University of Groningen

Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy

Guterres Marmitt, G; Pin, A; Ng Wei Siang, K; Janssens, G; Souris, K; Cohilis, M; Langendijk, J A; Both, S; Knopf, A; Meijers, A

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Original paper

Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy

G. Guterres Marmitt[a,*], A. Pin[b], K. Ng Wei Siang[a], G. Janssens[b], K. Souris[c], M. Cohilis[c], J.A. Langendijk[a], S. Both[a], A. Knopf[a], A. Meijers[a]

[a] *Department of Radiation Oncology, University Medical Center of Groningen, University of Groningen, Groningen, The Netherlands*
[b] *Ion Beam Applications, Louvain-la-Neuve, Belgium*
[c] *Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium*

ABSTRACT

For radiation therapy, it is crucial to ensure that the delivered dose matches the planned dose. Errors in the dose calculations done in the treatment planning system (TPS), treatment delivery errors, other software bugs or data corruption during transfer might lead to significant differences between predicted and delivered doses. As such, patient specific quality assurance (QA) of dose distributions, through experimental validation of individual fields, is necessary. These measurement based approaches, however, are performed with 2D detectors, with limited resolution and in a water phantom. Moreover, they are work intensive and often impose a bottleneck to treatment efficiency. In this work, we investigated the potential to replace measurement-based approach with a simulation-based patient specific QA using a Monte Carlo (MC) code as independent dose calculation engine in combination with treatment log files. Our developed QA platform is composed of a web interface, servers and computation scripts, and is capable to autonomously launch simulations, identify and report dosimetric inconsistencies. To validate the beam model of independent MC engine, in-water simulations of mono-energetic layers and 30 SOBP-type dose distributions were performed. Average Gamma passing ratio 99 ± 0.5% for criteria 2%/2 mm was observed. To demonstrate feasibility of the proposed approach, 10 clinical cases such as head and neck, intracranial indications and craniospinal axis, were retrospectively evaluated via the QA platform. The results obtained via QA platform were compared to QA results obtained by measurement-based approach. This comparison demonstrated consistency between the methods, while the proposed approach significantly reduced in-room time required for QA procedures.

1. Introduction

The number of proton therapy centers is further growing, permitting the delivery of highly conformal dose distributions through the use of pencil beam scanning (PBS) [1,2]. For a PBS treatment plan, the weight of several thousand of pencil beams is iteratively optimized to achieve a conformal high dose region while sparing organs at risk [3,4]. Multi-field optimization resulting in intensity modulated proton therapy (IMPT) plans is seen as state-of-the-art. The achievement of homogeneous target dose distribution with minimum and optimally balanced normal tissue doses for IMPT plans generally leads to highly complex in-homogeneous, per-field target dose distributions [4]. Sub-optimal treatment plans were shown to help account for the uncertainties during these optimizations [5].

Treatment planning systems (TPS) that calculate such plans are complex software systems [6], which makes comprehensive testing, commissioning and quality assurance inevitable. In addition to the optimized fluence map, the delivery of a PBS treatment plan requires at least two more transformations. In the first step, it needs to be converted into machine readable files and in a second step, these files have to be correctly interpreted and delivered by the treatment machine. Both of these transformations are potential sources of errors which may be difficult to detect, especially given the complexity of the treatment plans.

As such, patient specific quality assurance (PSQA) of absolute dose distributions, through experimental validation of individual fields, is currently necessary and commonly done. Multiple experimental approaches for patient specific QA have been reported [7–10]. Excepting few 3D measurement approaches [11], these measurement are performed with 2D detectors, with limited resolution and in a solid water

---

* Corresponding author.
  E-mail address: g.guterres.marmitt@umcg.nl (G. Guterres Marmitt).

phantom. Moreover, they are work intensive and often impose a bottleneck for the throughput of a treatment room, or limit the ability to adapt a treatment plan in a timely manner.

In order to decrease the PSQA measurement beam-time, work has been done for the use of delivery system control files (hereafter referred to as treatment log files) instead [12–14]. Meier et al. [15] have shown the use of log files for independent dose calculation systems, with the intention of detecting problems or differences in TPS dose computations. In order to achieve a greater independence, the Monte Carlo dose engine used for QA dose calculation should be based on independent algorithms with completely separated code bases.

After each delivery, files containing details of the machine parameters are generated by the Proton Therapy System (PTS). Treatment log files can either be obtained prior to the start of a treatment course by performing a dry-run irradiation or will be generated inherently during each delivery of a fraction. In order to apply this method successfully, log files must contain information on the delivered spot position, dose and energy recorded. This information may then be used to create a plan and reconstruct the dose that was actually delivered, which then could be compared to the prescribed planning dose.

The aim of this work is to describe the implementation of a platform for the execution of PSQA workflows, and present an extended validation of its many components. Such a platform should require minimal human intervention, relying on automated simulations when data is available. Also, it should be flexible enough in order to fit future applications, such as adaptive planning and 4D dose accumulation. Initially, two workflows were designed and implemented: TPS-plan-based QA, which uses an independent Monte Carlo engine for checking of the TPS dose calculation; and Log-based QA, which reconstructs the dose based on the machine logs. Dose calculation by QA platform is performed on patient's geometry using the planning CT.

2. Materials and methods

In order to integrate incoming data processing, computation, visualization and report, a software platform following a server-client architecture was developed. The main building blocks of the application are part of the OpenPATh initiative [16] created to support open-source software applications for research in proton-therapy. Open-source enables researchers to reuse and build upon existent code to avoid rewriting from scratch. In addition, a multi-party contribution to the development and usage of said applications improves the robustness and the trustworthiness of research software for proton-therapy [17]. The open-source modules used in this research include OpenREGGUI, Orthanc, MCsquare, and CAPTAIN and are presented in the following subsections.

2.1. OpenREGGUI

OpenREGGUI [18] is an image processing software featuring various registration methods, filtering methods, segmentation tools and other radiotherapy dedicated functions such as dose volume histogram computation and others. It is a powerful application interface that helps clinicians to monitor patient information, and to compare planned treatments with actual measurements when running clinical studies in research projects.

The use of OpenREGGUI requires MATLAB [19]. It offers a graphical interface to visualize DICOM images and to operate many image processing functions. It allows defining complex workflows that can be triggered directly from the MATLAB command line as well. The toolkit also provides many desired functionalities: a) a wrapper function that formats DICOM files to the input files required by MCsquare; b) Interpolation of dose maps to guarantee matching grid sizes between TPS and MCsquare dose simulations; c) Evaluation of clinical goals after dose computation (DVH computation). In this work, such workflows were used as data processing libraries.

2.2. Orthanc

The project known as Orthanc [20] was used as a standalone DICOM server. What makes Orthanc a compelling choice is the fact that it provides a comprehensive Application Programming Interface (API), making it possible to access it from any computer language. Orthanc receives a request for data download each time a task requires it, and uploads the resulting data when such process finishes.

2.3. MCsquare

Monte Carlo dose recalculations were performed using MCsquare [21]. It is an open-source fast multipurpose Monte Carlo algorithm, optimized for exploiting massively parallel central processing unit (CPU) architectures. Simulations were performed with 12 calculation threads, in an Intel Xeon server with 48 processing units. The 64 GB of RAM available are shared when multiple simulations are launched simultaneously, each allocating approximately 10 GB.

MCsquare was configured to run all simulations with a MC statistics of 1×10^8 particles, which is equivalent to a standard deviation between 1 and 2% calculated inside the 50% higher dose region for all clinical plans tested. Its inputs are the DICOM files for the plan and the CT coming from the TPS. The method described by Schneider et al. [22] was used to convert from HU to human tissues, which includes elemental composition, weights and densities. The elastic and inelastic nuclear interactions are sampled from cross sections in the ICRU 63 report. In order to be able to compare it with the TPS dose, the dose-to-medium exported MCsquare is later converted into dose-to-water by applying the appropriate Stopping Power Ratio to each voxel in the dose map [23].

MCsquare uses different algorithms, code base and physical tables from RayStation MC. However, they were shown to have similar accuracy in simulation and experimental validations [24].

2.4. CAPTAIN

The CAPTAIN project [18] was built with a series of industry standard web technologies and is a Free and Open Source Software (FOSS) released under the Apache 2 license. In short, CAPTAIN is an automated workflow manager. Its main feature is the autonomous launching of computation workflows without human intervention. CAPTAIN is developed based on Node.js [25] technology.

Fig. 1 shows a scheme of the platform, which can be separated into three layers:

1. The user interface composed of a web-site with access to workflow configuration and results.
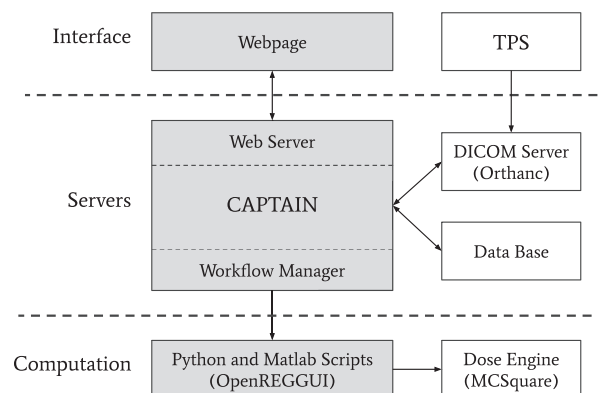2. The servers, composed of three isolated processes: CAPTAIN main

Fig. 1. Scheme of the platform architecture, its core units (in grey) and external components (in white).

server including the Workflow Manager, a dedicated DICOM server and a database server.

3. The computation layer, where a series of OpenREGGUI and Python scripts are called in order to complete the assigned tasks. These scripts are also responsible for launching the Monte Carlo simulations.

These layers are further discussed on the subsections below.

2.4.1. Interface

The user interface with CAPTAIN server is done through its website. It was written in Typescript and HTML5. It follows a centralized model where the client task is only interfacing with the servers.

The home page has an index of all patients, and from each patient entry one may configure workflow setups and preview results. Each workflow has its dedicated configuration page, where the default parameter values may be changed and/or new data may be uploaded. During a treatment, as new data become available, different configurations might be set for the workflows.

The result page of a workflow displays a list of all calculation tasks, their execution status, and links leading to reports. The report page has detailed information on a workflow run, such as configuration parameters and result values of the calculations. DICOM objects may also be downloaded from the report page, both from configuration and results fields.

2.4.2. Servers

The CAPTAIN main server is responsible for the data exchange occurring between clients, data storage servers and the computation scripts, introduced below. It was generated with the Angular Full-Stack Generator [26], is managed by Gulp.js [27] and written in javascript.

The server offers four Application Programming Interfaces (APIs) to connect to it and act on the database (DB). These four APIs are:

- User API: entry point from the user interface to configure the user access to the system.
- Dicom-server API: exchange of data between the CAPTAIN server and the DICOM server Orthanc.
- patient API: entry point from the user interface that is used to access patient data, to configure the workflows parameters.
- patientResults API: manages the access from the user-interface to the workflow results.

Apart from the dedicated DICOM server, described at Sub-Section 2.2, a second database server is responsible for holding state information of the platform. MongoDB [28] is a FOSS cross-platform document-oriented database that runs on NoSQL and uses JSON-like documents. This DB contains three collections: users, patients and results. These collections hold documents with information on user authorization and permissions, patient meta-data, and workflow result values or DICOM meta-data, respectively.

It is imperative that both data servers are kept synchronized to each other. To that end, a system of triggers and parser were implemented. Upon the arrival of new data at the DICOM Server, a signal is sent to CAPTAIN server requesting parsing of the received patient data. Meta-data is extracted from the DICOM objects, and then saved as a Patient document in the DB. These documents are lightweight structure data, holding only descriptions and links pointing to relevant instances inside the DICOM Server.

2.4.3. Computation

In this work, *workflow* denotes a complete computation chain, starting at a set of initial parameters and finalizing at a set of result values. The computation between these two states is divided in small chunks, here called *tasks*, which perform more specific calculations. Therefore, a workflow is the recipe of what tasks to execute, in which

order and with what parameters. It is defined by the following objects:

- Check function, which verifies if all necessary parameters are available, and queries the DB for previously calculated results.
- List of Tasks to be performed, each completing a specific computation step on the available data.
- Task input recipe, encoding which parameters to use in each of the corresponding tasks.
- Task output recipe, encoding which results to save from each of the corresponding tasks.

A single task may be used by multiple workflows, which lower complexity and increases consistency between different workflows. Examples of implemented tasks include: the conversion of treatment log files into a RT plan, independent dose recalculation and evaluation of Gamma analysis between two dose maps. Each task is composed of the following steps:

- Preparation, which creates temporary folders for computation, queries the DICOM server for data and the DB for other input parameters.
- Launching, responsible for the start of the computational step, which may include multiple script launches and/or external calls.
- Exporting, which saves the output data to the DB and/or DICOM objects to the DICOM server.

In order to function autonomously, the Workflow Manager receives a notification when new patient data is received by the DICOM Sever. The Manager then triggers the start of all available workflows, starting with the Check for available data. Many of the required input parameter are automatically set, such as the plan, structure set and dose received from TPS. Others may be adjusted, such as gamma analysis distance and dose tolerances.

2.5. Workflows for QA

For an envisioned automatic plan QA, two workflows were created: (A) a TPS-plan-based patient QA and (B) a Log-based patient QA workflow. A scheme of these workflows is showed in Fig. 2.

In (A) TPS-based plan QA, a secondary Monte Carlo dose calculation is executed with the same plan input as the TPS but with a completely independent implementation (using a different programming language, different algorithms, different physical models, and with the code being written by a different developer). Hence, this allows for a redundant check of the treatment planning dose calculation.

In (B) Log-based plan QA, a secondary Monte Carlo dose calculation is made, taking the field information from a different source than the TPS. A new plan is created from treatment log files, which become available after each beam delivery. This can be used as a consistency check between the expected TPS calculated dose and the PTS effective delivery dose.

In clinical routine, these workflows are foreseen to be executed consecutively. Planning is performed with RayStation's Monte Carlo algorithm, which generates the original (TPS) dose map. After plan approval in TPS, data is exported to a dedicated DICOM server that triggers the QA workflow (A). An independent dose recalculation automatically takes place and Gamma analysis results - comparing original and recalculated dose maps - become available at the website.

After review, a dry-run is performed by delivering the plan in air in order to generate treatment log files. The upload of the logs triggers the Log-based plan QA workflow (B). A Log-based plan is reconstructed from the retrieved treatment log files, and results become automatically available for review.

2.5.1. Log to plan conversion

The structure and content of treatment log files are vendor specific

TPS-Plan QA

Log-Plan QA

Log to plan conversion

→ PTS log files   Log-Plan →
→ TPS-Plan

MCsquare Dose Calculation

→ Planning CT   MC2-Dose →
→ TPS-Plan

MCsquare Dose Calculation

→ Planning CT   MC2-Dose →
→ Log-Plan

Clinical goals evaluation

→ Structure set   Goals pass/fail →
→ MC2-Dose

Clinical goals evaluation

→ Structure set   Goals pass/fail →
→ MC2-Dose

Dose gamma analysis

→ TPS-Dose   Gamma Map →
→ MC2-Dose   Pass ratio →

Dose gamma analysis

→ TPS-Dose   Gamma Map →
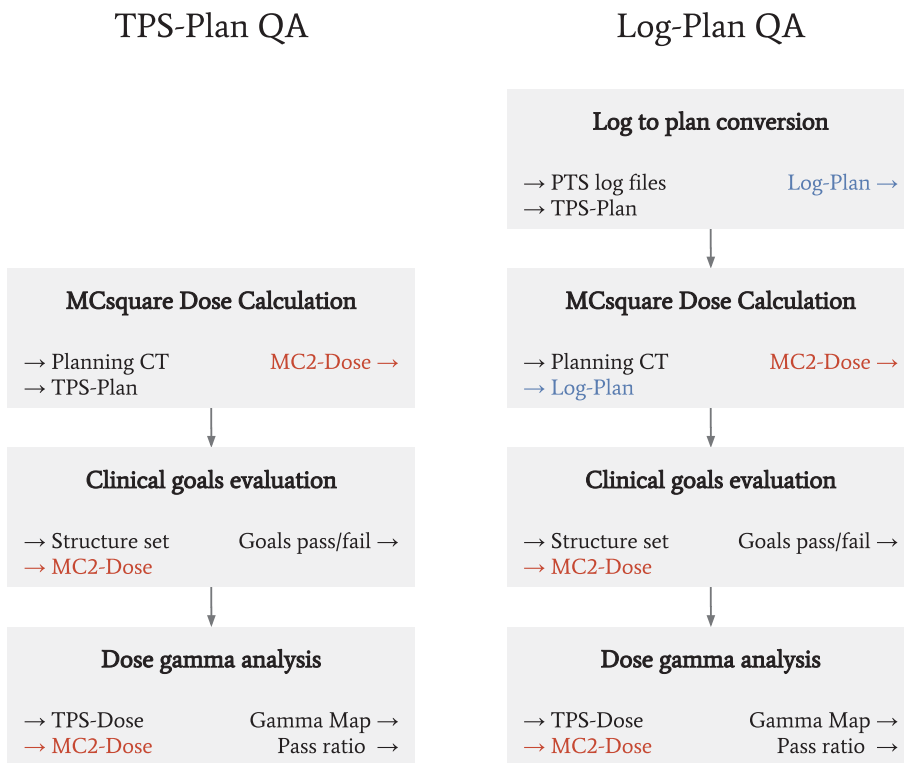→ MC2-Dose   Pass ratio →

Fig. 2. Schemes of the TPS- and Log-based Plan QA workflows. Each task has its input parameters showed on the left, and output on the right. The outputs of one task may be linked to the input of another inside a workflow chain, such as the MCsquare recomputed dose map (MC2-Dose, in red) and the log reconstructed plan (Log-Plan, in blue).

or even equipment model specific, however generally treatment log files will contain chronological list of events, which were registered by therapy control system during the delivery of specific treatment prescription, and a list of readouts that were acquired by sensors, which are integral part of the delivery control system. Sensor readouts may contain such information as potentiometer positions, hall probe readouts, set points on power supplies, charge on strips or wires of ionization chambers, etc. Although the readouts from the sensors may not directly be meaningful in any way to what type of prescription has been delivered to the patient, this information may be used to recontact what prescriptions have been delivered into clinically meaningful way. After all, therapy control system as an input uses clinical prescription (treatment plans), to define state of various components of the system to achieve delivery of a specific prescription. By reversing this process clinical prescription itself can be reconstructed based on the state of the machine, which is indicated by the output of various built-in sensors. Currently, the availability of such log files are subjected to specific agreements with the manufacturer of the delivery system.

Log file interpretation script retrieves from the treatment log files a set of spots that has been delivered during the specific session and assigns their position in X and Y directions at the isocenter plane by using readouts from the strips of the in-nozzle ionization chamber, energy based on the position of degrader wheel and MU based on readout from the integral pads of the in-nozzle ionization chambers. Additionally, a set of corrections, such as, virtual source-axis-distance (VSAD) and IC-to-isocentre distance correction, temperature and pressure correction, etc, needs to be applied.

After the content of the delivered prescription is reconstructed, it is written into a DICOM ion plan object and log-based plan is created. Compatibility to DICOM standard significantly eases usability of the data for other purposes; not only being limited to the use for independent dose re-calculation within the scope of patient QA platform, but also enables use cases like import and re-calculation in TPS, etc.

The quality of the reconstructed plan is dependent on the accuracy of log files recorded parameters, therefore constant validation of its performance is required. The procedure adopted for log files validation in this work is discussed in Sub-Section 2.6.1.

2.5.2. Structures overrides

Clinical plans often have overridden structures created in the TPS. Their main uses are: (a) cover patient support devices, (b) create regions with uniform water phantoms and (c) define boundaries for dose calculations within external contour. In order to perform accurate calculations, these exceptions are handled by python scripting. It automatically identifies override tags in the structure set DICOM header, such as the presence of a Material ID in the structure description or an Interpreted Type of 'External'. For (a) and (b) the density of the linked material is converted to Hounsfield Units (HU) by consulting the CT calibration curve, then overwriting the structure space inside the CT with its interpolated value. For (c), the volume outside the external structure is overwritten with air. Therefore, a new CT image with all overrides applied to it is created and then used as input for the MC simulations.

2.5.3. Clinical goals evaluation

In clinical practice treatment plans are commonly assessed on basis of clinical goals. Clinical goals are defined as a set of dosimetric criteria that should be met to achieve intended clinical outcome while risks of developing complications are maintained reasonably low. Clinical goals usually are either defined by user in TPS and checked automatically or are manually checked by the user during the plan review based on dose-volume histogram (DVH).

In case of RayStation (RaySearch, Sweden) TPS, clinical goal templates can be defined in TPS and in our clinical practice they are commonly used during plan review process. A dedicated script was developed that can be executed in TPS on patient specific basis to extract a list of defined patient-specific clinical goals. Afterwards the list is written in a specific format to a JSON file and may be imported into the patient QA platform via dedicated interface. By using the imported file, a set of clinical goals is populated in the QA platform.

This makes it possible to assess independently re-calculated dose distribution against clinical goals in a similar way as it is done during plan review in TPS. One of the main reasons to use clinical goals in combination with more commonly employed Gamma analysis for assessment of QA dose distribution is because often it is not exactly

straight forward to interpret Gamma analysis results in a clinically meaningful way. In other words, failing pixels or voxels in Gamma analysis cannot be easily interpreted in the sense of their clinical relevance. In some cases, localization of failing Gamma analysis points may be highly important clinically. The expectation is that evaluation of the QA dose distribution against clinical goals will help to identify these situations, even if Gamma passing ratio, especially globally, would not seem alarming.

2.5.4. Gamma analysis

In this study, dose distributions are compared using a 3D version of the global Gamma analysis [29,30]. The calculations are performed by an external Python script, which is based on the *npgamma* library [31]. Both reference (TPS) and evaluated (MCsquare) dose maps are Monte Carlo simulations, which may skew passing rate evaluations due to the statistical uncertainties [32]. For this reason, MC statistical error is kept less than 2 % in all calculations presented here.

2.6. Validation and testing

Due to complexity of the platform an extensive testing plan was put in place to ensure expected functionality. Testing and validation effort may be mainly split in four sections:

2.6.1. Log recording of plan delivery

Treatment log files from QA tests were collected at the IBA proton therapy system in the Groningen Proton Therapy Center and then used to reconstruct the delivered plans, here called log-based plans. These are plans from standard daily Morning QA. The plan is composed of 1580 spots, from which the recorded spot positions and MU were used to retrospectively evaluate agreement between log content (output) and delivery prescription (input). Following the deliveries through half a year period provided information on the accuracy and consistency of the delivery system. Split between the two delivery rooms, 60 log files sets were analyzed in this manner.

Additionally, independent external measurements from standard monthly QA procedure with a Lynx detector were used to assess the errors during delivery. The plans consisted of 5 spots per field, delivered with energies between 70 and 225 MeV and measured in gantry angles between 50 and 315°. Log files from these deliveries were collected and analyzed. The relative position of the spots in reference to a central spot was then calculated for measurement and log files recordings.

2.6.2. Low level component testing

As introduced earlier, workflows make use of multiple lower level computational modules in order to generate results. Performance of such modules was tested by performing calculations under controlled conditions, where data input is well defined and expected output is known or can be easily predicted. Testing of this type was applied to such modules, and the procedure varied per component:

- 3D Gamma analysis module was tested by introducing errors of known magnitude in the synthetic dose object. For the testing purpose geometrical and dose errors were introduced.
- The log to plan converter was tested in two steps: (a) by reconstructing spot energy, position and MU and comparing it directly to the planned ones and (b) by importing log-based DICOM plans back into the TPS, using TPS dose engine to recalculate the dose and comparing it to the original planned dose.
- REGGUI's implementation of clinical goals evaluation was validated against the planning TPS evaluation. A set of 5 plans were calculated with both methods, and the clinical goals evaluations were grouped into 4 categories: target volume coverage in CTV; Mean dose in ROIs; maximum and minimum dose in ROIs above 100 cGy cutoff; maximum and minimum dose in ROIs below 100 cGy cutoff.

2.6.3. Beam model validation

Accuracy of the independent Monte Carlo engine and quality of the beam model are crucial to the proposed QA workflow; therefore, particular attention was paid to validation of this component. Validation of the beam model was performed in several tests, where the complexity of the testing method gradually increases. Initially in-water calculations were performed for several mono-energetic layers. The energy of these layers was varied between 70 and 225 MeV. The main objective of these calculations was to determine range calculation accuracy in-water. Further a set of 30 SOBP-type fields, which was a sub-set of validation data that was earlier used for the purpose of TPS commissioning, was calculated in water. Range of SOBP fields was varied between 32 g/cm^2 and 4.1 g/cm^2 and modulation - between 2 and 4 cm. SOBP calculations performed by independent MC engine were compared to calculations performed by clinically used TPS, which has been already commissioned, by using 3D Gamma analysis with criteria of 2%/2 mm. The Gamma analysis was performed in absolute dose procedure which is sensitive to dose ratio discrepancies, and calculated range discrepancies were analyzed. Eventually, an experiment using animal tissues was set up to evaluate accuracy of MC calculations taking into account lateral and longitudinal heterogeneities. A vacuumed pig's head, positioned on top of solid water slabs, was scanned on a CT. Multiple treatment fields (SOBP and mono-energy layers) in anterior-posterior direction were prepared and a dose plane at the depth of 3.1 mm (WET) in solid water below the animal tissues was selected for comparison with the measurements. In the proton treatment room head was aligned using CBCT and measurements at the selected depth were performed with an ionization chamber array MatriXX PT (IBA dosimetry, Germany). Proton beam measurements were performed on a pigs head phantom using one SOBP and three mono-energetic beams with energies between 175 and 225 MeV, and an array of ionization chambers was used for measuring 2D dose distributions at different depths of 3 and 7 mm. Calculations were performed by TPS and MCsquare dose engines with MC statistic tuned for 0.5% uncertainty, adjusting the number of particles per plan accordingly. Measured dose planes were compared to calculations by using 2D Gamma analysis with the criteria of 3%/3 mm.

2.6.4. Functional workflow testing

To test the functionality of the QA platform in clinical setting 10 patient cases were evaluated by proposed QA method retrospectively. These clinical cases included such indications as head and neck, intracranial indications and cranio-spinal axis. Via functional testing full data flow was considered: beginning with data transfer from TPS to QA platform and ending with creation of the reports. As part of this testing phase timing of the workflows was performed.

The gamma analysis comparison for TPS- and Log-QA workflows are performed with an acceptance criteria of higher than 95% Passing Ratio (2 mm/2%) calculated in the 3D volume of the dose map. The criteria values were tuned for increasing the error sensitivity of the workflows. For validation purposes, its results are then compared to the standard measurement-based QA - which is currently performed with an acceptance criteria of higher than 95% Passing Ratio (3 mm/3%) in 2D dose maps measured and calculated at 3 different depths per field.

3. Results

The results are laid out in increasing complexity. We start from the validation of the log file recording consistency, followed by the validation of the MCsquare beam model and the application of our proposed PSQA workflows to a set of clinical plans. Then, some of the clinical plans are used for validation of the log to plan reconstruction algorithm and the OpenREGGUI implementation for clinical goals evaluation.

Table 1
Comparison between log files recordings and plan prescriptions over 60 deliveries split between two delivery rooms. Statistics for average and standard deviation of position and dose errors are given together with the maximum (max.) observed value in the data set.

Error description	Room 1 (30 deliveries)	Room 2 (30 deliveries)
Spot position shift in x (mm)	-0.0891 ± 0.476 max. 1.15	0.0213 ± 0.250 max. 1.15
Spot position shift in y (mm)	-0.0961 ± 0.347 max. 1.01	-0.439 ± 0.568 max. 1.12
Accumulated MU error (MU)	0.772 ± 0.316 max. 1.38	1.16 ± 0.259 max. 1.75

3.1. Log file consistency

Log-files consistency was validated, as per Sub-Section 2.6.1. When comparing the reconstructed spot position and MU to 60 morning QA plans prescriptions over half a year period, the observed average spot position error and standard deviation in X was -0.0339 ± 0.380 mm, and in Y was 0.268 ± 0.470 mm. The maximum accumulated MU error over one entire delivery was 1.75 MU, which corresponds to 0.4% of the prescribed dose (418 MU). A detailed list of the error analysis performed is presented in Table 1.

When analyzing the log reconstructed spot position to measurements from monthly QA plans, relative position errors were calculated. Considering deliveries in Room 1, the average and standard deviation in X was 0.0373 ± 0.221 mm and in Y was 0.0940 ± 0.217 mm; with a maximum position error of 0.514 mm. In Room 2 the results were similar, the average and standard deviation in X was 0.0727 ± 0.155 mm and in Y was 0.0494 ± 0.235 mm; with a maximum position error of 0.386 mm.

3.2. Beam model validation

As discussed in Sub-Section 2.6.3, the beam model was validated via MC-based calculations of treatment fields performed with MCsquare. The comparison of 30 SOBPs containing ranges between 4.1 and 32 g/cm^2 and modulation between 20 and 40 mm showed a 99% $\pm$ 0.5% Gamma passing ratio. For the full energy spectrum from 70 to 225 MeV range discrepancy in water was < 1 mm.

The dose calculation accuracy of the MC dose engine was also evaluated using heterogeneous real animal tissues [33]. Comparisons between 2D dose distributions from measurement and simulations with Gamma criterion of 3%/3 mm are provided in Table 2. Gamma pass ratios are approximately 95% or greater for all cases. Deviations are found at high density gradient regions (soft tissue/bones and air/tissue interfaces) and high dose gradients regions, which could be explained by the different material tables used to convert the CT image to chemical compositions in the two engines. MCsquare simulations

Table 2
Accuracy comparison for TPS and MCsquare dose calculations in heterogeneous animal tissue showing Gamma percentage pass ratios for different beam energies and depths. TPS dose maps and MCsquare dose maps are evaluated with 2D Gamma analysis (3%/3 mm).

Beam energy, solid-water depth	TPS dose (γ pass ratio %)	MCsquare dose (γ pass ratio %)
SOBP	98.2	94.9
225 MeV, 3 mm spacing	98.3	98.7
225 MeV, 7 mm spacing	97.9	98.9
200 MeV, 3 mm spacing	99.0	96.9
200 MeV, 7 mm spacing	98.0	94.6
175 MeV, 3 mm spacing	98.5	98.9
175 MeV, 7 mm spacing	99.2	98.3

Table 3
Gamma pass ratios for measurement-based plan QA, and the proposed TPS- and Log-based plan QA.

Patient	Measurement based (γ pass ratio %)	TPS-plan recalculation (γ pass ratio %)	Log-plan recalculation (γ pass ratio %)
1	100	97.43	97.71
2	99.81	98.75	96.94
3	100	98.60	91.56
4	100	98.43	98.59
5	98.31	95.37	93.30
6	98.55	96.05	95.54
7	99.44	99.05	97.98
8	99.12	99.14	98.99
9	99.52	99.30	99.14
10	99.31	96.02	94.97

performed with greater MC statistics (1×10^9 particles) showed no measurable improvement, indicating lower statistical noise compared to other sources of uncertainties.

3.3. Workflow testing

The feasibility of the TPS-based and Log-based QA workflows was tested against standard measurement-based QA, as referred in SubSection 2.6.4. A comparison between measurement-based, TPS-based and Log-based QA for 10 clinical cases, including craniospinal axis, intracranial and head and neck cases, is summarized in Table 3. An example of calculated dose distributions for a breast cancer case is shown in Fig. 3.

Independent MC calculations for these cases require 15–20 min calculation time per treatment plan. Patient specific QA according to the proposed TPS-based and Log-based methodology requires about 10 min of in-treatment-room time per patient for log file acquisition. That compares to 40 min in-treatment-room time per patient for measurement-based QA.

3.4. Component testing with clinical plans

Some of the aforementioned clinical plans were used for the validation of CAPTAIN's low-level components, as described in Sub-Section 2.6.2. Based on calculated DVH, no statistical meaningful differences between the two plans are found - example showcased in Fig. 4.

The average difference between the TPS and OpenREGGUI clinical goals evaluation is below 1% for all cases but for ROIs with low doses. For ROIs with <100 cGy, the difference is larger yielding to approximately 15%. In particular, average dose indicators are in very good agreement, with an average difference of 0.007%.

4. Discussion

The overall trend observed in the workflow validation is that TPS-based plan QA showed lower Gamma pass ratios than measurement-based QA. This is expected, since the former compares the whole 3D dose distribution volume with CT compositions and the later relies on 2D Gamma analysis at different depths of solid-water phantoms. In turn, the Log-based plan QA shows a Gamma analysis pass ratio marginally lower than the aforementioned cases. Since both TPS- and Log-based plan QAs use the same dose computation procedure, the difference can be attributed to discrepancies in delivered spot position or dose, which were recorded on the treatment log files and used during plan reconstruction.

The automated PSQA is managed and executed from inside the CAPTAIN server, a multipurpose and flexible platform. In order to provide specific functionalities, it integrates with other open source projects, such as OpenREGGUI and MCsquare. The use of these modules was crucial to speed up development and guarantee performance,
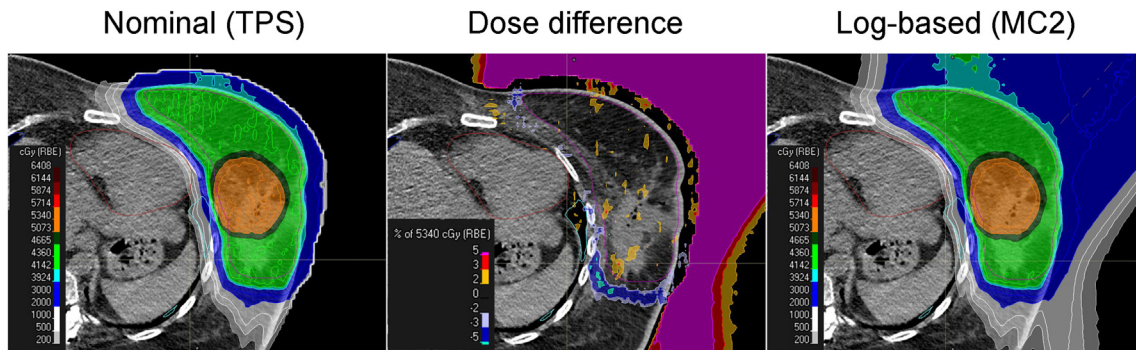
Nominal (TPS)   Dose difference   Log-based (MC2)

Fig. 3. Dose maps for the TPS calculated plan (left), the log-based recalculation (right) and the dose difference (center). For visualization purpose, recalculated dose maps have been imported in TPS. The dose maps are compared with Gamma analysis (2 mm, 2%).

however it also demands comprehensive testing of its components. First, MCsquare beam model profile was validated by comparing simulations for a wide array of SOBPs in solid water phantoms, and in inhomogeneous animal tissue phantoms. Over a wide range of energies and modulations, good agreement was found between MCsquare and TPS dose distributions. Secondly, OpenREGGUI evaluation of clinical goals was validated against the TPS evaluation method for a set of clinical plans. The results indicate that in ROIs subjected to doses higher than 100 cGy, the average difference between the two implementations was lower than 1%. Noticeably, the discrepancies are larger for small doses and small volumes; since different algorithms for DVH computation are used in TPS and OpenREGGUI, processes such as
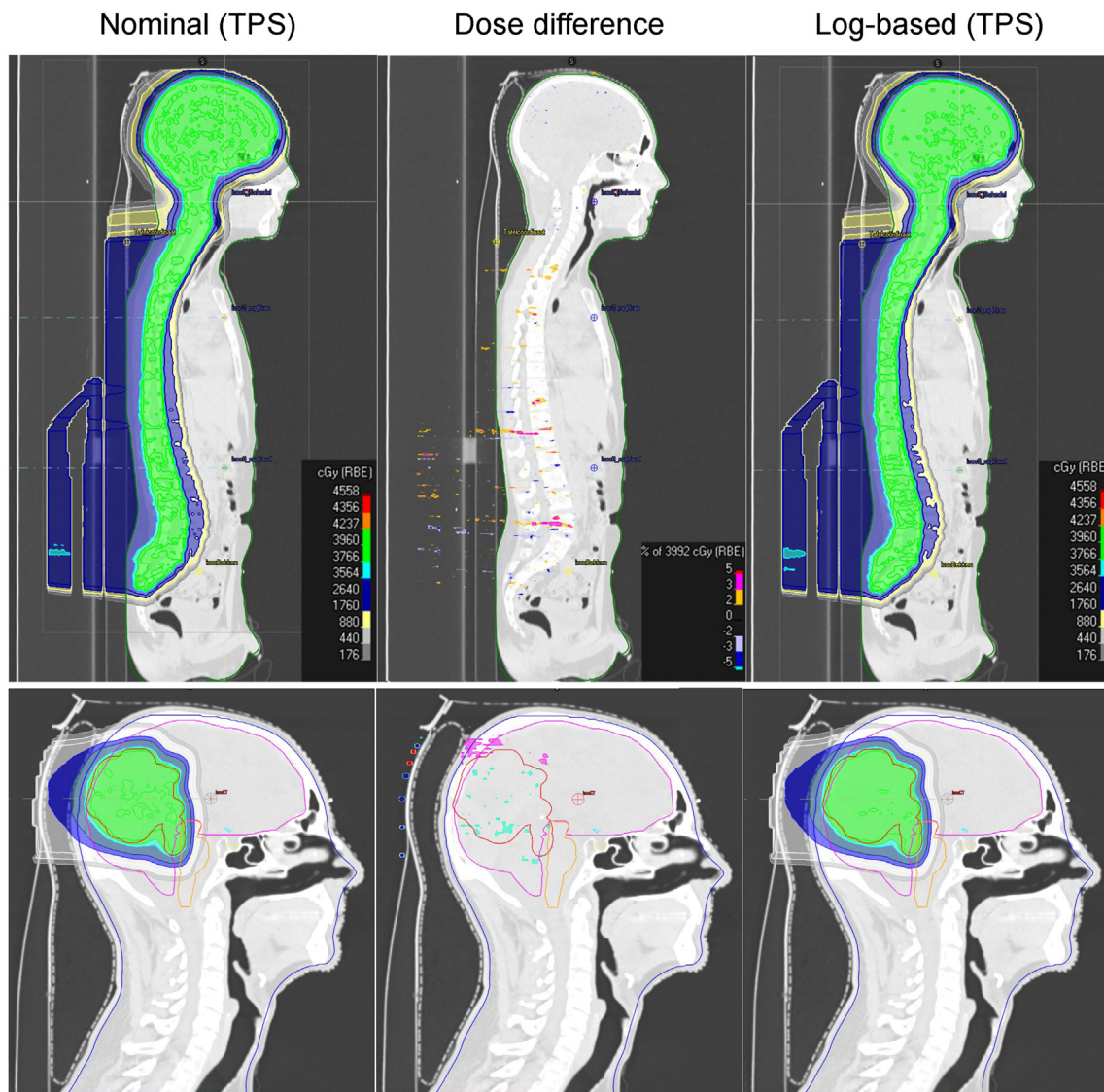
Nominal (TPS)   Dose difference   Log-based (TPS)

Fig. 4. TPS calculated dose maps: planning dose (left), log-based recalculation (right) and the dose difference (center). Isodose curves showed in color. The dose maps are compared with Gamma analysis (2 mm, 2%).

interpolation and voxelization play a greater role in these cases. The evaluation of clinical goals is dependent on the approach taken for calculation of DVHs. As there are different approaches possible, it may introduce bias in clinical goal evaluation. Consistency of the two methods is necessary for a complete re-evaluation of the reconstructed plan from treatment log files, where TPS evaluation of clinical goals are directly compared to OpenREGGUI's.

PSQA procedures based on independent dose calculations, such as ours, have already been introduced for PBS ion therapy in some facilities [12,34,35]. The main advantages are the reduced time required for the measurements and the high resolution 3D dose distributions provided in the patient geometry, which improves the verification procedure of both the planned and delivered dose. Recently, at PSI [15], a toolkit for independent dose calculations was developed, which allows for dose reconstructions at several points in the treatment workflow. Still, the implementation of such workflows in clinical environments is restricted to few examples. Our implementation of Log-based PSQA brings a reactive and automated platform for ease of use in clinical workflows. Since all presented here is open source, an interested reader should also be able to implement similar automated worflows in other clinics.

Due to its architecture, the CAPTAIN platform is capable to be extended for other purposes that would also benefit from its modularity. Workflows are simple computation recipes, and existing solutions are easily refitted into workflows' tasks. Based on fraction wise patient information this platform could accomplish automated daily dose reconstruction and accumulation. With an automated comparison of the accumulated dose against the expected dose, a request for adaptation could be automatically triggered in case of deviations. A concept for fraction-wise retrospective 4D dose reconstruction and accumulation was recently published [36]. Within a corresponding automated workflow in our introduced platform, supplied treatment log files, motion records and acquired repeated 4D CTs/CBCTs would trigger a dose reevaluation and would trigger an adaptation if significant treatment quality deviations are observed, such as dose discrepancy due to anatomical changes. We have plans to extend the platform to include such functionality in the future. In this case, the time required for the plan dry-run delivery and log files collection may become a hindrance for a fast online treatment tracking. Since adaptive workflow are generally complex and labor intensive, automatizing workflows as much as possible as proposed here will be essential for a clinical implementation of adaptive proton therapy.

The benefit from this kind of analysis mainly depends on the accuracy of the log-file values, therefore a preliminary assessment of the uncertainty in the recorded parameters of the scanning pencil beams is recommended. For example, Li et al. [37] have compared the planned and recorded values with dedicated measurements, to ensure that the monitor and the recording system work properly, and that the log-files are accurate enough to be used for evaluating the uncertainties in the delivered dose, caused by variations in the beam characteristics. Henceforth, similar routine machine QAs and further validation of CAPTAIN's components will be necessary for the clinical realization of the workflows here proposed.

5. Conclusions

A new PSQA workflow was developed using an automated web platform. Low level components were validated, such as the log to plan converter and clinical goals evaluation. MCsquare beam model was validated in solid water and animal tissue phantoms, displaying dose distributions comparable to others simulated by the Monte Carlo algorithm available in the TPS. The proposed patient specific automated QA shows consistency between the measurement- and Log-based QA for a wide range of clinical plans. This supports potential replacement of measurements with MC-based treatment plan QA in future. The use of this platform in clinical routine has the potential of significantly

reducing the required in-treatment-room time for PSQA. Furthermore, the implemented platform has the potential to also automatize other clinical procedures as for example fraction wise dose reconstruction and accumulation, which may provide input for decision support regarding plan adaptation.

References

[1] Mohan R, Grosshans D. Proton therapy – present and future. Adv Drug Deliv Rev 2017;109:26–44. Radiotherapy for cancer: present and future.

[2] Lomax A. What will the medical physics of proton therapy look like 10 years from now? A personal view. Med Phys 2018;45(11):e984–93.

[3] Lomax A. Intensity modulation methods for proton radiotherapy. Phys Med Biol 1999;44(1):185.

[4] Unkelbach J, Paganetti H. Robust proton treatment planning: physical and biological optimization. Seminars in Radiation Oncology 2018;28(2):88–96. Proton Radiation Therapy.

[5] Knopf AC, Lomax A. In vivoproton range verification: a review. Phys Med Biol 2013;58(15):R131–60.

[6] 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(2).

[7] 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(11):5831–7.

[8] 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(3):270–6.

[9] Lomax AJ, Böhringer T, Bolsi A, Coray D, Emert F, Goitein G, et al. Treatment planning and verification of proton therapy using spot scanning: initial experiences. Med Phys 2004;31(11):3150–7.

[10] 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(11):5998–6008.

[11] Henkner K, Winter M, Echner G, Ackermann B, Brons S, Horn J, et al. A motorized solid-state phantom for patient-specific dose verification in ion beam radiotherapy. Phys Med Biol 2015;60(18):7151.

[12] Zhu X, Li Y, Mackin D, Li H, Poenisch F, Lee A, et al. Towards effective and efficient patient-specific quality assurance for spot scanning proton therapy. Cancers 2015;06(7):631–47.

[13] 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(17):175001.

[14] 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(20):205014.

[15] 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(7):2819–36.

[16] Colaborators. The OpenPATh initiative; 2016. https://openpath.software/.

[17] Colaborators. Open-source; 2019. https://opensource.org/strategic.

[18] openREGGUI consortium. Image processing open-source platform for adaptive proton therapy in cancer treatment; 2016. https://openreggui.org/.

[19] MATLAB. The MathWorks Inc, Natick, MA, USA; 1994–2019. https://mathworks.com/.

[20] Jodogne S. Orthanc v1.4.2; 2012–2018. https://www.orthanc-server.com.

[21] Souris K, Lee JA, Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures. Med Phys 2016;43(4):1700–12.

[22] Bazalova M, Beaulieu L, Palefsky S, Verhaegen F. Correction of CT artifacts and its influence on Monte Carlo dose calculations. Med Phys 2007;34(61):2119–32.

[23] Paganetti H. Dose to water versus dose to medium in proton beam therapy. Phys Med Biol 2009;54(14):4399–421.

[24] Sorriaux J, Testa M, Paganetti H, de Xivry JO, Lee JA, Traneus E, et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous media. Phys Med 2017;38:10–5.

[25] Hammond S, Cantrill B. Node.js v10.11.0; 2009–2018. https://nodejs.org.

[26] contributors G. Angular Full-Stack generator v5.0.0; 2018. https://github.com/angular-fullstack/generator-angular-fullstack.

[27] Mao J, Schmitt M, Stryjewski T, Holt CC, Lubelski W. gulp.js v4.0.0; 2013–2018. https://github.com/gulpjs/gulp.

[28] Ittycheria D, Merriman D, Horowitz E. MongoDB v4.0.2; 2009–2018. https://www.mongodb.com.

[29] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25(5):656–61.

[30] Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. Med Phys 2003;30(9):2455–64.

[31] Biggs S, contributors. npgamma; 2015–2018. https://pypi.org/project/npgamma.

[32] Graves YJ, Jia X, Jiang SB. Effect of statistical fluctuation in Monte Carlo based photon beam dose calculation on gamma index evaluation. Phys Med Biol 2013;58(6):1839–53.

[33] Siang KNW. Validating Monte Carlo Calculations of Clinical Proton Beams in Animal Tissue Phantoms. Rijks Universitiet Groningen; 2019.

[34] Mackin D, Li Y, Taylor MB, Kerr M, Holmes C, Sahoo N, et al. Improving spot-scanning proton therapy patient specific quality assurance with HPlusQA, a second-check dose calculation engine. Med Phys 2013;40(12):121708.

[35] Molinelli S, Mairani A, Mirandola A, Freixas GV, Tessonnier T, Giordanengo S, et al. Dosimetric accuracy assessment of a treatment plan verification system for scanned proton beam radiotherapy: one-year experimental results and Monte Carlo analysis of the involved uncertainties. Phys Med Biol 2013;58(11):3837–47.

[36] Meijers A, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk JA, et al. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: implementation and proof-of-concept. Med Phys 2019.

[37] Li H, Sahoo N, Poenisch F, Suzuki K, Li Y, Li X, et al. Use of treatment log files in spot scanning proton therapy as part of patient-specific quality assurance. Med Phys 2013;40(2):021703.