Original paper

Experimental benchmarking of RayStation proton dose calculation algorithms inside and outside the target region in heterogeneous phantom geometries

Sirinya Ruangchan^{a,c}, Barbara Knäusl^{a,b,*}, Hermann Fuchs^{a,b}, Dietmar Georg^{a,b}, Monika Clausen^a

^a Division of Medical Radiation Physics, Department of Radiation Oncology, Medical University of Vienna, Vienna, Austria

^b Division of Medical Physics, MedAustron Ion Therapy Center, Wiener Neustadt, Austria

^c Division of Therapeutic Radiation and Oncology, Department of Radiology, King Chulalongkorn Memorial Hospital, Bangkok, Thailand

ARTICLE INFO

Keywords:

Pencil beam algorithm
Monte Carlo algorithm
Heterogeneity
Proton therapy

ABSTRACT

Purpose: The aim of the presented study was to complement existing literature on benchmarking proton dose by comparing dose calculations with experimental measurements in heterogeneous phantom. Points of interest inside and outside the target were considered to quantify the magnitude of calculation uncertainties in current and previous proton therapy practice that might especially have an impact on the dose in organs at risk (OARs). **Methods:** The RayStation treatment planning system (RaySearch Laboratories), offering two dose calculation algorithms for pencil beam scanning in proton therapy, i.e., Pencil Beam (PB) and Monte Carlo (MC), was utilized. Treatment plans for a target located behind the interface of the heterogeneous tissues were generated. Dose measurements within and behind the target were performed in a water phantom with embedded slabs of various tissue equivalent materials and 24 PinPoint ionization chambers (PTW). In total 12 test configurations encompassing two different target depths, oblique beam incidence of 30 degrees and range shifter, were considered. **Results:** PB and MC calculated doses agreed equally well with the measurements for all test geometries within the target, including the range shifter (mean dose differences $\pm 3\%$). Outside the target, the maximum dose difference of 9% (19%) was observed for MC (PB) for the oblique beam incidence and inserted range shifter. **Conclusion:** The accuracy of MC dose algorithm was superior compared to the PB algorithm, especially outside the target volumes. MC based dose calculation should therefore be preferred in treatment scenarios with heterogeneities, especially to reduce clinically relevant uncertainties for OARs.

1. Introduction

Dose calculation accuracy in the context of high precision radiotherapy, irrespective of beam quality, is one of the long-standing topics in medical physics research [1,2]. While pencil beam algorithms have become obsolete in high energy photon beam therapy for more than a decade, these algorithms are still widely used in particle beam therapy. Moreover, current clinical knowledge and experience is based on treatment planning with pencil beam algorithms [3]. Monte Carlo simulations serve as the gold standard in dose calculation since several decades. While powerful and versatile research Monte Carlo simulations are often used to benchmark dose calculation of commercial treatment planning system (TPS), their calculation speed is not up to the clinical needs. Therefore, assumptions are typically made to speed up the performance to be practical for clinical applications. In external beam therapy, Monte Carlo based dose calculation in TPS has become the standard for electron therapy for a

long time. In photon beam therapy, MC is integrated on several commercial TPS, and proton therapy is the next treatment technique to follow this trend. Recently, in a multicentric dosimetric study involving centers with different equipment, analytic dose calculation for proton lung treatments was demonstrated to be inadequate [4–6].

Several studies reported on the dosimetric validation of commercially implemented MC proton dose calculation algorithms [7–9]. The superior performance of MC over PB algorithms was reported in the presence of range shifter with large air gaps and oblique treatment beams as well as in the presence of inhomogeneities and different density interfaces [7–13]. Although the dose calculation accuracy of MC and PB has been investigated in different studies, most of them are based on measurements in homogeneous water which do not realistically mimic clinical cases with complex structures. When using more complex benchmarking geometries, e.g., with tissue heterogeneities, phantom materials might induce range changes in the beam path and are therefore a possible source of errors in the validation

* Corresponding author.

E-mail address: barbara.knaeusl@meduniwien.ac.at (B. Knäusl).

<https://doi.org/10.1016/j.ejmp.2020.07.010>

Received 12 February 2020; Received in revised form 3 July 2020; Accepted 7 July 2020

Available online 18 July 2020

1120-1797/ © 2020 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
Densities and elemental compositions used in this study.

Tissue	Density (g/cm ³)	Elemental compositions (percentage by weight)							
		H (1)	C (6)	N (7)	O (8)	P (15)	S (16)	Cl (17)	Ba (56)
Bone	1.820	3.6	28.8	1.1	32.0	10.8	0.1	–	23.3
Lung inhale	0.195	8.8	67.5	3.5	18.6	–	–	1.6	–
Soft tissue	0.960	10.0	71.3	1.8	16.4	–	–	0.2	0.3
PMMA	1.190	8.1	60.0	–	32.0	–	–	–	–
Water	1.000	11.2	–	–	88.8	–	–	–	–

of dose calculations algorithms. Therefore, the use of plastic-water phantoms is not recommended for reference dosimetry of proton beams [10], since the elemental compositions and interaction cross-sections in the plastic water are different from those in the water [11] and additional fluence corrections would be required, especially for high energy proton beams [12].

The aim of the presented study was to complement existing literature on benchmarking proton dose calculation algorithms by comparing dose calculations with dosimetric measurements in more complex test geometries. More specifically, test cases that encompass tissue heterogeneities, the presence of a range shifter or oblique beam incidence were considered to benchmark the Monte Carlo algorithm (MC) as well as the pencil beam algorithm (PB) of the commercial TPS RayStation (RaySearch Laboratories, Stockholm, Sweden). Besides the final dose calculation, also the spot positions and weight optimization were examined for both algorithms. For respective experiments, tissue-equivalent slabs (bone-lung or bone-soft tissue) enclosed in a watertight holder were placed in a water phantom in which the measurements were performed with an ionization chamber array inside as well as outside the treated volumes. Different distances with respect to the tissue heterogeneities were considered to quantify the magnitude of dosimetric uncertainties in current and previous proton therapy practice that might have an impact, especially on OAR doses.

2. Methods

2.1. Treatment planning system and dose calculation algorithms

To investigate the accuracy of proton dose calculations in complex irradiation geometries, two algorithms implemented in the commercial RayStation TPS (RaySearch, Sweden) were employed in this study, i.e., a Pencil Beam (PBv4.2) and Monte Carlo (MCv4.1) calculation algorithm which can be chosen independently for spot weight and position optimization as well as for the final dose calculation. A detailed study on the commissioning aspects of the TPS and the two dose calculation algorithms was published recently [13]. As mentioned earlier, this study goes beyond the existing work by focusing on tissue heterogeneities surrounded by water (see Section 2.3).

For the sake of completeness, a short algorithm description is given in the following. The PB is based on superposition and convolution methods, which calculate the dose distribution from an energy deposition kernel and energy fluence distribution. The kernel accounts for all possible radiation effects arising from the interaction of protons, such as scattering and absorption of secondary electron. Differently to the PB algorithm, the MC takes into account the physics of particle interactions on a particle by particle basis using theoretical models or experimental cross section data for electromagnetic and nuclear interactions [14–18]. The implemented version of the MC engine in the RaySearch TPS comes along with approximations for the transport mechanisms and electromagnetic processes. Compared to the general MC codes, the particle types, energy ranges and nuclear interactions in MC algorithm implemented in RayStation are based on the pre-computed cross section libraries which allow faster dose calculations. In addition, only the averaged energy deposition per voxel, without multiple scattering and straggling processes, is taken into account in calculations also saving an additional time [16,19]. Both beam engines of the

respective algorithms use the same beam model, which was derived from measured spot maps, integral depth dose distributions and absolute dose measurements. For beam model validation, absolute dose measurements in isocentric as well as non-isocentric conditions were performed as required by the vendor [13]. Besides in air-measurements, most basic beam data were acquired in the MP3-PL water phantom (PTW, Freiburg, Germany) in homogeneous conditions. The final dose calculation uncertainty, quantified at the TPS validation stage in homogenous conditions, was 3% for both algorithms. In the TPS RayStation, a range shifter is modeled as an external element in the beam path, which does not require separate measurements for model creation.

2.2. Treatment planning and test configurations

In the context of this study, a CT dataset of a cubic water phantom with inserted heterogeneities mimicking the materials used in the experiment was generated in MATLAB (2016b, Mathworks, USA) and imported into the TPS. The heterogeneities in the different test geometries, i.e., bone, lung, soft tissue, and PMMA were then overwritten in the TPS with the corresponding tissue densities and elemental composition information provided by the vendor (CIRS Inc., USA) of the heterogeneity slabs, see Table 1.

Target definition, heterogeneity positioning and tissue contouring was performed for this virtual phantom according to the test cases 1 to 7 described below. Dose calculations were performed in a $1 \times 1 \times 1 \text{ mm}^3$ grid for all phantom configurations, with 10^5 protons per spot considered for optimization for the MC algorithm [13,19]. The optimization aim was to deliver a uniform RBE-weighted dose of 1.1 Gy to a $4 \times 4 \times 4 \text{ cm}^3$ target, considering the scaling factor of 1.1 for biological effectiveness of protons. All measurements and dosimetric comparisons performed in this study were focused at the physical dose of 1 Gy. D_{mean} and the DVH value of $D_{95\%}$ were used to evaluate the target coverage. Three different approaches were tested. The spot positions and spot weights were optimized and the final doses were calculated with MCv4.1, this approach is labeled as “MC” throughout the whole manuscript. The second approach, labeled as “PB_{MC}”, employs the MCv4.1 when calculating spot position and spot weights; however, the final dose was calculated with PBv4.2. These two approaches were utilized for all test cases to explore the impact of the two different algorithms on the final dose calculation accuracy based on the same spot distribution. In the third approach, PB v4.2 was utilized for the optimization, to investigate differences in fluence patterns that were delivered during the experiments. The final dose was also calculated with the pencil beam algorithm and this approach is labeled “PB”. The last approach was employed in this study to follow a reasonable clinical practice. For MC dose calculation was the mean relative statistical uncertainty (MRSU) set to 0.5% [13]. The parameters used for the optimization and dose calculations were equivalent to the settings used for the beam model validation.

As stated above, the focus of this study was the dose calculation accuracy in and outside the target in the presence of tissue heterogeneities. More specifically, tissue interfaces bone-lung and bone-soft tissue representing the clinical scenarios for cranial, pelvic and thoracic malignancies were considered. Furthermore, the impact of the relative position between the inhomogeneous tissue slabs and the target was assessed by positioning the $4 \times 4 \times 4 \text{ cm}^3$ target at two different depths behind the

Fig. 1. Schematic representation of test cases. The red rectangles represent the target, while the yellow dashed rectangles indicate the measurement volume covered by the PinPoint IC. All proton beams are directed onto the phantom from left. Test case 1: Target directly behind the bone - lung inserts. Test case 2: Target at 8 cm depth behind the bone - lung inserts. Test case 3: Target directly behind the bone and soft tissue inserts. Test case 4: Target at 8 cm depth behind the bone and soft tissue inserts. Test case 5: Target directly behind the bone - soft tissue inserts with the range shifter. Test case 6: Target directly behind the bone - soft tissue inserts with 30 degrees beam angle. Test case 7: Target directly behind the bone - soft tissue inserts with 30 degrees beam angle and the range shifter.

heterogeneous inserts. The first target was created directly behind the tissue interfaces (SOBP1) while the center of the second target was placed at 8 cm from the inner surface of the phantom (SOBP2). Additionally, dedicated test cases with a range shifter in the beam path and an oblique beam incidence of 30 degrees were defined. In total 7 test cases are schematically illustrated in Fig. 1. The investigated region outside the target was located up to 3 cm distance behind the target for perpendicular beam incidence. For the oblique beam incidence this outside region was situated up to 3 cm laterally from the target and 1 cm behind the target, respectively. These areas are illustrated in Fig. 1 for all individual test geometries.

By changing the distances from the nozzle (air gaps of 16.1 cm, 36.1 cm and 66.1 cm) and using the MCv4.1 and the PBv4.2 for spot pattern and fluence determination, 12 test configurations were measured as summarized in Table 2 including the respective proton beam energy ranges.

2.3. Experimental measurements

All measurements were carried out at the synchrotron-based accelerator facility MedAustron with scanning pencil beam delivery and a horizontal beam line configuration [20]. Experimental dosimetry for all 12 test configurations was performed following the patient-specific QA procedure implemented at MedAustron utilizing an ionization chamber (IC) array consisting of 24 PinPoint ICs (0.03 cm³, T31015, PTW, Germany) connected to two Multidos electrometers (T10004, PTW, Germany). PinPoint ICs have a combined relative standard uncertainty of 1.9% [21]. The chambers were positioned in a 3D array holder (PTW, Germany), see right insert of Fig. 2. In this 3D array holder, always four chambers were arranged in a row (lateral distribution) with 1 cm distance between the chambers, at six depths along the beam path. The block position was designed so that the individual rows were shifted by 6 mm horizontally to avoid shielding of the ICs and

positioned along the falloff region of the SOBP. The block with the ICs was attached to a motorized arm of the water tank and the dose was determined at several positions along the beam path moving the 3D holder horizontally to cover the dose distribution inside and outside of the target. After treatment plan generation, an in-house developed script was used in the TPS to position the ICs and to calculate the dose for each of them [22]. The dose values were determined and exported for all setups taking into account the effective point of measurement of the cylindrical PinPoint ICs, which was determined for ion beams according to IAEA TRS-398 [23,24].

Dose measurements were performed in the water phantom with heterogeneous inserts, see Fig. 1. The beam entrance angles were perpendicular (test cases 1 to 5) and oblique (30 degrees, test case 6 and 7) with respect to the front window of the water phantom. Furthermore, in order to assess the overall measurement and setup uncertainties, repetitive measurements were performed in the water phantom without heterogeneities. The tissue equivalent plates (CIRS Inc., USA), each with dimensions of 30 × 15 cm² and a thickness of 1 cm, were inserted in pairs, i.e. side by side placed into a custom-made watertight holder, see upper left insert of Fig. 2. The holder consisting of 0.2 cm thick PMMA plates was fixed in a frame enclosing the tissue-equivalent materials and then inserted into the water phantom. The uniform distance between the holder and the inner wall of the water tank was kept constant to 2 cm for all experiments. Structures of PMMA plates enclosing the tissue-equivalent materials from the front and the backside were included in the CT dataset of the phantom. The positions of the 24 ICs received from the script were loaded into the PlanVerificator software V1.1 [22]. Once, the chambers were positioned in the water phantom, the treatment plan was irradiated.

2.4. Dosimetric evaluation and data analysis

The measured charge of the 24 ICs was read out, and the absorbed dose

Table 2
Overview of 12 test configurations used in this study.

Configuration	Material inserts	Target location	Range shifter	Beam angle	Energy-range (MeV)	Air gap (cm)	Algorithm
1	Bone - Lung	SOBP1	–	perpendicular	62.4 – 109.6	66.1	MC, PB _{MC}
2	Bone - Lung	SOBP2	–	perpendicular	83.9 – 129.0	66.1	MC, PB _{MC}
3	Bone - Soft tissue	SOBP1	–	perpendicular	67.5 – 109.6	66.1	MC, PB _{MC}
4	Bone - Soft tissue	SOBP1	–	perpendicular	67.5–112.3	36.1	MC, PB _{MC}
5	Bone - Soft tissue	SOBP1	–	perpendicular	67.5 – 109.6	66.1	PB
6	Bone - Soft tissue	SOBP2	–	perpendicular	90.4 – 129.0	66.1	MC, PB _{MC}
7	Bone - Soft tissue	SOBP1	in	perpendicular	97.4 – 134.4	16.1	MC, PB _{MC}
8	Bone - Soft tissue	SOBP1	in	perpendicular	97.4 – 134.4	36.1	MC, PB _{MC}
9	Bone - Soft tissue	SOBP1	in	perpendicular	97.4 – 134.4	66.1	MC, PB _{MC}
10	Bone - Soft tissue	SOBP1	in	perpendicular	97.4 – 134.4	16.1	PB
11	Bone - Soft tissue	SOBP1	–	30°	68.5 – 120.2	36.1	MC, PB _{MC}
12	Bone - Soft tissue	SOBP1	in	30°	95.9 – 144.2	36.1	MC, PB _{MC}

Fig. 2. Experimental setup using the watertight holder for tissue substitutes and 24 PinPoint IC exemplarily for bone and soft tissue inserts.

was determined as described previously [22]. The experimentally determined dose values for all test cases were then compared against the respective doses calculated with the different algorithms. Relative dose differences were determined and further used to create dose difference maps (1 chamber = 1 voxel) and histograms where appropriate. The mean dose differences (based on the physical dose) were always considered and calculated separately for the group of chambers located inside and outside the target. The dose differences were normalized to the maximum planned dose to provide a more realistic output for the dose distribution comparison [25].

$$\text{Percentage dose difference} = \frac{\text{TPS planned dose} - \text{Measured dose}}{\text{Maximum TPS planned dose}} \times 100$$

3. Results

The setup error assessment performed in the homogenous water phantom after each measurement with heterogeneities revealed no significant dose deviations and ensured the proper positioning of ICs relative to the water phantom. The mean point-to-point dose difference between the calculated dose (in water) and measurements performed with the 24 PinPoint ICs was $(1.1 \pm 0.1) \%$ and $(1.1 \pm 0.3) \%$ for MC and PB algorithms which were within 3% of final TPS dose calculation uncertainty [13].

3.1. Test case 1: Bone - lung with target directly behind the heterogeneities (SOBP1)

For test case 1 the agreement of the measurements with the MC calculated doses (mean dose difference of $(1.1 \pm 0.9) \%$) was comparable to the measurements in a homogenous water phantom. A maximum dose difference of 5.5% was observed at the heterogeneous interface for MC predictions, overestimating the dose towards the end of the target. Calculations performed with the PB_{MC} algorithm resulted in slightly larger dose differences than the MC within the target, with the mean dose difference of $(1.6 \pm 1.2) \%$ but a maximum dose difference up to 5.7%. Behind the target the dose difference for PB_{MC} increased up to 18.8% relative to the maximum planned dose, where the predictions underestimated the dose compared to the measurements. The comparison between the dose calculations and measurements for test case 1 is shown in Fig. 3a and the corresponding histograms are given in Fig. A.1a.

3.2. Test case 2: Bone - lung with target at 8 cm depth (SOBP2)

The dose differences between the calculated and measured doses were systematically smaller for this test case compared to test case 1. MC data showed a good agreement with the measurements with mean dose differences of $(0.6 \pm 0.4) \%$ within the target and maximum dose difference of 4.1% behind the target, respectively. The PB_{MC} algorithm underestimated the dose by 12.1% (maximum dose difference) behind the target, while differences within the target did not exceed 5% (Fig. 3b and Fig. A.1b).

3.3. Test case 3: Bone - soft tissue with target directly behind the heterogeneities (SOBP1)

For test case 3, MC and PB_{MC} plans were compared for two different air gaps, i.e. 66.1 cm and 36.1 cm. Further the scenario with 66.1 cm air gap was compared to the PB plan. For the SOBP1, a better agreement between measurement and dose prediction was observed when located behind the bone and soft tissue inserts compared to the bone-lung interface scenarios. This was expected as the density difference of the two materials was smaller.

Inside the target the agreement was very good for the air gap of 66.1 cm, with a mean dose difference of $(1.2 \pm 0.9) \%$ and $(1.3 \pm 1.1) \%$ for MC and PB_{MC}, respectively. These results were comparable to 36.1 cm air gap, for which dose differences of $(1.2 \pm 1.2) \%$ and $(1.3 \pm 1.5) \%$ were obtained. At the position behind the target for an air gap of 66.1 cm (Fig. 3c and Fig. A.1c), a maximum overestimation of the dose of 4.9% was observed for MC, while PB_{MC} underestimated the dose by maximal 10.4%. For the air gap of 36.1 cm, both algorithms overestimated the dose up to 7.0% and 13.1% for MC and PB_{MC}, respectively. For an air gap of 66.1 cm, the optimization of spot positions and weight was performed with the MCv4.1 as well as with the PBv4.2 algorithm. A mean dose difference of $(1.3 \pm 1.3) \%$ was found inside target for PB and agreed well with PB_{MC} ($1.3\% \pm 1.1\%$). The maximum dose underestimation of 7.3% and 10.4% was observed behind the target for PB and PB_{MC}, respectively. The histograms of the dose difference comparison for PB and PB_{MC} are shown in Fig. 4a.

Fig. 3. Dose differences between PB_{MC} and MC in the TPS are shown in the left column. Blue indicates PB_{MC} dose underestimation. Dose difference maps between planned and measured dose are shown for MC in the central column and for PB_{MC} in the right column; a) SOBP1 with bone-lung interface; b) SOBP2 with bone-lung interface; c) SOBP1 with bone-soft tissue interface; d) SOBP2 with bone-soft tissue interface. Negative values for the distance from the central axis were related to bone while positive values were obtained for lung or soft tissue slabs. The red squares indicate the measurement region, whereas the yellow dashed lines represent the distances from tissue interfaces to the target's center. Black vertical lines indicate the border between the target and non-target regions.

3.4. Test case 4: Bone – soft tissue with target at 8 cm depth (SOBP2)

Fig. 3d and Fig. A.1d show a comparison between the dose calculations and measurements with the bone-soft tissue inserts in the beam line for SOBP2. The dose differences were smaller than those in SOBP1 directly behind the interface (test case 3). MC and PB_{MC} showed a good agreement with mean dose differences of $(0.6 \pm 0.5) \%$ and $(1.0 \pm 0.9) \%$. Both algorithms predicted higher doses compared to measurements in this test configuration with the maximal dose difference of 8.4% (8.0%) for the MC (PB_{MC}) algorithm at the region behind the target.

3.5. Test case 5: Range shifter and air gaps (nozzle to isocenter distances)

Fig. 5 compares the dose difference maps for test case 5 derived from dose calculations and experimental measurements with three different air gaps (66.1 cm, 36.1 cm and 16.1 cm). The results from the two small air

gaps (16.1 cm and 36.1 cm) were comparable within the target for both algorithms. Also, when comparing the other configurations, similar maximal dose differences were found outside the target for MC and PB_{MC} for the air gap of 36.1 cm for SOBP1 (bone-soft tissue inserts) without the range shifter (Fig. A.3). The worst performance was observed for the largest air gap of 66.1 cm (Fig. A.2, Table A.1). Significant dose differences between dose calculation and measurements were observed between MC and PB_{MC} for all air gaps in the measurement regions outside the target (Fig. A.2), whereas the dose differences for PB_{MC} were more pronounced. The largest local differences behind the target were observed for PB_{MC} at the air gap of 66.1 cm, where the predictions overestimated the dose of up to 13.9% (maximum dose difference).

The results were comparable for the PB and PB_{MC} algorithms for the scenario without the range shifter. The difference was present for the configuration with the range shifter (air gap 16.1 cm), where the performance of the PB_{MC} algorithm was only a slightly better to the PB outcome,

Fig. 4. Histograms of dose differences between the calculations and measurements for PB and PB_{MC}; a) Bone and soft tissue inserts: SOBP1 without range shifter (air gap 66.1 cm); b) Bone and soft tissue inserts with range shifter (air gap 16.1 cm). Red vertical lines indicate clinically acceptable level (within 3%). PB > 3% and PB_{MC} > 3% are the number of measurement points that exceeded the limit of $\pm 3\%$.

especially within the target, Fig. 4b. The mean dose difference inside target of $(2.2 \pm 1.4)\%$ was observed for PB and $(2.1 \pm 1.4)\%$ for PB_{MC} with the investigated air gap.

3.6. Test case 6: Oblique beam incidence

The histograms representing oblique beam incidence with the air gap of 36.1 cm are shown in Fig. 6a. Good performance of both algorithms can be reported within the target with the mean dose difference of $0.8 \pm 0.6\%$ (MC) and $1.2 \pm 1.0\%$ (PB_{MC}) inside the target. Regarding the maximum dose difference, the largest local discrepancies of 2.1% and 8% for MC and 4.3% and 12.5% for PB_{MC} were found inside and outside the target, respectively.

3.7. Test case 7: Oblique beam incidence with a range shifter

MC based dose calculation agreed with the measurements within 3% in the target region. PB_{MC} clearly overestimated the dose and the mean dose difference of $(3.3 \pm 1.8)\%$ was obtained averaging over all measurement points (shown in Fig. 6b). The maximum differences of 9% and 19.2% (dose underestimation) for MC and PB_{MC} were obtained outside the target volume. These discrepancies were the largest for both algorithms when comparing all test configurations in this study.

A summary of all deviations between measurements and dose calculations performed with the two algorithms for seven different test cases (12 configurations) is additionally provided in Appendix (Table A.1). All three approached MC, PB_{MC} and PB presented a good agreement with the measurements inside the target, with an averaged dose difference well below 3%

for all configurations, except the configuration 12 ($3.3\% \pm 1.8\%$). Significantly higher dose discrepancies were found behind the target.

4. Discussion

In this study, the performance of two dose calculation algorithms, pencil beam and Monte Carlo, implemented in the state-of-the-art TPS for scanned proton therapy was quantitatively studied and benchmarked against dosimetric measurements in complex heterogeneous test geometries. In contrast to most previous studies, the dosimetry agreement was assessed in a water-filled phantom inside and outside the target regions.

The most pronounced dosimetric differences were observed for the test case with a bone-lung interface, where the PB_{MC} calculations underestimated the dose by up to 19%. As expected, the dose differences for the bone-soft tissue inserts were less pronounced than for the bone-lung inserts. In general, the observed limitations of a PB algorithm in the presence of lung heterogeneities is in line with previously published studies [6] and can be explained by differences in multiple Coulomb scattering between tissues with large density differences [7]. This disadvantageous feature of PB algorithms lies in its nature and cannot be optimized via improvements of the PB algorithms implementation. Similar uncertainties of 10–20% for PB algorithm was also reported Righetto et al. [17], however with the very dense materials, more specifically titanium. The MC calculations proved to be a reliable approach to compute the dose in patients with titanium implants.

On the other hand, there are other scenarios where dose calculation accuracy can be improved via implemented modifications. An example is the handling of a range shifter at the end of the nozzle exit, which was included

Fig. 5. Dose differences between PB_{MC} and MC in the TPS are shown in the left column. Blue indicates PB_{MC} dose underestimation. Dose difference maps between planned and measured dose with range shifter are shown for MC in the central column and for PB_{MC} in the right column; a. SOBP1 with bone-soft tissue interface: air gap = 16.1 cm; b. SOBP1 with bone-soft tissue interface: air gap = 36.1 cm; c. SOBP1 with bone-soft tissue interface: air gap = 66.1 cm. Negative values for the distance from the central axis were related to bone while positive values were obtained for lung or soft tissue slabs. The red squares indicate the measurement region, whereas the yellow dashed lines represent the distances from tissue interfaces to the target's center. Black vertical lines indicate the border between the target and non-target regions.

in test case 5 to mimic treatment geometries for cranial lesions located at the shallow depths. The spot sizes after the range shifter are larger along the beam path due to the multiple Coulomb scattering [26]. The observed dose calculation uncertainty for PB algorithms in various TPSs has been attributed to the lack of proper handling of secondary protons and the halo created in the range shifter and also to the particle transport over the air gaps [27,28]. In the present study, a good agreement ($\leq 3\%$) was found within the target in the presence of the range shifter with 16.1 cm, 36.1 cm and 66.1 cm air gaps for MCv4.1 and PBv4.2 based on an MC determined spot and fluence pattern. These results are also in line with Saini et al. [9], where the dose differences below 5% were reported in the SOBP using the recent version of RayStation TPS (version 4.2). According to the author's experience, this is a considerable improvement for the PB algorithm compared to those implemented in the previous versions of the TPS (PBv3.5 and older), achieved by the vendor. The newer versions of PB algorithm (from PBv4.1 onwards) introduce the spot subdivisions at the patient surface. For the calculations with the range shifter, less overlap of sub-spots is thus achieved at the patient surface compared to the previous version where sub-spots were generated at the range shifter [19]. However, the improvements of PB algorithm did not solve this issue completely [13]. In the presented study, the dose deviations for the test configurations with the range shifter are clearly higher compared to the test configurations without the range shifter. The MC algorithm is systematically underestimating the dose, and the dose calculation accuracy of MC is similar to PB_{MC} in the presence of range shifter, especially in the target regions. This

is related to the fact that the range shifter, located in the nozzle, is modeled as a passive element in the TPS. The precision of MC is, however, superior to PB_{MC}, and these differences are further increasing with larger air gaps.

Two target locations, i.e., SOPB1 and SOPB2, with respect to heterogeneities were investigated to quantify the performance of dose calculations for a tumor volume situated at different distance relative to the heterogeneity interface. The performance of both algorithms in the target was good with the maximal dose difference of 5%, but clearly worse for the shallow target (SOBP1) located directly behind the heterogeneities. These findings are in line with theory, wherein for the shallow targets, are the cross-sections for nuclear scattering higher at lower proton energy, which is more difficult to model and hence a greater angular spread contributes to higher dose deviations at the shallow depths [6,15].

The projected cross-section of the inhomogeneous region is broader along the beam path for the oblique beam incidence when comparing with other test cases. This enables data collection in high dose gradient with more PinPoint ICs positioned in the distal fall-off of the SOBP as well as some IC located outside of the target. Our results demonstrate a good performance of both algorithms within the target for the oblique beam without the range shifter. The combination of the oblique beam with the range shifter, however, presented the worst-case scenario for algorithms for all tested geometries. In case of RS-PB both mean dose differences within ($3.3 \pm 1.8\%$) and outside the target ($3.6 \pm 3.9\%$) were above the uncertainties of 3% reported during the TPS commissioning stage [13].

Fig. 6. Histograms of the dose differences compared to the measurements for MC and PB_{MC} for oblique beam incidence; a) Bone and soft tissue inserts: SOBPI without range shifter (air gap 36.1 cm); b) Bone and soft tissue inserts: SOBPI with range shifter (air gap 36.1 cm). Red vertical lines indicate clinically acceptable level (within 3%). MC > 3% and PB_{MC} > 3% are the number of measurement points that exceed the limit of $\pm 3\%$.

Looking at different spot and fluence pattern optimization strategies, the algorithm performance is, within this study, comparable for PB and MC optimizations for the investigated configurations and bone-soft tissue inserts. Higher discrepancies for PB are expected for scenarios with larger density differences, such as “lung regions” or treated sites with air cavities.

The limitation of detectors needs to be reported. Bone and soft tissue cases are characterized by dose deviations in rather small volumes behind the target with regard to the finite size of the IC array. Therefore, detectors with smaller active volumes, such as diamond detectors, would have been beneficial for experimental measurements. However, there is no diamond detector array commercially available at present, and the drawback of point detector experiments is the obviously lengthy experimental procedure.

Throughout the study, the dose predicted by the MC algorithm was in good agreement with the experimental results, with the mean dose differences clearly below 3% inside and outside the target for all tested configurations. As stated above, the PB_{MC} algorithm performed comparably well within the target except for the oblique beam incidence in combination with the range shifter. The differences between the algorithms were more pronounced outside the target, where the pencil beam final dose calculation exhibits higher discrepancies, especially when looking at the maximal dose differences in small regions distal to the interface location of materials with different densities. Discrepancies up to 19% were obtained outside the target in two test configurations: shallow target (bone-lung inserts) and oblique beam incidence with the range shifter (bone-soft tissue inserts). These results can be used for benchmarking and extrapolated to clinical settings in treatment planning procedures. As presented in this study, the performance of dose calculation algorithms depends on the influencing factors such as inserted range shifter, the treated air gap, or oblique beam incidence, particularly outside the treated volume. These factors shall be considered before performing the dose calculations to reduce the clinically relevant

uncertainties in OARs.

5. Conclusion

Our results demonstrate that the final dose calculation with MC is more accurate compared to the PB algorithm for the 12 investigated test configurations with bone-lung and bone-soft tissue inserts. Averaged maximal dose differences of 4% (6%) within the target and 7% (12%) outside the target were obtained for the MC (PB_{MC}) algorithm. The systematic decrease in accuracy was observed for both algorithms for regions behind the target for the configurations with the range shifter, especially with increasing air gap. The largest dosimetric deviations of 19% were obtained for positions outside the target volumes when using the PB for final dose calculation (PB_{MC}). The maximum dose difference of 9% observed for MC outside the target was found for the test case with the target located directly behind the heterogeneities (shallow target) for the bone-lung inserts. These dose calculation uncertainties can translate into unfavorable clinical results. MC based spot optimization and final dose calculation showed significantly better performance over the PB algorithm for the target as well as non-target region and should be therefore preferred in treatment scenarios with heterogeneities.

Acknowledgments

The medical physics team of MedAustron providing in-house developed software for measurements is gratefully acknowledged. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A

a. Bone-lung interface: SOBP1 (air gap =66.1cm) **b. Bone-lung interface: SOBP2 (air gap =66.1cm)** **c. Bone-soft tissue interface: SOBP1 (air gap =66.1cm)** **d. Bone-soft tissue interface: SOBP2 (air gap =66.1cm)**

Fig. A.1. Histograms of the dose differences compared to the measurements for MC and PB_{MC} for different materials insertion and target depths. Red vertical lines indicate clinically acceptable level (within 3%). MC > 3% and PB_{MC} > 3% are the number of measurement points that exceed the limit of $\pm 3\%$.

Fig. A.2. Histograms of the dose differences compared to the measurements for MC and PB_{MC} for range shifter with different air gaps (16.1 cm, 36.1 cm and 66.1 cm). Red vertical lines indicate clinically acceptable level (within 3%). MC > 3% and PB_{MC} > 3% are the number of measurement points that exceed the limit of $\pm 3\%$.

a. Bone-soft tissue interface: SOBP1 with range shifter (air gap = 36.1 cm)

b. Bone-soft tissue interface: SOBP1 without range shifter (air gap = 36.1 cm)

Fig. A.3. Dose differences between PB_{MC} and MC in the TPS for 36.1 cm air gap are shown in the left column. Blue indicates PB_{MC} dose underestimation. Dose difference maps between planned and measured dose are shown for MC in the central column and for PB_{MC} in the right column; a. SOBP1 with bone-soft tissue interface with range shifter; b. SOBP1 with bone-soft tissue interface without range shifter. Negative values for the distance from the central axis were related to bone while positive values were obtained for lung or soft tissue slabs. The red squares indicate the measurement region, whereas the yellow dashed lines represent the distances from tissue interfaces to the target's center. Black vertical lines indicate the border between the target and non-target regions.

Table A1

Percent of mean dose differences $\pm$ SD and maximum dose differences inside and outside the target for all measurement scenarios reported in this study.

Configuration	Target location	Algorithm	Range shifter	Beam angle	Air gap (cm)	Inside the target				Outside the target			
						Mean dose difference $\pm$ SD (%)		Maximum dose difference (%)		Mean dose difference $\pm$ SD (%)		Maximum dose difference (%)	
						MC	PB _{MC} /PB	MC	PB _{MC} /PB	MC	PB _{MC} /PB	MC	PB _{MC} /PB
Bone – lung													
1	SOBP1	MC, PB _{MC}	–	–	66.1	1.1 $\pm$ 0.9	1.6 $\pm$ 1.2	3.5	5.7	0.9 $\pm$ 1.5	2.4 $\pm$ 4.1	5.5	18.8
2	SOBP2	MC, PB _{MC}	–	–	66.1	0.6 $\pm$ 0.4	0.9 $\pm$ 0.7	1.6	4.4	1.2 $\pm$ 1.2	2.4 $\pm$ 2.6	4.1	12.1
Bone – soft tissue													
3	SOBP1	MC, PB _{MC}	–	–	66.1	1.2 $\pm$ 0.9	1.3 $\pm$ 1.1	3.1	5.3	0.6 $\pm$ 1.2	1.1 $\pm$ 2.5	4.9	10.4
4	SOBP1	MC, PB _{MC}	–	–	36.1	1.2 $\pm$ 1.2	1.3 $\pm$ 1.5	6.4	8.4	0.8 $\pm$ 1.6	1.1 $\pm$ 2.3	7.0	13.1
5	SOBP1	PB	–	–	66.1	–	1.3 $\pm$ 1.3	–	5.8	–	0.8 $\pm$ 1.5	–	7.3
6	SOBP2	MC, PB _{MC}	–	–	66.1	0.6 $\pm$ 0.5	1.0 $\pm$ 0.9	2.2	5.6	1.5 $\pm$ 2.0	1.6 $\pm$ 2.1	8.4	8.0
7	SOBP1	MC, PB _{MC}	in	–	16.1	2.0 $\pm$ 1.8	2.1 $\pm$ 1.4	7.5	6.6	0.9 $\pm$ 1.5	0.9 $\pm$ 1.3	6.6	5.3
8	SOBP1	MC, PB _{MC}	in	–	36.1	2.0 $\pm$ 1.2	2.1 $\pm$ 1.9	4.9	7.1	1.5 $\pm$ 2.4	1.9 $\pm$ 3.0	8.2	12.7
9	SOBP1	MC, PB _{MC}	in	–	66.1	2.7 $\pm$ 1.3	2.4 $\pm$ 1.8	6.8	6.9	0.9 $\pm$ 1.4	2.0 $\pm$ 3.4	5.5	13.9
10	SOBP1	PB	in	–	16.1	–	2.2 $\pm$ 1.4	–	6.5	–	0.7 $\pm$ 1.2	–	4.7
11	SOBP1	MC, PB _{MC}	–	30°	36.1	0.8 $\pm$ 0.6	1.2 $\pm$ 1.0	2.1	4.3	1.1 $\pm$ 1.4	1.4 $\pm$ 2.0	8.0	12.5
12	SOBP1	MC, PB _{MC}	in	30°	36.1	2.2 $\pm$ 0.5	3.3 $\pm$ 1.8	3.0	7.1	2.2 $\pm$ 1.8	3.6 $\pm$ 3.9	9.0	19.2

References

- [1] Krämer M, Jäkel O, Haberer T, Kraft G, Schardt D, Weber U. Treatment planning for heavy-ion radiotherapy: physical beam model and dose optimization. *Phys Med Biol* 2000;45:3299–317.
- [2] Ahnesjö A, Aspradakis MM. Dose calculations for external photon beams in radiotherapy. *Phys Med Biol* 1999;44. <https://doi.org/10.1088/0031-9155/44/11/201>.
- [3] Alshaikh J, Doolan PJ, D'Souza D, Holloway SMG, Amos RA, Royle G. Impact of varying planning parameters on proton pencil beam scanning dose distributions in four commercial treatment planning systems. *Med Phys* 2019;46:1150–62. <https://doi.org/10.1002/mp.13382>.
- [4] Maes D, Saini J, Zeng J, Rengan R, Wong T, Bowen SR. Advanced proton beam dosimetry part II: Monte Carlo vs. pencil beam-based planning for lung cancer. *Transl Lung Cancer Res* 2018;7:114–21.
- [5] Maes D, Bowen SR, Fung A, Saini J, Bloch C, Egan A, et al. Dose Comparison between Proton Pencil Beam and Monte Carlo Dose Calculation Algorithm in Lung Cancer Patients. *Int J Radiat Oncol Biol Phys* 2016;99:E694. <https://doi.org/10.1016/j.ijrobp.2017.06.2275>.
- [6] Taylor PA, Kry SF, Followill DS. Pencil beam algorithms are unsuitable for proton dose calculations in lung. *Radiat Oncol Biol Phys* 2017;99:750–6. <https://doi.org/10.1016/j.ijrobp.2017.06.003>.
- [7] Sorriaux J, Testa M, Paganetti H, Orban de Xivry J, Lee JA, Traneus E, et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous

- media. *Phys Medica* 2017;38:10–5. <https://doi.org/10.1016/j.ejmp.2017.04.020>.
- [8] Lin L, Huang S, Kang M, Hiltunen P, Vanderstraeten R, Lindberg J, et al. A benchmarking method to evaluate the accuracy of a commercial proton monte carlo pencil beam scanning treatment planning system. *J Appl Clin Med Phys* 2017;18:44–9. <https://doi.org/10.1002/acm2.12043>.
- [9] Saini J, Maes D, Egan A, Bowen SR, St James S, Janson M, et al. Dosimetric evaluation of a commercial proton spot scanning Monte-Carlo dose algorithm: comparisons against measurements and simulations. *Phys Med Biol* 2017;62:7659–81. <https://doi.org/10.1088/1361-6560/aa82a5>.
- [10] Seuntjens J, Olivares M, Evans M, Podgorsak E. Absorbed dose to water reference dosimetry using solid phantoms in the context of absorbed-dose protocols. *Med Phys* 2005;32:2945–53. <https://doi.org/10.1118/1.2012807>.
- [11] Lourenço A, Shipley D, Wellock N, Thomas R, Bouchard H, Kacperek A, et al. Evaluation of the water-equivalence of plastic materials in low- and high-energy clinical proton beams. *Phys Med Biol* 2017;62:3883–901. <https://doi.org/10.1088/1361-6560/aa67d4>.
- [12] Al-Sulaiti L, Shipley D, Thomas R, Owen P, Kacperek A, Regan PH, et al. Water equivalence of some plastic-water phantom materials for clinical proton beam dosimetry. *Appl Radiat Isot* 2012;70:1052–7. <https://doi.org/10.1016/j.apradiso.2012.02.002>.
- [13] Carlino A, Böhlen T, Vatsnitsky S, Grevillot L, Osorio J, Dreindl R, et al. Commissioning of pencil beam and Monte Carlo dose engines for non-isocentric treatments in scanned proton beam therapy. *Phys Med Biol* 2019;64:17NT01. <https://doi.org/10.1088/1361-6560/ab3557>.
- [14] RaySearch Americas Inc. Proton monte-carlo dose calculation in raystation; 2017.
- [15] RaySearch Laboratories AB. Raystation dose calculation algorithms; 2017.
- [16] Saini J, Traneus E, Maes D, Regmi R, Bowen SR, Bloch C, et al. Advanced Proton Beam Dosimetry Part I: review and performance evaluation of dose calculation algorithms. *Transl Lung Cancer Res* 2018;7:171–9. <https://doi.org/10.21037/tlcr.2018.04.05>.
- [17] Righetto R, Clemens LP, Lorentini S, Fracchiolla F, Algranati C, Tommasino F, et al. Accurate proton treatment planning for pencil beam crossing titanium fixation implants. *Phys Medica* 2020. <https://doi.org/10.1016/j.ejmp.2020.01.003>.
- [18] Durante M, Paganetti H. Nuclear physics in particle therapy: A review. *Reports Prog Phys* 2016. <https://doi.org/10.1088/0034-4885/79/9/096702>.
- [19] RaySearch Laboratories AB. Raystation 7 Manual Reference; 2018.
- [20] Stock M, Georg D, Ableitinger A, Zechner A, Utz A, Mumot M, et al. The technological basis for adaptive ion beam therapy at MedAustron: Status and outlook. *Z Med Phys* 2018;28:196–210. <https://doi.org/10.1016/j.zemedi.2017.09.007>.
- [21] Carlino A, Stock M, Zagler N, Marralle M, Osorio J, Vatsnitsky S, et al. Characterization of PTW-31015 PinPoint ionization chambers in photon and proton beams. *Phys Med Biol* 2018. <https://doi.org/10.1088/1361-6560/aadd39>.
- [22] Grevillot L, Stock M, Palmans H, Osorio Moreno J, Letellier V, Dreindl R, et al. Implementation of dosimetry equipment and phantoms at the MedAustron light ion beam therapy facility. *Med Phys* 2018. <https://doi.org/10.1002/mp.12653>.
- [23] IAEA. IAEA-TECDOC-1455 “Implementation of the International Code of Practice on Dosimetry in Radiotherapy (TRS 398): Review of testing” results. Trs 398 2005:101.
- [24] Palmans H. Perturbation factors for cylindrical ionization chambers in proton beams. Part I: Corrections for gradients. *Phys Med Biol* 2006. <https://doi.org/10.1088/0031-9155/51/14/014>.
- [25] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys* 2018;45:e53–83. <https://doi.org/10.1002/mp.12810>.
- [26] Gottschalk B, Koehler AM, Schneider RJ, Sisterson JM, Wagner MS. Multiple Coulomb scattering of 160 MeV protons. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms* 1993;74:467–90. [https://doi.org/10.1016/0168-583X\(93\)95944-Z](https://doi.org/10.1016/0168-583X(93)95944-Z).
- [27] Langner UW, Mundis M, Strauss D, Zhu M, Mossahebi S. A comparison of two pencil beam scanning treatment planning systems for proton therapy. *J Appl Clin Med Phys* 2018. <https://doi.org/10.1002/acm2.12235>.
- [28] Shen J, Liu W, Anand A, Stoker JB, Ding X, Fatyga M, et al. Impact of range shifter material on proton pencil beam spot characteristics. *Med Phys* 2015. <https://doi.org/10.1118/1.4908208>.