Received:
10 March 2020

Revised:
27 May 2020

Accepted:
02 July 2020

<https://doi.org/10.1259/bjr.20200228>

Cite this article as:

Aitkenhead AH, Sitch P, Richardson JC, Winterhalter C, Patel I, Mackay RI. Automated Monte-Carlo re-calculation of proton therapy plans using GEANT4/GATE: implementation and comparison to plan-specific quality assurance measurements. *Br J Radiol* 2020; **93**: 20200228.

FULL PAPER

Automated Monte-Carlo re-calculation of proton therapy plans using GEANT4/GATE: implementation and comparison to plan-specific quality assurance measurements

^{1,2}ADAM H AITKENHEAD, PhD, ¹PETER SITCH, PhD, ^{1,2}JENNY C RICHARDSON, PhD, ^{1,2}CARLA WINTERHALTER, PhD, ¹IMRAN PATEL, MSc and ^{1,2}RANALD I MACKAY, PhD

¹Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, UK

²Division of Cancer Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester, UK

Address correspondence to: Dr Adam H Aitkenhead

E-mail: adam.aitkenhead@christie.nhs.uk; adamaitkenhead@hotmail.com

Objectives: Software re-calculation of proton pencil beam scanning plans provides a method of verifying treatment planning system (TPS) dose calculations prior to patient treatment. This study describes the implementation of AutoMC, a GEANT4 v10.3.3/GATE v8.1 (GATE-RTion v1.0)-based Monte-Carlo (MC) system for automated plan re-calculation, and presents verification results for 153 patients (730 fields) planned within year one of the proton service at The Christie NHS Foundation Trust.

Methods: A MC beam model for a Varian ProBeam delivery system with four range-shifter options (none, 2 cm, 3 cm, 5 cm) was derived from beam commissioning data and implemented in AutoMC. MC and TPS (Varian Eclipse v13.7) calculations of 730 fields in solid-water were compared to physical plan-specific quality assurance (PSQA) measurements acquired using a PTW Octavius 1500XDR array and PTW 31021 Semiflex 3D ion chamber.

Results: TPS and MC showed good agreement with array measurements, evaluated using γ analyses at 3%, 3 mm with a 10% lower dose threshold: >94% of fields calculated by the TPS and >99% of fields calculated by MC had $\gamma \leq 1$ for >95% of measurement points within the plane. TPS and MC also showed good agreement with chamber measurements of absolute dose, with systematic differences of <1.5% for all range-shifter options.

Conclusions: Reliable independent verification of the TPS dose calculation is a valuable complement to physical PSQA and may facilitate reduction of the physical PSQA workload alongside a thorough delivery system quality assurance programme.

Advances in knowledge: A GATE/GEANT4-based MC system is thoroughly validated against an extensive physical PSQA dataset for 730 clinical fields, showing that clinical implementation of MC for PSQA is feasible.

INTRODUCTION

Background

Pre-treatment verification of radiotherapy treatment plans, often called plan- (or patient-) specific quality assurance (PSQA), is a legal requirement in the UK. The “Towards Safer Radiotherapy” report¹ makes recommendations on what plan verification should aim to check and how verification fits into the wider context of delivering safe radiotherapy. However, it makes no specific recommendations on the details of how a plan verification should be performed, leaving that to be defined by each radiotherapy centre’s own protocols.

There are two main approaches to plan verification, each of which validates different aspects of the planning and delivery process:

Physical verification: The plan is delivered to a physical phantom containing suitable dosimeters, and the physical measurements are compared to the predicted dose calculated by the treatment planning system (TPS) (e.g.,^{2,3}). This approach validates the TPS calculation within the phantom geometry, and the delivery of the plan. However, it can only validate delivery at the limited set of points where detectors are located, and it does not validate the TPS calculation within the patient geometry. Also, there are inherent uncertainties in the measurement of absolute dose using non-water phantoms or ionisation chambers⁴ which may limit the accuracy of physical dose measurements.

Software verification: The treatment plan is re-calculated in an independent system for comparison against the original

TPS calculation.^{5,6} This approach validates the TPS calculation throughout the 3D image of either the phantom or the patient.

Depending on the workflow, software verification may or may not validate the plan transfer from the TPS to the delivery system. One process that a software verification alone cannot check is the delivery of a plan. However, a centre could decide that its commissioning of the technique alongside a thorough machine quality assurance (QA) programme is a sufficient check of delivery.

Many of the pros and cons of physical and software verification for photon IMRT have been discussed.⁷ Similar advantages and disadvantages exist for radiotherapy delivered by proton pencil beam scanning (PBS), with the additional observation that physical plan QA is often substantially more time-consuming since measurements are typically performed at more than one depth. In addition, physical PSQA may not detect all clinically relevant errors.⁸

At The Christie NHS Foundation Trust, a proton therapy facility was opened in December 2018. For PSQA, physical verification is used to validate all plans prior to treatment. A software verification system has also been implemented to enable all plans to be re-calculated using a Monte-Carlo (MC) dose calculation, independent of the TPS. As yet, clinical decisions are based only on the physical verification, but all plans are checked using the software system to form a baseline dataset with a view to reducing the physical verification workload in future and placing more emphasis on software verification.

Monte-Carlo for proton PBS

TPSs for proton PBS typically provide some form of MC dose calculation.^{9,10} However, software verification must be performed within a system that is independent of the TPS. Such systems are commonly developed or implemented in-house, based on widely available general purpose MC environments such as GEANT4,¹¹ MCNPX¹² and FLUKA.¹³ For GEANT4, several wrappers are available which provide the user with accessible methods of configuring and controlling simulations without the need to write or compile code. Examples of these are GATE,¹⁴ TOPAS¹⁵ and GAMOS.¹⁶

Various groups have previously reported implementations of MC dose calculations for proton PBS which have been demonstrated on a limited number of clinical plans, such as^{17–20} and.²¹

Study aims

This study describes the MC system implemented at The Christie for re-calculation of proton PBS plans. Data are presented for the 153 patients treated within the first year of the clinical service, covering the full range of clinical indications eligible within the NHS proton service, as detailed by NHS policies for paediatric, teenage and young adults²² and adults.²³ All physical PSQA data for these clinical plans are compared to TPS and MC calculations, with the aims of validating the MC system's suitability as an independent check of the TPS calculation and facilitating a reduction of the physical PSQA workload.

METHODS

Overview of proton therapy at The Christie

The proton therapy facility at The Christie is a Varian (Palo Alto, CA, USA) ProBeam system with 3 matched 360° gantries. The deliverable energy range is 70–245 MeV, and three range-shifter options have been commissioned: 2, 3 and 5 cm physical thickness of Lexan, corresponding to 2.3, 3.4, 5.7 cm water-equivalent thickness. The ProBeam system features a movable nozzle, allowing the range-shifter to be moved towards/away from isocentre. This allows the air-gap to the patient surface to be controlled, minimising the lateral spread of the beam due to scattering in the range-shifter. Alternatively, fields may be delivered with no range-shifter.

The TPS used for all plans was Varian Eclipse v13.7, with Varian's proton-convolution-superposition (PCS) v13.7.16 analytical dose calculation algorithm.²⁴

Within Eclipse/ProBeam, if a range-shifter is applied for any field, it is in place for all layers within that field. All clinical plans used a single range-shifter for all fields.

Patient cohort details

Clinical plans were created for 153 patients, comprised of the following indications:

- Chordoma (21) and chondrosarcoma (14) in the base of skull and spine;
- Rhabdomyosarcoma (21) and Ewings (16) in the body, head and neck;
- Other sarcoma (10);
- Ependymoma (16), craniopharyngioma (12), and assorted glioma and other central nervous system tumours (25);
- Head and neck tumours (11);
- Craniospinal irradiations, predominantly medulloblastoma (7).

78 of the patients were treated using multiphase plans, and in terms of clinical workflow each phase was considered as a separate plan. Re-plans were needed on 21 occasions (often for multiple phases) due to patient anatomy changes. As a consequence, 250 plans were generated in total, consisting of 730 individual fields.

Dosimetric plan-specific quality assurance at The Christie

Physical verification of each plan was performed prior to delivery of the first fraction as follows:

- (1) Within the TPS, the plan was copied onto a CT image of a PTW (Freiburg, Germany) RW3 solid-water block phantom. Layer energies, spot positions and weights were identical to the original plan, but all gantry angles were set to 0° (*i.e.*, directing the beam vertically downwards).
- (2) The plan was delivered to the phantom, and measurements were made using the following dosimeters:
 - PTW Octavius 1500XDR array, positioned at a defined depth within the phantom and aligned using the in-room lasers. The chambers within the array have dimensions

4.4×4.4×3.0 mm and are arranged in a 5 mm pitch chequer-board grid, covering an area of 27×27 cm. Comparisons of the measured relative dose distribution against the TPS dose grid (exported in DICOM format) were done using a γ analysis.²⁵ Test criteria were 3% (relative to the local dose), 3 mm, with a 10% lower dose threshold. If >5% of points had $\gamma > 1$, the field was reviewed by a physicist.

◦ PTW 31021 Semiflex 3D ion chamber, positioned at defined locations within the solid-water phantom. The measured absolute dose was compared against the point dose reported by the TPS. Any dose-differences > 5% were reviewed by a physicist.

For the first 3 months (15 patients), array and chamber PSQA were repeated in multiple treatment rooms to check room-to-room consistency, and at three depths for each field. Depths were chosen to sample the shallow, central and deep parts of the high-dose region. After 3 months, PSQA was performed only in one treatment room. In months 4–9 (85 patients), array and chamber PSQA were performed at two depths for each field, sampling the shallow and deep parts of the high-dose region. From month 10 onwards (53 patients), array and chamber PSQA were performed only at the central depth of the high-dose region.

Overview of AutoMC

An in-house system (called AutoMC) for automated MC re-calculation of clinical treatment plans was implemented. AutoMC has the following key features:

- It is based mainly on free software:
 - Runs on a cluster running a GNU/Linux operating system.
 - Runs within an Octave²⁶ environment.
 - Uses GATE¹⁴ v8.1/GEANT4¹¹ v10.3.3 as the MC calculation engine, matching the versions used in GATE-RTION v1.0. GATE-RTION is a stable and long-term GATE release, provided to facilitate clinical use in the field of Light Ion Beam Therapy.²⁷
- It uses a modular approach:
 - Beam-models and CT calibrations are modular and interchangeable.
- It is automated:
 - Designed for routine use by clinical staff without the need for previous MC experience or programming skills.
 - Minimises risk of errors in user input. Plan data are imported in DICOM format, and beam-model/CT calibrations are imported from the relevant modules.
 - Minimises time required to prepare simulations and process results.
 - Includes automated evaluation (using 3D γ analysis) of the MC calculated dose in comparison with the original TPS calculated dose.
- It allows calculation of:
 - Dose-to-medium or dose-to-water, calculated as described by.²⁸
 - Dose-averaged or track-averaged linear energy transfer (LET).

- Dose deposited by specific particles (proton, electron, neutron, etc.)

In terms of hardware, we use a cluster consisting of 1 master CPU and 10 slave CPUs. Each CPU is a quad-core Intel Xeon E3-1240 v6 @ 3.70 GHz with 16 GB RAM. When running a simulation, each core runs a single GATE instance, so 40 GATE instances can run in parallel.

Physics list selection

We followed the recommendation of²⁹ and used the QGSP_BIC physics list. We applied electron cuts of 0.1 mm within the patient CT or phantom, and 5 mm within the world and range-shifter.

CT calibration

GATE natively contains features which facilitate the definition of custom CT calibrations. AutoMC added/modified several features required for clinical use at The Christie:

- Automatic application of CT overrides as defined in the treatment plan. For example, to correct CT artefacts; to simulate air- or fluid-filled cavities (such as sinus, bowel); or to control the CT numbers used for implants (such as carbon-fibre or titanium). In AutoMC, this was implemented as follows:
 - Details of CT overrides are contained within the DICOM RTSTRUCT file written by Eclipse not in terms of Hounsfield units (HU), but in terms of relative electron density.
 - Both Eclipse and AutoMC contain identical HU-to-relative electron density lookup tables as part of their CT calibrations. Overrides are converted to HU value before being applied to the image.
- Automatic cropping of the CT image to minimise the quantity of data used by the simulation.
- Control over the binning of the CT image and the number of materials used in the simulation. Rather than binning the CT image by physical density (as in the GATE implementation), AutoMC keeps the image in the original HU space throughout.

The CT calibration was created following the stoichiometric approach.³⁰ Generation of the CT calibrations for the TPS and MC was performed using the same scanner parametrisation, set of reference tissues³¹ and set of elemental ionisation potentials. This ensured that equivalent CT calibrations were implemented in the TPS and in AutoMC, although the format of each differed: HU to physical density and relative stopping power being required for the analytical dose calculation in Eclipse, and HU to physical density and elemental composition for the MC dose calculation.

Elemental ionisation potentials were taken from,³² or from³³ if not defined in.³² The Bragg additivity rule was used to compute ionisation potentials for all composite materials and tissues, including water. This led to an I-value of 75.3 eV for water, based on I-values of 19.2 eV and 106.0 eV for hydrogen and oxygen, respectively.

Commissioning measurement data

Table 1 summarises the measurement data obtained during clinical commissioning of the facility that was used for generation

Table 1. Details of the commissioning measurement data used for generation of the beam model in AutoMC.

Parameter	Integrated Depth Dose (IDD) curves	Absolute dose	Spot-size
Range-shifter	0 cm	0 cm	0 cm
Energy	70–245 MeV in 10 MeV steps	70–245 MeV in 10 MeV steps	70–245 MeV in 10 MeV steps
Medium	Water-tank	Water-tank	Air
Field shape	Single spot on central axis	10×10 cm array of spots	Single spot on central axis
Detector	PTW 34070 Bragg Peak chamber (Diameter 84 mm)	PTW 34001 Roos chamber	IBA Lynx
Location	Along central axis	At 2 cm deep on the central axis	Plane perpendicular to central axis at a range of positions (-40, -20, -10, 0, 10, 20, 40 cm) up-/down- stream of isocentre.

of the MC beam model. Absolute integrated depth dose (IDD) curves were measured as follows and used to derive the energy lookup table component of the MC beam model.

1. Delivery of a single spot travelling along the central axis into a water tank, with relative dose measured as a function of depth using a Bragg Peak chamber. (*i.e.*, small field; large chamber.)
2. Delivery of a uniformly weighted 10×10 cm array of spots to a water-tank, with absolute dose measured at a depth of 2 cm using a PTW 34001 Roos chamber (*i.e.*, large field; small chamber.)
3. Re-normalisation of the relative IDDs (from step 1) using the absolute dose from step 2. This provides absolute IDDs where the dose per monitor unit (MU) is known.
4. At this stage, the IDDs were scaled by a factor of 1.1 to take the relative biological effect (RBE) of protons into account. As a consequence, a fixed RBE is built into both the TPS and MC beam models.

Beam modelling

The beam model was derived from the same measurement data used to commission the TPS. It included modelling of divergent beam optics and divergent spot steering.

The beam model was based on the definition of the phase-space within the nozzle, and therefore beam transit through the gantry itself was not modelled.¹⁸ The location of the phase-space plane was matched to the location of the primary ionisation chamber, ensuring that the MC and the physical delivery system quantified the number of delivered protons at the same location to facilitate calculation of absolute dose.

The beam model consists of three parts (Figure 1):

- (1) Energy lookup table. This defines the mean and standard deviation energy and the number of protons per MU to be used in GATE for each nominal planned energy.
- (2) Beam optics table. This defines the initial angular spatial distribution (*i.e.*, the phase-space) of the beam at the primary ionisation chamber, as a function of energy.

Figure 1. Structure of the AutoMC system. All AutoMC functions run within Octave, an environment amenable to processing and analysis of 3D datasets. These functions (1) prepare the inputs to the GATE simulations, (2) co-ordinate the submission of the GATE jobs to the cluster scheduler and (3) on completion of the GATE simulations evaluate the simulation output and record the results in a database. Steps 1–3 are automated, such that after launching AutoMC no further user interaction is involved.

Figure 2. (a) Diagram showing the Bragg peak features used for tuning of the energy lookup table. (1) The $R_{80\%}$ point, defining the end of range. (2) The distance between the proximal and distal 80% points on the IDD. (3) The ratio between the maximum dose at the Bragg peak and the dose at the 2cm reference point. (b) Residual differences between the measured and the tuned MC IDD, as a function of distance from R_{80} . Results are shown for energies of 70 (red), 80, 90, ... 230, 240, 245 (blue) MeV.

(3) Range-shifter table. This defines the material composition, physical density and position relative to the nozzle for each range-shifter.

The commissioning IDD was used for tuning of the energy lookup table. For each IDD, GATE simulations of a single spot in water were performed to tune the energy required to match the IDD. Dose was scored only within the effective diameter of the Bragg peak chamber (84 mm). On each iteration, the energy (mean E_{mean} , and standard deviation E_{std}) within GATE was adjusted to match the simulated IDD to the measured IDD. Three metrics were assessed on each iteration (Figure 2a):

- (1) The $R_{80\%}$ point, defining the end of range. This allows E_{mean} to be tuned.
- (2) The distance between the proximal and distal 80% points on the IDD. This allows E_{std} to be tuned.
- (3) The ratio between the maximum dose at the Bragg peak and the dose at the 2 cm reference point. This allows E_{std} to be tuned.

Tuning was stopped for each energy when $R_{80\%}$ was within 0.1 mm of the measured value, and when E_{std} converged to the point where metrics 2 and 3 conflicted on the direction of tuning. The number of protons-per-MU was calculated by normalising the area under the curve of the simulated IDD to that of the re-normalised measured IDD. Figure 2b shows the residuals for each tuned energy, and Figure 3 shows the parameters in the energy lookup table as a function of nominal beam energy.

Beam optics were derived from spot size measurements in air for non-range-shifted spots. For each energy, the spot size as a function of position relative to isocentre was parametrised using a hyperbolic fit. These fit parameters formed another lookup table, allowing spot-size, emittance and divergence to be defined for each nominal beam energy. Figure 4 compares measured and simulated spot sizes in air at 150 MeV.

The range-shifter file defined the material (Lexan, physical density 1.21 g/cm^3), thickness (2, 3, or 5 cm) and position relative to the nozzle of each range-shifter according to the specification provided by Varian. No tuning of the range-shifter definition was performed. Figure 4 compares measured and simulated spot sizes in air at 150 MeV for each range-shifter option. The decision to implement the range-shifter as a material object within the simulation was made to ensure that interactions within the range-shifter were taken into account,³⁴ aiding calculation of absolute dose.

Validation of the beam model in terms of absolute dose

Box field plans for a variety of spread-out Bragg peak (SOBP) ranges and lengths were generated, sampling the full space of the beam model. Each field was delivered to a water tank and the dose at the centre of the SOBP was measured using a PTW Roos chamber. The measurements were used for two purposes:

- (1) To validate the MC beam model.

Figure 3. : Summary of the energy tuning results for the Varian ProBeam system at The Christie. (a) Difference between the nominal energy (i.e., as reported by the ProBeam control system) and the mean energy in the AutoMC beam-model. (b) The standard deviation energy in the AutoMC beam-model. (c) The number of protons delivered per MU in the AutoMC beam-model.

Figure 4. Spot size (σ) in air vs distance from isocentre for beam energy = 150 MeV and range-shifter thickness = 0, 2, 3, 5 cm. The points show measured spot widths orthogonal to the beam direction, and the lines show MC-simulated spot widths.

- (2) To perform the final tuning stage for the TPS beam model: A set of correction factors derived from the box field measurements were implemented to refine the TPS dose calculation as a function of SOBP range and length.

Plan re-calculation

The process for plan re-calculation is illustrated in Figure 1, which indicates the input files required at each stage of the process flow. In the final stage of the process, AutoMC stores the result of the MC vs TPS evaluation in a simple database for review by clinical staff.

We calculate plans on a $2 \times 2 \times 2$ mm dose grid to an uncertainty of 0.8% within the 90% isodose contour. Instead of parallelising by running identical simulations with independent seeds, we split each field by energy layer. This provides a number of advantages: it reduces the number of sources within each simulation, which reduces the initialisation time for each GATE instance. It allows each simulation to use the same seed. It is suited to simulations of fields where each layer uses independent beam modifiers (such as collimation). The disadvantage is that each layer takes a different length of time to simulate. As discussed later,

minimising the simulation time was not prioritised in this study. Following completion of a simulation, results for all layers are combined and an uncertainty grid is computed using the method described by.³⁵

RESULTS

Validation of the beam model in terms of absolute dose

Results for the box fields delivered to water are shown in Figure 5. MC calculated doses were systematically higher than measured doses by 1.0%, while the TPS was systematically 0.6% lower (pre-correction) and 0.1% higher (post-correction) than measurement.

All subsequent TPS dose calculations used the post-correction beam-model.

Comparison to TPS and PSQA data for a typical plan

Results are presented for a typical complex case treated at The Christie. This was a paediatric medulloblastoma patient planned

Figure 5. Validation of the beam-model in terms of absolute dose for box fields delivered to a water tank. (a) Geometry of each field: A spread-out-Bragg-peak (SOBP) of defined range (R) and length (L), with lateral dimensions of 10×10 cm. Measurements were made using a PTW Roos chamber at a depth of $R-L/2$. The prescribed dose for each field was 5 Gy. No range-shifter was used. (b) The relative difference between the calculated (TPS pre-correction; TPS post-correction; AutoMC) dose and the measured dose for different 27 box fields. The median and standard deviation of all fields is also shown.

Figure 6. Results for one field (gantry angle = 180°) of a four-field whole central nervous system plan. Row 1: Dose calculation in patient-CT: (a) TPS; (b) AutoMC; (c) γ (2%, 2mm, local, 10% threshold, Reference = MC, Evaluation = TPS). Row 2: Dose calculation in solid-water: (d) TPS; (e) AutoMC; (f) γ (2%, 2mm, local, 10% threshold, Reference = MC, Evaluation = TPS). Row 3: Plan-specific QA: (g) Array measurement; (h) γ (2%, 2mm, local, 10% threshold, Reference = Array, Evaluation = TPS); (i) γ (2%, 2mm, local, 10% threshold, Reference = Array, Evaluation = AutoMC). For the γ analyses: Green: $\gamma \leq 1$; Red: $\gamma > 1$ and Reference > Evaluation; Blue: $\gamma > 1$ and Reference < Evaluation. (Note that the definition of Reference and Evaluation datasets follows that of ³⁶)

for irradiation of the whole brain and spine. The treatment was split into two phases, with each phase effectively being considered as a separate plan. In Phase 1, 23.4Gy in 13 fractions was prescribed to the target volume, while in Phase 2, 30.6Gy in 17 fractions was prescribed to a boost volume. The plan for Phase 1 is presented here, which consisted of four fields which were optimised using a multifield optimisation approach.

The plan was re-calculated in AutoMC on both the patient CT and the solid-water phantom. Figure 6 illustrates one field of the plan calculated in the TPS and in AutoMC in both patient CT and solid-water along with a physical PSQA measurement taken in solid-water using the 2D array. Comparison of TPS to MC, TPS to array, and MC to array is shown in terms of γ analyses at 2%, 2 mm.

Comparison to TPS for 153 patients

Figure 7 summarises the comparison between MC and TPS for all 153 patients/250 plans/730 fields planned between December

2018 and December 2019. We separate the evaluation of dose distribution *shape* from that of *absolute dose*. For each field, shape was assessed by performing a γ analysis on the normalised MC and TPS dose distributions, while absolute dose was assessed by evaluation of the normalisation factors.

Results for γ analyses at 3%, 3 mm are shown in Figure 7a–b for calculations in solid-water and patient CT, respectively. All voxels within the patient and within the 10% isodose surface were included in the analyses. Agreement in solid-water was excellent for fields without range-shifter, with all 730 fields having $\gamma \leq 1$ for >95% of the comparison volume. While agreement was less good for fields with a range-shifter, all had $\gamma \leq 1$ for >90% of the comparison volume. Agreement was also less good for simulations in patient CT due to the increased complexity of the medium, where it is to be expected that MC and analytical dose calculations will differ. Similar observations can be made regarding the results for γ analyses at 2%, 2 mm (Figure 7c–d), where less good agreement was also seen with a range-shifter and in patient CT.

Figure 7. Comparison of MC to TPS dose calculations in (a,c,e,g) solid-water and (b,d,f,h) patient CT for $N = 730$ fields (belonging to 153 patients / 250 plans). Results for each range-shifter (0, 2, 3, 5 cm) are shown as separate series. (a,b): Histograms summarising γ analysis results (3%, 3 mm, local, 10% threshold, Reference = MC, Evaluation = TPS). The legend reports the proportion (%) of fields where <90%, <95% and <98% of the analysed voxels had $\gamma \leq 1$. (c,d): The same as (a,b), but for γ criteria: 2%, 2 mm. (e,f): Median absolute dose-difference between MC and TPS. Dose-difference $>0 \Rightarrow$ TPS $>$ MC. (g,h): MC simulation times (CPU.hours) per field. The CPU model was Intel Xeon E3-1240 v6 @ 3.70 GHz. Simulations were parallelised on a 40-core cluster, so cluster time per field was substantially less than the total CPU hours.

Figure 7e-f summarises the absolute dose-differences between MC and TPS, derived by evaluation of the normalisation factors obtained as part of the γ analyses. To normalise the dose distributions, all voxels within the 50% isodose surface in the MC dose distribution were taken, and the corresponding TPS dose was noted for each. The ratio of MC/TPS dose yields a normalisation factor for each voxel. The median of this set of factors was taken as the overall normalisation factor for the field, from which the dose-difference for the field as a whole can be calculated. Figure 7e-f summarises the resulting dose-differences observed between MC and TPS in solid-water and patient CT, showing that there is a dependence on range-shifter thickness. For fields without a range-shifter, MC tends to be $\sim 2.3\%$ (in solid-water) or

$\sim 1.8\%$ (in patient CT) colder than the TPS. For fields with a 5 cm range-shifter, MC tends to be $\sim 0.3\%$ (in solid-water) or $\sim 0.6\%$ (in patient CT) hotter than the TPS.

Figure 7g-h shows histograms of the total simulation time for the 730 fields calculated in both solid-water and patient CT. Simulations in patient CT take notably longer due to the increased complexity of the medium. Cluster time per simulation is dependent primarily on the simulation time of the largest layer in the field. For example, the longest total computation time for any field was 367.6 CPU h (for a field from a Ewing sarcoma plan), while the longest computation time for an individual layer within that field was 18.9 h.

Figure 8. Comparison of (a,c,e) MC and (b,d,f) TPS to PSQA measurements for $N = 1486$ array measurements and $N = 1290$ chamber measurements (for 153 patients/250 plans/730 fields). Results for each range-shifter (0, 2, 3, 5 cm) are shown as separate series. (a,b): Histograms summarising γ analysis results (3%, 3 mm, local, 10% threshold, Reference = MC or TPS in solid-water, Evaluation = 2D array). The legend reports the proportion (%) of array planes where $<90\%$, $<95\%$ and $<98\%$ of the points within the plane had $\gamma \leq 1$. (c,d): The same as (a,b), but for γ criteria: 2%, 2 mm. (e,f): Absolute dose-difference between ionisation chamber and MC or TPS.

Comparison to PSQA data for 153 patients

Figure 8 summarises the PSQA results for all 153 patients/250 plans/730 fields planned between December 2018 and December 2019, consisting of 1486 array measurements and 1290 ionisation chamber measurements. Figure 8a–d summarises results of γ analyses comparing both the TPS and MC to the array measurements at 3%, 3 mm and at 2%, 2 mm, showing that $>94\%$ of fields calculated by the TPS and $>99\%$ of fields calculated by MC had $\gamma \leq 1$ for $>95\%$ of measurement points within the plane. Figure 8e–f summarises the dose-differences observed between the chamber and TPS or MC, showing good agreement between chamber measurements and both TPS and MC, with systematic differences of $<1.5\%$ for all range-shifter options.

DISCUSSION

In Figure 2b, the residuals of the tuned IDDs contain non-random structure and a dependence on depth. The same structure appears for all energies, indicating that it arises not from within the measurement data (which is independent for each energy), but from within the MC simulation. Use of a different physics list during MC tuning leads to a change in the structure of the residuals, as well as affecting other features of the simulation, such as computation time. For the present application, the

choice of QGSP_BIC provides a sufficiently good match to the commissioning data, but a more optimal choice of physics list may provide improved residuals, leading to greater accuracy in calculations of absolute dose for clinical plans.

Figure 5 showed that MC dose calculations for box fields in water were systematically 1.0% hotter than PTW Roos chamber measurements at the centre of the SOBP. We note that the box field residuals (Figure 5b) were consistent with the residuals for the individual IDDs (Figure 2b), in that they can effectively be considered as the weighted sum of the residuals for the component energy layers. It is evident that the residuals show a trend with L: E.g. for the $R = 35$ cm SOBP, when L is small (3 cm) the SOBP error is dominated by the positive residual at the end of range of the component IDD. As L increases, the SOBP error becomes a mix of both positive and negative components, which brings the overall SOBP error closer towards zero. We also note that the residuals for fields in water are unaffected by the choice of I-value for water, as long as both the tuning and the validation simulations use the same I-value.

We did not use the PTW Octavius 1500XDR array for verification of absolute dose because the array read-out process influences

the measured dose. On each read-out cycle, the array discards the integrated charge below a preset threshold to reduce noise. However for PBS, the component of the total dose that is delivered by low-weighted spots, by the tails of spots, or by the nuclear halo is non-negligible and is dependent on spot-size, range-shifter thickness and depth of the array in the medium. To use the array for absolute dose measurement would require either modification of the readout process or a method of correcting for the discarded charge.

In the histograms of the residual differences between MC or TPS and PSQA chamber measurements (Figure 8e–f), the standard deviation of each histogram is between 1.7 and 2.6%. For comparison, the daily output for each of the three treatment rooms had a variability of 0.5% (1 standard deviation) measured using a Sun Nuclear (Melbourne, FL, USA) QA3 device. Repeating chamber PSQA measurements for the same field with fresh set-up exhibited a standard deviation of approximately 2%, suggesting that the variability is largely due to residual set-up errors. Currently, the chamber is positioned using the in-room lasers for alignment, and the lasers have a positional tolerance of ± 1.5 mm. Positional accuracy could be improved by using the in-room kV imaging system for chamber set-up, at a cost of increasing the time required for the PSQA check. For clinical purposes, this is hard to justify, as the current method provides sufficient accuracy for the purpose of checking plan safety.

Because a large number of PSQA measurements is accumulated over time, systematic differences between MC, TPS and measurement can be evaluated even when there is a large degree of variability, as was observed in the chamber data. The median dose-differences reported in Figure 8e–f demonstrate that both MC and TPS are systematically within 1.5% of the measured absolute dose for all range-shifter options, which is clinically acceptable. The results also show that in terms of absolute dose accuracy, the TPS had a dependence on range-shifter thickness, with non-range-shifted plans being systematically 1.1% hot while 5 cm range-shifted plans were systematically 0.6% cold. In contrast, MC absolute dose accuracy was almost independent of range-shifter, being systematically 1.4–0.9% cold across the four range-shifter options.

It is of note that the MC calculations of box fields in water were systematically 1.0% *hotter* than measurement (Figure 5b), while the MC calculations of clinical fields in solid-water were systematically 0.9–1.4% *colder* than measurement (Figure 8e). This could be a consequence of:

- (1) The use of different media (water vs solid-water), and the choices of I-value for each medium (and for their component elements).
- (2) Differences in the field geometry. Clinical fields are more modulated than box fields and were measured at a greater variety of positions within the high-dose region. PSQA data may therefore be more sensitive to the residual errors of the IDD tuning process at shallow depths, where the MC was typically colder than measurement (Figure 2b).
- (3) The use of different chambers (PTW Roos vs PTW Semiflex). However, routine chamber QA (regular linearity checks

and inter comparison checks against a common secondary standard) minimises the risk and impact of chamber differences.

It is worth emphasising that our beam-modelling process allows the MC system to calculate absolute dose distributions for patient treatment plans without applying any post-simulation correction factors. A key design choice was the placement of the phase-space at the same geometric location as the monitor unit ionisation chamber within the physical delivery system, so that the beam-line control system and the MC simulations both effectively count the number of delivered protons at the same location. Nevertheless, a residual systematic dose-difference (of 0.9–1.4%) was observed when comparing MC calculated dose to PSQA chamber measurements (Figure 8e) for 730 clinical fields. This is consistent with the differences between MC and measurement reported by³⁷ (for 104 fields) and³⁸ (for 33 fields).

Simulation times (Figure 7g–h) are adequate for our application, as plans re-calculated overnight are ready for review the next working day. We also use the system for other purposes, such as evaluation of LET for clinical plans,³⁹ where having flexible access to GATE/GEANT4 features is a higher priority than simulation speed. We have therefore not sought to minimise simulation time, but there are several straightforward routes to doing so. We currently simulate dose to an uncertainty of 0.8% in order to allow γ analyses to be performed using a 2% dose-difference criterion. If a dose-difference criterion of no lower than 3% was required then an uncertainty of $\sim 1\%$ would be sufficient, reducing the simulation time by a factor of approximately $(1/0.8)^2 \approx 1.6$. Similarly, enlarging the dose grid from $2 \times 2 \times 2$ mm to $2.5 \times 2.5 \times 2.5$ mm would reduce the simulation time by a further factor of $(2.5/2)^3 \approx 2$. Further speed improvements could also be made by selecting optimised range cuts and physics list, or by using a GPU Monte Carlo engine.⁶

CONCLUSIONS

This study presents an overview of the implementation of AUTOMC, a GEANT4/GATE-based system for automated re-calculation of clinical proton PBS plans. MC re-calculation of plans for 153 patients, consisting of a total of 730 fields, agree well with physical PSQA data acquired using a 2D array and an ionisation chamber. Results demonstrate that the system is capable of calculating absolute dose without the use of any post-simulation correction factors, with a systematic difference from chamber data of 0.9–1.4%. Reliable independent verification of the TPS dose calculation in patient CT is a valuable complement to physical PSQA and may facilitate reduction of the physical PSQA workload alongside a thorough beam delivery system QA programme.

ACKNOWLEDGMENT

The authors are grateful to Tony Lomax, Sairos Safai and Francesca Albertini of the Paul Scherrer Institut for advice when implementing the beam-model, and to Yannick Klopfenstein for initial work on the energy tuning process as part of his Masters degree at ETH Zurich.

FUNDING

This work was supported by Cancer Research UK (CRUK) via funding to the Manchester Cancer Research Centre [C147/A25254], and by the Advanced Radiotherapy Technologies

Network (ART-NET) [C309/A21993]. CW was supported by the Science and Technology Facilities Council (STFC) Advanced Radiotherapy Network [ST/N002423/1], and by the Engineering and Physical Sciences Research Council (EPSRC) [EP/R023220/1].

REFERENCES

- British Institute of Radiology, The Institute of Physics and Engineering in Medicine, The National Patient Safety Agency, The Society and College of Radiographers, The Royal College of Radiologists. *Towards Safer Radiotherapy*. 2008. 978-1-905034-25-3.
- Arjomandy B, Sahoo N, Ciangaru G, Zhu R, Song X, Gillin M. Verification of patient-specific dose distributions in proton therapy using a commercial two-dimensional ion chamber array. *Med Phys* 2010; **37**: 5831-7. doi: <https://doi.org/10.1118/1.3505011>
- Trnková P, Bolsi A, Albertini F, Weber DC, Lomax AJ. Factors influencing the performance of patient specific quality assurance for pencil beam scanning IMPT fields. *Med Phys* 2016; **43**: 5998. doi: <https://doi.org/10.1118/1.4964449>
- Lin L, Kang M, Solberg TD, Mertens T, Baumer C, Ainsley CG, et al. Use of a novel two-dimensional ionization chamber array for pencil beam scanning proton therapy beam quality assurance. *J Appl Clin Med Phys* 2015; **16**: 5323. doi: <https://doi.org/10.1120/jacmp.v16i3.5323>
- Meier G, Besson R, Nanz A, Safai S, Lomax AJ. Independent dose calculations for commissioning, quality assurance and dose reconstruction of PBS proton therapy. *Phys Med Biol* 2015; **60**: 2819-36. doi: <https://doi.org/10.1088/0031-9155/60/7/2819>
- Beltran C, Tseung HWC, Augustine KE, Bues M, Mundy DW, Walsh TJ, et al. Clinical implementation of a proton dose verification system utilizing a GPU accelerated Monte Carlo engine. *Int J Part Ther* 2016; **3**: 312-9. doi: <https://doi.org/10.14338/IJPT-16-00011.1>
- Siochi RAC, Molineu A, Orton CG. Point/Counterpoint. patient-specific QA for IMRT should be performed using software rather than hardware methods. *Med Phys* 2013; **40**: 070601. doi: <https://doi.org/10.1118/1.4794929>
- Matter M, Nenoff L, Meier G, Weber DC, Lomax AJ, Albertini F. Alternatives to patient specific verification measurements in proton therapy: a comparative experimental study with intentional errors. *Phys Med Biol* 2018; **63**: 205014. doi: <https://doi.org/10.1088/1361-6560/aae2f4>
- 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. doi: <https://doi.org/10.1002/acm2.12043>
- Schreuder AN, Bridges DS, Rigsby L, Blakey M, Janson M, Hedrick SG, et al. Validation of the RayStation Monte Carlo dose calculation algorithm using realistic animal tissue phantoms. *J Appl Clin Med Phys* 2019; **20**: 160-71. doi: <https://doi.org/10.1002/acm2.12733>
- Allison J, Amako K, Apostolakis J, Arce P, Asai M, Aso T, et al. Recent developments in GEANT4. *Nuclear Instruments and Methods in Physics Research A* 2016; **835**: 186-225.
- Pelowitz DB. MCNPX User's Manual. Version 2.7.0 Los Alamos National Laboratory, LA-CP-11-00438 2011;.
- Battistoni G, Bauer J, Boehlen TT, Cerutti F, Chin MPW, Dos Santos Augusto R, et al. The FLUKA code: an accurate simulation tool for particle therapy. *Front Oncol* 2016; **6**: 116. doi: <https://doi.org/10.3389/fonc.2016.00116>
- Jan S, Benoit D, Becheva E, Carlier T, Cassol F, Descourt P, et al. GATE v6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. *Phys Med Biol* 2011; **56**: 881-901. doi: <https://doi.org/10.1088/0031-9155/56/4/001>
- Pel J, Shin J, Schumann J, Faddegon B, Paganetti H. Topas: an innovative proton Monte Carlo platform for research and clinical applications. *Med Phys* 2012; **39**: 6818-37. doi: <https://doi.org/10.1118/1.4758060>
- Arce P, Lagares JJ, Harkness L, Desorgher L, de Lorenzo G, Abreu Y, et al. GAMOS: an easy and flexible way to use GEANT4. *IEEE Proc. Nuc. Sci. Symp. Conf. Rec* 2011; **2230**-7.
- Sawakuchi GO, Mirkovic D, Perles LA, Sahoo N, Zhu XR, Ciangaru G, et al. An MCNPX Monte Carlo model of a discrete spot scanning proton beam therapy nozzle. *Med Phys* 2010; **37**: 4960-70. doi: <https://doi.org/10.1118/1.3476458>
- Grevillot L, Bertrand D, Dessy F, Freud N, Sarrut D. A Monte Carlo pencil beam scanning model for proton treatment plan simulation using GATE/GEANT4. *Phys Med Biol* 2011; **56**: 5203-19. doi: <https://doi.org/10.1088/0031-9155/56/16/008>
- Verburg JM, Grassberger C, Dowdell S, Schuemann J, Seco J, Paganetti H. Automated Monte Carlo simulation of proton therapy treatment plans. *Technol Cancer Res Treat* 2016; **15**: 1-12. doi: <https://doi.org/10.1177/1533034615614139>
- Fracchiolla F, Lorentini S, Widesott L, Schwarz M. Characterization and validation of a Monte Carlo code for independent dose calculation in proton therapy treatments with pencil beam scanning. *Phys Med Biol* 2015; **60**: 8601-19. doi: <https://doi.org/10.1088/0031-9155/60/21/8601>
- Guterres Marmitt G, Pin A, Ng Wei Siang K, Janssens G, Souris K, Cohilis M, et al. Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy. *Phys Med* 2020; **70**: 49-57. doi: <https://doi.org/10.1016/j.ejmp.2019.12.018>
- NHS England Specialised Services Clinical Reference Group for Radiotherapy. Clinical Commissioning Policy: Proton Beam Therapy for Children, Teenagers and Young Adults in the treatment of malignant and non-malignant tumours. Available at: www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2019/07/Interim-Policy-PBT-for-CTYA-for-malignant-and-non-malignant-tumours.pdf (Accessed 26 May 2020). 2018.
- NHS England Specialised Commissioning Team. Proton beam therapy for cancer in adults. Available at: www.england.nhs.uk/wp-content/uploads/2018/07/Proton-beam-radiotherapy-high-energy-for-skull-base-tumour-treatment--NHS-overseas-programme-Adult.pdf (Accessed 26 May 2020). 2015.
- Sahoo N, Poenisch F, Zhang X, Li Y, Lii M, Li H, et al. 3D treatment planning system-Varian Eclipse for proton therapy planning.

- Med Dosim* 2018; **43**: 184–94. doi: <https://doi.org/10.1016/j.meddos.2018.03.006>
25. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys* 1998; **25**: 656–61. doi: <https://doi.org/10.1118/1.598248>
 26. Eaton J. W, Bateman D, Hauberg S, Wehbring R, Eaton J. W, Bateman D, et al. GNU Octave version 4.4.1 manual: a high-level interactive language for numerical computations. Available at: <https://www.gnu.org/software/octave/doc/v4.4.1/>. 2017.
 27. Grevillot L, Boersma DJ, Fuchs H, Aitkenhead A, Elia A, Bolsi M, et al. Technical note: GATE-RTion: a GATE/Geant4 release for clinical applications in scanned ion beam therapy. *Med Phys* 2020; **18** May 2020. doi: <https://doi.org/10.1002/mp.14242>
 28. Paganetti H. Dose to water versus dose to medium in proton beam therapy. *Phys Med Biol* 2009; **54**: 4399–421. doi: <https://doi.org/10.1088/0031-9155/54/14/004>
 29. Wright D., and Incerti S. A short guide to choosing physics Lists Available at: http://geant4.in2p3.fr/IMG/pdf_PhysicsLists.pdf (Accessed February). 2020.
 30. Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Phys Med Biol* 1996; **41**: 111–24. doi: <https://doi.org/10.1088/0031-9155/41/1/009>
 31. ICRP (International Commission on Radiation Protection) Publication 110: adult reference computational phantoms. *Annals of the ICRP* 2009; **39**.
 32. ICRU (International Commission on Radiation Units and Measurements) Report 49: stopping powers and ranges for protons and alpha particles. *Journal of the ICRU* 1993; **25**.
 33. Hubbell J. H., and Seltzer S. M. Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements Z = 1 to 92 and 48 additional substances of dosimetric interest: National Institute of Standards and Technology; Available at: <http://physics.nist.gov/PhysRefData/XrayMassCoef/cover.html> (Accessed February 2020). 1995.
 34. Winterhalter C, Aitkenhead A, Oxley D, Richardson J, Weber DC, MacKay RI, et al. Pitfalls in the beam modelling process of Monte Carlo calculations for proton pencil beam scanning. *Br J Radiol* 2020; **93**: 20190919. doi: <https://doi.org/10.1259/bjr.20190919>
 35. Chetty IJ, Rosu M, Kessler ML, Fraass BA, Ten Haken RK, Kong FM, et al. Reporting and analyzing statistical uncertainties in Monte Carlo–based treatment planning. *Int J Radiat Oncol Biol Phys* 2006; **56**: 1249–59. doi: <https://doi.org/10.1016/j.ijrobp.2006.03.039>
 36. Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys* 2003; **30**: 2455–64. doi: <https://doi.org/10.1118/1.1598711>
 37. Winterhalter C, Fura E, Tian Y, Aitkenhead A, Bolsi A, Dieterle M, et al. Validating a Monte Carlo approach to absolute dose quality assurance for proton pencil beam scanning. *Phys Med Biol* 2018; **63**: 175001. doi: <https://doi.org/10.1088/1361-6560/aad3ae>
 38. Elia A, Resch AF, Carlino A, Böhlen TT, Fuchs H, Palmans H, et al. A GATE/Geant4 beam model for the MedAustron non-isocentric proton treatment plans quality assurance. *Phys Med* 2020; **71**: 115–23. doi: <https://doi.org/10.1016/j.ejmp.2020.02.006>
 39. Smith EAK, Henthorn NT, Warmenhoven JW, Ingram SP, Aitkenhead AH, Richardson JC, et al. In silico models of DNA damage and repair in proton treatment planning: a proof of concept. *Sci Rep* 2019; **9**: 19870. doi: <https://doi.org/10.1038/s41598-019-56258-5>