Schneider W, Bortfeld T and Schlegel W 2000 Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions Phys. Med. Biol. 45 459
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
Schulte R W, Penfold S N, Tafas J T and Schubert K E 2008 A maximum likelihood proton path formalism for application in proton computed tomography Med. Phys. 35 4849–56
Segars W P, Tsui B M, Lalush D S, Frey E C, King M A and Manocha D 2002 Development and application of the new dynamic Nurbs-based Cardiac-Torso (NCAT) phantom
tessonnier T, Marcelos T, Mairani A, Brons S and Parodi K 2016 Phase space generation for proton and carbon ion beams for external users’ applications at the Heidelberg ion therapy center Front. Oncol. 5 297
Yang M, Zhu X R, Park P C, Titt U, Mohan R, Virshup G and Dong L 2012 Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration Phys. Med. Biol. 57 4095
Zhang R, Sharp G C, Lee K W, Cascio E W, Harms J M, Flanz J B and Lu H M 2018 Iterative optimization of relative stopping power by single detector based multi-projection proton radiography Phys. Med. Biol. 64 065022