Numerical and deep learning algorithms for automated quality assurance in proton therapy

Burlacu, T.

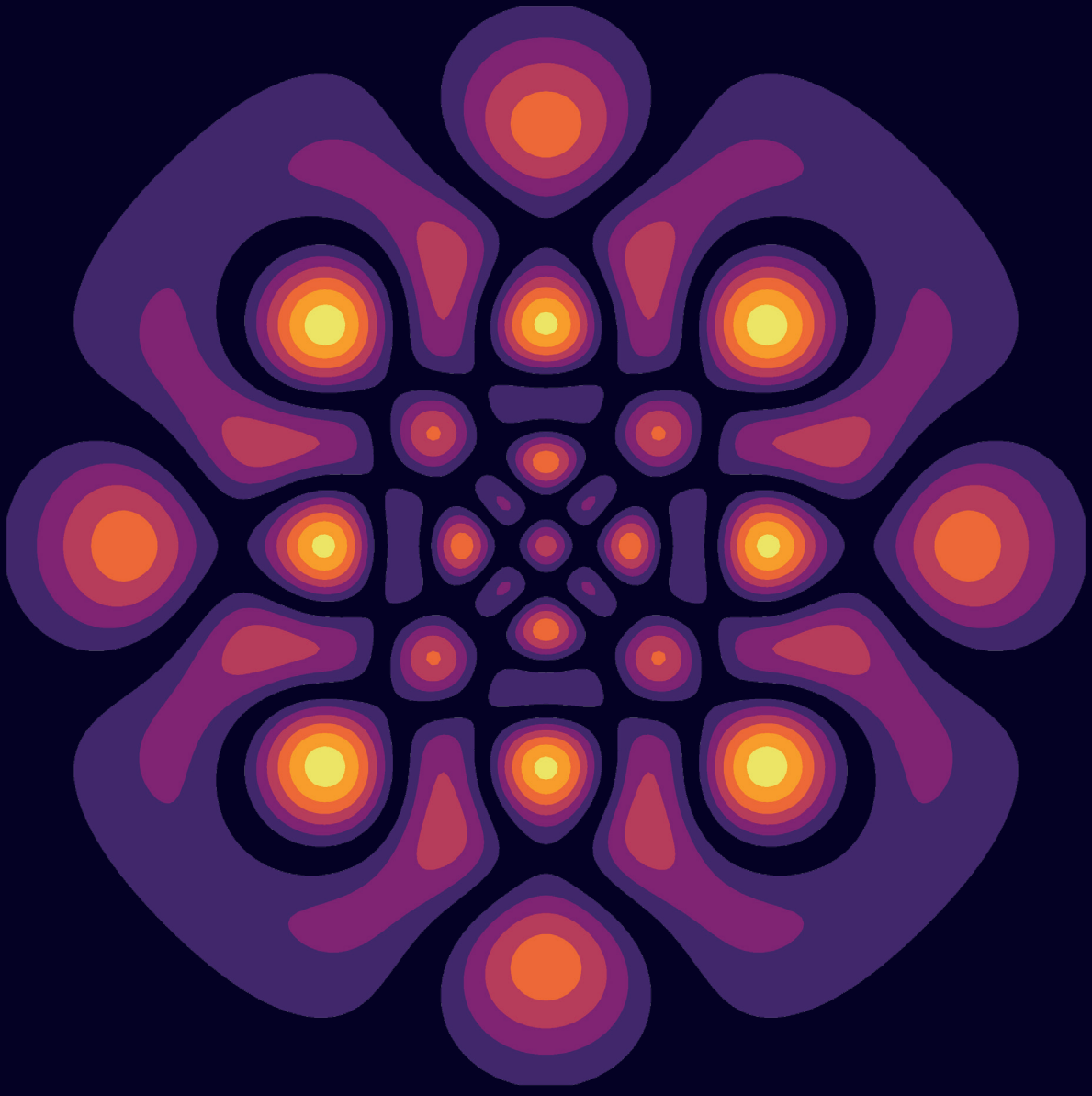
DOI
10.4233/uuid:52beee0c-8184-41c8-b61f-d074512d97cd

Publication date
2025

Document Version
Final published version

Citation (APA)
Burlacu, T. (2025). *Numerical and deep learning algorithms for automated quality assurance in proton therapy*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:52beee0c-8184-41c8-b61f-d074512d97cd

Numerical and deep learning algorithms for automated quality assurance in proton therapy



Tiberiu Burlacu

Numerical and deep learning algorithms for automated quality assurance in proton therapy

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 25 June 2025 at 17:30 o'clock

by

Tiberiu BURLACU

Master of Science in Applied Physics,
Delft University of Technology, Delft, the Netherlands
born in Dorohoi, Romania

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Dr. ir. D. Lathouwers	Delft University of Technology, NL, promotor
Dr. Z. Perkó	Delft University of Technology, NL, copromotor
Independent members	
Prof. dr. ir. M.B. van Gijzen	Delft University of Technology, NL
Prof. dr. ir. J.L. Kloosterman	Delft University of Technology, NL
Dr. D. Légrády	BME, HU
Prof. dr. ir. M. Staring	LUMC, NL
Prof .dr. ir. F.J.W. Verhaegen	Maastricht University, NL

Keywords: proton therapy, online, adaptive, numerical algorithms, deep learning, quality assurance

ISBN 978-94-6522-401-5
Printed by: Ridderprint, www.ridderprint.nl

The research described in this thesis was developed in the Medical Physics and Technology section of the Department of Radiation, Science and Technology of the Delft University of Technology, Delft, the Netherlands. The work described in this thesis has been financially supported by Varian, a Siemens Healthineers Company.

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Contents

Summary

External beam radiotherapy (EBRT) is a method for treating cancer in which the tumor is targeted by beams of radiation originating from the patient's exterior. The two main particles employed for EBRT are photons and protons, with electrons and carbon ions also being in use. Both photons and protons are capable of achieving adequate tumor coverage, but protons can theoretically achieve lower doses in the surrounding tissues (at the expense of increased economical costs). Regardless of the chosen modality, the radiotherapy (RT) workflow is similar. It consists of determining the patient anatomy via imaging, usually via computed tomography (CT) scans, contouring (delineating) the organs at risk (OARs) and the target, creating a treatment plan, performing quality assurance (QA) and delivering the plan safely. In classical (also called non-adaptive) RT this workflow is performed once and the treatment is delivered over several (around 30) daily sessions (also called fractions).

Theoretically, the best radiotherapy treatment is the one in which the tumor is completely eradicated, while the surrounding tissue is not irradiated at all. Given that this is physically impossible, due to the nature of photon and proton propagation and interaction with matter, the next best result is maximal tumor coverage and minimal radiation damage to OARs. As the patient anatomy changes on different time scales ranging from weeks (e.g., weight loss, tumor shrinkage) to days (e.g., day to day variations of cavity fillings or neck pose changes) to seconds (due to for example breathing and slight movements) it becomes apparent that the offline approach to RT is suboptimal. To improve on this, the radiotherapy workflow must be adjusted such that imaging, delineation and treatment planning are performed several times over the course of the treatment, resulting in adaptive radiotherapy (ART). ART results in better targeting of the tumor and lower OAR doses. If adaptation is performed without the patient on the treatment table, the process is called offline adaptation. The next time-scale is online, which refers to a daily adaptation regime where the patient remains online (on the treatment table) after imaging. In such a workflow, on a given day the patient is imaged and within a short time (from tens of seconds to several minutes) the complete offline workflow (contouring, treatment planning, quality assurance, safe delivery) is performed. The time between imaging and delivery should be as short as possible, in order to minimize inter-fractional and patient set-up errors and to maximize clinical output. The ideal scenario would be real-time adaptation, in which all the steps of the radiotherapy workflow (including imaging and irradiation adaptations) are performed in real-time.

The online adaptive RT workflow is challenging due to the short time available. For example, despite the progress achieved by deep learning based algorithms, OAR contouring still requires human intervention. Treatment planning is also challenging due to its high computational burden, traditionally requiring several hours to create. Another example is quality assurance, which usually refers to both plan QA and patient-specific QA. In plan QA, a physicist checks that the beams are correctly arranged, that they have the correct parameters and that the treatment planning system (TPS) correctly communicates with the beam delivery system (also called gantry). In patient-specific QA, the treatment plan is delivered to a water equivalent phantom and the delivered

dose is compared to the one from the TPS. Such a procedure is necessary despite the frequent (daily, weekly, monthly and yearly) checks of machine function (machine QA). Machine QA is a limited procedure, with only a select number of energies, spot positions and safety systems being checked. It is impossible for the machine QA programme to measure the patient-specific irradiation fields, therefore making the PSQA procedure critical. Clearly, neither manual plan QA nor the measurement-based PSQA are possible in a time-frame of tens of seconds to minutes in which the patient is seated on the treatment table.

To automate PSQA, independent dose calculations (IDCs) have been proposed. Independent dose computations should be performed not only with a different implementation of the TPS algorithm but, for maximal independence, with a methodologically different algorithm. IDCs can be employed at different points in the data pipeline that starts at the TPS and ends at the beam delivery system. The first point is employing IDCs as a software redundancy of the TPS computed dose. This would be similar to the hardware redundancy systems employed in gantries. Another point is the one where the TPS outputted treatment plan is converted into machine readable files (also called steering files). These are lower level files, that could be accessed pre-treatment and used for a-priori dose computations. Performing dose computations at this level would also indicate errors in the data transfer between the TPS and the delivery system, thereby performing part of the plan QA procedure. Lastly, the delivery system contains detectors that record for each treatment plan spot, the actually delivered position, the number of monitor units (which can be related to the number of protons) and the spot size. These records are stored in a logfile, that can be used a-posteriori to reconstruct the actually delivered dose to the patient. Performing such computations would be valuable not only for patient-specific QA purposes but also for machine QA purposes, as significant deviations would likely point to hardware failures. Ideally, logfiles would be available in real-time and used for real-time dose computations.

The current lack of computational resources and methods prohibit achieving full plan reoptimization in the online adaptive workflow. However, for plan QA automation we can reasonably assume that limited plan adaptation will soon be possible and a new - though not fully optimal - plan will be available in the online adaptive RT setting. Such a plan could be based, for example, on the previously generated plan. To assess the daily generated plan, one method could be to model the patient anatomy over the course of the treatment and construct a-priori optimal plans. Following this, a fast comparison in terms of dosimetric characteristics could be performed on the given day. Alternatively, a plan library approach could be used. In this approach, a suite of plans is generated based on the planning CT image and potentially on artificially generated ones.

This work aims to provide solutions for automating both PSQA and plan QA. For PSQA, a novel, truly TPS independent dose and dose change algorithm was developed. The algorithm, referred to as Yet anOther Dose Algorithm (YODA), is deterministic and adjoint based. Chapter 2, starts with the Linear Boltzmann Equation, that describes the proton transport in tissue. Thereafter it details the physics-based approximations that are applied to it to obtain a system of two partial differential equations (PDEs). One of the PDEs is the Fokker-Planck one-dimensional equation, and the other is the Fermi-Eyges equation. This approach has two main advantages. First, the Fokker-Planck equation is cheap to solve numerically and the Fermi-Eyges equation is analytically solvable. This combination makes YODA fast when compared to the golden standard Monte Carlo (MC) simulations and accurate when compared to empirical approximation based pencil beam algorithms (PBAs). Details about the solution methodology are provided in the Sections 2.4 and 2.5. The second main advantage of this approach, and a

unique feature when compared to other dose algorithms, is that the presence of the PDEs allows the adjoint mathematical framework to be applied. Adjoint theory is useful when approximations in a metric (e.g., the dose in a region in the patient, the tumor control probability or the normal tissue complication probability for an organ) are needed as a function of system parameters (e.g., the HU value of a voxel, or the number of MU values delivered to a spot). Adjoint theory enables YODA to avoid re-computations, making it applicable in a number of PSQA relevant scenarios. For example, in the online ART workflow, once the patient has been imaged and the organs are contoured, a decision must be made on whether or not the plan must be adapted. Plan evaluations can take up to several minutes due to having to re-compute the dose from the plan on the new anatomy. In this scenario, YODA can use the adjoint framework to approximate the dose to the new patient anatomy from yesterday's plan. Additional scenarios are possible and are explained in both Chapter 2 and 3. The chapter develops in Section 2.7 the mathematical formalism for the case in which the metric is the dose in the patient and the variable that changes is the patient anatomy (the HU values of the CT image voxels). The chapter ends in Section 2.8 by describing the data sources used for the algorithm.

Chapter 3 details the application of YODA to clinically realistic cases. The chapter starts by describing the development of a Gaussian beam splitting scheme, which is necessary as the YODA formalism is valid for in-depth heterogeneous and laterally homogeneous anatomies. Next, YODA is compared to TOPAS (the golden standard MC dose engine) in a variety of cases and for energies spanning the clinical energy spectrum (70 MeV to 230 MeV). Initial comparisons are done in a water tank, in which slabs of either bone or air are gradually slid into the beam path. This part consists of over 120 comparisons, the overview of which can be seen in Appendix C. Overall, YODA performs very well when compared to TOPAS in these synthetic tests. The lowest gamma pass rate (with 1 mm, 1 %, 10 percent cut-off criteria) was ≈95 % for a 160 MeV beam encountering an air slab that was offset by −2 mm off the beam axis. The average gamma pass rate was 96 % and the best pass rate was 100 %. YODA was also compared to TOPAS in realistic CT scans, where it again performed well. In a head and neck CT scan it achieved a pass rate of 99.85 %, in a prostate CT scan it achieved a pass rate of 99.58 % and in a lung CT scan it achieved a pass rate of 94.55 %. Given these results, and that a single spot takes hours to compute in TOPAS and 2 s to compute in YODA, it can be concluded that the dose engine of YODA achieves MC like accuracy in a fraction of the time. By further developing YODA, for example via a GPU implementation, this speed can be further improved, resulting in millisecond computation times that are competitive with other available commercial solutions.

The following suite of tests were designed to test the adjoint component by simulating an adaptation trigger system. In this system, the patient is online, having had an image acquired and the organs and target contoured. Following this, a quick assessment of the dose from yesterday's plan to today's anatomy is desired. The comparison was done between the dose based on re-computing yesterday's plan on today's CT image and the dose to today's CT image given by the adjoint computation. Two patients were tested and each patient had one robustly and one non-robustly optimized plans. For the case of the robustly optimized plans, the maximal error in the average dose to the tumor between re-computing and the adjoint computation was 4.8 % (i.e., 1.16 Gy versus 2.21 Gy), while for non-robustly optimized plans it was 5.7 % (i.e., 2.22 Gy versus 2.31 Gy). Given these results, it can be concluded that YODA's adjoint component is capable of being incorporated into an adaptation trigger system. The chapter ends with the commissioning procedure for YODA's dose engine and dose comparisons against the

MC dose engine within the TPS RayStation. YODA was commissioned based on a combination of physical measurements performed at the Holland Proton Therapy Center (HPTC) and on the RayStation MC dose engine itself (which was itself commissioned on experimental Bragg curves). Through this procedure, the boundary conditions that YODA should use to match RayStation doses are obtained. The commissioning procedure is general and can be applied for any other TPS and also for modelling beams with range shifters inserted. Comparisons were performed in simple water tanks (using simple Bragg peaks and spread out Bragg peaks) and for 4 patient plans generated in RayStation (3 of them with tumors in the head and neck region and 1 in the brain). In the single spot tests in the water tank, it was found that due to the lack of nuclear interactions in YODA, there is a consistent underdosing in the entrance region that increases with increasing beam energy. This underdosing results in greater discrepancy between YODA and RayStation when several spots are overlaid to create a spread out Bragg peak. For the patient plans, the gamma index pass rate with criteria of 2 mm, 2 %, 10 % cut-off did not drop below 90 %. Depending on the exact clinical scenario, for some situations this could be sufficient. By modelling nuclear interactions the gamma pass rate is expected to increase, therefore allowing YODA to act as a versatile IDC tool with a wide variety of applications in the online ART workflow.

Chapter 4 presents a probabilistic deep learning algorithm for predicting head and neck patient's anatomies throughout the treatment course. The probabilistic model approximates a joint probability distribution of the repeat CT image (rCT) and associated contours (rM) conditioned on the planning CT image (pCT) and its associated contours (pM). It takes as input a pair of pCT and pM and can be sampled to generate deformation vector fields (DVFs), which are in turn used to warp the inputs into rCTs and rMs. The performance of the model was assessed based on its capability to reconstruct the test set and to generate realistic CT images. In terms of test set reconstruction accuracy, the model achieved an overlap between generated and ground truth contours (DICE score) of 0.83 (the DICE score has values between 0 and 1) and an image similarity score (with values between 0 and 1) between generated and ground truth images of 0.60. In terms of generative performance, the volume loss in the left and right parotid glands, the spinal cord and the constrictor muscle and their center of mass shifts were assessed. By comparing the anatomical change distributions on the training, test and generated sets, it was concluded that the model produces volume change and COM shift ranges that are broad enough to encompass the training and test set ones, with means and medians in reasonable agreement. Moreover, the model is capable of generating realistic CT images, showing neck pose shifts and the expected flattening and medial shifts of the parotid glands. Given these findings, the model could be incorporated into a plan QA tool by having for each of the generated CT images a truly optimal plan stored in a database. This could be used on the given day, when the patient is online, to quickly compare and assure the quality of the adapted and refined plan.

Chapter 5 summarizes the main developments of this work, its findings and suggestions for future research.

Sammenvatting

Bestraling met externe straling (EBRT) is een methode voor kanker behandeling waarbij de tumor wordt bestraald met stralenbundels afkomstig van de buitenkant van de patiënt. De twee belangrijkste deeltjes die voor EBRT worden gebruikt zijn fotonen en protonen, maar er worden ook elektronen en koolstofionen gebruikt. Zowel fotonen als protonen kunnen een adequate tumorbedekking bereiken, maar protonen kunnen theoretisch lagere doses in de omliggende weefsels bereiken (ten koste van hogere economische kosten). Ongeacht de gekozen modaliteit is de workflow voor radiotherapie (RT) vergelijkbaar. Deze bestaat uit het bepalen van de anatomie van de patiënt via beeldvorming, meestal via computertomografie (CT), het contouren (afbakenen) van de risicoorganen (OAR's) en het doelwit, het maken van een behandelplan, het uitvoeren van kwaliteitsborging (QA) en het veilig afleveren van het plan. Bij klassieke (ook wel niet-adaptieve) RT wordt deze workflow eenmalig uitgevoerd en wordt de behandeling verdeeld over meerdere (ongeveer 30) dagelijkse sessies (ook wel fracties genoemd).

Theoretisch is de beste radiotherapiebehandeling de behandeling waarbij de tumor volledig wordt uitgeroeid, terwijl het omringende weefsel helemaal niet wordt bestraald. Aangezien dit fysiek onmogelijk is door de aard van de voortplanting van fotonen en protonen en de interactie met materie, is het volgende beste resultaat een maximale tumorbedekking en minimale stralingsschade aan de OARs. Aangezien de anatomie van de patiënt verandert op verschillende tijdschalen, variërend van weken (bijv. gewichtsverlies, krimpen van de tumor) tot dagen (bijv. dagelijkse variaties van vullingen in de holte of veranderingen in de houding van de nek) tot seconden (door bijvoorbeeld ademhaling en lichte bewegingen), wordt het duidelijk dat de offline benadering van RT suboptimaal is. Om dit te verbeteren moet de radiotherapieworkflow zodanig worden aangepast dat beeldvorming, afbakening en behandelplanning meerdere keren worden uitgevoerd in de loop van de behandeling, wat resulteert in adaptieve radiotherapie (ART). ART resulteert in een betere targeting van de tumor en lagere OAR-doses. Als de adaptatie wordt uitgevoerd zonder de patiënt op de behandeltafel, wordt het proces offline adaptatie genoemd. De volgende tijdschaal is online, wat verwijst naar een dagelijks adaptatieregime waarbij de patiënt online blijft (op de behandeltafel) na beeldvorming. In een dergelijke workflow wordt de patiënt op een bepaalde dag in beeld gebracht en wordt binnen een korte tijd (van tientallen seconden tot enkele minuten) de volledige offline workflow (contourvorming, behandelplanning, kwaliteitsborging, veilige aflevering) uitgevoerd. De tijd tussen beeldvorming en bestraling moet zo kort mogelijk zijn om interfractionele fouten en fouten bij de opstelling van de patiënt te minimaliseren en de klinische output te maximaliseren. Het ideale scenario is realtime aanpassing, waarbij alle stappen van de radiotherapieworkflow (inclusief beeldvorming en bestralingsaanpassingen) in realtime worden uitgevoerd.

De online adaptieve RT-workflow is een uitdaging vanwege de korte beschikbare tijd. Ondanks de vooruitgang die is geboekt met algoritmen op basis van deep learning, is er bijvoorbeeld nog steeds menselijke tussenkomst nodig om OAR-contouren te maken. De planning van de behandeling is ook een uitdaging vanwege de hoge computerbelasting, die traditioneel enkele uren in beslag neemt. Een ander voorbeeld is de kwaliteitsborging, die meestal verwijst naar zowel plan QA als patiënt-specifieke QA. Bij plan

QA controleert een fysicus of de bundels correct zijn opgesteld, of ze de juiste parameters hebben en of het behandelplanningssysteem (TPS) correct communiceert met het bundelafgiftesysteem (ook wel gantry genoemd). Bij patiëntspecifieke QA wordt het behandelplan afgegeven aan een waterequivalent fantoom en wordt de afgegeven dosis vergeleken met die van het TPS. Een dergelijke procedure is noodzakelijk ondanks de frequente (dagelijkse, wekelijkse, maandelijkse en jaarlijkse) controles van de machinefunctie (machine QA). Machine QA is een beperkte procedure, waarbij slechts een select aantal energieën, spotposities en veiligheidssystemen worden gecontroleerd. Het is onmogelijk voor het machine QA programma om de patiëntspecifieke bestralingsvelden te meten, waardoor de PSQA procedure cruciaal is. Het is duidelijk dat noch handmatige plan QA noch de op metingen gebaseerde PSQA mogelijk zijn in een tijdsbestek van tientallen seconden tot minuten waarin de patiënt op de behandeltafel zit.

PSQA kan worden geautomatiseerd met onafhankelijke dosisberekeningen (IDC's). Onafhankelijke dosisberekeningen moeten niet alleen worden uitgevoerd met een andere implementatie van het TPS-algoritme, maar voor maximale onafhankelijkheid ook met een methodologisch verschillend algoritme. IDC's kunnen worden gebruikt op verschillende punten in de gegevenspijplijn die begint bij de TPS en eindigt bij het bundeltoedieningssysteem. Het eerste punt is het toepassen van IDC's als softwareredundantie van de door de TPS berekende dosis. Dit zou vergelijkbaar zijn met de hardwareredundantiesystemen die in portalen worden gebruikt. Een ander punt is het punt waarbij het door TPS uitgevoerde behandelplan wordt omgezet in machineleesbare bestanden (ook wel stuurbestanden genoemd). Dit zijn bestanden van een lager niveau, die vóór de behandeling kunnen worden geopend en gebruikt voor dosisberekeningen a-priori. Het uitvoeren van dosisberekeningen op dit niveau zou ook fouten aangeven in de gegevensoverdracht tussen het TPS en het toedieningssysteem, waardoor een deel van de QA procedure van het plan wordt uitgevoerd. Tot slot bevat het afgiftesysteem detectoren die voor elke spot van het behandelplan de daadwerkelijk afgeleverde positie, het aantal monitoreenheden (dat kan worden gerelateerd aan het aantal protonen) en de spotgrootte registreren. Deze gegevens worden opgeslagen in een logbestand dat achteraf kan worden gebruikt om de werkelijk toegediende dosis aan de patiënt te reconstrueren. Het uitvoeren van dergelijke berekeningen zou niet alleen waardevol zijn voor patiënt-specifieke QA-doeleinden, maar ook voor machine QA-doeleinden, aangezien significante afwijkingen waarschijnlijk zouden wijzen op hardwarefouten. Idealiter zouden logbestanden in real-time beschikbaar zijn en gebruikt worden voor real-time dosisberekeningen.

Door het huidige gebrek aan computermiddelen en methoden is volledige heroptimalisatie van het plan in de online adaptieve workflow niet mogelijk. Voor plan QA automatisering kunnen we echter redelijkerwijs aannemen dat beperkte planaanpassing binnenkort mogelijk zal zijn en dat een nieuw - hoewel niet volledig optimaal - plan beschikbaar zal zijn in de online adaptieve RT setting. Zo'n plan zou bijvoorbeeld gebaseerd kunnen zijn op het eerder gegenereerde plan. Om het dagelijks gegenereerde plan te beoordelen, zou een methode kunnen zijn om de anatomie van de patiënt in de loop van de behandeling te modelleren en a-priori optimale plannen te construeren. Vervolgens kan op de gegeven dag een snelle vergelijking worden uitgevoerd in termen van dosimetrische kenmerken. Als alternatief kan een planbibliotheek worden gebruikt. In deze aanpak wordt een reeks plannen gegenereerd op basis van het CT-beeld van de planning en mogelijk op kunstmatig gegenereerde plannen.

Dit werk is gericht op het bieden van oplossingen voor het automatiseren van zowel PSQA als plan QA. Voor PSQA werd een nieuw, werkelijk TPS-onafhankelijk algoritme voor dosering en dosisverandering ontwikkeld. Het algoritme, dat Yet anOther Dose Algorithm (YODA) wordt genoemd, is deterministisch en gebaseerd op adjoint

theorie. Hoofdstuk 2 begint met de lineaire vergelijking van Boltzmann, die het protonentransport in weefsel beschrijft. Daarna worden de op fysica gebaseerde benaderingen beschreven die worden toegepast om een stelsel van twee partiële differentiaalvergelijkingen (PDEs) te verkrijgen. Een van de PDEs is de eendimensionale vergelijking van Fokker-Planck en de andere is de vergelijking van Fermi-Eyges. Deze benadering heeft twee belangrijke voordelen. Ten eerste is de Fokker-Planck vergelijking goedkoop om numeriek op te lossen en de Fermi-Eyges vergelijking is analytisch op te lossen. Deze combinatie maakt YODA snel in vergelijking met de gouden standaard Monte Carlo (MC) simulaties en nauwkeurig in vergelijking met empirische benaderingen op basis van pencil beam algoritmen (PBAs). Details over de oplossingsmethodologie worden gegeven in de paragrafen 2.4 en 2.5. Het tweede grote voordeel van deze aanpak, en een uniek kenmerk in vergelijking met andere doseringsalgoritmen, is dat de aanwezigheid van de PDEs het mogelijk maakt om het adjunct-wiskundige kader toe te passen. De adjoint-theorie is nuttig wanneer benaderingen in een metriek (bijv. de dosis in een gebied in de patiënt, de waarschijnlijkheid van tumorcontrole of de waarschijnlijkheid van complicaties in normaal weefsel voor een orgaan) nodig zijn als functie van systeemparameters (bijv. de HU-waarde van een voxel of het aantal MU-waarden dat aan een plek is toegediend). Dankzij de adjoint-theorie kan YODA herberekeningen vermijden, waardoor het toepasbaar is in een aantal PSQA-relevante scenario's. In de online ART-workflow bijvoorbeeld moet, zodra de patiënt in beeld is gebracht en de organen zijn gecontourd, worden besloten of het plan al dan niet moet worden aangepast. Planevaluaties kunnen enkele minuten duren omdat de dosis van het plan opnieuw moet worden berekend op de nieuwe anatomie. In dit scenario kan YODA het adjunct-kader gebruiken om de dosis voor de nieuwe anatomie van de patiënt te benaderen op basis van het plan van gisteren. Andere scenario's zijn mogelijk en worden uitgelegd in zowel hoofdstuk 2 als hoofdstuk 3. Het hoofdstuk ontwikkelt in paragraaf 2.7 het wiskundige formalisme voor het geval waarin de metriek de dosis in de patiënt is en de variabele die verandert de anatomie van de patiënt is (de HU-waarden van de voxels van het CT-beeld). Het hoofdstuk eindigt in paragraaf 2.8 met een beschrijving van de gegevensbronnen die voor het algoritme worden gebruikt.

Hoofdstuk 3 beschrijft de toepassing van YODA op klinisch realistische gevallen. Het hoofdstuk begint met het beschrijven van de ontwikkeling van een Gaussisch bundelsplitsingsschema, dat nodig is omdat het YODA-formalisme geldig is voor diep heterogene en lateraal homogene anatomieën. Vervolgens wordt YODA vergeleken met TOPAS (de gouden standaard voor MC-doseringen) in verschillende gevallen en voor energieën uit het klinische energiespectrum (70 MeV to 230 MeV). Initiële vergelijkingen worden gedaan in een watertank, waarin platen bot of lucht geleidelijk in het bundelpad worden geschoven. Dit deel bestaat uit meer dan 120 vergelijkingen, waarvan een overzicht te zien is in Appendix C. Over het algemeen presteert YODA zeer goed in vergelijking met TOPAS in deze synthetische tests. Het laagste gammapassagepercentage (met afkapcriteria van 1 mm, 1 % en 10 %) was ongeveer 95 % voor een bundel van 160 MeV die in aanraking kwam met een luchtplaat die een afwijking van -2 mm had ten opzichte van de bundelas. Het gemiddelde gammapassagetarief was ≥ 96 % en het beste passagetarief was 100 %. YODA werd ook vergeleken met TOPAS in realistische CT-scans, waarbij het opnieuw goed presteerde. Bij een CT-scan van hoofd en nek werd een slagingspercentage van 99.85 % gehaald, bij een CT-scan van de prostaat een slagingspercentage van 99.58 % en bij een CT-scan van de longen een slagingspercentage van 94.55 %. Gezien deze resultaten en het feit dat het uren duurt om een enkele spot te berekenen in TOPAS en 2 s om te berekenen in YODA, kan worden geconcludeerd dat de doseringsengine van YODA een MC-achtige nauwkeurigheid bereikt in een fractie van de tijd. Door YODA verder te ontwikkelen, bijvoorbeeld via een GPU-implementatie,

kan deze snelheid verder worden verbeterd, wat resulteert in rekentijden in milliseconden die concurrerend zijn met andere beschikbare commerciële oplossingen.

De volgende reeks tests is ontworpen om de adjoint component te testen door een adaptatietriggersysteem te simuleren. In dit systeem is de patiënt online, is er een beeld gemaakt en zijn de organen en het doel omlijnd. Hierna is een snelle beoordeling van de dosis van het plan van gisteren naar de anatomie van vandaag gewenst. Er werd een vergelijking gemaakt tussen de dosis op basis van het opnieuw berekenen van het plan van gisteren op het CT-beeld van vandaag en de dosis op het CT-beeld van vandaag op basis van de adjunctberekening. Twee patiënten werden getest en elke patiënt had een robuust en een niet-robuust geoptimaliseerd plan. Voor de robuust geoptimaliseerde plannen was de maximale fout in de gemiddelde dosis voor de tumor tussen opnieuw berekenen en de adjoint berekening 4.8 % (d.w.z. 1.16 Gy versus 2.21 Gy), terwijl dit voor niet-robuust geoptimaliseerde plannen 5.7 % was (d.w.z. 2.2 Gy versus 2.31 Gy). Gezien deze resultaten kan worden geconcludeerd dat de adjoint component van YODA kan worden opgenomen in een adaptatietriggersysteem. Het hoofdstuk eindigt met de inbedrijfstellingsprocedure van de YODA-doseringsmotor en dosisvergelijkingen met de MC-doseringsmotor binnen het TPS RayStation. De inbedrijfstelling van YODA is gebaseerd op een combinatie van fysische metingen die zijn uitgevoerd in het Holland Proton Therapy Center (HPTC) en op de MC-doseringsmotor van het RayStation zelf (die zelf in bedrijf is gesteld op basis van experimentele Bragg-curven). Door deze procedure worden de randvoorwaarden verkregen die YODA moet gebruiken om de RayStation-doses te evenaren. De inbedrijfstellingsprocedure is algemeen en kan worden toegepast op elke andere TPS en ook voor het modelleren van bundels met range shifters. Er werden vergelijkingen uitgevoerd in eenvoudige watertanks (met eenvoudige Bragg-pieken en uitgespreide Bragg-pieken) en voor 4 patiëntplannen gegenereerd in RayStation (waarvan 3 met tumoren in het hoofd-halsgebied en 1 in de hersenen). Bij de single spot tests in de watertank bleek dat door het ontbreken van nucleaire interacties in YODA er een consistente onderdosering is in het ingangsgebied die toeneemt met toenemende stralingsenergie. Deze onderdosering resulteert in een grotere discrepantie tussen YODA en RayStation wanneer meerdere spots over elkaar worden gelegd om een gespreide Bragg-piek te creëren. Voor de patiëntplannen daalde het slagingspercentage van de gamma-index met criteria van 2 mm, 2 %, 10 % cut-off niet onder 90 %. Afhankelijk van het precieze klinische scenario kan dit voor sommige situaties voldoende zijn. Door de modellering van nucleaire interacties wordt verwacht dat de gammapassmentsnelheid zal toenemen, waardoor YODA kan fungeren als een veelzijdig IDC-hulpmiddel met een breed scala aan toepassingen in de online ART-workflow.

Hoofdstuk 4 presenteert een probabilistisch deep learning-algoritme voor het voorspellen van de anatomie van hoofd- en halspatiënten gedurende de behandelingskuur. Het probabilistische model benadert een gezamenlijke waarschijnlijkheidsverdeling van de herhaalde CT-afbeelding (rCT) en bijbehorende contouren (rM) geconditioneerd op de plannings-CT-afbeelding (pCT) en de bijbehorende contouren (pM). Het neemt als invoer een paar pCT en pM en kan worden bemonsterd om vervormingsvectorvelden (DVF's) te genereren, die op hun beurt worden gebruikt om de invoer te vervormen tot rCT's en rM's. De prestaties van het model werden beoordeeld op basis van het vermogen om de testset te reconstrueren en realistische CT-afbeeldingen te genereren. In termen van nauwkeurigheid van de reconstructie van de testset bereikte het model een overlap tussen gegenereerde en grondwaarheidscontouren (DICE-score) van 0.83 (de DICE-score heeft waarden tussen 0 en 1) en een beeldgelijkenisscore (met waarden tussen 0 en 1) tussen gegenereerde en grondwaarheidsbeelden van 0.60. In termen van generatieve prestaties werden het volumeverlies in de linker- en rechterparotisklieren, het ruggenmerg en de constrictorspier en hun zwaartepuntverschuivingen beoordeeld.

Door de anatomische veranderingsverdelingen op de trainings-, test- en gegenereerde sets te vergelijken, werd geconcludeerd dat het model volumeverandering en COM-verschuivingsbereiken produceert die breed genoeg zijn om de trainings- en testsets te omvatten, met gemiddelden en medianen in redelijke overeenstemming. Bovendien is het model in staat om realistische CT-beelden te genereren, die nekhoudingverschuivingen en de verwachte afvlakking en mediale verschuivingen van de parotisklieren laten zien. Gezien deze bevindingen zou het model kunnen worden opgenomen in een plan QA-tool door voor elk van de gegenereerde CT-beelden een echt optimaal plan op te slaan in een database. Dit zou op de gegeven dag, wanneer de patiënt online is, kunnen worden gebruikt om snel de kwaliteit van het aangepaste en verfijnde plan te vergelijken en te verzekeren.

Hoofdstuk 5 vat de belangrijkste ontwikkelingen van dit werk, de bevindingen en suggesties voor toekomstig onderzoek samen.

Chapter 1

Introduction

1.1 A brief overview of radiotherapy history & techniques

Cancer is an umbrella term for a collection of diseases and can be characterized by self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, as well as tissue invasion and metastasis (Hanahan and Weinberg, 2000). Cancer is a significant public health concern, which accounted for over 1.1 million deaths, or 21.6 % of all deaths, in the European Union (EU) in 2021 (*Cancer Statistics* 2024). The number of cancer diagnoses has increased in 14 of the 24 reporting EU countries in the period between 2010 and 2022 and is foreseen to continue increasing over the next decade (Praagman et al., 2022). The increase in diagnoses is caused by multiple factors, the main two being the growth and the double aging of populations (there will be more elderly people that are, on average, getting older). Additional factors are changes in lifestyle and improved diagnostics and early detection (Praagman et al., 2022). For example, cancer types related to obesity, such as liver, bile duct and kidney cancer, are foreseen to notably increase over the next decade in The Netherlands (Praagman et al., 2022). Despite the number of cancer diagnoses increasing during the period between 2010 and 2022, in the same period the mortality has declined by over 10 % in all 27 EU countries. The trend of cancer mortality falling is a recent one, having only started as late as the 1990s (Hajdu, Vadmal, and Tang, 2015). While it is difficult to pinpoint the cause of this recent trend, improvements in awareness, screening programmes, diagnostic and treatment methodologies are likely explanations (*Cancer Statistics* 2024; Praagman et al., 2022).

Much effort has been dedicated towards the understanding of the causes of cancer and the (further) development of treatment modalities. The main treatment modalities for cancer are surgery, chemotherapy and radiation therapy. Surgery is one of the oldest treatment modalities, being employed by the Egyptians as early as 1500 BC (Hajdu, 2011a). Throughout time, alternative treatments such as different salts, boiled cabbage, solutions or pastes containing arsenic, iron or copper and bacterial toxin treatments (Hajdu, 2011b; Hajdu, 2012) have also been prescribed to cancer patients, with varying degrees of effectiveness. Despite this broad range of alternatives, the continuous progress in surgical techniques has established surgery as the main modality of treating cancer up until the second half of the 19th century.

Wilhelm Röntgen's accidental discovery of x-rays in 1895 (Röntgen, 1898) and their immediate application in treating inoperable breast cancer marked the beginning of radiation therapy (Hajdu, 2012, p. 4). Two other important developments for radiotherapy followed shortly after this. One is Becquerel's discovery of the natural radioactivity of a uranium sample. The other is Curies' chemical isolation of the radioactive elements of polonium and radium (Becquerel and Curie, 1901; Gianfaldoni et al., 2017). The discoveries of Röntgen, Becquerel and the Curies paved the way for the production of devices

meant to generate energetic beams for diagnostic and therapeutic purposes (Elaimy et al., 2021). At the turn of the century, such devices were capable of only treating skin cancer due to the low energy of the beams (Gianfaldoni et al., 2017). Next to the diagnostic and therapeutic purposes of radiation, an additional use of radiation was recognized by Marie Curie, namely the palliative one. She personally trained women to operate mobile x-ray vehicles and radiographed and assisted wounded soldiers on the frontlines in World War I (Curie, 1921; Reed, 2011; Loap, Huynh, and Kirova, 2021).

The 1920s brought several improvements to radiotherapy (Gianfaldoni et al., 2017), with one of the most important being the relationship between dose and cell survival. Coutard's "protracted-fractional method" showed that the administration of doses in fractions as opposed to singular treatment sessions is beneficial for tumor control and fewer side effects, a method that is still in use to this day (Coutard, 1934). The period between the 1930s and 1950s saw improvements focused on the treatment of deeper tumors via the use of brachytherapy and the supervoltage x-ray tubes (Gianfaldoni et al., 2017).

The significant advances during the 1950s, such as the introduction of megavoltage therapy, Co-60 therapy machines (Huh and Kim, 2020) and Wilson's proposal to use the proton beam (Wilson, 1946) mark the start of modern external beam radiotherapy (Hajdu and Vadmal, 2013). Wilson proposed the therapeutic use of proton beams due to the advantages their dose distribution has over conventional photon beams (Wilson, 1946). An overview of the characteristics of proton and photon beams can be seen in Figure 1.1. When compared to a photon dose-depth curve, the proton one shows simultaneously lower doses achievable in organs at risk (OARs) and an increased target dose conformality due to the presence of the Bragg peak (BP) (Paganetti, 2018).

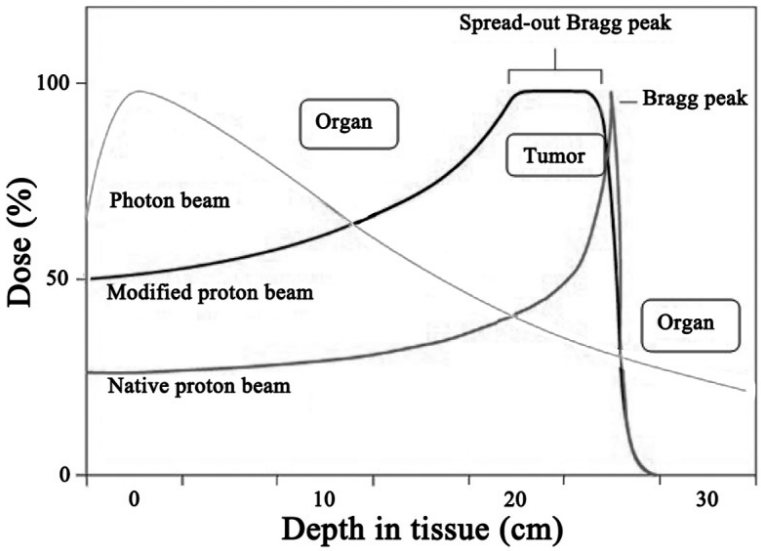


FIGURE 1.1: Overview of the dosimetric characteristics of photons and protons. The figure displays dose-depth curves for a photon beam, a native proton beam and a modified proton beam for different zones in the patient (organ and tumor). The figure was taken from (Tian et al., 2018).

By the 1960s, both conventional photon-based radiation therapy (RT) and proton therapy (PT) are established as capable of sparing the skin and tissues around the tumor and RT, in particular, as an useful adjuvant after surgery or chemotherapy for many cancers (Hajdu and Vadmal, 2013; Elaimy et al., 2021). Despite the theoretical

dosimetric advantages of protons, technical difficulties and large costs kept PT mainly as the subject of research laboratories (Elaimy et al., 2021) while photon based radiation therapy was the main modality in hospitals. This situation only changed in the 1990s, when the first PT center based in a hospital is established at the Loma Linda University Medical Center (Elaimy et al., 2021).

During the 1960s and the 1970s, RT was revolutionized by the inclusion of radiation images in the mostly two-dimensional treatment planning of the time (Webb and Evans, 2006). Specifically, single photon emission computed tomography and positron emission tomography were introduced in the 1960s and computed tomography (CT), which is currently the standard in treatment planning, was introduced during the 1970s (Huh and Kim, 2020). CT images use multiple projection images from different angles to construct a voxelized three-dimensional patient anatomy, with units of Hounsfield units (HU), that can be used for cross-sectional views (Taubmann et al., 2018). The introduction of CT images dramatically improved the accuracy of radiotherapy, by allowing physicians to define both tumor and organs at risk on the cross-sectional slices (Taylor and Powell, 2004).

During the 1980s, personal computers became fast enough to perform simple or heavily simplified treatment planning calculations in almost real time (Webb and Evans, 2006). An exception to this are Monte-Carlo (MC) calculations, which are currently considered the golden standard in dose computations and in their most accurate form (e.g. TOPAS (Perl et al., 2012)) are computationally expensive. Photon treatments were performed using either square block apertures (called standard therapy) or multi leaf collimators (MLCs) shaped to the target (called conformal therapy). In conformal therapy, treatment planning relied on the skills of the planner to decide the number, shape and orientation of the beams (Taylor and Powell, 2004), a process called forward planning. At the same time, during the 1980s Brahme (Brahme, Roos, and Lax, 1982) proposed intensity-modulated radiation therapy (IMRT), which rapidly developed in the 1990s and became widely available in the 2000s (Huh and Kim, 2020; Webb and Evans, 2006). IMRT, as opposed to conformal therapy, has radiation beams with non-uniform intensities and employs computerised inverse planning (Taylor and Powell, 2004). By dividing each beam into several beamlets, with each beamlet being allowed to have a different intensity from the others, complex dose patterns can be created. In inverse planning, as opposed to forward planning, the plan outcome is specified (in terms of tumor dose and organs at risk limits) (Taylor and Powell, 2004) together with the beam angles and an automated optimization procedure determines the appropriate beam intensities. Since its proposal, IMRT has established itself as the accepted standard for clinical practice in several sites such as prostate, liver, breast, lung and head and neck (Huh, Park, and Choi, 2019), due to improved target conformity, increased normal tissue sparing and the enablement of dose escalation (Taylor and Powell, 2004).

1.2 Intensity modulated proton therapy

The ideas developed in IMRT have also been translated to PT, in the form intensity modulated proton therapy (IMPT) (Moreno et al., 2019), which is expected to become the standard mode of proton therapy (Kooy and Grassberger, 2015; Mohan and Grosshans, 2017). Due to the differences in dose deposition and scattering characteristics of photons and protons, the techniques and algorithms used for IMRT planning, optimization and delivery are not directly applicable to IMPT. This section provides an overview of the technology and treatment planning techniques implemented in IMPT, and their respective challenges.

1.2.1 Technology

Prior to the development of IMPT, the lateral and depth extents of proton radiation fields around the three-dimensional volumes identified on the CT scan were mechanically controlled using scatterers, apertures and range compensators (Kooy and Grassberger, 2015). Once electronic controls were implemented, the well-collimated "pencil" beam that comes from the accelerator could be magnetically deflected laterally to the central beam axis. This delivery process is called pencil beam scanning (PBS). The treatment delivery system (TDS) that facilitates pencil beam scanning, can be subdivided into three components: the beam production system (BPS), the beam delivery system (BDS) and the patient monitoring system (PMS).

The BPS consists of the proton accelerator and the beam transport system and it produces a proton beam of certain requested properties (energy, intensity and spot size) at the entrance of the BDS. There are multiple choices possible in terms of proton accelerators, but the most common are the cyclotron and the synchrotron. A thorough overview of the differences between these two technologies can be found in the work of (Xiao et al., 2024). One important difference with regards to treatment delivery is that a cyclotron produces a beam of a fixed energy (varied downstream in the BPS via a degrader and an energy selection system) while a synchrotron produces beams of variable energies. Regardless of the chosen technology, both cyclotrons and synchrotrons have limitations in terms of the time required to switch energies (in the order of 1 s to 2 s) (Kang, Wilkens, and Oelfke, 2008). In the case of the cyclotron mechanical intervention is necessary, while in the case of the synchrotron a new proton bunch must be accelerated. The time for switching energies can add up to a significant fraction of the total treatment time for plans with many energy layers (as can be the case for IMPT) (Van de Water et al., 2015; Kang, Wilkens, and Oelfke, 2008), resulting in increased uncertainties and higher costs.

The BDS receives the treatment instructions (i.e., the plan) generated by the treatment planning system (TPS), converts it into settings for the different delivery components and monitors the safe delivery of the plan. An example of the hardware present in a BDS can be seen in Figure 1.2. When the pencil beam enters the BDS, the profile monitor records the center of the incident beam and the beam profile (spot size in the x and y directions) (Li et al., 2013). Next, the beam is deflected by two scanning magnets (x and y) using the settings (current values) derived from the desired lateral position of the spot in the treatment plan. Following this, the center location of the delivered spot is recorded by the spot position monitor and the number of protons (or equivalently the number of Monitor Units (MU)) is recorded by two dose monitors. In between the scanning magnets and the spot position monitor, a helium chamber is placed to reduce the lateral dispersion of the proton beam (Mohan and Grosshans, 2017).

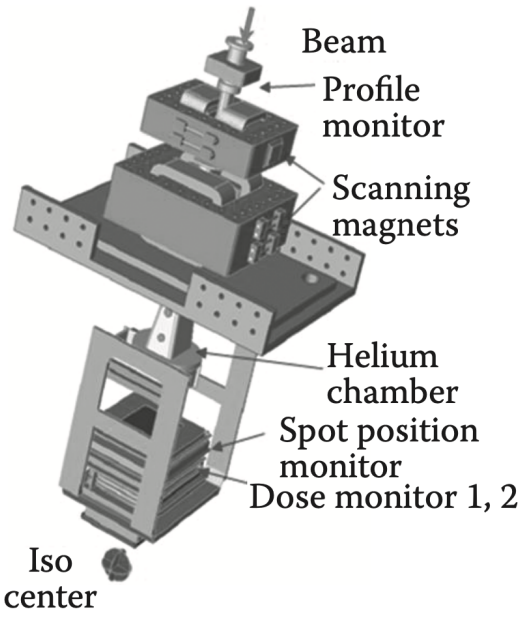


FIGURE 1.2: Example of hardware present in a beam delivery system. The figure displays a gantry, that contains, from top to bottom, a beam profile monitor, scanning magnets, a spot position monitor and two dose monitors before exiting towards the isocenter. Figure taken from (Paganetti, 2018)

1.2.2 Quality assurance

Quality assurance (QA) is a vital part of the safe operation of the BDS. Usually, QA can be divided into machine QA and patient-specific QA (PSQA). Machine QA is meant to ensure the correct function of the BDS for all treated patients. Patient-specific QA is meant to ensure that the planned and delivered dose for a specific treatment plan are within the accepted tolerances.

Machine QA consists of procedures that are performed daily, weekly, monthly and yearly (Poenisch et al., 2021). Daily QA verifies the dosimetric characteristic of the proton pencil beam, the proper functioning of the dose monitor chambers, the proper functioning of the spot position monitor, the proper functioning of the imaging system used for patient set-up and the correct functioning of the safety components of the treatment machine before treatment. The dosimetric checks consist of range and spot positioning checks. For example, at the Proton Therapy Center University of Texas MD Anderson Cancer Center both checks have an allowed tolerance of 1 mm. Due to the large number of energies that the BPS can produce, only a select number of them are tested. Similarly, only select spot positions (on the central axis and at symmetric offsets from it in both x and y directions) are tested. Additional dosimetric checks are volumetric dose constancy, with an allowed tolerance of 3 % and spot size constancy, with an allowed tolerance of 1 mm or 10 %. Safety interlock checks performed daily are the in-room beam stops, the facility beam stop, the beam delivery indicator light, the radiation monitors (inside and outside), audiovisual monitoring, the room clearance button and the room motion sensor. During weekly QA, the measurements taken during daily QA are reviewed by a certified medical physicist and the isocentricity of the couch rotation is verified. Monthly QA has the same goals as daily QA, namely the verification of the dosimetric accuracy, of the alignment of the the x-ray imaging system and of the safety systems, with a main difference being that it is performed by a certified medical physicist using different detector systems.

Patient-specific QA can be divided into two components. The first stage of PSQA involves a physicist that checks the plan to ensure that the beam arrangements and parameters are appropriate. The physicist also tests the correct transmission of information from the treatment planning system to the gantry (the structure that contains the BDS) (Poenisch et al., 2021). The flow of information is prone to error due to the conversion steps it entails. The fluence map that results from the plan optimization in the TPS is converted into machine readable files, which thereafter must be correctly interpreted and delivered by the machine (Meier et al., 2015). The second stage of PSQA involves the delivery of the treatment plan to a water-equivalent phantom. Here the differences between the measured and planned doses are compared against the allowed clinical limits. For example, the allowed clinical limits at the Proton Therapy Center University of Texas MD Anderson Cancer Center are the gamma pass rate being higher than 90 % using criteria of 3 mm and 3 % (Poenisch et al., 2021; Gottschalk, 2012). Such measurements are necessary, as the machine QA process is limited in scope. As already mentioned, only select energies and spot positions are tested, with the main goal being to verify the correct operation of the safety interlocks of the system (Poenisch et al., 2021). PSQA also acts as a form of machine QA, as failure to meet the necessary tolerances most likely indicate an issue in the BDS (Poenisch et al., 2021; Frank and Zhu, 2020). Given this, and the patient-specific nature of the treatment fields, measurements form an integral part of the safe operation of the delivery system.

1.2.3 Treatment planning

Both IMPT and IMRT make use of inverse planning, a process in which the prescription of the physician is converted into a treatment plan. The treatment plan is in turn converted into machine settings meant to produce a delivered dose distribution that is equivalent to the planned one.

The first step in treatment planning is the imaging of the patient and the delineation (also called contouring) of the relevant volumes of interest (VOIs). The most common imaging procedure is CT imaging which, as explained in Section 1.1, produces a voxelized three-dimensional patient anatomy on which the voxels are assigned to VOIs. The visible part of the tumor is identified as the gross tumor volume (GTV), while the microscopic visible and non-visible parts of it are identified as the clinical target volume (CTV), a region that surrounds the GTV. Organs to which dose should be limited are identified as organs at risk (OARs). In IMPT, the dose is delivered by pencil beams that are magnetically directed towards a location (spot). The dose from an individual pencil beam with index j to a voxel with index i is written as

$$d_i = \sum_j x_j D_{ij},$$

where D_{ij} is the dose influence matrix and x_j is the weight (i.e., the number of protons) of the pencil beam with index j.

The prescription of the physician consists of a set of constraints (e.g., minimum dose to the target of 70 Gy) and intents or objectives (e.g., minimize dose to the parotid glands). Mathematically, the objectives and their priorities are expressed in objective functions with corresponding weights. For example, an objective function that aims at minimizing the volume within a given OAR with index n that exceeds a maximum allowed dose $\mathrm{D}^{\max}$ can be written as

$$O_n(d) = \sum_{i \in OAR_n} H(d_i - D^{\max}) \, (d_i - D^{\max})^2 \,,$$

where $H(d)$ is the Heaviside step function. Constraints for a given VOI are mathematically specified as bounds, namely

$$D^{\min} \leq d_i \leq D^{\max}, \forall i \in \text{VOI}.$$

Thus, IMPT optimization seeks to find the weights of the individual pencil beams x_j (also called intensity or fluence maps) that minimize the sum of the weighted objective functions subject to the desired constraints (or bounds). Oftentimes, the clinician will have to consider several trade-offs due to the contradictory nature of the desired constraints and objectives. The field of multicriteria decision analysis (or multicriteria optimization) (MCO) formalizes the decision making in the presence of conflict stemming from multiple criteria (Kooy and Grassberger, 2015). The two main approaches to MCO are multilevel optimization and Pareto optimization.

Multilevel optimization (Long et al., 2012) (also called prioritized optimization or lexicographic ordering) proceeds in different stages by imposing an ordering in terms of importance of the different objectives and constraints. Therefore, each stage takes the optimized plan from the previous stage and attempts to optimize given additional criteria (e.g., the current stage attempts to minimize the dose to the parotids after the previous stage maximized the dose to the target).

In contrast, Pareto optimization treats all objectives equally and computes a set of Pareto optimal plans. Given a set of objectives and constraints, a plan is Pareto optimal if it is achievable and if there is no other achievable plan that is better with respect to one or more objectives and that is at least as good for the rest (Paganetti, 2018). The set of all Pareto optimal plans creates a Pareto surface, with intermediate plans being obtained by interpolation (Kooy and Grassberger, 2015). The clinician can traverse this space to assess the effect of different trade-offs (e.g., minimum dose to the target of 70 Gy versus 65 Gy).

Regardless of the chosen multicriteria optimization method, the doses that result in IMPT have a high degree of conformality (i.e., they have steep dose gradients) that is obtained by exploiting the finite range, sharp distal fall-off and scattering characteristics of protons over photons. The steep dose gradients are usually obtained by summing up doses from radiation fields that are highly inhomogeneous and composed of several energy layers. Due to this, small errors can cause hot or cold spots in the final dose distribution or shift part of the dose into tumor adjacent critical organs at risk. Some of the sources of uncertainties are the patient immobilization and setup in the treatment room, the tumor delineation strategy, the intrafraction patient motion, the anatomical changes that occur over the duration of the treatment, the conversion of HU values into proton stopping powers, CT artifacts and treatment delivery uncertainties. In general, the more conformal the total IMPT dose is, the more heterogeneous the fluence maps are and the more susceptible it is to uncertainties (Paganetti, 2018). Thus, the beneficial physical characteristics of protons over photons make IMPT doses more susceptible to errors (or less robust) when compared to IMRT doses (Mohan and Grosshans, 2017).

A plan that does not degrade in quality when uncertainties occur is called a robust plan (Paganetti, 2018). To achieve robustness in IMRT, margins are applied around the CTV, to create a new structure called the planning target volume (PTV). However, this concept is not applicable to IMPT, as it does not ensure coverage when hot or cold spots occur. To incorporate uncertainties in IMPT optimization, robust optimization and probabilistic approaches have been proposed. In these methods, the doses d are assumed to depend on an additional set $\lambda = \{\lambda_1, \ldots, \lambda_N\}$ of N uncertain parameters, namely $\mathbf{d} = \mathbf{d}(\mathbf{x}, \lambda)$, with $\mathbf{x}$ being as before the set of beam spot weights that are to be optimized. For robust optimization the values of the uncertain parameters are

assumed to be within some interval. This approach usually leads to an optimization problem in which the constraints have to be satisfied for every realization of the uncertain parameters (Paganetti, 2018). In the case of objectives, the robust optimization is a worst-case optimization problem. Thus, if the original objective was to minimize the maximum dose to the parotid glands, the robust optimization would seek to minimize the maximum dose to the parotid glands for any possible range or setup error, resulting therefore in a plan which is as good as possible for the worst case that was modelled to occur. In the probabilistic approach, a probability distribution $P(\lambda)$ that reflects the probability of the error scenarios occurring is included into the planning itself.

One limitation of robust optimization is that it creates an enlarged dose region that surrounds the target, just as margins do in IMRT, therefore delivering dose to surrounding healthy tissue. Another is that some uncertainties, such as weight loss over the duration of the treatment, are too complex to be included in robust optimization. Anatomical robust optimization, in which the possible anatomies of the patient are included in the optimization process has been proposed and found to be beneficial in increasing the robustness of plans (Van de Water et al., 2018) for a head and neck cohort. An additional approach to limit the enlarged dose region that surrounds the target and to increase conformality is adaptive therapy and is the subject of the next section.

1.3 Adaptive therapy

Adaptive radio therapy (ART) for photons, was initially proposed by Yan (Yan et al., 1997) in the 1990s. ART can be defined as a form of radiotherapy where "the delivered dose is monitored for clinical acceptability during the course of the treatment and modified as needed with the goal of improving clinical outcomes" (Green, Henke, and Hugo, 2019). Adapting the treatment plan to the daily anatomy, theoretically achieves simultaneously better tumor control and lower toxicity in OARs through the reduction of the necessary robustness settings to the minimum intra-fractional ones (Paganetti et al., 2021). Adaptation of the treatment plan is particularly important for PT, since IMPT doses are more susceptible to uncertainties than IMRT ones (Lomax, 2008a; Lomax, 2008b). There are three different time scales, namely offline, online and real-time, on which treatment plan adaptation could in principle be applied on (Green, Henke, and Hugo, 2019). An overview of the workflow in conventional, non-adaptive settings and of the workflow in generic adaptive settings can be seen in Figure 1.3. There are several differences between the different adaptive worfklows, many of them stemming from the available human, computational and clinical resources and hardware, but in general, the adaptive workflow consists, in one form or another, of re-imaging the patient and re-delineating the VOIs, assessing the need for adaptation, adapting the plan, ensuring the quality of the plan and thereafter safely delivering the plan. This section briefly discusses the different time-scales on which adaptation can occur and the challenges involved in achieving online adaptive proton therapy (APT).

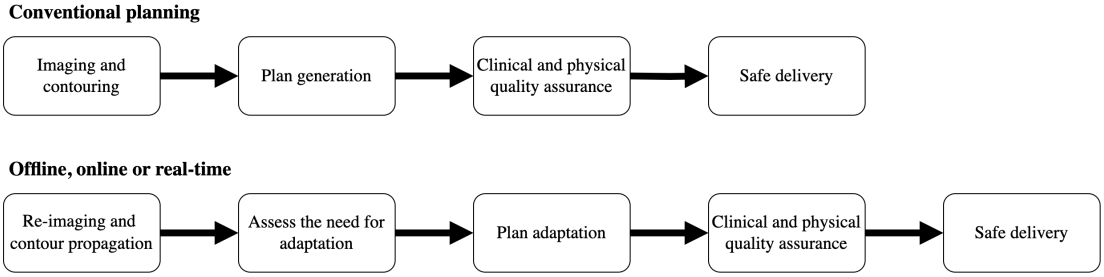


FIGURE 1.3: Overview of different RT and PT workflows. On the top row, the convetional one-plan workflow is illustrated while on the bottom row the generic adaptive workflow (offline, online and real-time) is illustrated.

Offline adaptation can happen at fixed time intervals or as a result of trigger signals (Paganetti et al., 2021). Offline ART is currently in clinical use in multiple variants and organ sites (Sonke and Belderbos, 2010; Chen et al., 2014; Feng et al., 2018; Green, Henke, and Hugo, 2019). In contrast, offline APT is less widely available, but is a reality in a couple of select clinics which monitor the anatomy of the patient throughout the treatment course (van de Schoot et al., 2016; Maeda et al., 2018; Placidi et al., 2017; Albertini et al., 2020; Visser et al., 2022). In the case of offline APT, the process usually takes several days between the acquisition of the new CT image and the delivery of the clinically approved new or adapted plan (Albertini et al., 2020). This is effective against slow inter-fractional anatomical changes, but is inadequate for fast ones. Ideally, the process would be performed on the time scale on which anatomical changes occur. Real-time adaptation is the most desired form of adaptation, as it would allow the smallest margins for photons and the lowest uncertainty incorporated in robust optimization for protons and therefore create the least toxicity.

In the online adaptation regime, a new CT image is acquired on the given day. Research is being undertaken to reduce the integral dose to the patient, by using low dose alternatives such as cone beam CT (CBCT), magnetic resonance (MR) imaging or low-dose CT protocols (Albertini et al., 2020). Following acquisition of a new anatomy, the contours from the previously obtained image are transferred onto it, usually via rigid or deformable registration or auto-segmenting software. The current methodology creates a bottleneck in the adaptive workflow, especially in the online and real-time regimes, as physicians need to manually correct the new contours. Deep learning algorithms based on prior information are expected to surpass the limitations of current methodologies and provide a workable solution (Albertini et al., 2020). Once a new anatomy is obtained and contoured, the need for adaptation is assessed. There is no broad consensus established for adaptation triggering (Albertini et al., 2020), and many clinics perform adaptation ad-hoc (Barragán-Montero et al., 2023). The work of Reiners (Reiners et al., 2023) used daily CBCTs, obtained during patient set-up, to create synthetic CTs that were thereafter used to establish protocols for adaptation triggers. This was found to improve the clinical workflow and allow for more frequent inter-fraction dose evaluations. Alternative systems, such as the traffic light system for photons (Kwint et al., 2014) have been described. In the traffic light system, the CBCT images obtained during set-up are categorized according to the severity of anatomical changes and adaptation is triggered if the system results in a "red" signal. Assuming that the need for adaptation has been established, a plan adaptation process follows. Here too, different approaches have been proposed. One is to optimize the initial plan (using the original constraints and objectives) on the new anatomy. This approach requires re-running the traditional planning pipeline and requires very fast dose computations, optimization as

well as full QA and is currently infeasible for online APT. The other approach is to change the original plan to restore the original plan quality or the original dose distribution. Several works reported plan adaptations within a couple of minutes, each with its own set of drawbacks (Paganetti et al., 2021). Depending on the chosen method, this approach may or may not require full QA.

One important limitation in the implementation of online APT is the manual nature of the patient-specific QA procedure. As explained before, patient-specific QA implies the delivery of the treatment plan to a water phantom in order to assess the differences between the planned and delivered dose. This procedure is infeasible, given that in online APT the patient must remain on the treatment table. Independent dose calculations (IDCs), in different forms, have been proposed as a form of software redundancy to the doses computed by the TPS and also as capable of replacing measurement based PSQA (Meier et al., 2015; Johnson et al., 2019). There are different degrees of independence in dose algorithms, ranging from low independence in the form of different implementations of the same algorithm that the TPS uses to full independence in the form of methodologically different dosimetric algorithms. Different sources of inputs are also possible, resulting in different types of QA. One solution is to tap into the data stream between the TPS and the gantry and extract the steering files. These files contain beam line settings required to provide the desired energy, the currents in the magnets for the required lateral deflection, the position of the patient couch and the number of monitor units (Meier et al., 2015). Combining steering files with the IDC results in an a-priori form of PSQA, where the patient can remain on the table.

Another alternative is to use the built-in measurements (specifically, the beam profile on the central axis, the spot position after lateral deflection and the monitor unit values) that the gantry detectors perform during the plan delivery. The measurements performed by the gantry, are compiled into a file called log file, which when combined with an IDC can be used a-posteriori to verify for each fraction the delivered dose to the patient (Meier et al., 2015; Scandurra et al., 2016). Ideally, for real-time ART, log-files would be extracted in real-time and the IDC would be fast enough (e.g., on the time-scale of energy layer switching) to provide real-time feedback of the actually delivered dose. This could be further used to modify the remaining spots or eventually halt highly erroneous deliveries.

1.4 Contents of the dissertation

This thesis investigated and developed software based methods for the automation of quality assurance processes in online APT. As discussed, a clinical implementation of the online or real-time adaptive workflows is impossible without this automation. Additionally, automation of QA processes will improve the clinical workflow in several ways. First, less time will be spent on time-consuming, repetitive processes, thereby allowing radiation technicians to focus their efforts on other relevant areas. Second, less time spent on QA measurements, implies more beam time would be available for patient treatment and therefore an increase in the needed patient throughput.

Chapter 2 presents the methodological framework of a TPS independent dose and dose change calculation algorithm, referred from here on as Yet anOther Dose Algorithm (YODA). The chapter starts with the Linear Boltzmann Equation (LBE) for the proton phase-space density, which is the same equation that MC methods solve. It thereafter details the physics-based approximations that are applied to it to reduce it to a system of two partial differential equations (PDEs). By employing physics-based approximations, the algorithm remains accurate with respect to the golden standard

MC methods. One of the PDEs obtained in this process is analytically solvable which makes the algorithm fast (the dose from a spot takes roughly 1 s to compute). Details about the solution methodology for both equations are provided and for how to convert the proton phase-space density into the dose in a region of interest (ROI) in the patient. The algorithm is truly independent from TPS based dose calculations, as it employs neither MC or pencil beam algorithm (PBA) methods. Next to the advantageous middle ground between speed and accuracy the algorithm achieves, another advantage is the presence of the PDEs that allows the application of adjoint theory. Adjoint theory is a mathematical formalism that approximates the change in a metric (e.g., the dose in a region of interest) as a function of a change in system variables (e.g., the HU values of the CT scan). This process is desirable when the number of system variables is large, as is the case in a CT scan with millions of voxels, as expensive re-computations of the chosen metric (e.g., the dose in a region) are avoided. Such a formalism is powerful, as it is applicable to a variety of metrics (e.g., tumor control probability, normal tissue complication probability, minimum or maximum dose in a region of interest) for a variety of changing parameters (e.g., the patient anatomy, a spot MU value, the energy spread of a spot). By avoiding re-computations, YODA can be employed in numerous time-constrained situations. One example of this is the triggering of daily adaptations, where the adjoint component could be used to quickly assess yesterday's dosimetric characteristics on today's anatomy. Thus, the chapter ends with an exposition of the adjoint formalism for the case when the metric of interest is the dose in the patient and the independent variables that change are the HU values of the CT scan.

In Chapter 3 the methodology developed in Chapter 2 is comprehensively tested first in water box with different slab inserts and thereafter in different patient (head and neck, prostate and lung) scans. To this end, Chapter 3 starts by developing a Gaussian beam splitting scheme. This is a necessary component of YODA, as the mathematical framework is valid for in-depth heterogeneous and laterally homogeneous geometries. Therefore, each treatment plan spot, taken as a symmetric two-dimensional Gaussian, is approximated by a summation of symmetric two-dimensional Gaussian sub-spots. The sub-spots are placed on concentric rings, with their amplitude, spreads and ring radii obtained in an optimization procedure. Using this, the dose algorithm of YODA is tested against the established golden standard MC algorithm TOPAS (Perl et al., 2012) in a variety of situations, ranging from challenging synthetic tests to real CT scans. Following this, the adjoint component's capability of computing yesterday's plan on today's anatomy is assessed for different patients and plans (both robustly and non-robustly optimized ones). This allows YODA to work in an adaptation trigger system. Lastly, a commissioning procedure is described, that when performed matches YODA's dose engine to any other dose engine (e.g., RayStation (Bodensteiner, 2018)). This is necessary as different dose engines use different stopping power data, which leads to range disagreements between the algorithms. Following this procedure, YODA could be used as an independent dose calculation tool. The commissioning procedure details the extraction of the beam optical parameters (either from actual measurements or from simulations in the desired dose engine), the mean energy, the energy spread and the MU to number of protons conversion that YODA should use to match the given dose engine. Following this, a comparison between the dose engine of RayStation and YODA is done.

Chapter 4 details a deep learning algorithm for predicting anatomical changes that occur over the course of treatment in head and neck patients. Such an algorithm can be used in an automatic plan QA pipeline. Specifically, a given patient can have a multitude of realistic future anatomies predicted and for each future anatomy an optimal plan is created a-priori to the treatment and stored in a database. On the given day, when

only a quickly adapted plan is available for the anatomy of the day, a plan comparison can be performed between the adapted plan and the corresponding truly optimal plan. Another use for such an algorithm is in anatomical robust optimization, to enhance the robustness of plans in the face of the predicted anatomical changes. The chapter starts by describing the probabilistic framework of the algorithm that can be sampled to predict deformation vector fields (DVFs), based on the planning CT (pCT) image and corresponding RT structures. The DVFs are in turn used to warp the pCTs and masks into repeat CT (rCT) images and masks. Next, details about the variational autoencoder architecture of the model and training details are provided. To assess the performance of the algorithm, a literature study of the observed anatomical changes in head and neck patients is presented. Lastly, the performance of the algorithm is assessed, both in terms of the changes expected from literature studies but also in terms of a comparable deep learning based model.

Chapter 5 concludes the thesis, summarizing the main findings and the possible applications of the presented models and provides recommendations for future research.

Chapter 2

A deterministic adjoint-based formalism for fast response changes

2.1 Novelty and own contributions

This chapter details the physical and mathematical formalism behind a novel dose and dose change computation engine, which entailed a number of own contributions. First, the idea to numerically solve the Fokker-Planck equation as opposed to approximating it, as is done in the work of (Gebäck and Asadzadeh, 2012), leads to a more accurate dose engine that remains fast enough for practical purposes. Second, a new density dependent adaptive stepping scheme was developed to speed up the tracking of the proton flux through tissue in the algorithmic implementation of the numerical solution to the Fokker-Planck equation. Third, the analytical solution to the Fermi-Eyges equation was fully derived in Appendix B, which resulted in a general solution capable of handling asymmetric pencil beams, as opposed to the symmetric ones from the work of (Gebäck and Asadzadeh, 2012). Last, the adjoint formalism was mathematically derived for this particular system of PDEs and the chosen metric (the average dose). The algorithmic solution to the adjoint equations was implemented and tested, showing that the adjoint method is capable of avoiding expensive re-computations thereby speeding up the assessments of dose changes caused by anatomical changes during patient-specific quality assurance.

2.2 Introduction

To perform fast, TPS independent and log-file based dose computations the interactions between the proton beam and the patient must be modelled, ideally not only using a different implementation of the TPS dose engine but also using a different methodology altogether. The two methods that are likely to be employed by a TPS are the Monte Carlo (MC) method and the analytical PB method. The MC method (e.g., TOPAS (Perl et al., 2012)) trades fast computation times for high computational precision (Zheng-Ming and Brahme, 1993) by solving the in-tissue proton balance equation (i.e. the Linear Boltzmann Equation) using statistical sampling methods. The analytical PB method (e.g., Bortfeld's model (Bortfeld, 1997)) trades high precision for fast computation times by employing a series of approximations and fits to obtain the dose in the tissue of interest. PB methods are still routinely used in TPS (Trnková et al., 2016) despite their limitations being well documented (Soukup, Fippel, and Alber, 2005).

This chapter is based on the publication of Tiberiu Burlacu, Danny Lathouwers, and Zoltán Perkó (Jan. 2023b). "A Deterministic Adjoint-Based Semi-Analytical Algorithm for Fast Response Change Computations in Proton Therapy". In: *Journal of Computational and Theoretical Transport* 52.1, pp. 1–41. ISSN: 2332-4309. DOI: 10.1080/23324309.2023.2166077. (Visited on 08/30/2023)

A methodologically different approach is based on a deterministic solution of the Linear Boltzmann Equation (Burlacu, Lathouwers, and Perkó, 2023b). This approach, which will henceforth be referred to as Yet anOther Dose Algorithm (YODA), is a hybrid numerical and analytical solution to a physics motivated approximation of the same equation that MC methods solve. The method strikes a balance in terms of accuracy versus speed. It is accurate with respect to MC methods due to the physical modelling of the interactions between the proton beam and the patient and it is fast due to the partly analytical solution. An additional advantage of this approach is the ease of applying the adjoint method. Given planning and repeat CT images with delineated structures and a treatment plan, the adjoint method computes an approximation of the change in dose caused by delivering the treatment plan to the repeat CT image, thereby avoiding expensive re-computations.

This chapter details the mathematical formalism behind YODA. In Section 2.3 the approximations employed to reduce the Linear Boltzmann Equation to two simpler to solve partial differential equations are detailed. Sections 2.4 and 2.5 show the methods through which the Fokker-Planck and Fermi-Eyges equations are solved while Section 2.6 shows how to compute the dose in the patient using the solution to the previously mentioned partial differential equations. Thereafter, Section 2.7 develops the adjoint formalism for dose changes as a function of anatomical changes. Lastly, the data sources used for the coefficients that appear in the previously mentioned equations are discussed in Section 2.8.

2.3 Approximating the LBE

The physical system under consideration is given by a proton beam irradiating the patient. This system can be characterized through the (steady-state) LBE, the validity of which for PT has been discussed by Börgers (Börgers, 1999). The LBE describes the proton balance in an arbitrary volume. Its derivation is obtained by equating all the gain and loss mechanisms for protons at position $\mathbf{r} \in \mathbb{R}^3$ with a certain energy E in $\mathrm{d}E$ and direction given by the unit vector $\hat{\mathbf{\Omega}} = \mathbf{v}/|\mathbf{v}|$ (with $\mathbf{v}$ the velocity vector of the protons) in $\mathrm{d}\hat{\mathbf{\Omega}}$ in an arbitrary volume V with a boundary denoted by ∂V as outlined by Duderstadt & Hamilton (Duderstadt and Hamilton, 1991). The equation is an integro-differential equation for the proton flux ($\varphi = vn$) with v the proton speed and $n(\mathbf{r}, E, \hat{\mathbf{\Omega}})$ the angular proton density,

$$\hat{\mathbf{\Omega}} \cdot \boldsymbol{\nabla}\varphi + \Sigma_t(\mathbf{r}, E)\varphi(\mathbf{r}, E, \hat{\mathbf{\Omega}}) = \int_{4\pi} \mathrm{d}\hat{\mathbf{\Omega}}' \int_0^\infty \mathrm{d}E'\, \Sigma_s(E' \to E, \hat{\mathbf{\Omega}}' \to \hat{\mathbf{\Omega}})\varphi(\mathbf{r}, E', \hat{\mathbf{\Omega}}') \quad (2.1)$$

$$\text{BC: } \varphi(\mathbf{r}_s, E, \hat{\mathbf{\Omega}}) = 0 \text{ if } \hat{\mathbf{\Omega}} \cdot \hat{\mathbf{e}}_s < 0 \text{ with } \mathbf{r}_s \in \partial V, \quad (2.2)$$

where BC is a boundary condition of the non-reentrant type, $\mathbf{r}_s$ denotes a vector on the boundary surface ∂V of the volume V, $\hat{\mathbf{e}}_s$ is the unit outward pointing normal vector to the boundary ∂V at $\mathbf{r}_s$, Σ_t is the total macroscopic cross section, Σ_s is the macroscopic double differential scattering cross section.

Currently, the LBE in its form is computationally expensive to solve. A first step is to divide the total Σ_t and scatter Σ_s cross sections according to the main interactions that a proton undergoes as it propagates through the medium, namely $\Sigma_t = \Sigma_a + \Sigma_e + \Sigma_{in}$ where Σ_a is the catastrophic (absorption) scatter cross section, Σ_e is the elastic scatter cross section between the incident protons and the nuclei of tissue, Σ_{in} is the inelastic scatter cross section between the incident protons and atomic electrons. By doing so,

Equation 2.1 can be written as

$$\begin{aligned}\hat{\mathbf{\Omega}} \cdot \boldsymbol{\nabla}\varphi = &\int_{4\pi} \mathrm{d}\hat{\mathbf{\Omega}}' \int_{E}^{\infty} \mathrm{d}E' \Sigma_a(E' \to E, \hat{\mathbf{\Omega}}' \to \hat{\mathbf{\Omega}})\varphi(\mathbf{r}, E', \hat{\mathbf{\Omega}}') - \Sigma_a(\boldsymbol{r}, E)\varphi(\mathbf{r}, E, \hat{\mathbf{\Omega}}) \\ &+ \int_{4\pi} \mathrm{d}\hat{\mathbf{\Omega}}' \Sigma_e(\mathbf{r}, E, \hat{\mathbf{\Omega}}' \to \hat{\mathbf{\Omega}})\varphi(\mathbf{r}, E, \hat{\mathbf{\Omega}}') - \Sigma_e(\boldsymbol{r}, E)\varphi(\mathbf{r}, E, \hat{\mathbf{\Omega}}) \\ &+ \int_{0}^{\infty} \mathrm{d}Q \Sigma_{in}(\boldsymbol{r}, E + Q \to E, \hat{\mathbf{\Omega}})\varphi(\mathbf{r}, E + Q, \hat{\mathbf{\Omega}}) - \Sigma_{in}(\boldsymbol{r}, E)\varphi(\mathbf{r}, E, \hat{\mathbf{\Omega}}),\end{aligned} \tag{2.3}$$

with Q defined as the amount of energy transferred during an interaction. In this splitting it is assumed that the energy transfer in Coulomb elastic scatter interactions is negligible and that the angular deflection in Coulomb inelastic scatter interactions is negligible (Zheng-Ming and Brahme, 1993). The next step is to apply approximations to each of the collision integrals in Equation 2.3.

The inelastic scatter integral is approximated using the Continuous Slowing Down Approximation (CSDA) and the Energy-loss Straggling (ELS) approximation (Zheng-Ming and Brahme, 1993). Given the difference between the proton and electron mass, the energy loss of a proton beam in each such individual collision is small. Thus, the stopping process can effectively be approximated by a continuous energy loss process with a mean (called the stopping power) and a deviation around the mean (called the straggling coefficient).

Therafter, we orient the beam along the z-direction and apply the small angle (i.e., $\Omega_z = 1$ in Cartesian angular coodinates) Fokker-Planck approximation to the elastic scatter angular integral. In doing so, $\hat{\mathbf{\Omega}}$ is redefined as $\hat{\mathbf{\Omega}} = (\Omega_x, \Omega_y) \in \mathbb{R}^2$. Similarly to the CSDA process, due to the small angular deflection that the proton beam suffers through its Coulomb interactions with the atom the scattering process is approximated as a continuous diffusion term in the lateral angular plane. Moreover, in the elastic scattering cross section $\Sigma_e(E, \hat{\mathbf{\Omega}} \cdot \hat{\mathbf{\Omega}}')$ the energy is replaced by the depth-dependent mean energy $E_a(z)$ (Gebäck and Asadzadeh, 2012; Zheng-Ming and Brahme, 1993).

The catastrophic inscatter integral is neglected completely with only the absorption catastrophic scatter cross section term remaining. Applying these approximations to the LBE reduces the integro-differential equation to the following PDE

$$\begin{aligned}\frac{\partial\varphi}{\partial z} + \Omega_x\frac{\partial\varphi}{\partial x} + \Omega_y\frac{\partial\varphi}{\partial y} - \frac{\partial S(\boldsymbol{r}, E)\varphi}{\partial E} - \frac{1}{2}\frac{\partial^2 T(\boldsymbol{r}, E)\varphi}{\partial E^2} \\ + \Sigma_a(\boldsymbol{r}, E)\varphi - \frac{\Sigma_{tr}(z)}{2}\left(\frac{\partial^2\varphi}{\partial\Omega_x^2} + \frac{\partial^2\varphi}{\partial\Omega_y^2}\right) = 0,\end{aligned} \tag{2.4}$$

where $S(\boldsymbol{r}, E)$ is the stopping power (the mean energy loss per unit path of the proton), $T(\boldsymbol{r}, E)$ is the straggling coefficient (the deviation of the energy loss around its mean value), Σ_a is the absorption cross section (the removal of protons from the beam due to nuclear interactions) and Σ_{tr} is the transport cross section (the rate at which protons diffuse in the lateral angular plane). The resulting PDE is linear in the dependent variable φ which in turn depends on the six independent system variables $\mathbf{r}, \hat{\mathbf{\Omega}}, E$.

We generalize the work of Gebäck and Asadzadeh (Gebäck and Asadzadeh, 2012) by considering a laterally homogeneous, in-depth heterogeneous geometry and write the flux as

$$\varphi = \varphi_{FE}(\boldsymbol{r}, \hat{\mathbf{\Omega}}) \cdot \varphi_{FP}(z, E). \tag{2.5}$$

To simplify notation, the dependence of the fluxes $\varphi_{FE}(\boldsymbol{r}, \hat{\boldsymbol{\Omega}})$ and $\varphi_{FP}(z, E)$ on their respective independent variables $(\boldsymbol{r}, \hat{\boldsymbol{\Omega}})$ and (z, E) will be suppressed in the rest of this text. Substitution in Equation 2.4 results in

$$\Upsilon(\varphi_{FE}) \cdot \varphi_{FP} + \varphi_{FE} \cdot \mathrm{1DFP}(\varphi_{FP}) = 0, \tag{2.6}$$

where $\Upsilon(\varphi_{FE})$ is the Fermi-Eyges equation and $\mathrm{1DFP}(\varphi_{FP})$ is the one-dimensional Fokker-Planck equation. In order to avoid the trivial solution both of these equations are set to zero (for a proof of separability see Appendix A), yielding

$$\Upsilon(\varphi_{FE}) = \frac{\partial \varphi_{FE}}{\partial z} + \Omega_x \frac{\partial \varphi_{FE}}{\partial x} + \Omega_y \frac{\partial \varphi_{FE}}{\partial y} - \frac{\Sigma_{tr}(z)}{2}\left(\frac{\partial^2 \varphi_{FE}}{\partial \Omega_x^2} + \frac{\partial^2 \varphi_{FE}}{\partial \Omega_y^2}\right) = 0 \tag{2.7}$$

and

$$\mathrm{1DFP}(\varphi_{FP}) = \frac{\partial \varphi_{FP}}{\partial z} - \frac{\partial S(z,E)\varphi_{FP}}{\partial E} - \frac{1}{2}\frac{\partial^2 T(z,E)\varphi_{FP}}{\partial E^2} + \Sigma_a(z,E)\varphi_{FP} = 0. \tag{2.8}$$

Searching for the solution in the split form defined by Equation 2.5 is a usual mathematical trick for the separation of variables, ensuring that the solutions of Equations 2.7 and 2.8 yield the exact solution of Equation 2.4. However, such a split of the proton flux also has strong physics foundations, even in the more general setting. Since catastrophic inscatter interactions are rare, the process mostly responsible for energy change is the Coulomb inelastic scatter. The energy loss of protons is therefore primarily determined by the stopping power and range straggling properties of the materials they traverse through. Due to the laterally homogeneous (or at least not too inhomogeneous) geometry on the scale of a highly focused beam (typically only 2-3 mm in clinical proton beams) and the strong forward scattering, the materials along the traversed through path till any given depth z are very similar for protons travelling under slightly different angles $\hat{\boldsymbol{\Omega}}$, resulting in strong coupling between the energy spectrum of the beam and the depth. Moreover, since the elastic Coulomb scatter mostly responsible for the angular spread of the beam causes negligible energy change, the energy spectrum of protons having slight deviations $\hat{\boldsymbol{\Omega}}$ from the main beam direction $\Omega_z = 1$ and that of protons with the original un-collided direction $\hat{\boldsymbol{\Omega}} = (0, 0, 1)^T$ is similar. These observations provide strong reasoning for searching for the solution in the form of Equation 2.5 – with a depth dependent energy spectrum $\varphi_{FP}(z, E)$ that is independent from the spatially dependent angle distribution $\varphi_{FE}(\boldsymbol{r}, \hat{\boldsymbol{\Omega}})$ – even in more general cases.

2.4 Solving the Fokker-Planck equation

The one-dimensional Fokker-Planck equation is a convection-diffusion equation in energy whose character is well suited for Discontinuous Galerkin (DG) methods. Consequently, its semi-discrete form was obtained using the Symmetric Interior Penalty Galerkin (SIPG). The main advantages of the SIPG method over other finite element methods (FEM) are the relative ease with which the approximating polynomial can be changed on different mesh elements, the fact that the method allows unstructured or adaptive meshes, and that the method satisfies a local energy balance (as opposed to the global energy balance satisfied by continuous Galerkin methods) (Rivière, 2008). The semi-discrete form was solved using the SDIRK3 method, which is a third order

accurate implicit finite difference method[1].

The one-dimensional Fokker-Planck equation can be written in a more standard convection-diffusion form

$$\frac{\partial \varphi_{FP}}{\partial z} - \frac{\partial S^*(z,E)\varphi_{FP}}{\partial E} - \frac{\partial}{\partial E}\left(T^*(z,E)\frac{\partial \varphi_{FP}}{\partial E}\right) + \Sigma_a(z,E)\varphi_{FP} = 0, \tag{2.9}$$

where the modified stopping power $S^*(z,E) = S(z,E) + \frac{1}{2}\frac{\partial T(z,E)}{\partial E}$ and the modified straggling coefficient $T^*(z,E) = T(z,E)/2$ are introduced. To simplify notation, from here on the stars will be dropped and the explicit (z,E) dependence of the stopping power, the straggling coefficient and the absorption cross section on the depths z and energy E will only be shown if necessary. Moreover, it is Equation 2.9 that will from now on be referred to as the one-dimensional Fokker-Planck equation. The Fokker-Planck equation can also be written in a simpler form as $L\varphi_{FP} = r$, where L is the differential operator acting on the Fokker-Planck flux φ_{FP} and r is the source of the equation, which in this case is equation to 0.

2.4.1 Domain definition and discretization

The computational domain of the equation is given as $\mathscr{D} = (0, z_{max}) \times (E_{min}, E_{max})$, $\mathscr{D} \subset \mathbb{R}^2$. The solution of the one-dimensional Fokker-Planck equation is the Fokker-Planck flux $\varphi_{FP}(z,E) \in \mathscr{H}$ where $\mathscr{H} = \mathscr{L}_2(\mathbb{R}^2)$ is a real Hilbert space of square integrable functions with an associated inner product defined as

$$\langle f, g \rangle = \int_0^\infty \mathrm{d}z \int_0^\infty \mathrm{d}E f g.$$

To ensure a unique solution to Equation 2.9 boundary conditions must be imposed, namely

$$\text{BCE: } \varphi_{FP}(z,E)\Big|_{E=E_{max}} = 0, \frac{\partial \varphi_{FP}(z,E)}{\partial E}\Big|_{E=E_{max}} = 0. \tag{2.10}$$

$$\text{BCS: } \varphi_{FP}(0,E) = Ae^{-\left(\frac{E-E_0}{\sigma_E}\right)^2}. \tag{2.11}$$

The boundary conditions in energy (BCE) are homogeneous Dirichlet and Neumann conditions while the boundary condition in space (BCS) is given by a Gaussian function in energy. Gerbershagen (Gerbershagen et al., 2017) showed that this is a realistic energy spectrum for a proton beam that has suffered energy degradation. The amplitude of the Gaussian is given by A, the mean by E_0 and the energy spread is given by σ_E. In line with usual practice, a rigorous proof of the existence and uniqueness of the solution to Equation 2.9 and its associated boundary conditions is not given and these properties are assumed to be true.

The energy component of the domain $\mathscr{D}$ is discretized into a number NG of groups with each group having the same width. The minimum and maximum energy of the domain are chosen to encompass the standard clinical proton energies range of 1 MeV to 201 MeV. In a given group g the high energy boundary is denoted by $E_{g-1/2}$, the low energy one by $E_{g+1/2}$ and the center value by E_g. Thus, $E_{max} = E_{1/2}$ and $E_{min} = E_{NG+1/2}$. An illustration of this discretization can be seen in Figure 2.1.

[1]When compared to the Crank-Nicholson method in our earlier work (Burlacu, Lathouwers, and Perkó, 2023b), this method increased the accuracy of the Fokker-Planck fluxes without degrading the speed of the algorithm.

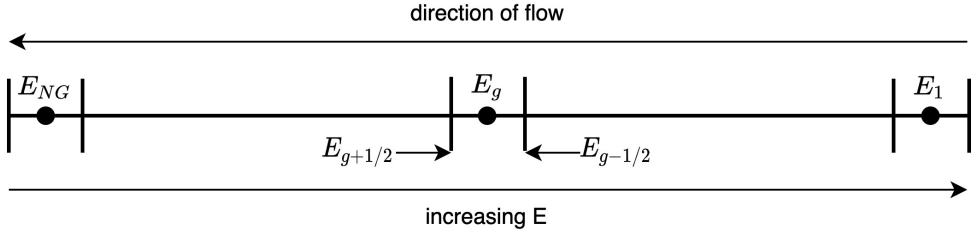


FIGURE 2.1: Energy domain discretization

The spatial part of the domain $\mathscr{D}$ is discretized into a number of steps N_s with the interval length Δz allowed to vary on a per step basis and the start and end points of the spatial domain are given by $z_0 = 0$ and $z_{Ns} = z_{max}$.

Equation 2.9 can also be written in a short-hand form as

$$L(\boldsymbol{\alpha})\varphi_{FP} = 0$$

where the vector of system parameters $\boldsymbol{\alpha}$ and the differential operator $L(\boldsymbol{\alpha})$ acting on the flux are introduced as

$$L(\boldsymbol{\alpha})(\cdot) = \frac{\partial}{\partial z}(\cdot) - \frac{\partial S^*(\cdot)}{\partial E} - \frac{\partial}{\partial E}\left[T^*\frac{\partial(\cdot)}{\partial E}\right] + \Sigma_a(\cdot),$$
$$\text{and } \boldsymbol{\alpha} = (S^*(z,E), T^*(z,E), \Sigma_a(z,E)).$$

The stopping power, energy straggling and absorption cross sections are all approximated as continuous, piece-wise linear functions in the NG energy groups. Since these properties are unique for each nuclide, we typically need as many material datasets as many different HU voxels (defined as a cubic element in the CT scan) the beam traverses, as the CT HU units are mapped to different material compositions. Denoting the space of univariate polynomials with real coefficients and degree at most k as $\mathbb{P}^k$, and the mesh in the energy variable as $\mathcal{T} = \{EI_i\}_{i\in 1,\ldots,NG}$, with $EI_i = [E_{i+1}, E_i]$ being the continuous energy interval in the i^{th} group, all our material data can represented by the vector space of continuous, piecewise linear functions defined as $P^1_{\mathcal{T}} = \left\{v_{\mathcal{T}} \in \mathcal{C}^0(R^+)\,|\forall i \in \{1,\ldots,NG\},\ v_{\mathcal{T}}|_{I_i} \in \mathbb{P}^1\right\}$. Thus, the space in which α resides is the tensor product constructed from the individual spaces $P^1_{\mathcal{T}}$ to which the stopping powers, straggling coefficients and absorption cross sections of all material domains belong.

2.4.2 Semi-discrete variational formulation

The first step to obtain an approximation to the solution of Equation 2.9 and its associated boundary conditions 2.10, 2.11 is to obtain the semi-discrete variational formulation. To do so, several quantities must be defined. First, the jump and the average of the flux at the edges of an energy group are defined as (Rivière, 2008)

$$[\varphi] = \varphi(E_j^-) - \varphi(E_j^+),$$
$$\text{and } \{\varphi\} = \frac{1}{2}\Big(\varphi(E_j^-) + \varphi(E_j^+)\Big),$$

where $j = \frac{1}{2}, \ldots, NG + \frac{1}{2}$ and with $E_j^- = \lim_{\epsilon \downarrow 0}(E_j - \epsilon)$ and $E_j^+ = \lim_{\epsilon \downarrow 0}(E_j + \epsilon)$. Special cases are defined at the boundary of the energy domain where

$$[v(E_{NG+1/2})] = -v(E^+_{NG+1/2}),\ \{v(E_{NG+1/2})\} = v(E^+_{NG+1/2}), \text{ and}$$
$$[v(E_{1/2})] = v(E^-_{1/2}), \qquad \{v(E_{1/2})\} = v(E^-_{1/2}).$$

Second, the penalty term is defined as (Rivière, 2008)

$$J_0(v, w) = \sum_{j=1/2}^{NG+1/2} \frac{\sigma^0}{h_{j-1,j}}[v(E_j)][w(E_j)]$$

where $h_{j-1,j} = \max(\Delta E_{j-1}, \Delta E_j)$ and σ^0 is a real and nonnegative number bounded from below. The role of this term is to penalize the jumps in the solution.

By multiplying Equation 2.9 with a test function v, integrating over one group, thereafter summing over all energy groups and making use of the definitions of the jump and the average, the semi-discrete variational formulation is found to be

$$\int_{E_{min}}^{E_{max}} \mathrm{d}E \frac{\partial \varphi_{FP}}{\partial z} v + a_{SIPG}(\varphi_{FP}, v) + \int_{E_{min}}^{E_{max}} \mathrm{d}E \left[-\frac{\partial S^* \varphi_{FP}}{\partial E} v + \Sigma_a \varphi_{FP} v \right] = 0, \quad (2.12)$$

where the SIPG bilinear a_{SIPG} is (Rivière, 2008)

$$a_{SIPG}(\varphi_{FP}, v) = \int_{E_{min}}^{E_{max}} T \frac{\partial \varphi_{FP}}{\partial E} \frac{\mathrm{d}v}{\mathrm{d}E} \mathrm{d}E$$
$$+ \sum_{\Gamma_i} -\left\{ T \frac{\partial \varphi_{FP}}{\partial E} \right\} \cdot [v] - [\varphi_{FP}] \cdot \left\{ T \frac{\mathrm{d}v}{\mathrm{d}E} \right\} + \frac{\sigma^0}{h_{i-1,i}} [\varphi_{FP}][v], \quad (2.13)$$

where Γ_i denotes the interior boundary faces of the energy domain. Following Hillewaert's work (Hillewaert, 2013), the penalty parameter was set as a function of the maximum polynomial degree $\max(\deg(p_g^i))$ of the basis functions, namely

$$\sigma^0 = \frac{(\max(\deg(p_g^i)) + 1)^2}{2}. \quad (2.14)$$

Both a coercivity analysis and the proof of equivalence between the semi-discrete variational formulation from Equation 2.12 and the model problem 2.9 with its associated boundary conditions 2.10 and 2.11 are beyond the scope of this work and can be found in the work of Hillewaert and Riviere respectively (Hillewaert, 2013; Rivière, 2008).

2.4.3 Basis functions

The first three group-centered Legendre polynomials[1]

$$p_g^i(E) \equiv P_i\left(\frac{2}{\Delta E_g}(E - E_g) \right), i = 0, 1, 2 \quad (2.15)$$

[1]Initially, the algorithm used first order basis functions. However, the resulting fluxes for coarse energy and spatial grids resulted in unphysical negative values. The addition of the third quadratic basis function reduced the behaviour to negligible values.

were used as the basis functions for the expansion of the flux in the computational domain as

$$\varphi_{FP}(z,E) = \sum_{g=1}^{NG}\sum_{i=0}^{2}\varphi_g^i(z)p_g^i(E). \tag{2.16}$$

Introducing the expansion from Equation 2.16 into the semi-discrete variational formulation from Equation 2.12 and sequentially replacing the function v with the chosen basis functions $p_g^i(E)$ yields a system of linear equations. This system can be written as

$$M\frac{\mathrm{d}\mathbf{\Phi}}{\mathrm{d}z} + G\mathbf{\Phi} = 0, \tag{2.17}$$

where $\mathbf{\Phi}$ is a vector with dimension $(1+\max(\deg(p_g^i)))\times NG$ and its elements are given by the unknown coefficients $\varphi_g^i(z)$ from Equation 2.16, the mass matrix M is a diagonal matrix that in a given group g has elements $\int \mathrm{d}E p_g^i(E)p_g^i(E)$ with $i=0,1,2$ along the diagonal and G is the system matrix which receives contributions from the continuous slowing down (CSD), straggling and absorption discretization.

This resulting system is discretized in space using either the Crank-Nicholson method or the three-stage, third-order accurate Singly Diagonally Implicit Runge-Kutta (SDIRK) method (Kennedy and Carpenter, 2016). Depending on the chosen number of groups the size of the resulting system is on the order of 10^3. This relatively small size of the system of equations implies that direct solution methods are comparable in computational time to iterative ones. To this end, the banded system solver DGBSV from the LAPACK library (Anderson et al., 1999) was used.

2.5 Solving the Fermi-Eyges equation

This section provides the main results of the analytical solution to the Fermi-Eyges equation and the steps taken to implement it. A full derivation of this solution can be found in Appendix B. This solution is based on refinements brought to Fermi's original theory on the distribution of charged particles undergoing multiple elastic scattering in their passing through matter. Authors such as Eyges, Brahme and Asadzadeh (Eyges, 1948; Brahme, 1975; Gebäck and Asadzadeh, 2012) have brought the theory into its form presented here. As shown in Section 2.3 the Fermi-Eyges equation takes the form

$$\Upsilon(\varphi_{FE}) = \frac{\partial\varphi_{FE}}{\partial z} + \Omega_x\frac{\partial\varphi_{FE}}{\partial x} + \Omega_y\frac{\partial\varphi_{FE}}{\partial y} - \frac{\overline{\Sigma_{tr}}(z)}{2}\left(\frac{\partial^2\varphi_{FE}}{\partial\Omega_x^2} + \frac{\partial^2\varphi_{FE}}{\partial\Omega_y^2}\right) = 0, \tag{2.18}$$

and can be solved by separating the x and y directions, namely $\varphi_{FE}(\boldsymbol{r},\Omega_x,\Omega_y) = H(z,x,\Omega_x)\cdot H(z,y,\Omega_y)$. This results in two separate PDEs for each direction

$$\frac{\partial H(z,\xi,\omega)}{\partial z} + \omega\frac{\partial H(z,\xi,\omega)}{\partial\xi} - \frac{\overline{\Sigma_{tr}}(z)}{2}\frac{\partial^2 H(z,\xi,\omega)}{\partial\omega^2} = 0, \tag{2.19}$$

where ξ stands for one of x, y and ω stands for one of Ω_x, Ω_y. For simplicity, the resulting PDEs have the same boundary condition imposed, namely

$$H(0,\xi,\omega) = C\exp\left(-\left(a_1\xi^2 + a_2\xi\omega + a_3\omega^2\right)\right), \tag{2.20}$$

with $a_i\in\mathbb{R}, \forall i = 1,2,3$ and $C>0$.

The solution of Equation 2.19 is found by artificially extending the usual domain of Ω_x and Ω_y from $\{(\Omega_x, \Omega_y) : |\hat{\mathbf{\Omega}}| = 1\}$ to $(\Omega_x, \Omega_y) \in [-\infty, \infty]^2$, applying two-dimensional Fourier transforms in ξ and ω and accounting for the Gaussian initial condition. In doing so the solution to Equation 2.19 is found to be

$$H(z, \xi, \omega) = \frac{C}{\sqrt{D}} \frac{1}{\sqrt{B}} \exp - \frac{\left(A_0 \xi^2 - 2A_1 \xi\omega + A_2 \omega^2\right)}{2B} \tag{2.21}$$

where $B = A_0 A_2 - A_1^2$, $D = 4a_1 a_3 - a_2^2$ and the FE coefficients A_0, A_1 and A_2 are the moments of the $\overline{\Sigma_{tr}}$ transport cross section and are given by

$$A_0(z) = \frac{2a_1}{D} + \int_0^z \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta \tag{2.22a}$$

$$2A_1(z) = \frac{4a_1 z - 2a_2}{D} + 2 \int_0^z (z - \eta) \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta \tag{2.22b}$$

$$A_2(z) = \frac{2a_1 z^2 - 2a_2 z + 2a_3}{D} + \int_0^z (z - \eta)^2 \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta. \tag{2.22c}$$

Gottschalk (Gottschalk, 2012) showed that the FE coefficients $A_0(z), A_1(z), A_2(z)$ can be intepreted as the variances of the angular direction, the lateral position and the covariance of the lateral position and the angular direction respectively.

The quantity $\overline{\Sigma_{tr}}$ is the depth-dependent energy spectrum (i.e., φ_{FP}) averaged[1] macroscopic transport cross section Σ_{tr}, namely

$$\overline{\Sigma_{tr}}(z) = \int \mathrm{d}E \varphi_{FP}(z, E) \Sigma_{tr}(z, E) \Big/ \int \mathrm{d}E \varphi_{FP}(z, E), \tag{2.23}$$

with the macroscopic transport cross Σ_{tr} computed using the macroscopic elastic scatter cross section Σ_s via

$$\Sigma_{tr}(z, E) = \int_{-1}^{1} \mathrm{d}\mu_0 \, \Sigma_s(z, E, \mu_0)(1 - \mu_0), \text{ with } \mu_0 = \cos\left(\hat{\mathbf{\Omega}} \cdot \hat{\mathbf{\Omega}}'\right).$$

Multiplying the x and y directions gives the general FE flux as

$$\varphi_{FE}(z, \boldsymbol{\rho}, \hat{\mathbf{\Omega}}) = \frac{C^2}{DB(z)} \exp - \frac{1}{2B(z)} \left(A_0 |\boldsymbol{\rho}|^2 - 2A_1 \boldsymbol{\rho} \cdot \hat{\mathbf{\Omega}} + A_2 |\hat{\mathbf{\Omega}}|^2\right) \tag{2.24}$$

where $\boldsymbol{\rho} = (x, y)$, $\hat{\mathbf{\Omega}} = (\Omega_x, \Omega_y)$.

Next to its analytical nature an important feature of the Fermi-Eyges solution from Equation 2.24 is that it is a Gaussian function in both the spatial and angular directions with coefficients that are determined by the depth-dependent beam energy spectrum $\varphi_{FP}(z, E)$ and the elastic scatter cross section from Equation 2.23 corresponding to that energy.

[1] In the original formalism, Σ_{tr} depends on the average depth-dependent beam energy $E_a(z)$. It was found that weighing Σ_{tr} with the depth-dependent energy spectrum yields more accurate lateral profiles that better match MC results.

2.5.1 Solution method

The boundary condition $H(0,\xi,\omega)$ from Equation 2.20 is chosen to be a two-dimensional normal distribution, namely

$$H(0,\xi,\omega) = \frac{1}{2\pi\sigma_\xi\sigma_\omega\sqrt{1-\rho_\xi^2}}\exp\frac{-1}{2(1-\rho_\xi^2)}\left(\frac{\xi^2}{\sigma_\xi^2} - 2\rho_\xi\frac{\xi}{\sigma_\xi}\frac{\omega}{\sigma_\omega} + \frac{\omega^2}{\sigma_\omega^2}\right).$$

Using this, the coefficients $C, a_1, a_2,$ and a_3 are identified as

$$\begin{aligned} C &= \frac{1}{2\pi\sigma_\xi\sigma_\omega\sqrt{1-\rho_\xi^2}}, \\ a_1 &= \frac{1}{2(1-\rho_\xi^2)\sigma_\xi^2},\ a_2 = \frac{-\rho_\xi}{(1-\rho_\xi^2)\sigma_\xi\sigma_\omega},\ a_3 = \frac{1}{2(1-\rho_\xi^2)\sigma_\omega^2}, \\ D &= \frac{1}{(1-\rho_\xi^2)\sigma_\xi^2\sigma_\omega^2}, \end{aligned}$$

where ϱ is the correlation coefficient between the spatial dimension ξ and the angular dimension ω, σ_ξ standard deviation in ξ and σ_ω standard deviation in ω. The a_i coefficients are thereafter used to initialize the values of the FE coefficients from Equations 2.22

To compute the FE coefficients at a given depth z_i the beam energy spectrum at that depth must be known together with $\Sigma_{tr}(z,E)$ at z_i. The μ_0 integral from $\Sigma_{tr}(z,E)$ is computed using the QAGE routine from the QUADPACK library (Piessens et al., 1983). As z increases in the integrals from Equations 2.22 so do the integrands and the computational expense of these integrals. We chose to approximate Σ_{tr} in a given step as the average of its values at the start and endpoint of the step thereby employing the trapezoidal integration rule. In doing so, the integrals could be re-written to depend only on the previous value as shown in Appendix B. Given the fact that the segments over which Σ_{tr} is integrated are small ($\leq 0.01\,\text{cm}$) the trapezoidal integration scheme is sufficiently accurate for our purposes.

2.5.2 The angular integral

In the metric computation the angle integrated FE flux is needed, namely

$$\Psi_{FE}(\boldsymbol{r}) = \int_{4\pi} \mathrm{d}\hat{\boldsymbol{\Omega}}\,\varphi_{FE} = \int_{-\infty}^{\infty} \mathrm{d}\Omega_x H(z,x,\Omega_x) \int_{-\infty}^{\infty} \mathrm{d}\Omega_y H(z,y,\Omega_y).$$

where the integration domain was extended to $(-\infty,\infty)$ as was done in the Fermi-Eyges solution. As shown in Appendix B, by using the solution from Equation 2.21 the angular integral takes the following simple form

$$\Psi_{FE}(\boldsymbol{r}) = \frac{1}{2\pi A_2}\exp -\frac{|\boldsymbol{r}|^2}{2A_2}. \tag{2.25}$$

2.6 Metric definition

The 6-dimensional phase-space density resulting from the individual solutions to the FP and FE equations can be used to obtain all clinically relevant metrics. For example,

let Ψ_{FE} be the angular integral of φ_{FE}, namely

$$\Psi_{FE}(\boldsymbol{r}) = \int_{4\pi} \mathrm{d}\hat{\boldsymbol{\Omega}} \varphi_{FE}(\hat{\boldsymbol{\Omega}}, \boldsymbol{r}), \tag{2.26}$$

and let Ψ_{FP} be

$$\Psi_{FP}(z) = \int_0^\infty \mathrm{d}E\, E\left[-\frac{\partial S(z,E)\varphi_{FP}}{\partial E} - \frac{1}{2}\frac{\partial^2 T(z,E)\varphi_{FP}}{\partial E^2} + \Sigma_a(z,E)\varphi_{FP}\right]. \tag{2.27}$$

Then, if the CT image volume is given by the union of all of its N_v voxels (i.e., $\mathscr{V} = \bigcup V_k, k = 1, \ldots, N_v$ where V_k is the volume of one voxel), the energy E_k deposited by the proton beam in a voxel V_k is given by

$$E_k = \int_{V_k} \mathrm{d}V \Psi_{FE}(\mathbf{r}) \Psi_{FP}(z).$$

The dose D_k in the same voxel k is given as

$$D_k = \frac{E_k}{m_k} = \frac{1}{\Delta V} \int_{V_k} \mathrm{d}V \frac{\Psi_{FE}(\mathbf{r})\Psi_{FP}(z)}{\rho_k}, \tag{2.28}$$

where $\Delta V = \Delta x \Delta y \Delta z$ is the volume of voxel k (constant for all voxels in the CT image) and ρ_k is the mass density of voxel k. Thus, the total dose in a certain region of interest (ROI) of the CT image, identified by the union of its corresponding voxels, is the sum of D_k over all k in the ROI.

2.6.1 Lateral density scaling

A scaling to better account for lateral heterogeneities is introduced in the dose in a voxel V_k from Equation 2.28. Specifically, the energy density in a voxel k is scaled by the ratio of the density ρ_{ck} on the central beam axis at a depth that corresponds to the voxel k and the density ρ_k of the voxel itself, namely

$$E_k = \int_{V_k} \mathrm{d}V \frac{\rho_k}{\rho_{ck}} \Psi_{FE}(\mathbf{r}) \Psi_{FP}(z). \tag{2.29}$$

Using this scaling, the dose in voxel k becomes

$$D_k = \frac{1}{\Delta V} \int_{V_k} \mathrm{d}V \frac{\Psi_{FE}(\mathbf{r})\Psi_{FP}(z)}{\rho_{ck}}. \tag{2.30}$$

Thus, a pencil beam distributes laterally a dose proportional to the density along the central beam axis. It is Equation 2.30 that defines D_k and the one that will be used for both dose computations and the development of the adjoint formalism.

2.7 Dose changes via the adjoint method

Next to its dose computation capabilities, an advantage of YODA is the ease of applying the adjoint method. This general mathematical framework approximates to first order the change in a metric as a function of the change in all independent variables. Examples of possible independent variables are HU values in the CT image or treatment plan spot

characteristics such as mean energy, energy spread, position, MU value (or equivalently the number of protons), angular spread and the spot size. Examples of metrics are the mean dose to an OAR or NTCP values. The adjoint method is useful when the number of independent variables is large (so that re-computing the metric for each new variable becomes prohibitively expensive) and their change is relatively small (so that the first order adjoint approximation is accurate). Examples of applications are computing dose or NTCP differences caused by differences between planned and delivered spot MU values or isocenter positions or by delivering yesterday's treatment plan on today's CT image. Since CTs typically have millions of voxels this is likely always the case in radiotherapy. This section illustrates the main details of the adjoint method for the case when the independent variables that change are the HU values of the CT image and the metric considered is the dose in a voxel V_k. Larger regions of clinical interest are trivial generalizations of this case.

A given change in the HU values of the CT image implies two distinct changes in the deposited dose D_k in the voxel k. One is a direct change, since a HU change in the voxel k implies, among others, a stopping power change which can be directly inputted in the D_k change via Equation 2.27. The other is an indirect change, as a stopping power change somwhere along the proton beam path implies a proton flux change in the considered voxel k. This change can only be known by re-solving for φ from the FP and FE equations with the new HU values. Thus, the change in D_k is written as,

$$\delta D_k = \delta D_{k,dir} + \delta D_{k,indir}\,, \tag{2.31}$$

where δ denotes a variation, $\delta D_{k,dir}$ denotes the part of δD_k that can be directly computed and $\delta D_{k,indir}$ denotes the part that would have to be re-computed.

2.7.1 Adjoint source derivation

Formally, taking the Gateaux-differential of D_k gives δD_k as

$$\begin{aligned}
\delta D_k(\mathbf{e^0},\mathbf{h}) = {} & \frac{1}{\Delta V}\int\limits_{V_k} \mathrm{d}V\, \frac{-h_{\rho_{ck}}}{\rho_{ck}^2}\Psi_{FE}(\mathbf{r},\mathbf{e^0})\Psi_{FP}(z,\mathbf{e^0}) && \text{(direct effect)}\\
& + \frac{1}{\Delta V}\int\limits_{V_k} \mathrm{d}V\, \frac{1}{\rho_{ck}}\,\delta\Psi_{FE}(\mathbf{r},\mathbf{e^0},\mathbf{h})\,\Psi_{FP}(z,\mathbf{e^0}) && \text{(direct and indirect effect)}\\
& + \frac{1}{\Delta V}\int\limits_{V_k} \mathrm{d}V\, \frac{1}{\rho_{ck}}\Psi_{FE}(\mathbf{r},\mathbf{e^0})\,\delta\Psi_{FP}(z,\mathbf{e^0},\mathbf{h})\,. && \text{(direct and indirect effect)}\\
& = \delta D_{k1}(\mathbf{e^0},\mathbf{h}) + \delta D_{k2}(\mathbf{e^0},\mathbf{h}) + \delta D_{k3}(\mathbf{e^0},\mathbf{h})\,,
\end{aligned}$$

where the explicit dependence on a vector $\boldsymbol{e}^0$ of nominal or base parameters has been indicated and $\mathbf{h}$ denotes the change in a quantity. In the case of this work, the nominal parameters are S, T, Σ_a, Σ_s on the planning CT together with the spot optical and energy parameters in the original plan.

The first contribution to the direct effect is found to be

$$\delta D_{k,dir} = \frac{1}{\Delta V}\int\limits_{V_k} \mathrm{d}V\, \frac{-h_{\rho_{ck}}}{\rho_{ck}^2}\Psi_{FE}(\mathbf{r},\mathbf{e^0})\Psi_{FP}(z,\mathbf{e^0}). \tag{2.32}$$

The remaining terms have to be separated into their respective direct and indirect components. The $\delta\Psi_{FE}$ term can be computed by making use of the previously found

Ψ_{FE} definition,

$$\Psi_{FE}(\boldsymbol{r}) = \frac{1}{2\pi A_2} \exp -\frac{|\boldsymbol{r}|^2}{2A_2}. \tag{2.33}$$

Taking the Gateaux-differential of Ψ_{FE} gives

$$\delta\Psi_{FE} = \Psi_{FE}\, \delta A_2 \left[\frac{-1}{A_2} + \frac{|\boldsymbol{r}|^2}{2}\frac{1}{A_2^2}\right]$$

which contains both direct and indirect effect components. Continuing the process of taking Gateaux-differentials gives for A_2

$$\delta A_2 = \int_0^z (z - z')^2\, \delta\overline{\Sigma_{tr}}\,(z')\mathrm{d}z',$$

where, for simplicity, variations in the Fermi-Eyges initial condition were ignored. By including these variations, which lead to straightforward direct effects, perturbations in initial spot spatial spread, angular spread and divergence can be computed.

The energy-flux averaged transport cross section is given by

$$\overline{\Sigma_{tr}}(z, \mathbf{e^0}) = \frac{1}{N_p(z)} \int \mathrm{d}E \varphi_{FP}(z, E)\Sigma_{tr}(z, E)$$

where

$$N_p(z, \varphi_{FP}) = \int \mathrm{d}E \varphi_{FP}(z, E).$$

The Gateaux-differential of $\overline{\Sigma_{tr}}(z, \mathbf{e^0})$ is found to be

$$\begin{aligned}
\delta\overline{\Sigma_{tr}}\,(z, \mathbf{e^0}, \mathbf{h}) = &\int \mathrm{d}E h_\varphi(z, E)\frac{N_p(z)\Sigma_{tr}(z, E) - \Sigma_\varphi(z)}{N_p(z)^2} && \text{(indirect effect)}\\
&+ \int \mathrm{d}E \varphi_{FP}(z, E)\frac{h_{\Sigma_{tr}}(z, E)}{N_p(z)} && \text{(direct effect)}
\end{aligned}$$

where

$$\Sigma_\varphi(z, \mathbf{e^0}) = \int \mathrm{d}E \varphi_{FP}(z, E)\Sigma_{tr}(z, E).$$

Introducing this back into δA_2 gives

$$\begin{aligned}
\delta A_2 = &\int_0^z \mathrm{d}z'(z - z')^2 \int \mathrm{d}E h_\varphi(z', E)\frac{N_p(z')\Sigma_{tr}(z', E) - \Sigma_\varphi(z')}{N_p(z')^2} && \text{(indirect effect)}\\
&+ \int_0^z \mathrm{d}z'(z - z')^2 \int \mathrm{d}E \varphi_{FP}(z', E)\frac{h_{\Sigma_{tr}}(z', E)}{N_p(z')} && \text{(direct effect)}\\
&= \delta A_{2,indir} + \delta A_{2,dir}
\end{aligned}$$

At this point, the direct effect contribution from δD_{k2} can be written as

$$\delta D_{k,dir} \mathrel{+}= \frac{1}{\Delta V}\int_{V_k} \mathrm{d}V\, \frac{1}{\rho_{ck}}\Psi_{FE}(\mathbf{r})\Psi_{FP}(z)\left[\frac{-1}{A_2} + \frac{|\boldsymbol{r}|^2}{2}\frac{1}{A_2^2}\right]$$

$$\cdot \int_0^z \mathrm{d}z'(z-z')^2 \int \mathrm{d}E\varphi_{FP}(z',E)\frac{h_{\Sigma_{tr}}(z',E)}{N_p(z')}. \tag{2.34}$$

The indirect part $\delta A_{2,indir}$ is re-written such that the integral over z' covers the semi-infinite region of $(0,\infty)$, namely

$$\delta A_{2,indir} = \int_0^\infty \mathrm{d}z' \int \mathrm{d}E h_\varphi(z',E)H(z-z')(z-z')^2\frac{N_p(z')\Sigma_{tr}(z',E)-\Sigma_\varphi(z')}{N_p(z')^2}.$$

Using this, the indirect part from δD_{k2} is given as

$$\begin{aligned}\delta D_{k,indir} = &\frac{1}{\Delta V}\int_{V_k}\mathrm{d}V\,\frac{1}{\rho_{ck}}\Psi_{FE}(\mathbf{r})\Psi_{FP}(z)\left[\frac{-1}{A_2}+\frac{|\boldsymbol{r}|^2}{2}\frac{1}{A_2^2}\right]\\ &\cdot\int_0^\infty \mathrm{d}z' \int \mathrm{d}E h_\varphi(z',E)H(z-z')(z-z')^2\frac{N_p(z')\Sigma_{tr}(z',E)-\Sigma_\varphi(z')}{N_p(z')^2}.\end{aligned}$$

This can be re-arranged as

$$\begin{aligned}\delta D_{k,indir} = &\int_0^\infty \mathrm{d}z' \int \mathrm{d}E\, h_\varphi(z',E)\frac{N_p(z')\Sigma_{tr}(z',E)-\Sigma_\varphi(z')}{N_p(z')^2}\\ &\cdot\frac{1}{\Delta V}\int_{V_k}\mathrm{d}V\,\frac{1}{\rho_{ck}}\Psi_{FE}(\mathbf{r})\Psi_{FP}(z)\left[\frac{-1}{A_2}+\frac{|\boldsymbol{r}|^2}{2}\frac{1}{A_2^2}\right]H(z-z')(z-z')^2\\ &=\langle h_\varphi(z',E), r_2^\dagger(z',E)\rangle.\end{aligned}\tag{2.35}$$

The last term to be separated into direct and indirect contributions is the $\delta\Psi_{FP}$ one. By employing the SIPG formalism and using E as a basis function, Ψ_{FP} is written as

$$\Psi_{FP}(z) = ES\varphi_{FP}\Big|_{E_{min}} + \int_{E_{min}}^{E_{max}}\mathrm{d}E\left(S\varphi_{FP}+T\frac{\partial\varphi_{FP}}{\partial E}+E\Sigma_a\varphi_{FP}\right)+\sum_{\Gamma_i}-[\varphi]T.$$

This can be re-written under one E integral by using the property of the Dirac-delta function, namely

$$\begin{aligned}\Psi_{FP}(z) = \int_{E_-}^{E_+}\mathrm{d}E\,\varphi_{FP}\Bigg[&ES\delta(E-E_-)+S-T\delta(E-E_-)-\frac{\mathrm{d}T}{\mathrm{d}E}\\ &+\left(\sum_{\Gamma_i}\left(-\delta(E-E_i^-)+\delta(E-E_i^+)\right)T(E_i)\right)+E\Sigma_a\Bigg].\end{aligned}$$

where E_{min} was be replaced by E_- and E_{max} by E_+ to reduce the footprint of equations. The Gateaux-differential of this is given as

$$\delta\Psi_{FP} = \int_{E_-}^{E_+}\mathrm{d}E h_{\varphi_{FP}}\Bigg[ES\delta(E-E_-)+S-T\delta(E-E_-)-\frac{\mathrm{d}T}{\mathrm{d}E}$$

$$+\left(\sum_{\Gamma_i}\big(-\delta(E-E_i^-)+\delta(E-E_i^+)\big)T(E_i)\right)+E\Sigma_a\Bigg]$$

(indirect effect)

$$+\int_{E_-}^{E_+}\mathrm{d}E\varphi_{FP}\Bigg[\,Eh_S\delta(E-E_-)+h_S-h_T\delta(E-E_-)-\frac{\mathrm{d}h_T}{\mathrm{d}E}$$

$$+\left(\sum_{\Gamma_i}\big(-\delta(E-E_i^-)+\delta(E-E_i^+)\big)h_T(E_i)\right)+Eh_{\Sigma_a}\Bigg].$$

(direct effect)

Thus, the last contribution to the direct effect is found to be

$$\begin{aligned}\delta D_{k,dir}\,+= &\frac{1}{\Delta V}\int_{V_k}\mathrm{d}V\;\frac{1}{\rho_{ck}}\Psi_{FE}(\mathbf{r})\\ &\cdot\int_{E_-}^{E_+}\mathrm{d}E\,\varphi_{FP}\Bigg[Eh_S\delta(E-E_-)+h_S-h_T\delta(E-E_-)-\frac{\mathrm{d}h_T}{\mathrm{d}E}\\ &\qquad+\left(\sum_{\Gamma_i}\big(-\delta(E-E_i^-)+\delta(E-E_i^+)\big)h_T(E_i)\right)+Eh_{\Sigma_a}\Bigg].\end{aligned}\tag{2.36}$$

The indirect part $\delta\Psi_{FP,indir}$ is written into an inner product between $h_{\varphi_{FP}}$ and a quantity denoted by $r_1^\dagger$ by splitting the voxel volume integral into a depth and lateral part, namely

$$\int_{V_k}\mathrm{d}V=\int_{z_-}^{z^+}\mathrm{d}z\iint_{V_{k,xy(z)}}\mathrm{d}x\mathrm{d}y.$$

Using this, $\delta D_{k,indir}$ is incremented by

$$\begin{aligned}&\int_{E_-}^{E_+}\mathrm{d}E\int_{z-}^{z+}\mathrm{d}z\,h_{\varphi_{FP}}\Bigg[ES\delta(E-E_-)+S-T\delta(E-E_-)-\frac{\mathrm{d}T}{\mathrm{d}E}\\ &\qquad+\left(\sum_{\Gamma_i}\big(-\delta(E-E_i^-)+\delta(E-E_i^+)\big)T(E_i)\right)+E\Sigma_a\Bigg]\\ &\qquad\cdot\frac{1}{\Delta V}\int_{V_{k,xy(z)}}\mathrm{d}V\;\frac{1}{\rho_{ck}}\Psi_{FE}(\mathbf{r})=\langle h_\varphi(z,E),r_1^\dagger(z,E)\rangle.\end{aligned}\tag{2.37}$$

Thus, the variation in D_k from Equation 2.31 has been expressed into two parts. One is the directly computable part $\delta D_{k,dir}$ which is the summation of Equations 2.32, 2.34 and 2.36. The other is the indirectly computable part $\delta D_{k,indir}$ which is the summation of Equations 2.35 and 2.37.

2.7.2 Adjoint system derivation

The adjoint method removes from δD_k the part $\delta D_{k,indir}$ that would have to be re-computed and in this process computes a first order approximation to δD_k. This is done by expressing $\delta D_{k,indir}$ as an inner product between two quantities as was done in Equations 2.35 and 2.37. One is the change in the proton flux h_φ caused by the change in the HU values and the other is a vector denoted by $r^\dagger$, namely $\delta D_{k,indir} = \langle h_\varphi, r^\dagger \rangle$. The vector $r^\dagger$ is identified as the right-hand side of a new system called the adjoint system. This system is written as $L^\dagger \varphi^\dagger = r^\dagger$ and its solution is called the adjoint flux $\varphi^\dagger$. Using this, together with the properties of the adjoint operator gives

$$\delta D_{k,indir} = \langle h_\varphi, r^\dagger \rangle = \langle h_\varphi, L^\dagger \varphi^\dagger \rangle. \tag{2.38}$$

The last inner product can be found by starting with

$$\langle \varphi^\dagger, L h_\varphi \rangle = \int_0^\infty \mathrm{d}z \int_0^\infty \mathrm{d}E \varphi^\dagger \left[\frac{\partial h_\varphi}{\partial z} - \frac{\partial S h_\varphi}{\partial E} - \frac{\partial}{\partial E}\left(T \frac{\partial h_\varphi}{\partial E} \right) + \Sigma_a h_\varphi \right]. \tag{2.39}$$

At this point we extend φ_{FP} and consequently h_φ to the whole $\mathbb{R}^2$ plane with the condition that these quantities are zero everywhere outside of the computational domain $\mathscr{D}$. Through partial integration along the z-direction for the first term, and along the E-direction for the stopping power, range straggling terms and absorption terms, Equation 2.39 is found to be equal to

$$\begin{aligned} \langle \varphi^\dagger, L(\boldsymbol{\alpha}) h_\varphi \rangle &= \int_0^\infty \mathrm{d}E \varphi^\dagger(0,E) h_\varphi(0,E) + \left\langle -\frac{\partial \varphi^\dagger}{\partial z} + S\frac{\partial \varphi^\dagger}{\partial E} - \frac{\partial}{\partial E} T \frac{\partial \varphi^\dagger}{\partial E} + \Sigma_a \varphi^\dagger, h_\varphi \right\rangle \\ &= \int_0^\infty \mathrm{d}E \varphi^\dagger(0,E) h_\varphi(0,E) + \left\langle L^\dagger \varphi^\dagger, h_\varphi \right\rangle. \end{aligned} \tag{2.40}$$

In the process of deriving Equation 2.40 the adjoint operator $L^\dagger$ together with its associated boundary conditions were found to be

$$L^\dagger \varphi^\dagger = -\frac{\partial \varphi^\dagger}{\partial z} + S\frac{\partial \varphi^\dagger}{\partial E} - \frac{\partial}{\partial E}\left(T \frac{\partial \varphi^\dagger}{\partial E} \right) + \Sigma_a \varphi^\dagger \tag{2.41}$$

$$\text{BCE: } \varphi^\dagger(z, E_{min}) = 0, \left. \frac{\partial \varphi^\dagger}{\partial E} \right|_{E=E_{min}} = 0, \tag{2.42}$$

$$\text{BCS: } \varphi^\dagger(z_{max}, E) = 0. \tag{2.43}$$

Making use of Equation 2.40 in Equation 2.38 gives

$$\delta D_{k,indir} = \langle \varphi^\dagger, L h_\varphi \rangle - \int_0^\infty \mathrm{d}E \varphi^\dagger(0,E) h_\varphi(0,E) \tag{2.44}$$

The effect of the operator L acting on the perturbation in the FP flux can be found by taking the Gateaux-differential of the Fokker-Planck equation, namely

$$\delta(L\varphi) = 0 \iff \delta L\, \varphi = -L h_\varphi$$

Using this, gives the final form of the indirect part of D_k as

$$\begin{aligned}\delta D_{k,indir} &= -\langle \varphi^\dagger, \delta L\, \varphi \rangle - \int_0^\infty \mathrm{d}E \varphi^\dagger(0,E) h_\varphi(0,E) \\ &= \int_0^\infty \mathrm{d}E \int_0^\infty \mathrm{d}z\, \varphi^\dagger \left[\frac{\partial}{\partial E}(h_S \varphi_{FP}) + \frac{\partial}{\partial E}\left(h_T \frac{\partial \varphi_{FP}}{\partial E} \right) - h_{\Sigma_a} \varphi_{FP} \right] \\ &\quad - \int_0^\infty \mathrm{d}E \varphi^\dagger(0,E) h_\varphi(0,E) \end{aligned} \tag{2.45}$$

Thus, the goal of the adjoint method has been reached. The indirect change in D_k has been replaced by the inner product shown in Equation 2.45. For the purpose of this work, the perturbation in the initial condition of the FP flux h_φ has been ignored. In future work, this can easily be included, to model perturbations in the mean energy, energy spread or the number of protons of a treatment plan spot.

2.8 Data sources

In order to obtain the solution to the two PDEs and the response, the stopping power, straggling coefficient, absorption cross section and elastic scatter cross section must be known as a function of energy and tissue composition. The CT scan HU values were converted to density and fractional compositions according to Schneider's method (Schneider, Bortfeld, and Schlegel, 2000). The density and fractional composition were used to interpolate nuclide specific tables of the stopping power versus energy. The tables were extracted from TOPAS (Perl et al., 2012) using an adapted extension distributed on the TOPAS forum. The stopping power for protons in water versus energy can be seen in Figure 2.2.

The straggling coefficient represents the statistical variation around the mean of the energy loss of a proton in a material. The consequence of energy straggling is the spreading of the energy spectrum of an initially mono-energetic beam (Noshad and Bahador, 2012). The equation that was used for the straggling coefficient is (Williams and Bragg, 1932)

$$T(E, N_{\mathbb{A}}(z)) = \sum_{i \in \mathbb{A}} \frac{1}{(4\pi\epsilon_0)^2} N_i(z) 4\pi e^4 Z_i(z) \left(1 + \frac{4 I_i(z)}{3 m_e v_p^2} \ln \frac{2 m_e v_p^2}{I_i(z)} \right), \tag{2.46}$$

where $N_{\mathbb{A}}$ is the set of atomic densities corresponding to the set of atoms atoms $\mathbb{A}$ that were considered to constitute human tissue, namely $\mathbb{A} = \{$H, C, N, O, Na, Mg, P, S, Cl, Ar, K, Ca$\}$. Moreover, Z_i is the atomic number of the target atom i with $i \in \mathbb{A}$, ϵ_0 is the vacuum permitivity constant, e is the elementary charge, m_e is the electron mass, v_p is the proton speed, I_i is the mean atomic excitation energy of atom i. The straggling coefficient for protons versus energy in water can be seen in Figure 2.3.

The elastic scatter cross section can be found by considering the deflection that a proton suffers due to the Coulomb field of the nucleus. A derivation of this can be found in the work of Goldstein (Goldstein, Poole, and Safko, 2002) who gives the microscopic elastic scatter cross section for protons incident on a target nucles $t, t \in \mathbb{A}$ with atomic

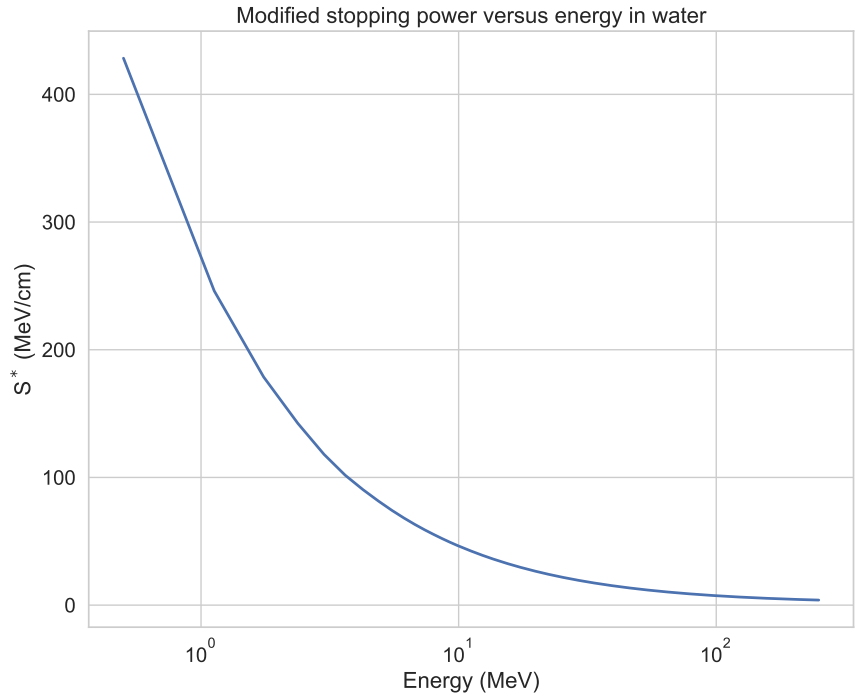


FIGURE 2.2: Water stopping power versus proton energy.

number Z_t and atomic mass numbers A_t as

$$\sigma_{s,t}(E,\mu,z) = \frac{\left(1+\frac{2\mu}{A_t(z)}+\frac{1}{A_t(z)^2}\right)^{3/2}}{1+\frac{\mu}{A_t(z)}}\left(\frac{Z_t(z)e^2}{4\pi\epsilon_0 m_0 v_p^2}\right)^2 \frac{1}{(1-\mu+2\eta(z))^2}, \tag{2.47}$$

where m_0 is the reduced mass which is defined by

$$\frac{1}{m_0} = \frac{1}{m_p} + \frac{1}{m_t(z)}$$

with m_p the mass of the proton and m_t the mass of the target nucleus, v_p is the incident speed of the proton, ϵ_0 is the vacuum permittivity, e is the elementary charge and

$$\eta(z) = \Theta_{min}^2(z) = \left(\frac{Z_t^{1/3}(z)\alpha m_e c}{p}\right)^2$$

with m_e the electron mass, α the fine structure constant, c the speed of light and p the momentum of the incident proton. Equation 2.47 is used to define the macroscopic scatter cross section as

$$\Sigma_s(\mu,E,N_{\mathbb{A}}(z)) = \sum_{i\in\mathbb{A}} N_i(z)\sigma_{s,i}(E,\mu,z) \tag{2.48}$$

with $N_i, i \in \mathbb{A}$ the individual atomic density in the material under consideration. This cross section can be seen in Figure 2.4 for a range of energies.

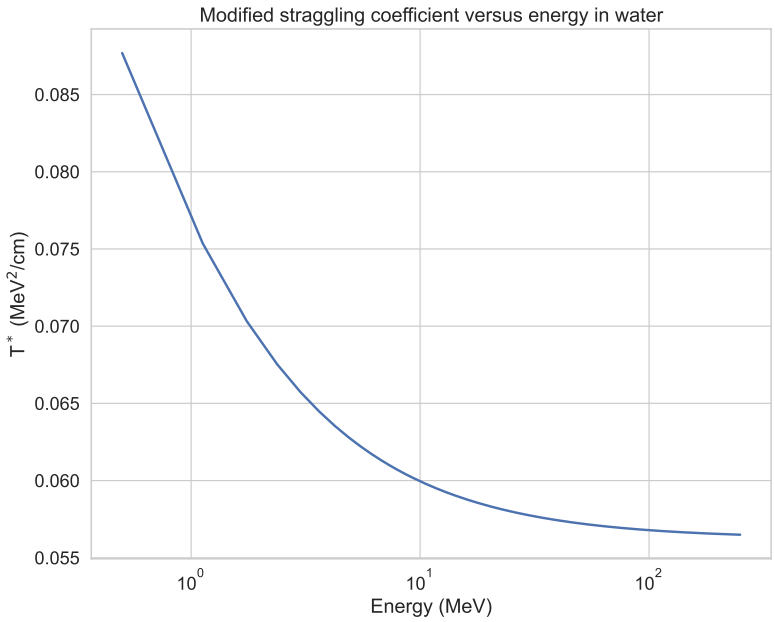


FIGURE 2.3: Water straggling coefficient versus proton energy.

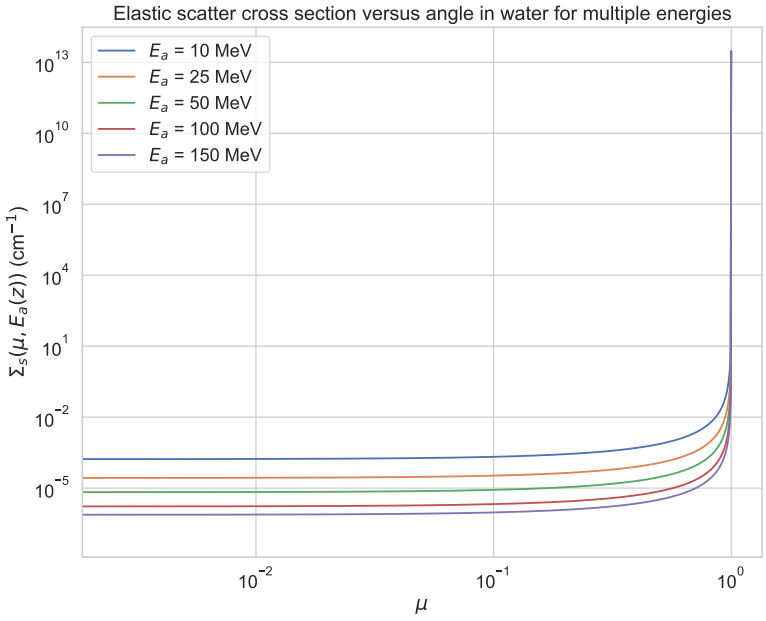


FIGURE 2.4: Water elastic scatter cross section for protons versus angle for multiple proton energies.

Chapter 3

Dose and dose change computations using YODA

3.1 Novelty and own contributions

This chapter details the extensive validation of the dose and dose change computation engine developed in Chapter 2 in realistic scenarios using patient CT scans, which required a number of own contributions. First, by generalizing the work of (Yang et al., 2020) a novel beam-splitting scheme was developed. This enabled YODA to handle lateral heterogeneities, overcoming a key limitation of the original Fermi-Eyges formalism requiring homogeneous lateral heterogeneities. Second, the conversion from spots in clinical DICOM treatment plans into inputs for YODA (or any other dose engine that requires a proton flux boundary condition) was implemented into a novel data processing pipeline, which allows YODA to be directly used in any proton therapy clinic in the world. Third, by changing the objective function and the specific constraints in the commissioning procedure of MCSquare (*MCsquare - Commissioning Procedure* 2024), better agreement and an improved speed of convergence in the optimization was achieved. Last, a comprehensive assessment of both YODA's dose and dose change computation capabilities was performed, thereby demonstrating YODA is capable of performing patient-specific quality assurance tasks.

3.2 Introduction

This chapter details the application of the mathematical formalism developed in Chapter 2 to the computation of doses and dose changes caused by changing HU values. In order to compute doses in laterally heterogeneous geometries, a Gaussian beam splitting scheme is developed in Section 3.3. Next, Section 3.4 presents and discusses the accuracy of YODA in computing doses and dose changes. Dose computation accuracy is assessed by taking TOPAS as a reference and the three-dimensional gamma index passing rate as the accuracy metric. First, water boxes with varying slab inserts are irradiated with energies spanning the whole clinical energy spectrum. Next, realistic CT scans are irradiated with either one or two spots, also with different energies. For the dose change computation accuracy, a treatment plan is created for a certain VOI and thereafter delivered to two different anatomies. The accuracy is defined as the error between re-computing the average dose (to the VOI) on the new anatomy, as opposed to computing it using the adjoint method. The test starts with a simple water tank in which slabs of

This chapter is based on the publication of Tiberiu Burlacu, Danny Lathouwers, and Zoltán Perkó (July 2024). "Yet anOther Dose Algorithm (YODA) for Independent Computations of Dose and Dose Changes Due to Anatomical Changes". In: *Physics in Medicine & Biology* 69.16, p. 165003. ISSN: 0031-9155. DOI: 10.1088/1361-6560/ad6373. (Visited on 08/16/2024)

different compositions are inserted in the beam path. Next, two patients had two plans (one robustly optimized and one non-robustly optimized) generated in RayStation. The chapter continues in Section 3.5 with an exposition of the procedures needed to read clinical TPS outputted treatment plans in YODA. Specifically, Subsection 3.5.1 details how two-dimensional treatment plan spots defined on the isocenter plane are converted into three-dimensional points on the surface of the CT cube, taking into account the physical measurements of the BDS. Next, Subsection 3.5.2 describes a "commisioning" procedure for YODA. This procedure ensures that a spot in YODA of a given energy produces a similar dose distribution to a spot in the desired TPS dose engine. Thus, Subsection 3.5.2 describes how to match the optical parameters of a pencil beam in YODA to the ones of the chosen gantry (HPTC gantry in this case), how to match the energy and energy spread of a spot so that the range between the two dose engines is in agreement and lastly how to calibrate the number of protons to MU conversion (a quantity that is clinic specific). Following the "commissioning" procedure, Subsection 3.5.3 contains comparisons between plans, for different sites and patients, computed in RayStation and YODA. The chapter ends with a conclusion presented in Section 3.6

3.3 Optimized Gaussian beam splitting

On the boundary of the computational domain, the lateral dependence of the six-dimensional phase-space density is described by

$$\Psi_{FE}^{z=0}(x,y) = \int_{4\pi} \varphi_{FE}(x,y,z=0,\Omega_x,\Omega_y)\mathrm{d}\hat{\mathbf{\Omega}} = \frac{1}{2\pi\sigma_s^2}\exp\left(-\frac{(x^2+y^2)}{2\sigma_s^2}\right), \tag{3.1}$$

where σ_s is the spatial standard deviation or spread of the x and y symmetric Gaussian. For the purpose of lateral beam splitting the original spot's central axis is placed at the origin of a 2D lateral grid. Given the radial symmetry of the Gaussian, placing sub-spots or beamlets on $N_r + 1$ concentric rings with radii r_i around the original spot location was chosen, in a similar manner to Yang's method (Yang et al., 2020). On a given ring i the beamlets share the same weight w_i and spread σ_i. The zeroth ring has a radius equal to zero and a single beamlet that is placed at the origin of the 2D lateral grid. Thus, the approximated fluence Ψ_{FE}^{a} is written as

$$\begin{aligned}\Psi_{FE}^{a}(x,y) &= \sum_{i=0}^{N_r}\sum_{k=1}^{n_i}\frac{w_i}{2\pi\sigma_i^2}\exp\left(-\frac{(x-x_{ik})^2+(y-y_{ik})^2}{2\sigma_i^2}\right), \\ x_{ik} &= r_i\cos\left(\frac{2\pi k}{n_i}+\alpha_i\right), y_{ik} = r_i\sin\left(\frac{2\pi k}{n_i}+\alpha_i\right),\end{aligned} \tag{3.2}$$

with n_i being the number of sub-spots placed on ring i, (x_{ik}, y_{ik}) are the coordinates of a sub-spot with index k on ring i and α_i is a ring-dependent angular offset (meant to improve coverage for consecutive rings with the same number of beamlets). Prior to the optimization the number of rings N_r, the number of points on each ring n_i and the ring offsets α_i are specified. As opposed to Yang's (Yang et al., 2020) approach this formalism and implementation is not restricted to a number of pre-defined schemes. In principle any number of beamlets per ring and number of rings can be optimized. The optimization parameters (weights, spreads and ring radii) are collected in a vector denoted by $\boldsymbol{\theta} \in \mathbb{R}^{3(N_r+1)}$ with a structure of $\boldsymbol{\theta} = (\ldots, w_i, r_i, \sigma_i, \ldots)$. The objective

function of the optimization problem is defined as

$$J(\boldsymbol{\theta}) = \iint_{-10\sigma_s}^{10\sigma_s} \mathrm{d}x\mathrm{d}y\,(\Psi_{FE}^{a} - \Psi_{FE}^{z=0})^2 \Big/ \iint_{-10\sigma_s}^{10\sigma_s} \mathrm{d}x\mathrm{d}y(\Psi_{FE}^{z=0})^2,$$

and is input into a SciPy implementation of a trust-region constrained algorithm (Virtanen et al., 2020; Lalee, Nocedal, and Plantenga, 1998). The weights w_i are bound constrained to be in the unit interval, namely $0 \leq w_i \leq 1, \forall i = 0, \ldots, N_r$ and are constrained such that

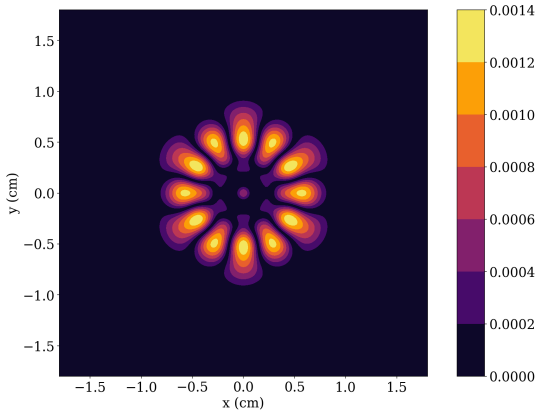
$$\sum_{i=0}^{N_r} w_i n_i = 1,$$

in order to ensure particle number conservation. To further guide the highly degenerate solution space towards useful splitting schemes, the ring radii are bound according to the initial spatial spread of the 2D Gaussian σ_s such that $0 \leq r_i \leq r_{i+1} \leq 2\sigma_s$. This evenly distributes the rings in $[0, 2\sigma_s]$ and avoids optimal but less useful configurations where all the rings are placed close to one another and the origin. Similarly, the spreads of the rings σ_i are bound such that $0.3\sigma_s \leq \sigma_i \leq \sigma_{i+1} \leq 0.8\sigma_s$. The first ring should have the smallest spread so that errors coming from the central axis are limited. In the case of a spot with an initial spread of $\sigma_s = 0.3\,\mathrm{cm}$ Figure 3.1 shows for three different splitting schemes the absolute difference between $\Psi_{FE}^{z=0}(x, y)$ and $\Psi_{FE}^{a}(x, y)$ in the left column and the actual positions of the beamlets on the concentric rings together with the optimized spreads (indicated by the circle radii) around each spot in the right column.

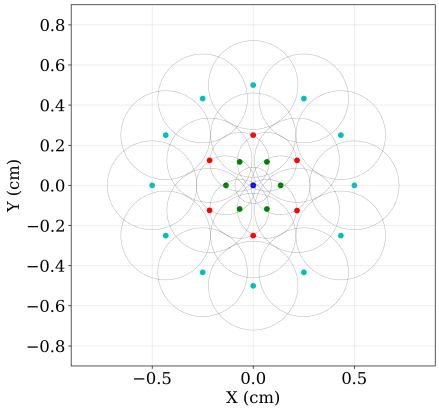
3.4 Dose computations and discussion

3.4.1 Dose engine performance

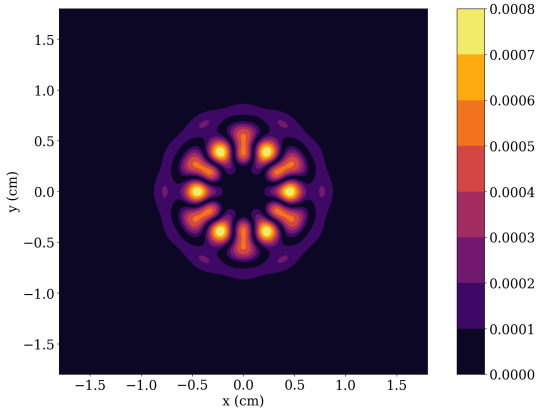
The dose engine in YODA was benchmarked against TOPAS in several irradiation test-cases such as homogeneous and hetereogeneous water tanks, head and neck (H&N), prostate and lung CTs. TOPAS simulations were performed using the em-opt4 physics list which is the most accurate modelling of electromagnetic interactions available within TOPAS. Nuclear interactions were excluded from this comparison as YODA does not currently account for nuclear interactions. In all TOPAS simulations the number of protons per spot was set to 10^8 and the maximum number of available cores (48) was used. Using this physics list and number of cores, the run-times of TOPAS were in the order of hours. In all test cases, a YODA spot was split according to a $1 + 6 + 6 + 12 + 12 + 24$ Gaussian beam splitting scheme as this was found to yield accurate results when compared to TOPAS. For this splitting scheme on average one spot takes 2 s to compute. Additional speed-ups could be achieved in two ways. One is to address the main speed limitation (memory access bandwidth) by implementing the algorithm on a graphics processing unit card. The second is to implement an adaptive energy grid on a per sub-spot level. Currently the energy grid is divided into a fixed number of groups which results in the majority of the groups and thereby the system solved at each step being empty. By adapting the energy grid to be finely discretized in the locations in energy where the flux has significant values and coarse everywhere else significant speed-ups can be expected.



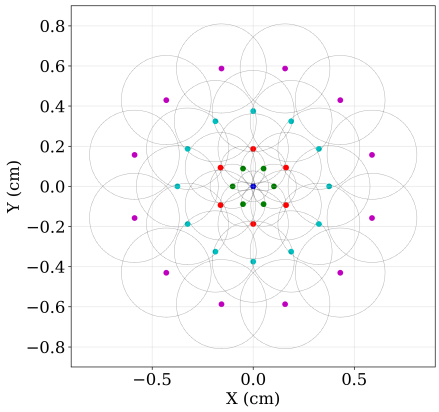
(A) Configuration 1+6+6+12



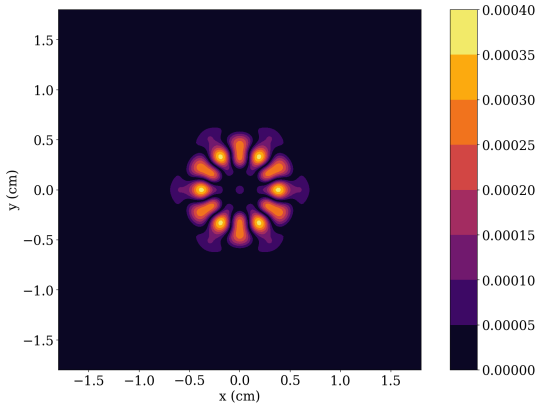
(B) Configuration 1+6+6+12



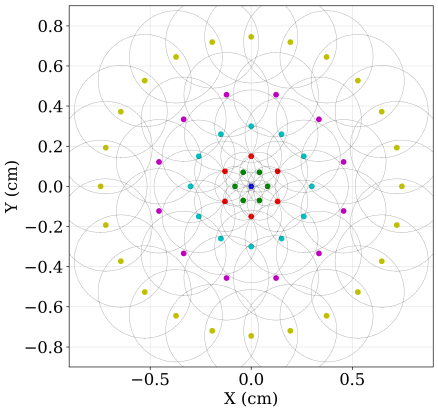
(C) Configuration 1+6+6+12+12



(D) Configuration 1+6+6+12+12



(E) Configuration 1+6+6+12+12+24



(F) Configuration 1+6+6+12+12+24

FIGURE 3.1: Left column: absolute normalized difference between Ψ^0_{FE} and Ψ^a_{FE} using the optimized parameters. Right column: corresponding physical optimized positions of the individual beamlets in the lateral plane. Points with the same color are on the same ring, circle radii are σ_i on ring i.

Simplified tank geometries

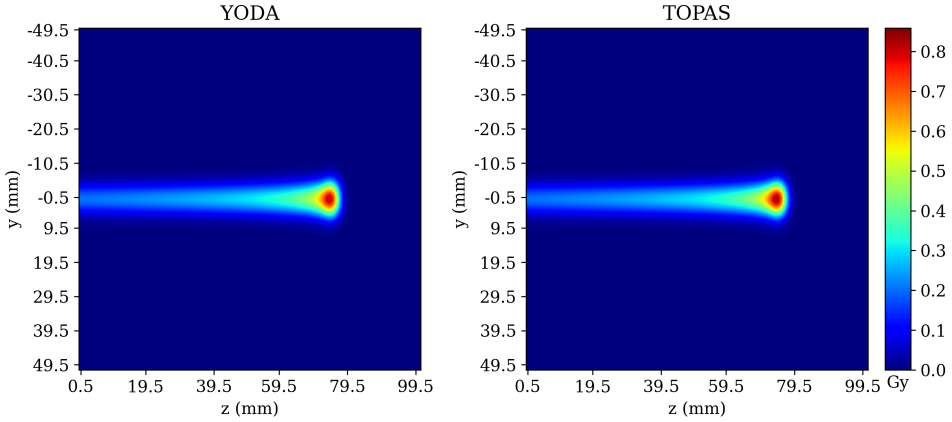
First, three tank-based tests, that are typical for benchmarking pencil beam algorithms, were performed. In all three cases, a tank (of dimensions of $10 \times 10 \times 10$ cm^3) was irradiated with a spot with nominal energy 100 MeV, an energy spread of 1 MeV, a spot size of 0.3 cm, an angular spread of 1×10^{-8} rad and a correlation of 0. The first case, denoted by (a), is the one in which the tank is composed homogeneously of water (0 HU). In the other two cases, a half-plane slab is introduced in the tank between 2 and 3 cm in depth in the upper-half of the x-y plane (with z being the depth). This is usually one of the most challenging geometries for pencil beam algorithms. In one case, denoted by (b), the slab was composed of bone-like tissue of 1000 HU and in the other, denoted by (c), it was composed of air-like tissue of −1000 HU. The tank was created using an in-house DICOM CT scan writer and was composed of $100 \times 100 \times 100$ voxels with a voxel size of $0.1 \times 0.1 \times 0.1$ cm^3. Two-dimensional slices of the dose distributions of YODA and TOPAS can be seen in Figure 3.2. Integrated depth doses (IDDs) and lateral profiles at different depths along the original spot axis can be seen in Figure 3.3.

For these simple test cases, the visual agreement is excellent, as illustrated by both Figure 3.2 and Figure 3.3. This is also reflected in the 3D gamma index pass rates shown in Table 3.1 under the columns denoted by −1000 HU, 0 HU and 1000 HU. The worst passing rate using the strict 1 mm, 1 %, 10 % dose cutoff is 98.22 %. All passing rates presented can be further improved by fine tuning the splitting scheme. One way of doing so is to increase the number of rings. Another, is to take advantage of the underlying CT grid in the case of this perpendicular propagating spot. If in the lateral beam eye view grid, one beamlet is placed per voxel and the spread is contained to the voxel lateral dimensions, the error is bound to decrease without much increase in computational cost.

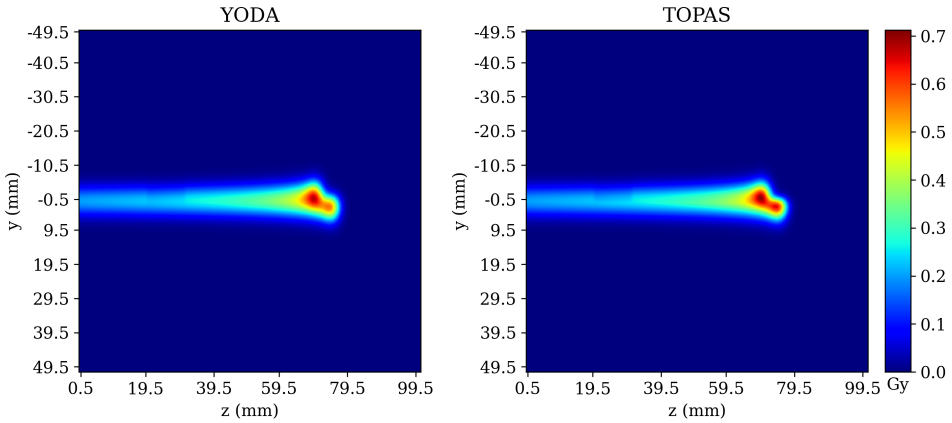
Table 3.1: Gamma index passing rates for different criteria and test cases.

Gamma index									
Criteria			Passing rates (%) for						
mm	%	% - cutoff	-1000 HU	0 HU	+1000 HU	H&N	Prostate	Lung 1	Lung 2
1	1	0	99.96	100	99.99	100	100	100	99.99
1	1	10	98.22	99.93	99.45	99.85	99.58	95.62	94.55
2	2	0	100	100	100	100	100	100	100
2	2	10	99.61	99.95	99.78	99.99	99.99	99.72	98.09
3	3	0	100	100	100	100	100	100	100
3	3	10	99.73	100	99.85	99.99	100	99.86	99.12

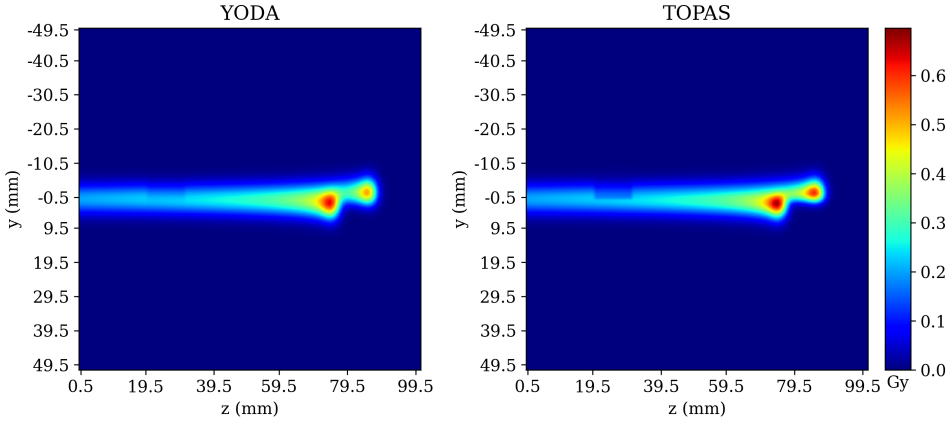
Next to the simple geometries, a more challenging sliding slab experiment was also performed. In this experiment, the slab is moved with respect to the central axis of the beam from −4 mm to 4 mm in increments of 2 mm. As in the previous tests, the composition of the slab is set to either -1000 HU or 1000 HU and its depth is kept between 2 cm and 3 cm. To assess the accuracy of the dose algorithm across the clinical energy spectrum, beam energies of 70 MeV, 160 MeV, 190 MeV and 230 MeV are tested. The beam spread is set to 1.0 in TOPAS, which implies spreads of 0.7 MeV, 1.6 MeV, 1.9 MeV and 2.3 MeV. The remaining beam characteristics are kept identical to the previous test cases, i.e., a spread of 0.3 cm, an angular spread of 1.0×10^{-8} rad and a correlation of 0. Figures 3.4, 3.5 and 3.6 show the best, average and the worst cases of these tests.



(A) Comparison in a homogeneous 0 HU tank

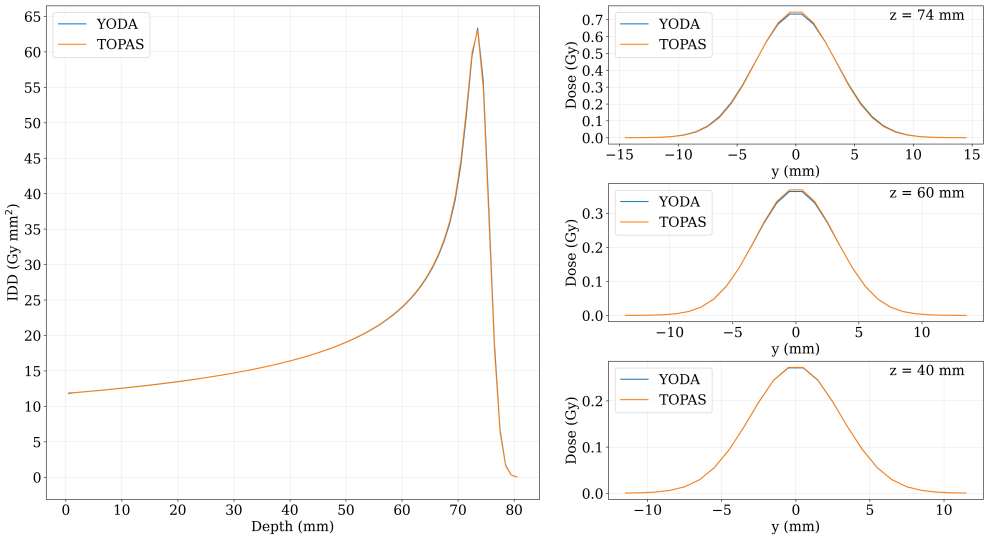


(B) Comparison when a slab of +1000 HU is introduced between 2 and 3 cm

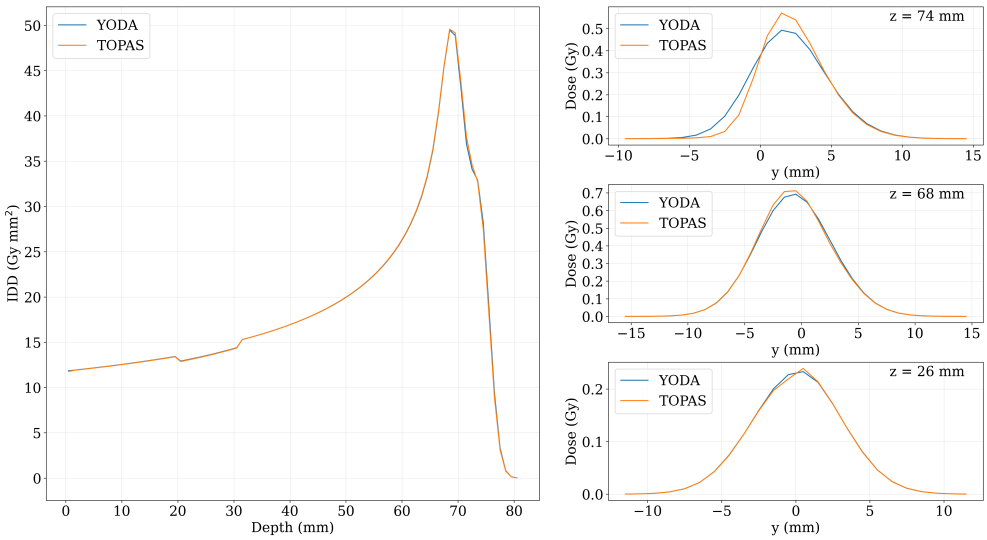


(C) Comparison when a slab of -1000 HU is introduced between 2 and 3 cm

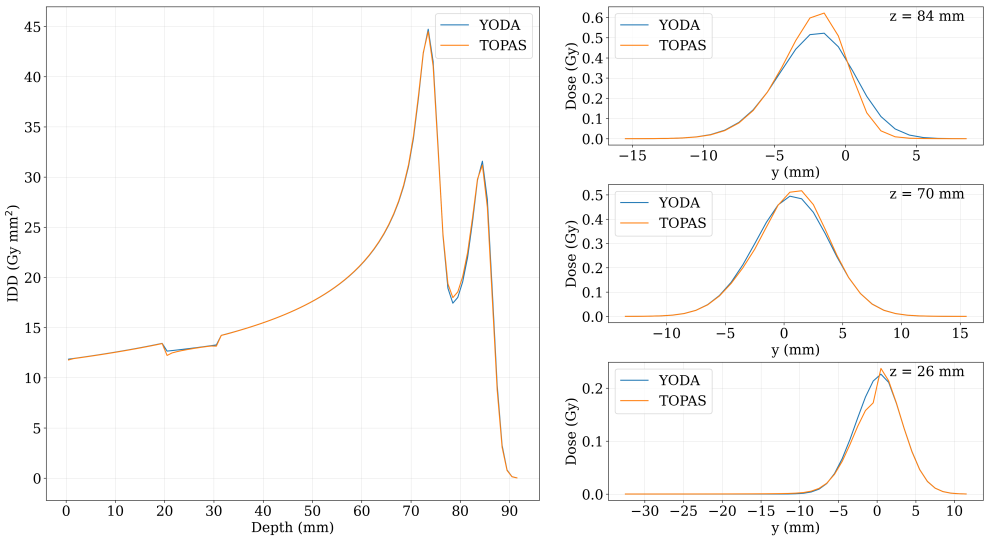
FIGURE 3.2: Dose comparison between YODA and TOPAS in the three tank-based tests. The figures illustrate 2D dose slices along the central beam axis for YODA in the left column and for TOPAS in the right column.



(A) Comparison in a homogeneous 0 HU tank



(B) Comparison when a slab of +1000 HU is introduced between 2 and 3 cm



(C) Comparison when a slab of -1000 HU is introduced between 2 and 3 cm

FIGURE 3.3: Dose comparison between YODA and TOPAS in the three tank-based tests. The figures illustrate IDDs and lateral profiles along the central beam axis for both codes.

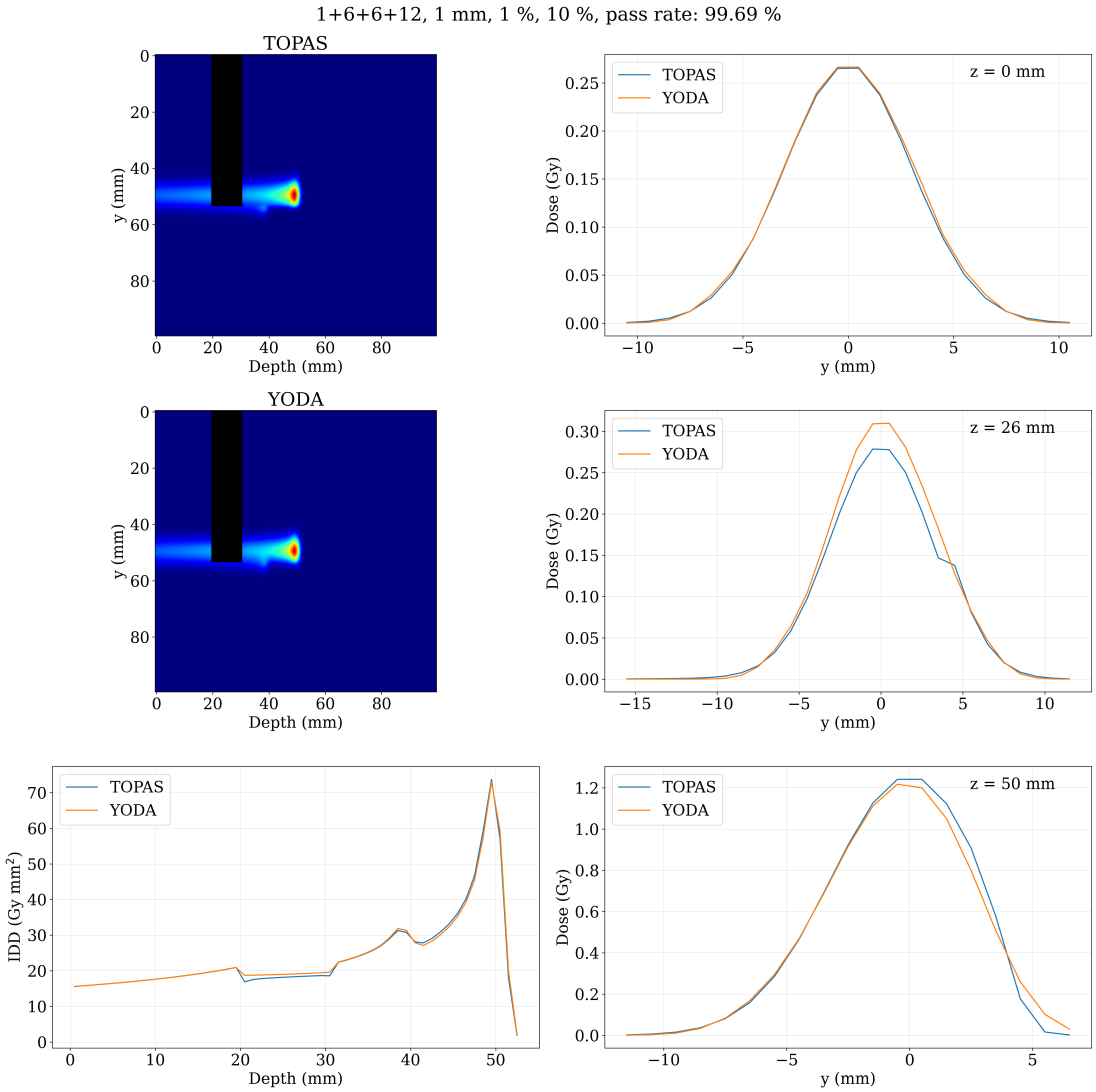


FIGURE 3.4: Best performing case for the sliding slab experiment. The slab, composed of 1000 HU, was positioned at 4 mm off the central beam axis and the beam energy was 70 MeV. The figure displays in the left column, for both YODA and TOPAS, two dimensional dose cuts along the central beam axis and IDDs and in the right column lateral profiles at the entrance, in the slab and in the Bragg peak.

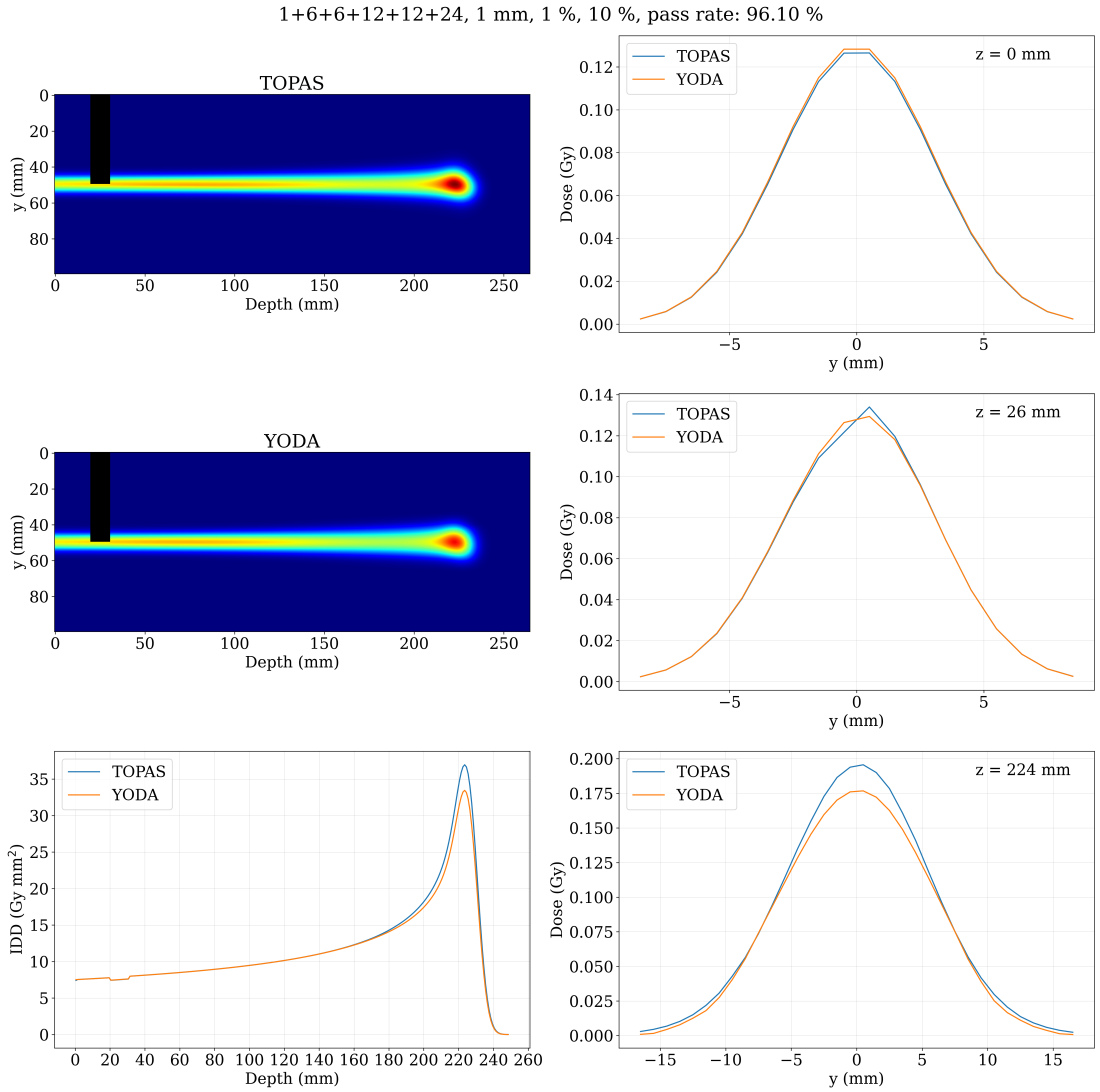


FIGURE 3.5: Average performing case for the sliding slab experiment. The slab, composed of 1000 HU, was positioned at 0 mm off the central beam axis and the beam energy was 190 MeV. The figure displays in the left column, for both YODA and TOPAS, two dimensional dose cuts along the central beam axis and IDDs and in the right column lateral profiles at the entrance, in the slab and in the Bragg peak.

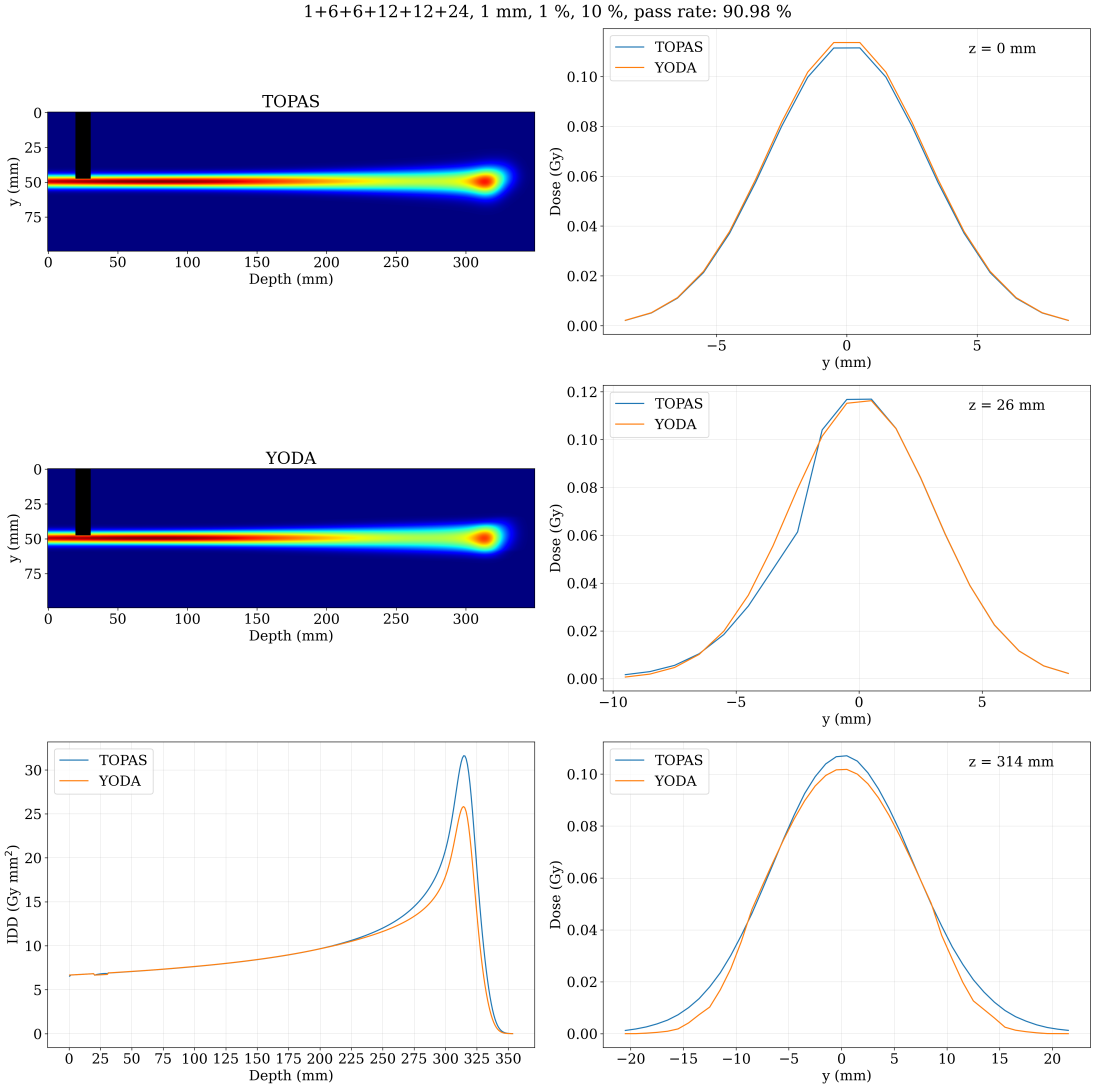


FIGURE 3.6: Worst performing case for the sliding slab experiment. The slab, composed of -1000 HU, was positioned at −4 mm off the central beam axis and the beam energy was 230 MeV. The figure displays in the left column, for both YODA and TOPAS, two dimensional dose cuts along the central beam axis and IDDs and in the right column lateral profiles at the entrance, in the slab and in the Bragg peak.

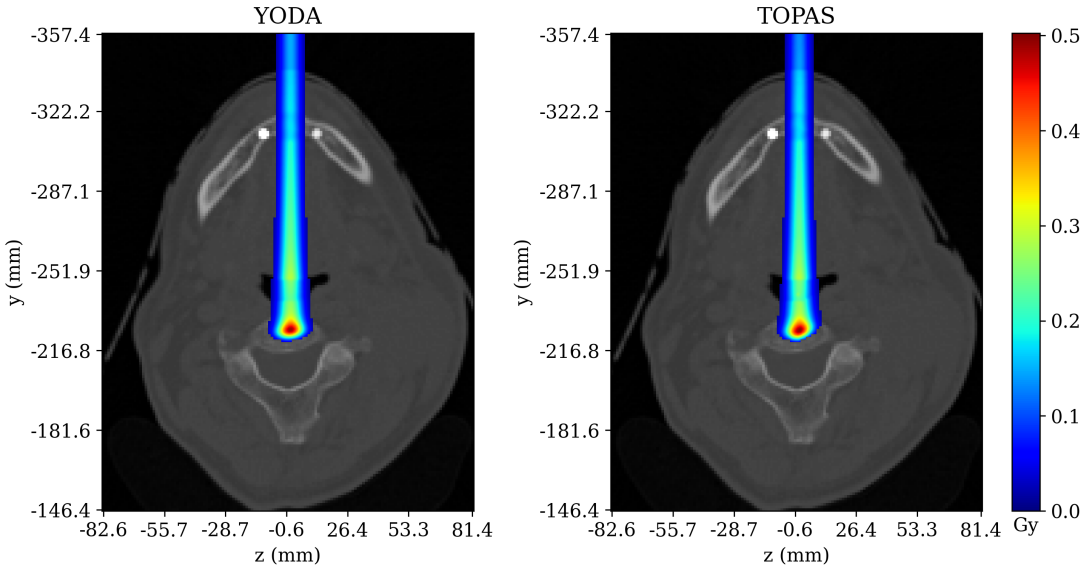
A full overview of the 120 test cases is presented in the tables from Appendix C. Despite the broad range of energies and challenging lateral heterogeneities encountered, the experiment shows that YODA computes doses with high passing rates using the strictest gamma index criteria. In addition to the conclusions drawn from Table 3.1, the tables presented in Appendix C provide three main findings. One, more beams do not necessarily imply a more accurate result. This is supported by Table C.8, where a simpler beam splitting scheme of 1+6+6+12 often performs better than the more complex 1+6+6+12+12+24 one. This is likely due to the interplay of several factors. The chosen beam size, the underlying CT grid size (and the projection of this in the beam-eye view coordinate system), the specific location on the rings of the optimized sub-spots and the location of the heterogeneity itself all play a role in the accuracy of a given scheme. Given this, and the fact that in a realistic treatment plan spots are placed in close proximity one to another, it is likely that YODA will result in accurate and quickly computed doses using simpler beam splitting schemes. Second, as already illustrated in Table 3.1 and further illustrated in Tables C.3 and C.4 the accuracy of YODA with respect to TOPAS is slightly worse when air gaps are placed in the beam path. Such small differences can arise due to the inherent limitations of the Fermi-Eyges modelling and the different modelling of Coulomb elastic scattering in air between the two codes. Third, there is a slight degradation of accuracy occuring towards the high part of the energy spectrum (effect that was not observed towards the low part of the energy spectrum). This could be explained by a number of factors. The first one comes from differences in the underlying data that the two codes use. The stopping powers were extracted from TOPAS using increasingly coarse steps in energy towards the high side of the energy domain. This can cause slight range differences, especially as the stopping power is linearly interpolated in the energy groups, which in turn can result in range differences. Differences in stopping power imply differences in lateral scattering which contribute to further differences. Moreover, the straggling coefficient is computed using an analytical equation that could prove inaccurate for the high energy part of the domain. Despite these differences, the gamma index pass rate using the strictest criteria and a splitting scheme of 1+6+6+12 does not fall below $\approx$ 95 % for energies above 190 MeV for all the tested geometries.

CT based anatomies

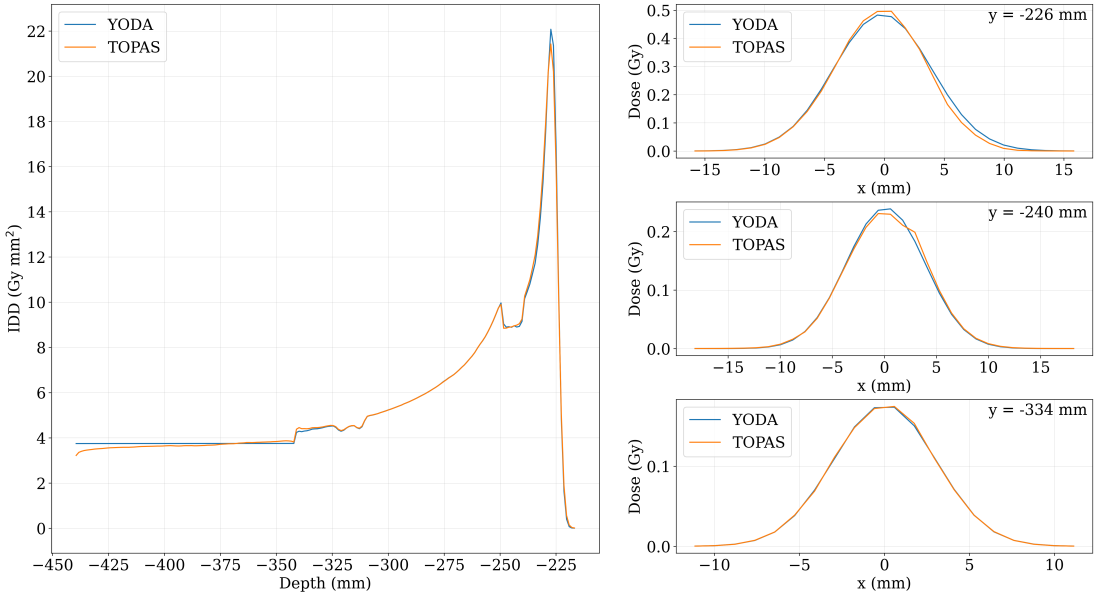
In addition to the tank-based tests, three real CT images were also tested. The H&N scan was taken from the CORT dataset (Craft et al., 2014), the prostate scan was taken from the cancer imaging archive (Yorke et al., 2019) and the lung scan was taken from the Holland Proton Therapy Center (Pastor Serrano, 2023). The used isocenter locations and gantry angles are not meant to be clinical and were chosen only due to their simplicity of set-up in TOPAS.

In the H&N case, one spot was irradiated with the beam impinging along the y axis (i.e., at a gantry angle of 0°) with a nominal beam energy of 125 MeV with the isocenter being the center of the CT scan volume. The two dimensional dose profile can be seen on the top row of Figure 3.7 and the IDD and lateral profiles at three depths can be seen in at the bottom of Figure 3.7. Good agreement is observed, as the 99.85 % gamma index pass rate from the H&N column of Table 3.1 also shows. Figure 3.7 shows a discrepancy in the air region between −440 mm and −340 mm. This is also the case for the lung and prostate cases. Two possible reasons are differences in the modelling of air between the two algorithms or a slight mismatch in the positioning of the beams with respect to the CT grid caused by the placement of the beam at the interface of

voxels. Given that the agreement is good in the clinically relevant region of the scan, this discrepancy is deemed acceptable.



(A) 2D dose slices along the central beam axis for YODA, in the left column, and TOPAS, in the right column.

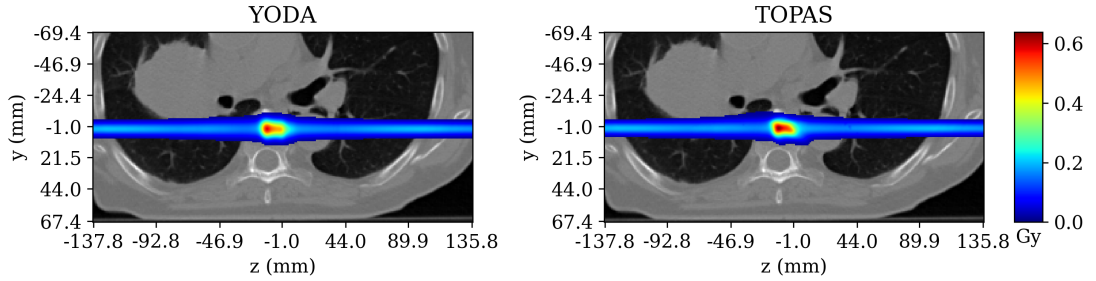


(B) IDDs and lateral profiles at three different depths for both YODA and TOPAS.

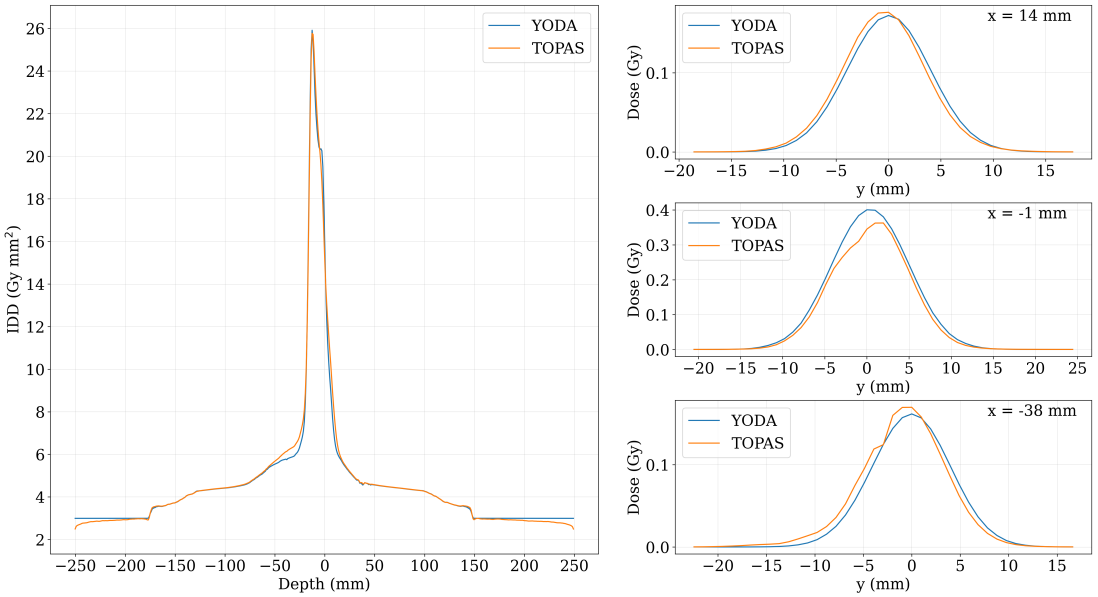
FIGURE 3.7: H&N test case dose comparisons. The scan is irradiated by one spot of 125 MeV.

The lung scan was irradiated with two spots where one beam goes from $-x$ to $+x$ and the other in the opposite direction (i.e., at 90° and 270° gantry angles respectively). Both spots had a mean energy of 125 MeV, energy spread of 1 MeV, a spot size of 0.3 cm, an angular spread of 1.0×10^{-8} rad and a correlation of 0. This cases is denoted by Lung 1. Given the challenging anatomy, the results from Figure 3.8 together with the passing rate of 95.62 % from the Lung 1 column of Table 3.1 are very good.

To further test YODA's dose engine performance, a second test for the lung was performed where the Bragg peak was moved towards a more heterogeneous area by changing the beam energies. In this case, one beam had an energy of 105 MeV with a



(A) 2D dose slices along the central beam axis for YODA, in the left column, and TOPAS, in the right column.

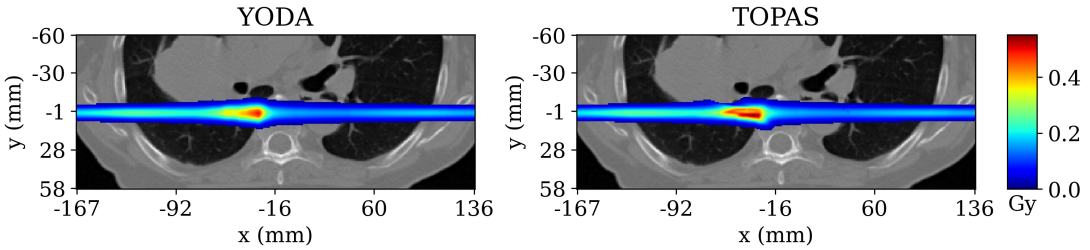


(B) IDDs and lateral profiles at three different depths for both YODA and TOPAS.

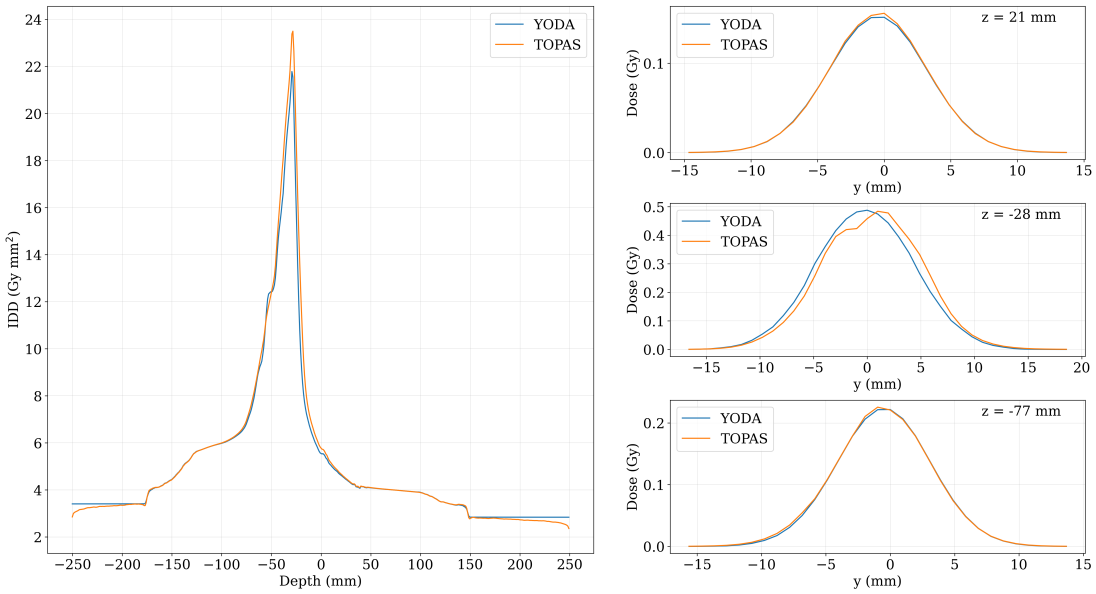
FIGURE 3.8: Lung case 1 dose comparisons. The scan is irradiated by 2 opposing spots with mean energies of 125 MeV.

spread of 0.84 MeV and the other an energy of 135 MeV with a spread of 1.08 MeV. This case was denoted by Lung 2. Here too, despite the challenging heterogeneous anatomy, YODA performs well given that the worst gamma index passing rate is 94.55 % (as seen in column Lung 2 in Table 3.1). Two dimensional profiles, IDD comparisons and lateral profiles can be seen in Figure 3.9. The lateral profiles from Figures 3.8 show a consistent lateral shift between YODA and TOPAS at the 14 mm, −1 mm and −38 mm depths. A reason for this could be the initial location of the Gaussian split sub-spots on the CT scan surface. The spots are generally not aligned with the CT grid (as such alignment is only possible in cases of perfectly perpendicular beams) and therefore slight asymmetries could arise if spots are placed exactly at the interface of voxels. The accuracy can be improved by fine-tuning the Gaussian beam splitting scheme in several ways. One is to include the number of rings and the number of beamlets per ring into the optimization procedure itself. Another is to consider alternative, non-concentric sub-spot arrangements. A metric for lateral heterogeneity could help in guiding the optimization towards sparsely placing beamlets in areas of low heterogeneity and more densely covering areas with high heterogeneity. Lastly, a progressive splitting scheme could also be employed, whereby once a threshold of lateral heterogeneity has been reached, the beamlets encountering it are re-consolidated and a new (finer) split occurs. Given that the parameters of such schemes can be pre-optimized and tabulated the

computational increase of such an approach could be kept minimal.



(A) 2D dose slices along the central beam axis for YODA, in the left column, and TOPAS, in the right column.



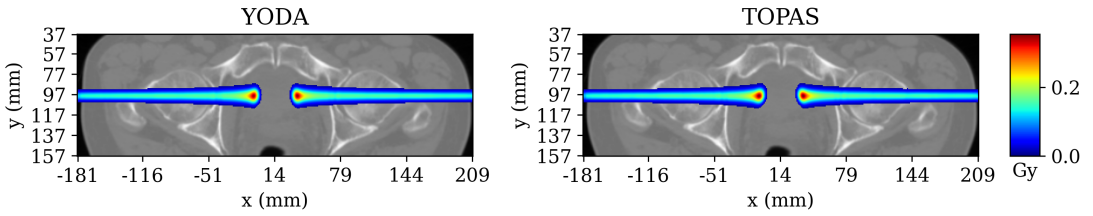
(B) IDDs and lateral profiles at three different depths for both YODA and TOPAS.

FIGURE 3.9: Lung case 2 dose comparisons. The scan is irradiated by 2 opposing spots with mean energies of 105 MeV and 135 MeV.

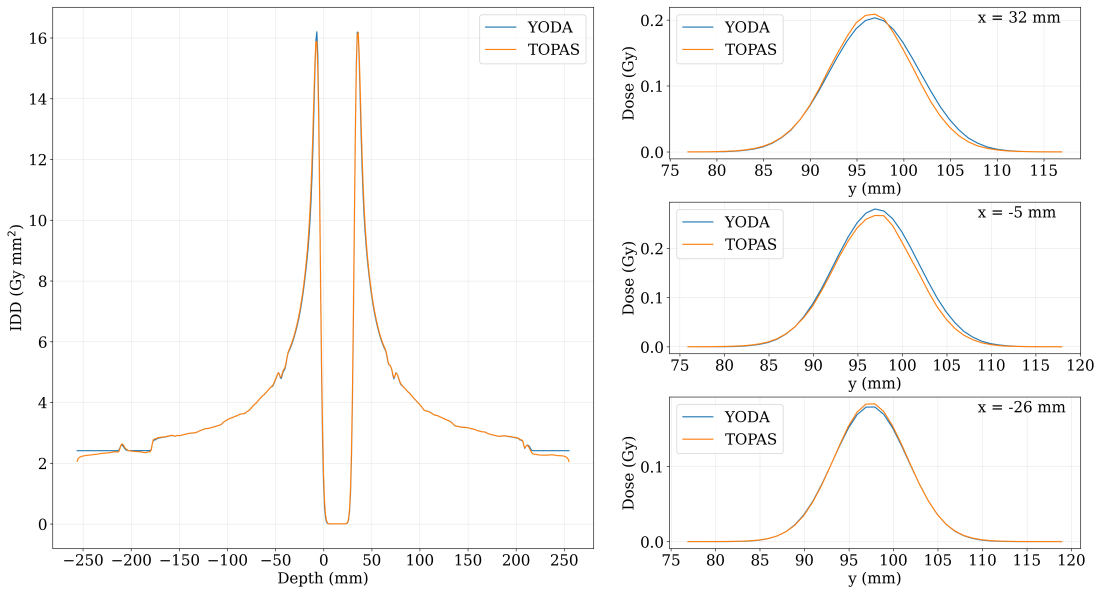
The prostate case set-up was identical to that of the lung with the only difference being the spot mean energy of 165 MeV and the spot energy spread of 0.825 MeV. Here again the agreement is very good as seen in Figure 3.10 and by the high passing rate of 99.58 % from the prostate column of Table 3.1.

3.4.2 Dose change computations

In addition to the dose engine performance, the performance of the average dose change computation was also benchmarked. Given a specific volume within the CT scan denoted as ROI, the adjoint component is able to cheaply and accurately compute the change in the dose deposited in the ROI (for small enough anatomical perturbations). The speed of such an operation far exceeds that of plain re-computation as effectively, the only computation necessary comes in the form of vector inner products. This could be employed in an online re-adaptation trigger system where YODA assesses the effect of delivering yesterday's plan on today's anatomy. For all the test cases, the metric computed was the average dose to the ROI as a function of the HU values of CT scan. The benchmark starts with the same simplified tank test-cases and thereafter moves toward more realistic cases using RT plans for clinical RT structures on CT images.



(A) 2D dose slices along the central beam axis for YODA, in the left column, and TOPAS, in the right column.



(B) IDDs and lateral profiles at three different depths for both YODA and TOPAS.

FIGURE 3.10: Prostate case dose comparisons. The scan is irradiated by 2 opposing spots with mean energies of 165 MeV.

Simplified tank geometries

In the case of the simple tank geometries, the adjoint component used a ROI defined as everything past the depth of 60 mm in the tank. The composition of the half-slab was varied from −1000 HU to 1000 HU. The mean dose deposited in the ROI was computed for each new geometry using two methods: re-computations and adjoint computations. Figure 3.11 shows the mean dose deposited in the ROI as a function of the HU composition of the slab. The two lines are close one to another around the value of 0 HU which was considered the base case and they start to diverge towards the edges of the HU domain. The maximal relative error of 2.2 % occurs at the −1000 HU end of the HU domain. Based on these results, it can be concluded that the adjoint component is capable of cheaply and accurately computing the change in the deposited dose in the ROI for this test case.

Treatment plan tests

Four treatment plans for the gross tumor volume (GTV) were generated for two H&N patients (patients 1 and 2) in RayStation (Bodensteiner, 2018). For both patients, one robustly and one nonrobustly optimized plans were created. The plans are not clinical and are only used for the purpose of creating conformal doses around the target. Patient 1 had plans that contained roughly 1000 spots and patient 2 had plans with roughly 300 spots. Both plans were split according to a $1 + 6 + 6 + 12$ Gaussian beam splitting

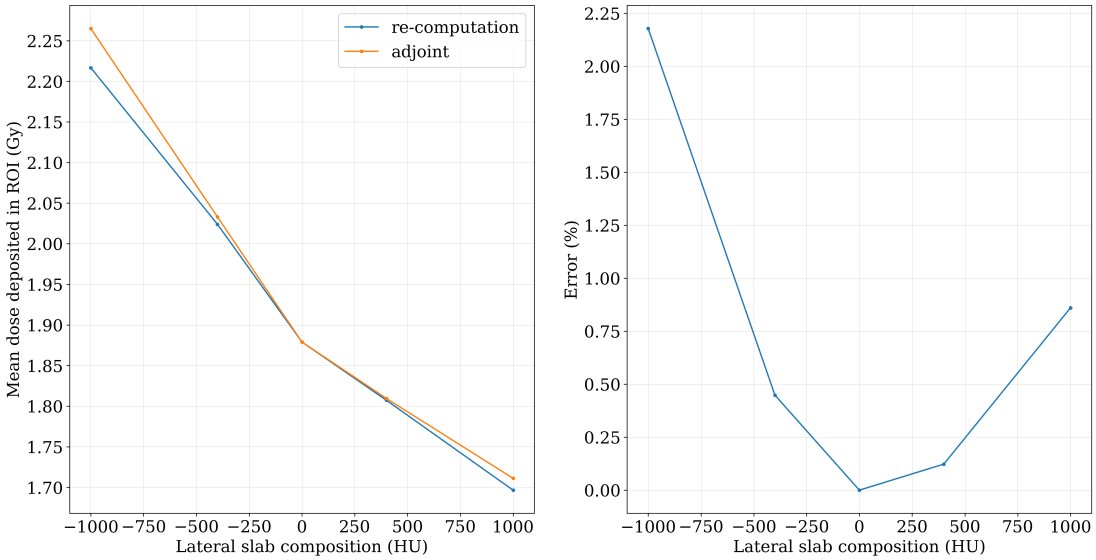


FIGURE 3.11: On the left the re-computed versus adjoint computed doses for the simplified tank geometries are displayed. On the right, the error, in percentages, between these two quantities is displayed.

scheme. Each patient had multiple repeat CTs (rCTs) which were registered to the planning CT (pCT) using the simple-itk library (Beare, Lowekamp, and Yaniv, 2018). The adjoint component computed the change in the GTV dose caused by the new CT image. This is meant to simulate the situation of a daily re-adaptation trigger system where the effect of yesterday's plan is assessed on today's anatomy. As long as the anatomical changes between the planning and repeat CT images are not too large, the adjoint component is accurate and fast as it does not require re-computing the original plan on the new image.

Figures 3.12 and 3.13 show each of the CT images for patient 1 (image number 0 is the planning image), a 2D dose slice of the re-computed average dose distribution on the CT image, the GTV dose computed via re-computation and via the adjoint component and the relative error between these two results. In the case of a non-robustly optimized plan, the adjoint component attains a maximal error of 5.5 % as presented in Figure 3.12. In the case of the robustly optimized plan, the adjoint component attains a maximal error of 4.8 %. Thus, whether the plan is robustly or non-robustly optimized, the adjoint component is capable of avoiding an expensive re-computation attaining an acceptable error.

Figures 3.14 and 3.15 shows each of the CT images for patient 2 (image number 0 is the planning image), a 2D dose slice of the re-computed dose distribution on the CT image, the GTV dose computed via re-computation and via the adjoint component and the relative error between these two results. In the case of the non-robustly optimized plan, the adjoint component attains a maximal error of 5.7 % and in the case of the robustly optimized plan the adjoint component attains a maximal error of 1.3 %. Here too the adjoint component computes the dose to the GTV with acceptable errors and is thus capable of avoiding expensive re-computations.

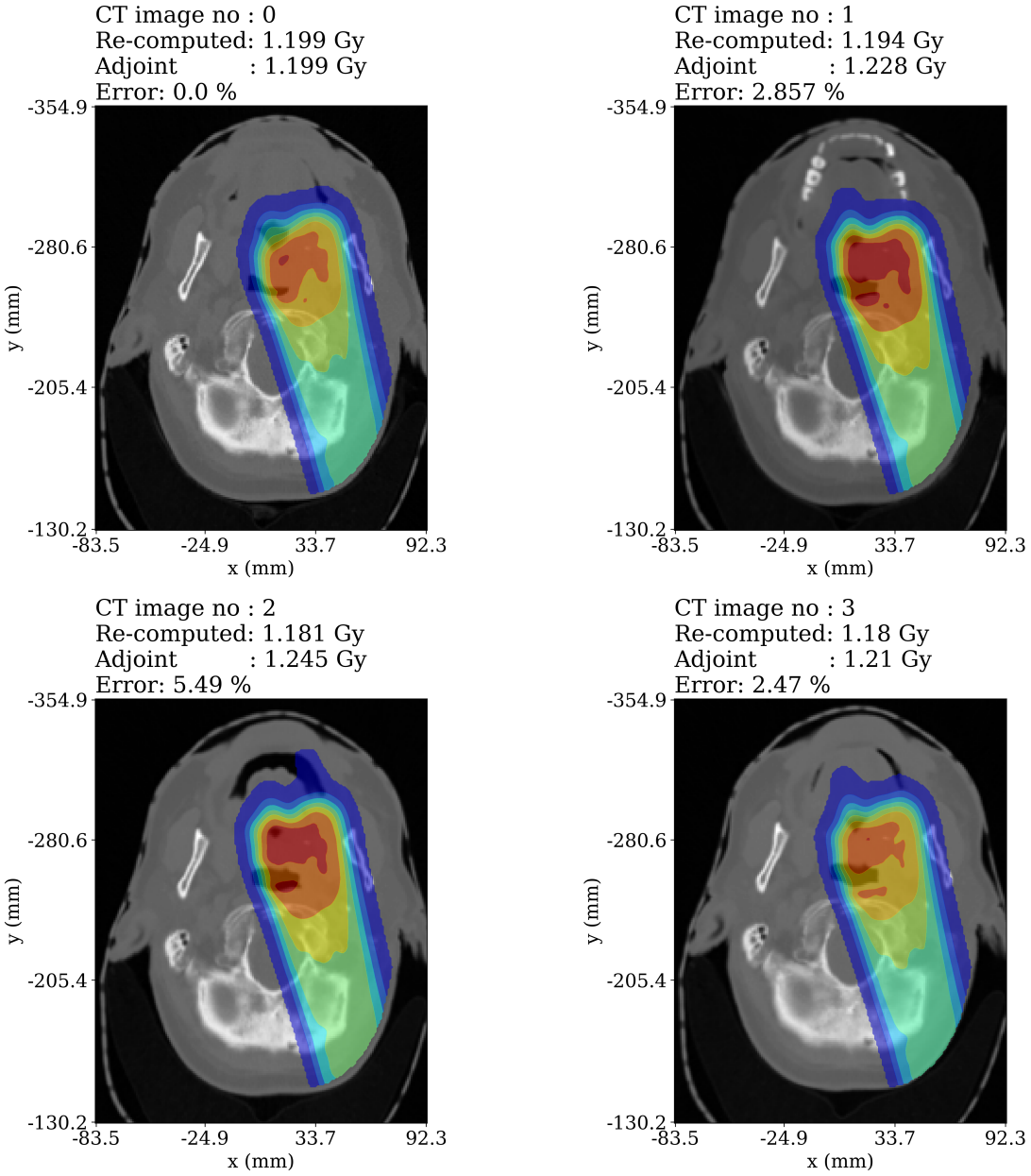


FIGURE 3.12: Color-coded are the re-computed dose distributions from a non-robustly optimized plan, overlaid on the CT images of Patient 1. Above each image, the mean dose to the GTV is stated based on the recomputation and on the adjoint method. The relative difference is also given.

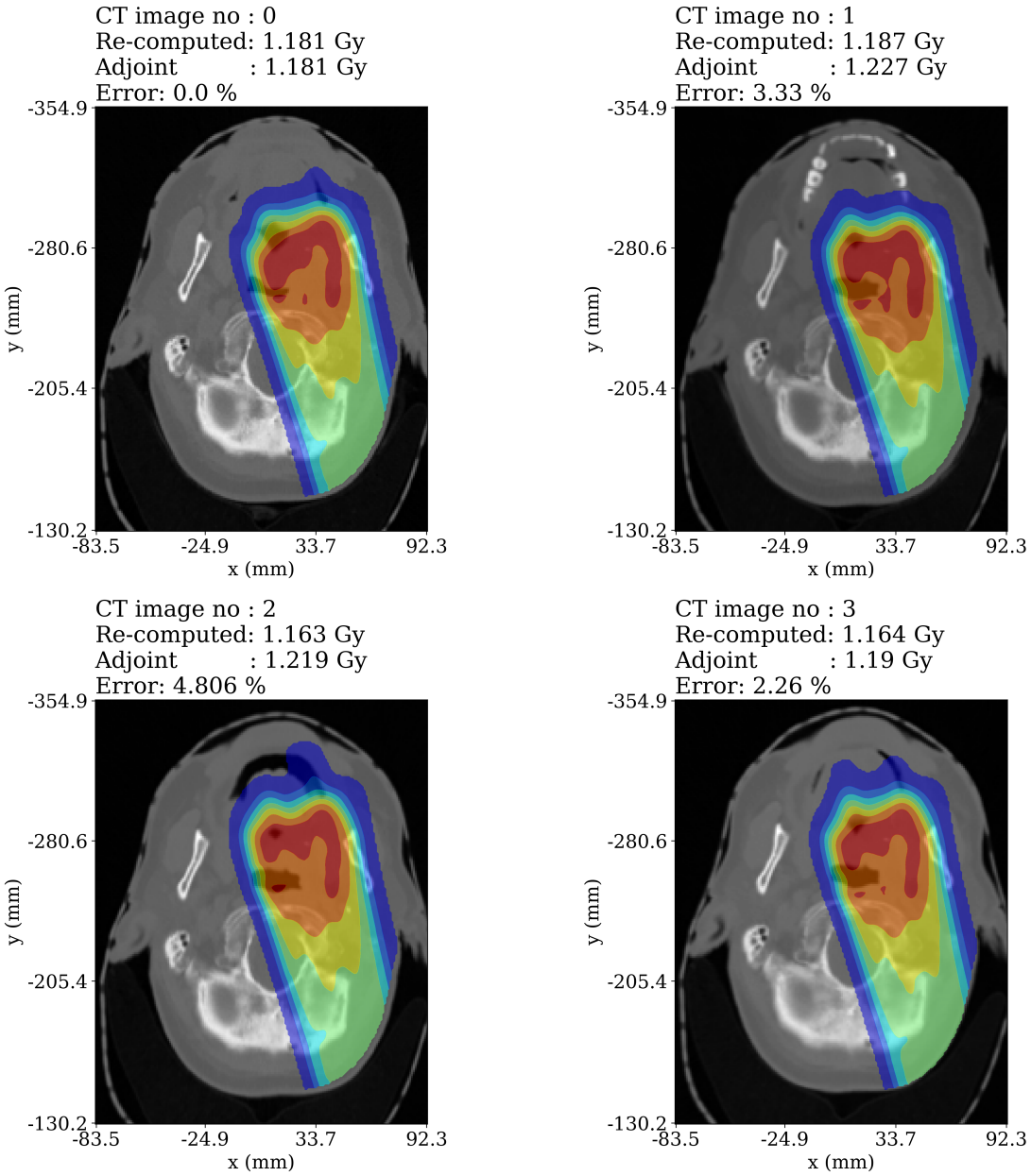


FIGURE 3.13: Color-coded are the re-computed dose distributions from a robustly optimized plan, overlaid on the CT images of Patient 1. Above each image, the mean dose to the GTV is stated based on the recomputation and on the adjoint method. The relative difference is also given.

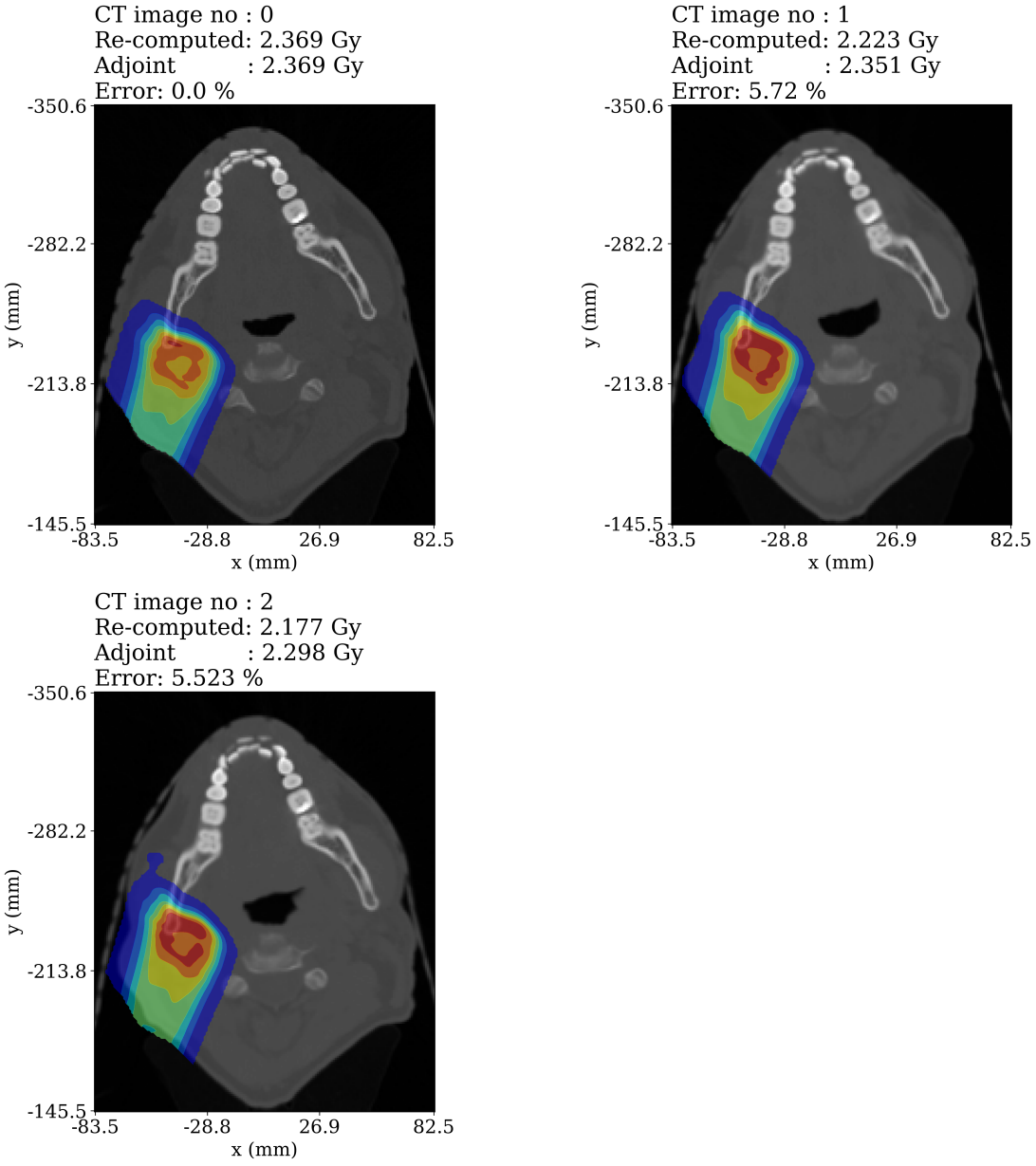


FIGURE 3.14: Color-coded are the re-computed dose distributions from a non-robustly optimized plan, overlaid on the CT images of Patient 2. Above each image, the mean dose to the GTV is stated based on the recomputation and on the adjoint method. The relative difference is also given.

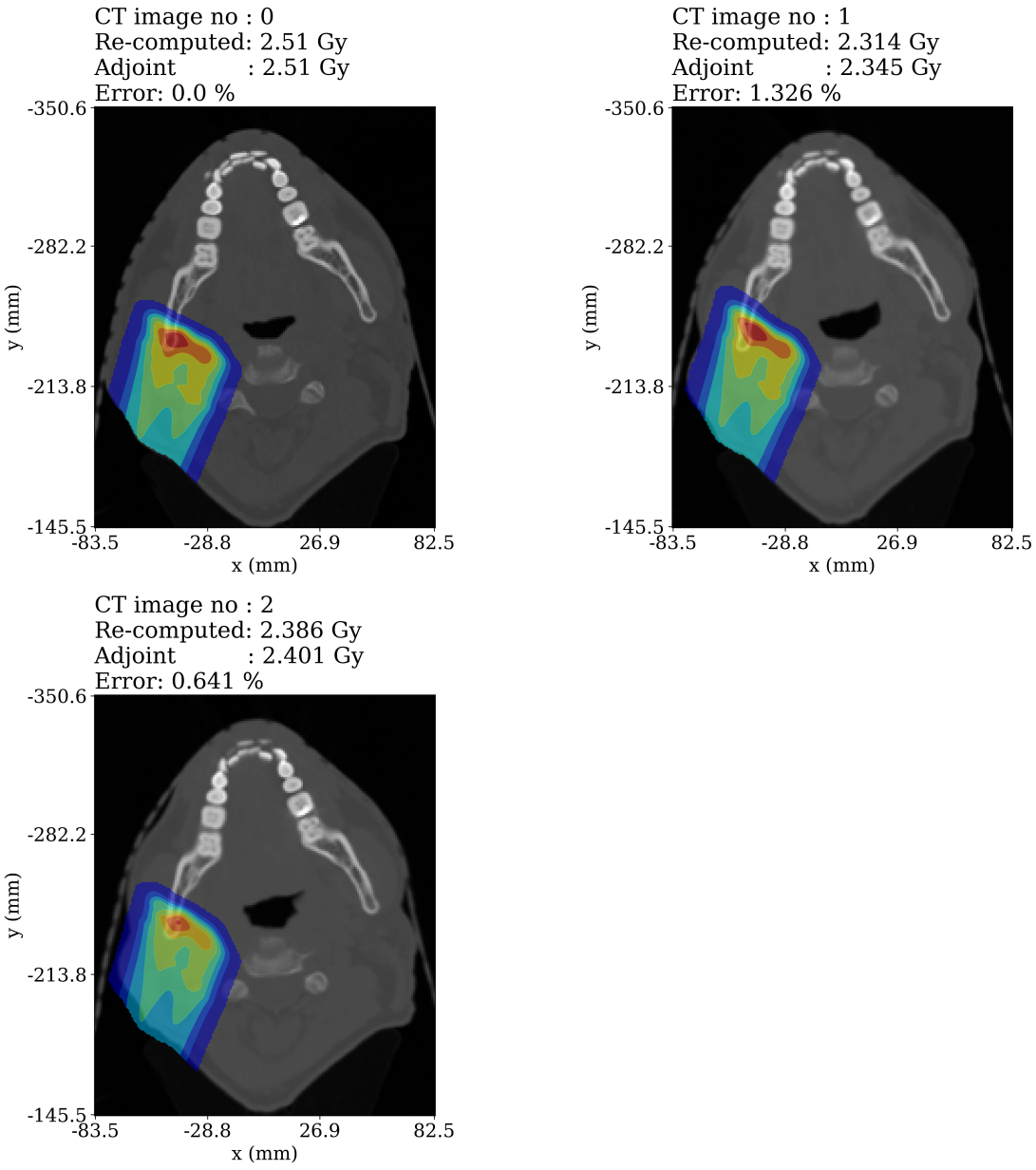


FIGURE 3.15: Color-coded are the re-computed dose distributions from a robustly optimized plan, overlaid on the CT images of Patient 2. Above each image, the mean dose to the GTV is stated based on the recomputation and on the adjoint method. The relative difference is also given.

3.5 Dose engine commissioning

This section details the procedures necessary to compute clinical treatment plans, given by a TPS in the DICOM RT plan format, with YODA. First, a brief overview of the inputs that YODA requires to define beams and the data contained within a DICOM RT plan file is given. Thereafter, two procedures, that allow the desired plan evaluation to be performed, are defined and further explained in Subsections 3.5.1 and 3.5.2. Lastly, Subsection 3.5.3 presents and discusses several dosimetric comparisons between YODA and the RayStation TPS.

A pencil beam in YODA, on the boundary of the computational domain, is characterized by the product of two two-dimensional Gaussians (one in the x direction and one in the y direction) and by a Gaussian in energy. Thus, a pencil beam is characterized by its mean energy E_0 and the energy spread s_E, the number of protons n_p (i.e., the amplitude of the energy distribution), the spatial spread s_x, the angular spread s_{ω_x} and the correlation ρ_x between the spatial and angular directions (assumed identical for both x and y directions). Moreover, the spot has associated two three-dimensional points, referred to as start $\mathbf{S} \in \mathbb{R}^3$ and end $\mathbf{E} \in \mathbb{R}^3$, between which a pencil beam propagates. Thus, the intersection between the line parametrized by the unit vector $\hat{\mathbf{l}} = \mathbf{SE}/|\mathbf{SE}|$ (or the pencil beam central axis) and the CT cube results in the voxels (and therefore the materials) and the length of the segments that the pencil beam encounters as it passes through the CT cube. In contrast, a DICOM RT plan file, produced by a TPS, defines a structure called beam, with each beam having multiple energy layers and each energy layer having multiple spots. A beam is characterized by its gantry and patient support angles, the virtual source to isocenter axis distances between the scanning magnet x and the scanning magnet y (the two magnets are located at different positions along the beam line), denoted by sm_x and sm_y respectively, the isocenter position $\mathbf{r}_{\mathrm{is}}$, the snout position (or rather distance from the isocenter) d_{sn} and, if needed, additional information on the type of range shifter in use. An energy layer is characterized by its nominal energy, the corresponding spot size in air at the isocenter and if used, by the range shifter type. The spot size in air at isocenter is a function of energy alone, and is usually extracted from fluence measurements during beam line commissioning. Lastly, a treatment plan spot is characterized by its two-dimensional $(x_{\mathrm{is}}, y_{\mathrm{is}})$ position on the isocenter plane, and the number of MUs delivered to that spot.

Thus, two procedures are necessary to convert a DICOM RT plan into inputs for YODA. The first is a geometrical conversion of the two-dimensional spot position $(x_{\mathrm{is}}, y_{\mathrm{is}})$ on the isocenter plane into the $\mathbf{S}$ and $\mathbf{E}$ points. This procedure is not TPS dependent, depends only on the specific snout and scanning magnet distances of a given gantry and the gantry and patient support angle and is detailed in Subsection 3.5.1. The second procedure is an optimization procedure for determining the necessary inputs for YODA to best match the TPS dose calculation. This is necessary because the DICOM RT plan file does not specify, for example, the spread and correlation of a spot in the plan. It is also necessary because, due to the proprietary nature of most TPS systems, the used material properties (stopping power, straggling coefficient formulation, etc.) cannot be readily extracted. A mismatch in material data results in a mismatch in ranges, despite the two algorithms having the same inputs. This procedure is detailed in Subsection 3.5.2.

3.5.1 Geometrical conversions

The first step in converting the two-dimensional isocenter spot position into the **S** and **E** points is to define the necessary coordinate systems. On the left of Figure 3.16 the IEC61217 gantry coordinate system can be seen.

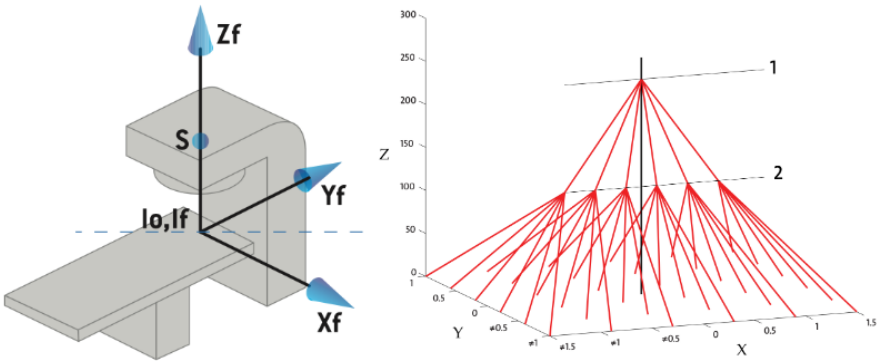


FIGURE 3.16: On the left is an overview of the gantry fixated coordinate system with its isocenter Io, If and the three axes Xf, Yf and Zf. The figure also shows the point S, which is the virtual source of protons. On the right is an overview divergent grid created by the scanning magnet x (at position 1) and y (at position 2). The left figure was taken from the RayStation manual (RayStation, 2019) and the figure on the right was taken from the Eclipse ProBeam manual (Way, 2015).

This coordinate system is fixed to the gantry (hence the f identifier for its axes) and is identified by its origin or isocenter at Io, If, and the three axes Xf, Yf and Zf. The Zf axis increases towards the source of protons, denoted by S in this figure. The figure also displays the patient support table and the gantry with the source of protons S. Within the gantry system, in the case of pencil beam scanning two scanning magnets are present. The scanning magnets are used to steer the pencil beam in the x and y directions, are at different locations along the beam line and therefore create two virtual sources with distances to the isocenter denoted by sm_x and sm_y. This is illustrated on the right side of Figure 3.16. The figure shows the scanning x magnet at location 1 with the multiple deflections it creates for the x coordinate of a spot and the scanning magnet y position at location 2 with the multiple deflections it creates for the y coordinate. At the end of the gantry (on the side closest to the isocenter) is the snout plane. Thus, a pencil beam ray (or central axis) starts at the first scanning magnet at the virtual source position, traverses the snout plane and subsequently the CT cube, as illustrated for the x direction in Figure 3.17. The figure also shows the points **S** and **E**, identified as the first and last intersection of the ray with the CT cube.

The independence of steering in the x and y dimensions allows treating them separately and similarly. Thus, given the isocenter spot coordinates $(x_{\mathrm{is}}, y_{\mathrm{is}})$ in the gantry coordinate system, the corresponding spot coordinates in the gantry coordinate system on the snout plane are written as

$$x_{\mathrm{sn}} = x_{\mathrm{is}} \cdot \frac{\mathrm{sm}_x - d_{\mathrm{sn}}}{\mathrm{sm}_x},$$
$$y_{\mathrm{sn}} = y_{\mathrm{is}} \cdot \frac{\mathrm{sm}_y - d_{\mathrm{sn}}}{\mathrm{sm}_y},$$

and the three-dimensional points in the gantry coordinate system are given by

$$\mathbf{r}_{\mathrm{g,sn}} = (x_{\mathrm{sn}}, y_{\mathrm{sn}}, d_{\mathrm{sn}}),$$

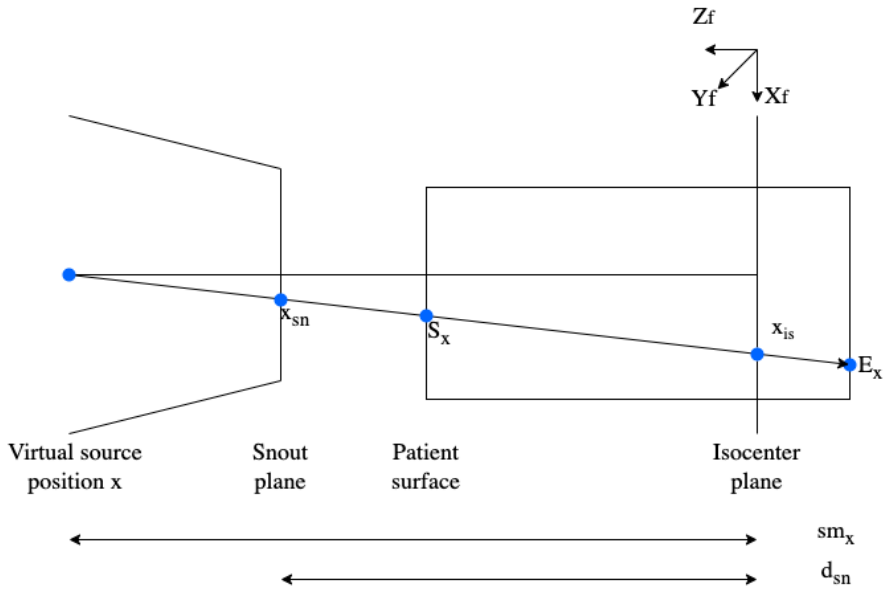


FIGURE 3.17: Diagram of the relevant components necessary for determining the start and end point of a treatment plan spot in the CT cube. The figure illustrates, for the x component, the virtual source position x, the snout plane with the point x_{sn} at which the spot crosses the plane, the patient surface, with the point S_x at which the spot crosses it, the isocenter plane, with the point x_{is} at which the spot crosses it and the end of the patient cube, with the point E_x at which the spot crosses it. The figure also displays the distance sm_x from the virtual source to the isocenter plane, the distance d_{sn} from the snout plane to the isocenter plane and the fixed gantry coordinate system with axes Xf, Yf and Zf.

$$\mathbf{r}_{\text{g,is}} = (x_{\text{is}}, y_{\text{is}}, 0).$$

In general, both the gantry and patient support table can be rotated. Rotating a vector $\mathbf{v}$ around a unit vector $\hat{\mathbf{k}}$ by an angle of θ can be achieved using the Rodrigues formula (Dai, 2015), namely

$$\mathbf{v}_{\text{rot}} = \mathbf{v}\cos\theta + (\hat{\mathbf{k}} \times \mathbf{v})\sin\theta + \hat{\mathbf{k}}(\hat{\mathbf{k}} \cdot \mathbf{v})(1 - \cos\theta).$$

Figure 3.18 illustrates on the left that gantry rotations are identified as rotations around the Yg axis of the gantry coordinate system (identified by the g letter). Thus, to account for gantry rotations, the two vectors $\mathbf{r}_{\text{g,sn}}$ and $\mathbf{r}_{\text{g,is}}$ are rotated around $\hat{\mathbf{k}} = (0, 1, 0)$. In the middle of Figure 3.18, rotations of the patient support are illustrated as rotations around the Zs axis of the patient support coordinate system (identified by the s letter). Thus, to account for rotations in the patient support, the previously obtained vector is translated to the location where the axis of rotation is located (Ie in Figure 3.18, not necessarily identical to the point Io, Is) and thereafter a rotation is performed around the vector $\hat{\mathbf{k}} = (0, 0, 1)$. The last transformation necessary is to convert from the gantry coordinate system shown in Figure 3.16 into the head first supine (HFS) DICOM patient coordinate system, shown on the right of Figure 3.18. These systems are related to each other by a 90° rotation around the Xp axis, i.e. $\hat{\mathbf{k}} = (1, 0, 0)$, and by incrementing the result with the location of the isocenter in the patient coordinate system.

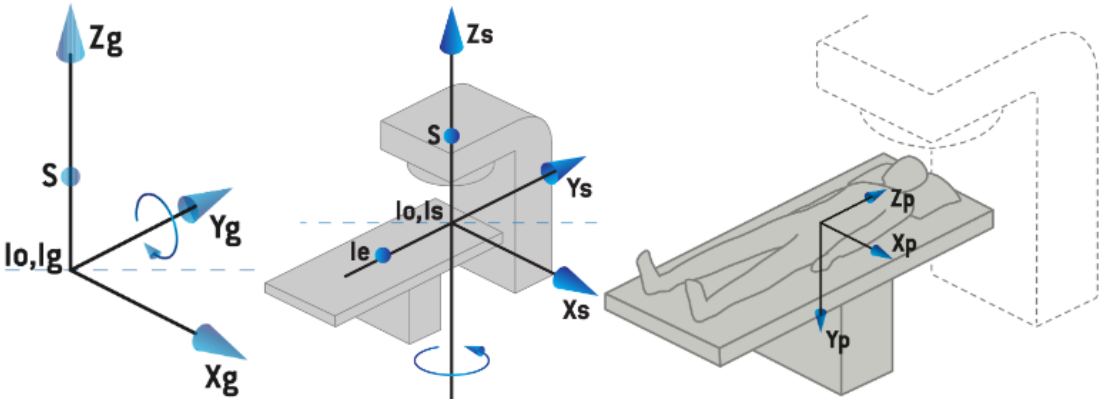


FIGURE 3.18: Overview of the different necessary rotations. On the left, the gantry coordinate system with axes Xg, Yg and Zg and the source S of protons is illustrated. Rotations of the gantry are illustrated as rotations around the Yg axis. In the middle, the patient support coordinate system with axes Xs, Ys and Zs with its origin at Io, Is is illustrated. Rotations of the patient support table are illustrated as rotations around the Zs axis around the point Ie. On the right, the HFS DICOM patient coordinate system with axes Xp, Yp and Zp is illustrated. Figures taken from (RayStation, 2019).

3.5.2 YODA input optimization

The input optimization procedure broadly follows the commissioning procedure used by MCSquare for a given gantry (*MCsquare - Commissioning Procedure* 2024) and results in a file called the Beam Data Library (BDL). The procedure can be divided into three parts which are performed sequentially.

First, the optical beam parameters in air at the isocenter are extracted, either from commissioning fluence measurements or from simulations. The optical parameters are the spatial spread, the angular spread and the correlation between the two at the isocenter. They are extracted as a function of beam energy, for every 10 MeV of the available energy spectrum of the delivery system. The procedure is similar for the x and y directions, but as YODA assumes symmetry in these directions, the procedure is described only for the x direction. The mathematical formalism for separation in the x and y directions was presented in Appendix B and its implementation should improve YODA's dose accuracy due to the inherent asymmetry observed in clinical practice. Once YODA's algorithm is generalized to include asymmetry, generalizing the commissioning procedure for the y direction is trivial. Given a set beam energy, several fluence measurements (at least 3) are performed at several distances (before, after and including) from the isocenter in air. From these measurements, the spatial spread (standard deviation) as a function of the distance from the isocenter $\sigma_x(z_i)$ is extracted. The isocenter is identified as the 0 point and the z axis runs positive towards the snout. Next, a one-dimensional fit to the Courant-Snyder formula (Courant and Snyder, 2000),

$$\sigma_x^2(z) = \sigma_x^2(0) - 2\rho_{x\omega_x}(0)\sigma_x(0)\sigma_{\omega_x}(0)z + \sigma_{\omega_x}^2(0)z^2,$$

results in the spatial spread at the isocenter $\sigma_x(0)$, the angular spread at the isocenter $\sigma_{\omega_x}(0)$ and the correlation between the x and ω_x direction at the isocenter $\rho_{x\omega_x}(0)$. This process is repeated for each new set beam energy. The isocenter optical parameters are used to find the optical parameters at other surfaces (such as the snout plane or the patient surface). The Courant-Snyder formula already gives the spatial spread at an arbitrary distance z from the isocenter. The remaining optical parameters, namely the

angular spread and correlation at any distance z from the isocenter, are found using

$$\rho_{x\omega_x}(z) = \frac{\rho_{x\omega_x}(0)\sigma_x(0) - \sigma_{\omega_x} z}{\sigma_x(z)},$$
$$\sigma_{\omega_x}(z) = \sigma_{\omega_x}(0).$$

Second, the mean energy E_0 and spread s_E that a YODA spot should use to match the IDDs of a TPS dose engine are computed. To this end, single spots are sequentially placed at the position of $(0, 0)$ in a water tank, in steps of 10 MeV. For each energy, the IDD of both YODA and the TPS dose engine is computed and normalized by their respective area. Next, an objective function is calculated based on the square difference between the IDDs, computed only at the points $z_i, i = 1, \ldots, K$ in the middle of the voxels situated between the proximal R_{50} and the end of the Bragg peak, namely

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{K} \left(\mathrm{IDD}_{YODA}(z_i, \boldsymbol{\theta}) - \mathrm{IDD}_{TPS}(z_i) \right)^2, \forall z_i \in (R_{50}, R_0),$$

where $\boldsymbol{\theta} = (E_0, s_E)$ are the mean energy and energy spread that YODA should use to match the RayStation IDDs. This objective function was found to yield faster and improved agreements compared to the weighted objective function suggested by MCSquare (*MCsquare - Commissioning Procedure* 2024). To help the optimization quickly converge, the mean energy is guessed (denoted by $E_{0,g}$) from the RayStation IDD by using Bortfeld's relationship between energy and range (Bortfeld, 1997), namely

$$E_{0,g} = \left(\frac{R_{80}}{\alpha} \right)^{\frac{1}{p}}$$

where $p = 1.77$ and $\alpha = 0.0022$. Furthermore, the energy spread is initialized to 1 MeV and the optimization is bounded in such a way that the mean energy is bound between $0.9 \cdot E_{0,g}$ and $1.2 \cdot E_{0,g}$ and the spread is bound between 0.1 MeV and 4.0 MeV.

Third, the number of protons per delivered MU is calibrated. This is done by re-using the single spot targeted towards the isocenter simulations performed for the second calibration procedure. These simulations were all done by setting the MU value to 100 in the TPS. The objective function used is the square difference between the maximums of the IDDs, and the guess for the optimization is obtained by converting the TPS IDD into energy and dividing the total deposited energy by the initial energy guess $E_{0,g}$. Here too the optimization is bounded between 0.5 and 1.5 times the initial number of protons per MU guess.

3.5.3 RayStation comparisons

To test the commissioning procedure, several dose comparisons between YODA and RayStation were performed. First, a comparison was performed across the energy spectrum for the situation in which a single spot is placed at $(0, 0)$ on the isocenter plane in a water tank. These simulations can be seen in Figure 3.19. The 3D gamma pass rate, with criteria of 1 mm, 1 %, 10 % cut-off, progressively degrades as the planned energy in RayStation is increased, with the lowest passing rate achieved being 97 %. Moreover, it can be seen that the agreement in the proximal region of the dose-depth curve progressively degrades as the energy is increased. This effect is likely due to the lack of nuclear interactions modelling in YODA.

Next, a plan containing a SOBP was generated and computed using the Monte Carlo engine of RayStation (with accuracy of 0.1 %) and independently computed with YODA.

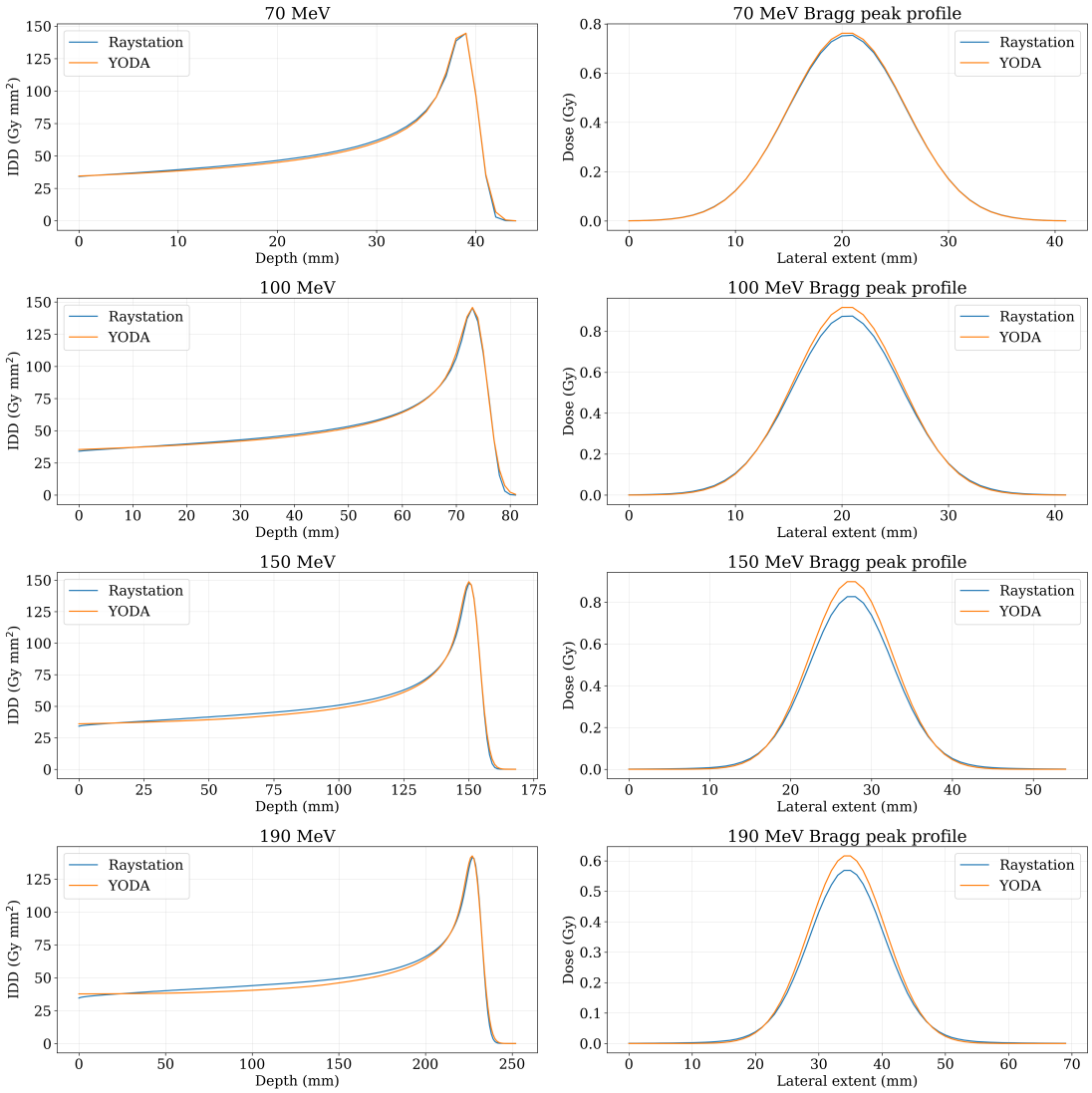


FIGURE 3.19: Dose comparisons between YODA and RayStation from single spots in a water tank. The left column displays IDDs for different planned energies in RayStation (given in the individual plot title), while the right column displays Bragg peak profiles.

The results can be seen in Figure 3.20. The SOBP consisted of 7 pencil beams, with energies ranging from 143 MeV to 158 MeV in steps of roughly 2.5 MeV. This yielded a gamma passing rate of $\sim 97\,\%$ with criteria of 1 mm, 1 % and 10 % cut-off, which is below the gamma passing rate for individual spots in a water tank. Additionally, systematic differences in the entrance region can be observed between YODA and RayStation (that are likely attributed to nuclear interactions) which add up when the SOBP is composed. This results in a systematic slanting of the flat SOBP that RayStation computes, as can be seen in Figure 3.20.

To understand the cause of this result, several tests were performed. First, the correct positioning of the two-dimensional spots from the treatment plan into the three-dimensional CT cube volume was confirmed by generating plans in which spots were placed off the central beam axis with symmetric offsets and for different gantry angles. Next, it was confirmed that the energy resolution in the BDL file of 10 MeV does not impact results for treatment plans that contain non-integer energy values, by generating a BDL file every 5 MeV. Next, the effect of nuclear interactions was investigated. For

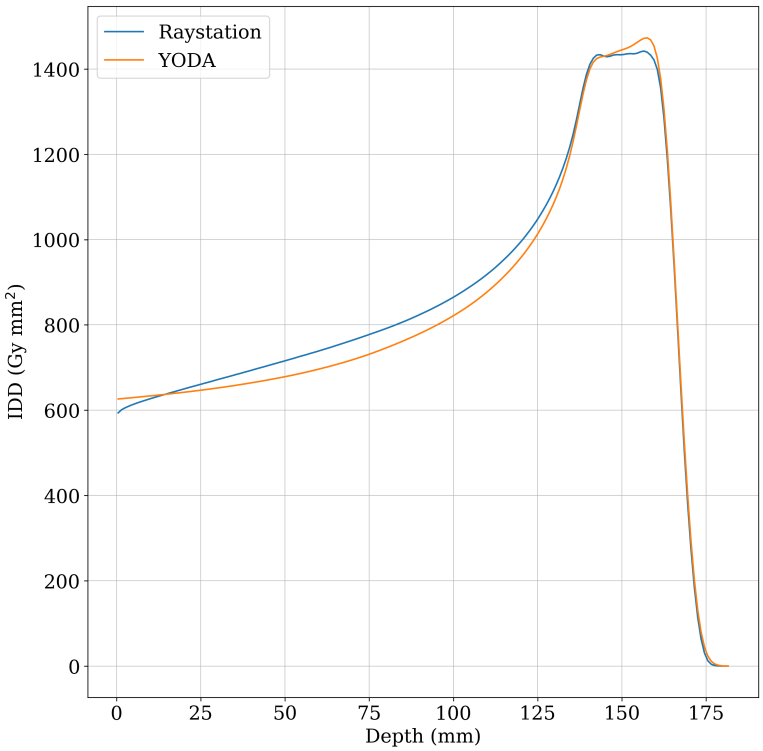


FIGURE 3.20: Dose comparison between YODA and RayStation when a 7 pencil beam SOBP is created in a water tank.

this purpose, the head and neck CT scan from Figure 3.7 was re-irradiated, this time computing doses with and without nuclear interactions in both TOPAS and YODA. To better account for nuclear interactions in YODA, Equation 2.27 for Ψ_{FP} was adapted by including a beam-energy dependent term γ, namely

$$\Psi_{FP}(z) = \int_0^\infty \mathrm{d}E\, E\left[-\frac{\partial S(z,E)\varphi_{FP}}{\partial E} - \frac{1}{2}\frac{\partial^2 T(z,E)\varphi_{FP}}{\partial E^2} + \gamma\Sigma_a(z,E)\varphi_{FP}\right]. \tag{3.3}$$

The effect of this term is to locally deposit only a fraction γ of the energy released in nuclear interactions that occur between the proton beam and the patient tissue. As opposed to Bortfeld's constant γ value of 0.6 (Bortfeld, 1997), this work set γ to be beam average energy dependent, with the energy dependence of γ obtained from the publication of (National Institute of Standards and Technology, 1993). Figure 3.21 displays the IDD comparison between YODA and TOPAS with and without nuclear interactions, using the same computational set-up as the one in Figure 3.7. When nuclear interactions are accounted for, the agreement between YODA and TOPAS is 95.63 % using gamma index criteria of 1 mm, 1 % and 10 % cut-off. This result illustrates the need for YODA to model nuclear interactions.

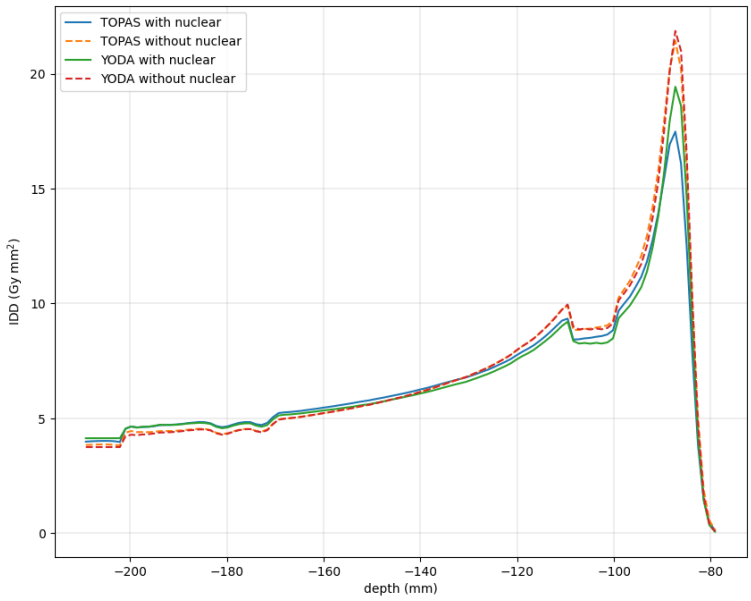


FIGURE 3.21: Comparison between YODA and TOPAS with and without nuclear interactions in the head and neck CT scan from Figure 3.7.

To assess the current capabilities of YODA to act as an independent dose calculation method, 4 patient plans were generated in RayStation, output to DICOM RT plans and re-computed with YODA. The results from the three-dimensional gamma index comparison using criteria of 2 mm, 2 % and 10 % cut-off can be seen in Figures 3.22, 3.23, 3.24 and 3.25

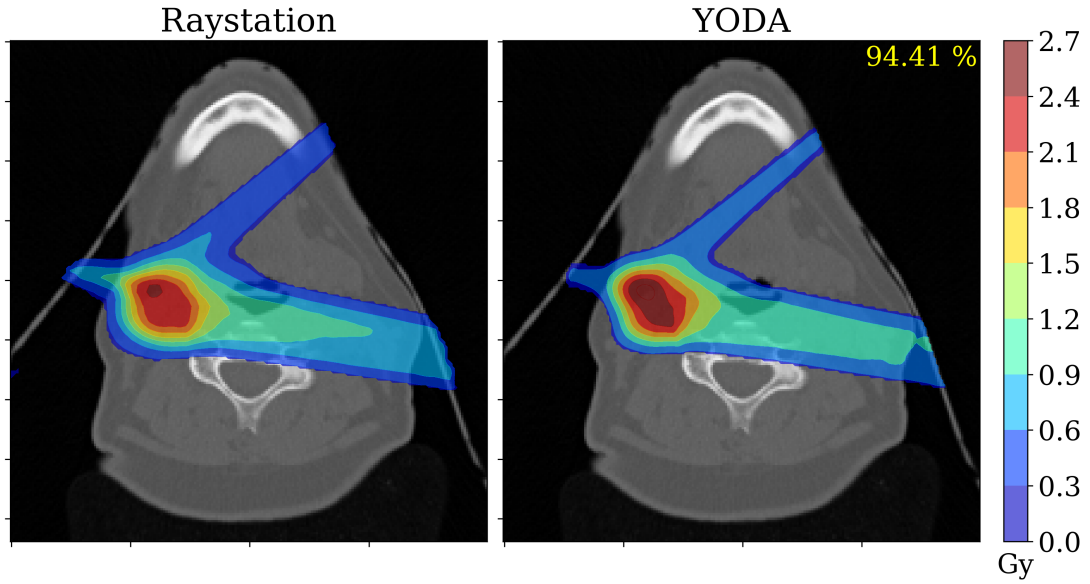


FIGURE 3.22: Dosimetric comparison between RayStation and YODA for the Patient 1. The figure displays RayStation (on the left) and YODA (on the right) two-dimensional dose cuts, overlaid on the corresponding CT slice. The figure displays in the top right corner of the YODA plot, the three-dimensional gamma index passing rate using 2 mm, 2 % and 10 % cut-off.

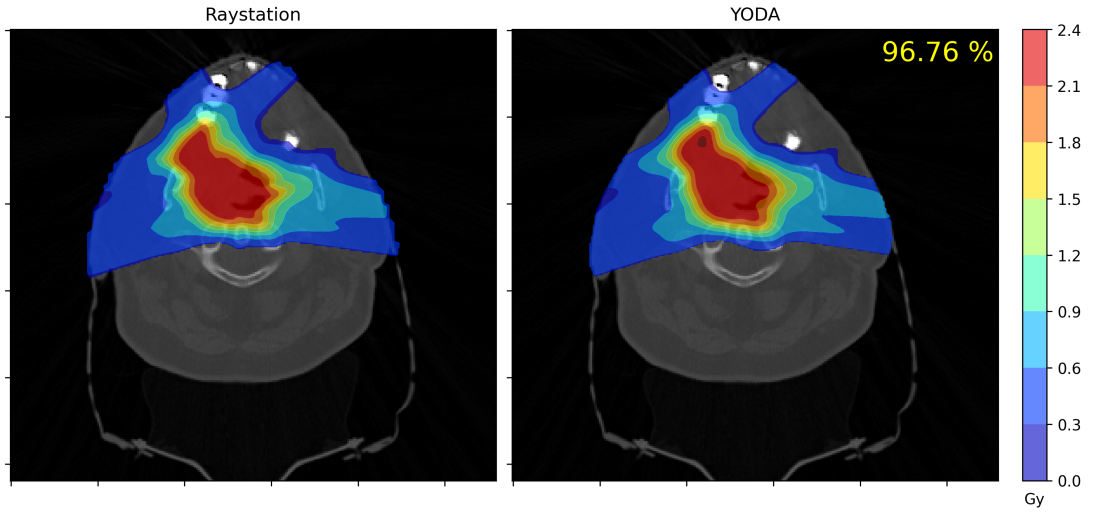


FIGURE 3.23: Dosimetric comparison between RayStation and YODA for the Patient 2. The figure displays RayStation (on the left) and YODA (on the right) two-dimensional dose cuts, overlaid on the corresponding CT slice. The figure displays in the top right corner of the YODA plot, the three-dimensional gamma index passing rate using 2 mm, 2 % and 10 % cut-off.

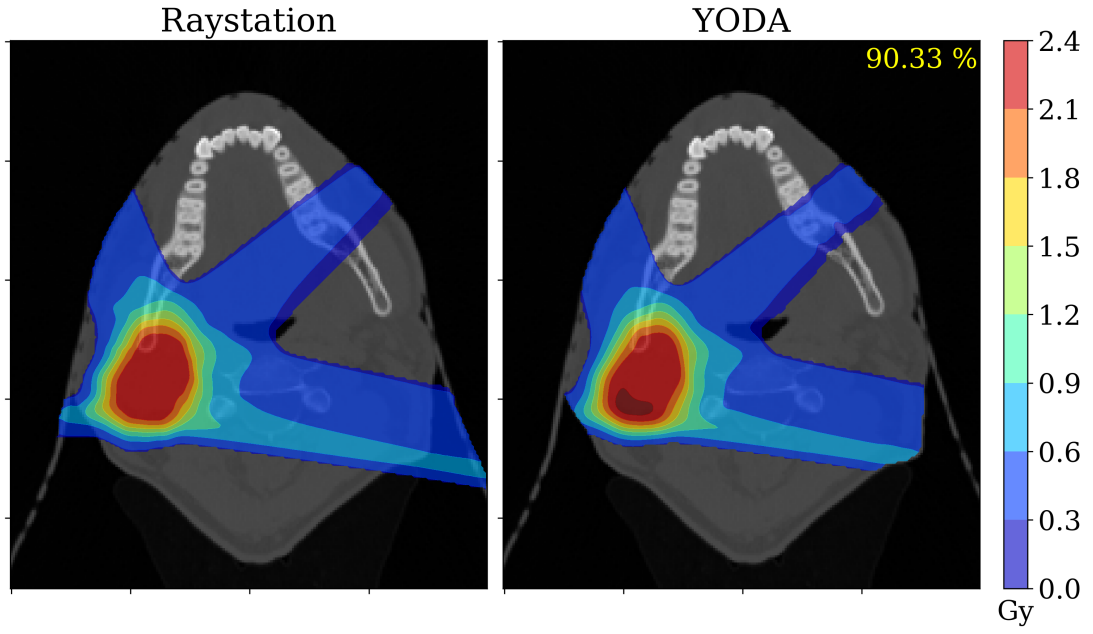


FIGURE 3.24: Dosimetric comparison between RayStation and YODA for the Patient 3. The figure displays RayStation (on the left) and YODA (on the right) two-dimensional dose cuts, overlaid on the corresponding CT slice. The figure displays in the top right corner of the YODA plot, the three-dimensional gamma index passing rate using 2 mm, 2 % and 10 % cut-off.

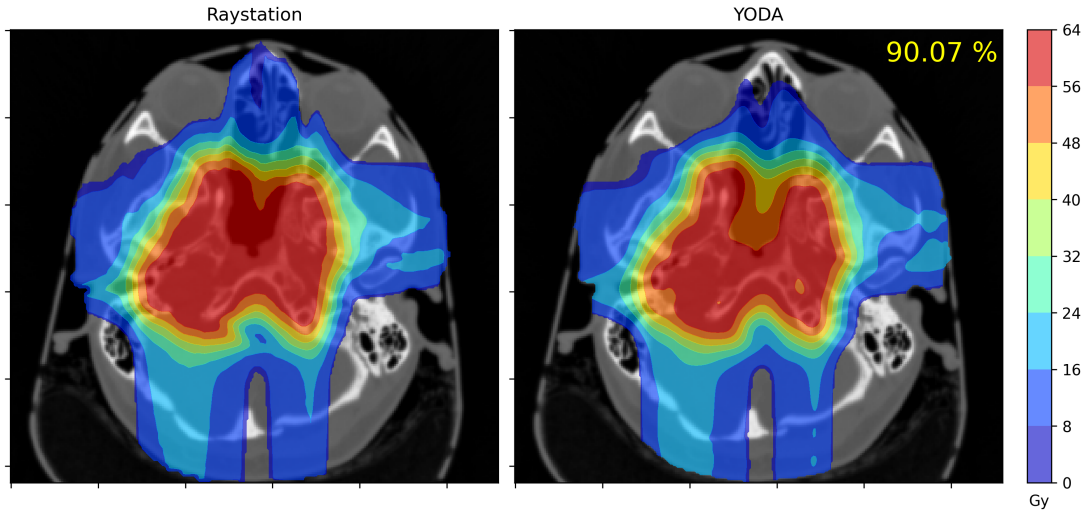


FIGURE 3.25: Dosimetric comparison between RayStation and YODA for the Patient 4. The figure displays RayStation (on the left) and YODA (on the right) two-dimensional dose cuts, overlaid on the corresponding CT slice. The figure displays in the top right corner of the YODA plot, the three-dimensional gamma index passing rate using 2 mm, 2 % and 10 % cut-off.

All figures show that the doses that YODA re-computes, based on the RayStation generated plans, display similar characteristics (range, shape and conformality) as the RayStation doses. Some general observations are that YODA sometimes overestimates the doses in certain regions (likely due to lack of nuclear interactions modelling) and that YODA doses are somewhat more smeared (due to the Gaussian beam splitting and the specific chosen scheme). The lowest passing rate observed is 90 % which, depending on the clinic, may or may not be acceptable. These figures further illustrate that while the features of the dose distribution (thus the placement of the beams and the ranges accounted for by commissioning are correct), the current modelling must be adapted to account for nuclear interactions. Given these results it can be concluded that the commissioning procedure for YODA has been successful with the observed differences being likely attributable to the lack of nuclear interactions modelling.

3.6 Conclusion

This chapter presented the performance of YODA in dose and dose change computations and a commissioning procedure for YODA. YODA uses a hybrid approach to solve a physics-based approximation to the same equations that MC methods solve. This approach enables YODA to achieve TOPAS like performance with a significant speed-up. The lowest three dimensional gamma index passing rates achieved using the strict criteria of 1 mm, 1 %, 10 % cut-off is 94.55 % in the Lung 2 case. YODA computes a treatment plan spot in 2 s while the same spot takes hours in TOPAS. An adjoint computation depends on the size of the structure but is generally in the order of milliseconds to tens of milliseconds on a single CPU. A typical commercial TPS plan re-evaluation is in the order of minutes to tens of minutes on multicore CPUs or GPUs (Nystrom, Jensen, and Nystrom, 2020). If YODA's speed would be further improved (e.g., via a GPU implementation), YODA could be used as a patient-specific quality assurance tool by tapping into the data stream between the TPS and the delivery machine to quickly re-construct the dose to be delivered. Alternatively, the logfiles could

be used after treatment to re-construct the actually delivered dose to the patient. A multi-treatment site patient cohort study is necessary to validate the accuracy of YODA versus commercial TPS calculations in a wide variety of settings. Additionally, nuclear interactions must be accounted for, as illustrated in the comparison to RayStation doses, where despite the commissioning procedure, the lowest passing rate is 90 %. However, given that the dose engine contained in Eclipse (AcurosPT) is accurate with criteria of 2 mm, 2 % in hetereogeneous cases (De Martino et al., 2021) and the various speed and accuracy improvements still achievable in YODA it can be concluded that this engine could compete with/replace other commercial dose algorithms and is certainly capable of TPS independent dose calculations.

Next to performing TPS independent dose calculations, YODA can leverage the adjoint component to accurately compute dose changes caused by appropriately small anatomical changes. Such a feature, to the best of the authors' knowledge, has not been integrated into a dose algorithm before. This component could be used in a time constrained re-adaptation trigger system where on the given day YODA avoids re-computing the old treatment plan on the new CT image if the CT image is deemed anatomically close enough to the original one. This performance was illustrated via four treatment plans where a maximal error of 5.7 % was achieved for a non-robustly optimized plan and 4.8 % for a robustly optimized plan. Alternatively, if log-files were available during treatment delivery YODA would be capable of halting erroneous deliveries in near real-time (i.e., below energy layer switching times) by converting spot position differences into anatomical changes and ultimately into dosimetric changes via the adjoint component.

Chapter 4

A deep learning model for inter-fraction head and neck changes

4.1 Novelty and own contributions

This chapter details a deep learning algorithm for modelling anatomical changes that occur in head and neck proton therapy patients. While much of the algorithmic implementation is based on the original work of (Dalca et al., 2019) with slight modifications brought by (Pastor-Serrano et al., 2023), a number of own contributions were required to complete this study. First, a novel GPU based data processing (image registration, visualization, cropping, interpolation) pipeline was created to speed up the training set creation and result evaluation phases. Second, the architecture of the model was modified to allow for a variable number of organs at risk, as opposed to the fixed one used in (Pastor-Serrano et al., 2023). Third, the performance of the model was improved by implementing a grid search hyperparameter optimization technique. Fourth, a literature study was performed in order to conglomerate the reported anatomical changes that occur in head and neck RT patients and to characterize the degree to which the training set is representative of the broader population. Last, the results were statistically analyzed and interpreted and the performance was compared with the state of the art deep learning model of (Smolders et al., 2024), thereby demonstrating that this model achieves similar performance to other state of the art deep learning models.

4.2 Synthetic CT uses in Proton Therapy

Proton Therapy (PT) has desirable dose characteristics, such as similar target coverage and lower organs at risk (OAR) doses, when compared to traditional photon based radiotherapy (RT) (Chen et al., 2023). However, the increased dose conformality implies an increased susceptibility to dose degradation by uncertainties such as setup errors, range errors and anatomical changes over the course of the typically month long treatment duration (van Kranen et al., 2009). To diminish the dose degradation, robust optimization and evaluation (Unkelbach and Paganetti, 2018) with isotropic setup and range settings (Liu et al., 2013) and offline adaptive replanning (Deiter et al., 2020) is performed in clinical practice. This results in a high dose region that surrounds the target, which in the case of the head and neck (H&N) region where OARs are in close

This chapter is based on the publication of Tiberiu Burlacu et al. (Mar. 2025). "A Deep Learning Model for Inter-Fraction Head and Neck Anatomical Changes in Proton Therapy". In: *Physics in Medicine & Biology* 70.6, p. 065011. ISSN: 0031-9155. DOI: 10.1088/1361-6560/adba39. (Visited on 05/07/2025)

proximity to the target, could result in high chances of side effects. Moreover, there are certain anatomical changes (e.g., tumor shrinkage (Cubillos-Mesías et al., 2019)) that are not effectively accounted for by robust optimization only taking setup and range errors into account. One proposed option (Van de Water et al., 2018) is the inclusion of additional (synthetic) computed tomography (CT) images in the (anatomical) robust optimization process. While this provided increased target coverage and lower OAR doses for the specific H&N patients in the cohort, compared to conventional robust optimization, it still created a high dose region surrounding the target.

To reduce this region to its minimum and counter long and short-term inter-fraction occurring anatomical variations, online adaptive proton therapy (OAPT) has been proposed. In this workflow, a new CT is acquired for each fraction and within a short time a new fully re-optimized plan is generated. The resulting plan would only need minimal robustness settings to counter the effects of range uncertainties, machine related setup uncertainties and remaining intra-fraction uncertainties. The short time available and the limited computational resources imply that fully robust reoptimization in the online setting still requires research (Oud et al., 2024) and is not feasible clinically. The plan library (PL) approach was proposed as an intermediate solution (Oud et al., 2022; van de Schoot et al., 2016). This approach used the planning CT image to generate multiple plans with varying robustness settings. On the given day, it administers an appropriately chosen plan, therefore resulting in NTCP reductions or sometimes in increased robustness that ensures adequate target coverage. In this approach, synthetic CT images could be used to expand the pre-compiled library of plans, by generating optimal plans for the future patient anatomies predicted by the model. An additional use case for synthetic CT images could be for plan QA, in the scenario in which an adapted or refined (e.g., by using yesterday's optimal plan) is generated with the patient on the table. Specifically, several CT images with associated truly optimal plans, could be generated a priori. On the given day, a fast dosimetric check can be performed between the adapted and refined plan and the truly optimal pre-generated plan.

Thus, models of inter-fractional anatomical changes have applications in several PT related workflows such as anatomical robust optimization, plan quality assurance in OAPT or expanding the plan library approach. Multiple approaches to synthetic CT generation have been employed, such as principal component analysis (PCA) or deep learning. An overview of the different possible approaches is given by the work of (Smolders et al., 2024). Deep learning models have been shown to outperform PCA based ones in the case of prostate anatomies (Pastor-Serrano et al., 2023) and denoising diffusion probabilistic models (Smolders et al., 2024) were successfully applied for artificial CT generation for the H&N site where they were additionally shown to increase robustness to anatomical changes. This work builds upon the previous publication of (Pastor-Serrano et al., 2023) on a generative deep learning daily anatomy model (DAM) for prostate inter-fractional anatomical changes. The model architecture and the data processing pipeline are changed and thereafter applied to a H&N radiotherapy cohort. The model is referred to from here on as DAM_{HN}. Section 4.3 details the probabilistic framework of the model. Section 4.4 provides details on the dataset generation and the specific architecture configuration used for training. Section 4.5 contains the results and their discussion. The performance of the model was assessed via several tests. The results of a reconstruction accuracy test are shown in Subsection 4.5.1. The generative performance was assessed in terms of the model's capability to predict realistic anatomical changes. To this end, an overview of the typical changes in head and neck patients reported by literature studies is given in Subsection 4.5.2. The anatomical changes present on the training set are discussed in Subsection 4.5.3. Subsection 4.5.4 presents and discusses the anatomical changes predicted by the model. Subsection 4.5.5

compares these anatomical changes with the ones presented in the recently published denoising diffusion probabilistic models DiffuseRT model (Smolders et al., 2024). Lastly, a latent space analysis is presented in Subsection 4.5.6. Section 4.6 concludes this work and discusses some improvement points.

4.3 Model architecture

This section provides only the main details of this model's architecture. An in-depth exposition can be found in (Pastor-Serrano et al., 2023). The patient anatomy at a certain point in time is described by the CT image and the associated RT structures (masks), which are both taken as random variables. On the planning CT image (pCT), an image with N voxels is denoted by $\mathbf{x} \in \mathbb{R}^N$ (defined as floats due to the need to normalize the data prior to further processing) and the corresponding structures (pM) are denoted by $\mathbf{s}_x \in \mathbb{R}^N$. On the repeat CT images (rCTs), the image is denoted by $\mathbf{y} \in \mathbb{R}^N$ and the corresponding masks (rM) by $\mathbf{s}_y \in \mathbb{R}^N$. Generally, pCTs and rCTs do not have the same dimensionality and to achieve this, the images are resampled and cropped.

The presence of anatomical deformations over the course of treatment, e.g., the systematic medial translation of the lateral regions of the parotid glands, the shrinkage of the parotid and submandibular glands (Fiorentino et al., 2012), the change in the parotid shape from convex to flat or concave (Santos et al., 2020) and the center of mass (COM) shifts towards the medial side (Vásquez Osorio et al., 2008) motivates the existence of an unknown generative joint conditional probability distribution $P^*(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ of the voxel CT HU values $\mathbf{y}$ and the structure masks $\mathbf{s}_y$ conditioned on the planning CT $\mathbf{x}$ and structures $\mathbf{s}_x$. If such a distribution was known, given a new pCT and pM, it could be sampled to generate future possible anatomies, denoted by $\mathbf{y}$ and $\mathbf{s}_y$. In general it is impossible to find such a distribution, and a good approximation $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x) \approx P^*(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ is sought instead. The distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ is parametrized by a vector of parameters $\boldsymbol{\theta}$ that is learned during training.

The dataset $\mathbb{D}$ consists of elements $\mathbf{s}^i \in \mathbb{R}^{4N}$, which are the concatenation of a given pCT and rCT and their associated structures, i.e., $\mathbb{D} = \{\boldsymbol{\tau}_i = (\mathbf{x}_i, \mathbf{s}_{x_i}, \mathbf{y}_i, \mathbf{s}_{y_i}) \,|\, i = 1, \ldots, N_D\}$ with N_D the number of elements in the dataset. Moreover, the dataset $\mathbb{D}$ is assumed to be independent and identically distributed (i.i.d.). As the dataset $\mathbb{D}$ is i.i.d., the log-probability assigned to the data is

$$\log P_{\boldsymbol{\theta}}(\mathbb{D}) = \sum_{\boldsymbol{\tau} \in \mathbb{D}} \log P_{\boldsymbol{\theta}}(\boldsymbol{\tau}). \tag{4.1}$$

The framework of Maximum Likelihood (ML) searches for the parameters $\boldsymbol{\theta}$ that maximize the sum, or equivalently the average, of the log-probabilities assigned to the data by the model in Equation 4.1 (Kingma and Welling, 2019).

As most explicitly parametrized generative distributions are too simplistic to model inter-fractional anatomical variations, implicitly parametrized distributions are considered instead. Therefore, a joint conditional probability distribution, denoted by $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x)$, that also depends on latent variables $\mathbf{z}$ is constructed. Latent variables are variables that are not observed, and therefore they are not part of the dataset of images and associated structures. They are meant to encode (represent in a lower dimensional space) the information between the pCT and the rCT. The marginal distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ over the observed variables $\mathbf{y}, \mathbf{s}_y$ is recovered by marginalizing,

namely

$$P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x) = \int \mathrm{d}\mathbf{z}\, P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x) \tag{4.2}$$
$$= \int \mathrm{d}\mathbf{z}\, P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{z}, \mathbf{x}, \mathbf{s}_x) P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}, \mathbf{s}_x).$$

This is also referred to as the (single datapoint) marginal likelihood, or model evidence, when taken as a function of $\boldsymbol{\theta}$ (Ghojogh et al., 2022). The distribution $P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)$ is called the prior distribution, which in the case of this work is taken as a multivariate Normal distribution with mean and variance that depend on the pCT and pM and on the vector of learned parameters $\boldsymbol{\theta}$, namely

$$P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}, \mathbf{s}_x) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}_x), \Sigma_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}_x)). \tag{4.3}$$

The dependence of the parameters of the prior distribution on the pCT and pM results in a different distribution for each patient (insofar as a patient is identified with a single image). The mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}_x)$ and the covariance matrix $\Sigma_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}_x)$ are computed in the down-sampling part of a U-net neural network and the parameters $\boldsymbol{\theta}$ of the prior are the weights of the encoder, as illustrated in Figure 4.1.

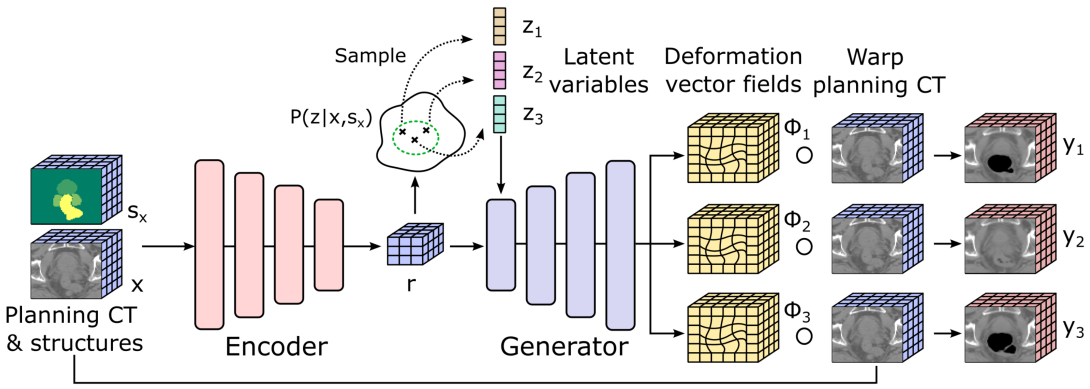


FIGURE 4.1: The proposed variational autoencoder model. The Encoder represents the down-sampling part of a U-net that computes the parameters of $P(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)$. The up-sampling part of the U-net, denoted by Generator, takes samples from $P(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)$ together with a reduced representation of the inputs $\mathbf{x}, \mathbf{s}_x$ and computes artificial CT images $\mathbf{y}, \mathbf{s}_y$. Figure reproduced with permission from (Pastor-Serrano et al., 2023).

The up-sampling part of the U-net, denoted by Generator in Figure 4.1, outputs a deformation vector field (DVF) $\Phi : \mathbb{R}^{N\times 3} \to \mathbb{R}^{N\times 3}$ used to map coordinates $\mathbf{p} \in \mathbb{R}^3$ between images. The DVF Φ is used to obtain the prediction of the model $\mathbf{y} = \Phi \circ \mathbf{x}$ (Jaderberg et al., 2016). Based on work by (Krebs et al., 2019), the distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{z}, \mathbf{x}, \mathbf{s}_x)$ (referred to as the likelihood) is taken as a function of the normalized cross-correlation (NCC) between the ground truth image $\hat{\mathbf{y}}$ and the predicted image $\mathbf{y}$ with an additional scaling factor $w_{NCC} \in \mathbb{R}^+$, namely

$$P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{z}, \mathbf{x}, \mathbf{s}_x) = \exp\left(-w_{\mathrm{NCC}}\mathrm{CC}(\mathbf{y}, \hat{\mathbf{y}})\right), \tag{4.4}$$

where the CC term is defined as

$$\mathrm{CC}(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{\mathbf{p}\in\Omega} \frac{\left[\sum_{i=1}^{n^3} \left(\hat{\mathbf{y}}(\mathbf{p}_i) - \hat{w}(\mathbf{p})\right)\left(\mathbf{y}(\mathbf{p}_i) - w(\mathbf{p})\right)\right]^2}{\left[\sum_{i=1}^{n^3} \left(\hat{\mathbf{y}}(\mathbf{p}_i) - \hat{w}(\mathbf{p})\right)\right]\left[\sum_{i=1}^{n^3} \left(\mathbf{y}(\mathbf{p}_i) - w(\mathbf{p})\right)\right]}, \tag{4.5}$$

and $w(\mathbf{p})$ and $\hat{w}(\mathbf{p})$ are the local mean over a small cube Ω with side length n voxels of the generated and true images, namely

$$w(\mathbf{p}) = \frac{1}{n^3}\sum_{j=1}^{n^3} \mathbf{y}(\mathbf{p}_j), \text{ and } \hat{w}(\mathbf{p}) = \frac{1}{n^3}\sum_{j=1}^{n^3} \hat{\mathbf{y}}(\mathbf{p}_j).$$

The vector of parameters $\boldsymbol{\theta}$ of the likelihood distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{z}, \mathbf{x}, \mathbf{s}_x)$, that stores in part of it the weights of the Encoder network, also stores the weights of the Generator network.

The main difficulty of this proposed framework is that the marginal probability of the data, or the model evidence, given in Equation 4.2 is intractable due to not having an analytic solution or an efficient estimator. In turn, this makes optimization of such a model computationally expensive.

4.3.1 Learning the optimal parameters

To overcome the previously mentioned intractability of the framework, the posterior distribution $P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)$ is approximated by a multivariate Normal distribution $Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)$ parametrized by a vector of parameters $\boldsymbol{\psi}$ with mean and variance that depend on both the planning and repeat images and masks, namely

$$Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{\boldsymbol{\psi}}(\mathbf{x}, \mathbf{s}_x, \mathbf{y}, \mathbf{s}_y), \Sigma_{\boldsymbol{\psi}}(\mathbf{x}, \mathbf{s}_x, \mathbf{y}, \mathbf{s}_y)). \tag{4.6}$$

The parameters $\boldsymbol{\psi}$ are the weights of the down-sampling part of a U-net, referred to as Inference network at the top of Figure 4.2.

Regardless of the choice of the approximating posterior distribution $Q_{\boldsymbol{\psi}}$, the log-likelihood of the data can be written as

$$\begin{aligned}
\log P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x) &= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)\right] \\
&= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log \frac{P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x)}{P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)}\right] \\
&= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log \frac{P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x)}{Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)} \frac{Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)}{P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)}\right] \\
&= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log \frac{P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x)}{Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)}\right] && (4.7) \\
&\quad + D_{KL}\left(Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)||P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x)\right). && (4.8)
\end{aligned}$$

The D_{KL} term in Equation 4.8 defines the Kullback-Leibler divergence between the approximated posterior distribution and the true posterior distribution. The term is non-negative, measures the distance between the shapes of the two distributions, and

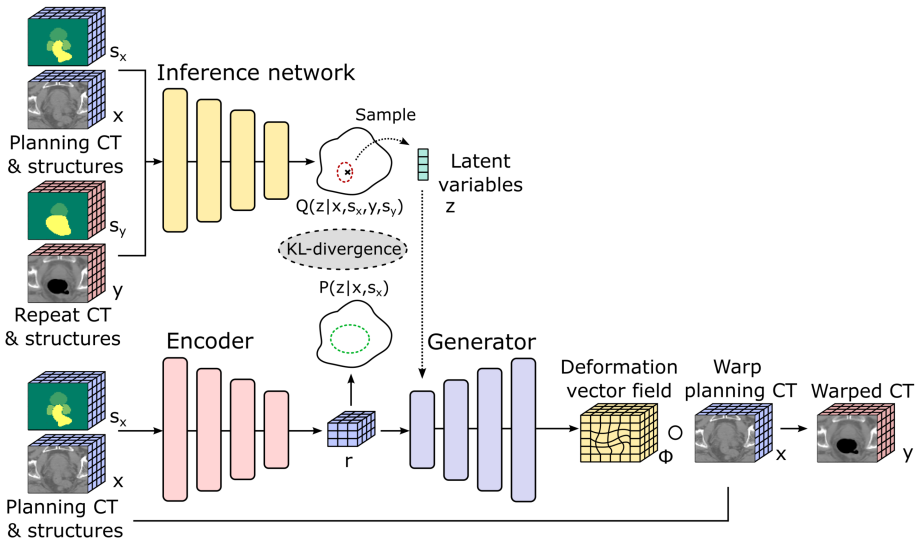


FIGURE 4.2: Architecture for finding the optimal parameters $\boldsymbol{\theta}$, $\boldsymbol{\psi}$ of the network. Figure reproduced with permission from (Pastor-Serrano et al., 2023).

is zero if, and only if, the approximated posterior equals the true posterior. The expectation term in Equation 4.7, defines the evidence lower bound (ELBO) as

$$\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}} = \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log P_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{s}_y,\mathbf{z}|\mathbf{x},\mathbf{s}_x) - \log Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x)\right],$$

which can also be re-written as

$$\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}} = \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log P_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{s}_y|\mathbf{z},\mathbf{x},\mathbf{s}_x)\right] - D_{KL}(Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x)||P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x},\mathbf{s}_x)). \tag{4.9}$$

As the D_{KL} term is non-negative, it is clear that the ELBO is a lower bound on the log-likelihood of the data, i.e.,

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}} &= \log P_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{s}_y|\mathbf{x},\mathbf{s}_x) - D_{KL}\left(Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x)||P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x)\right)\\ &\leq \log P_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{s}_y|\mathbf{x},\mathbf{s}_x).\end{aligned}$$

Thus, by maximizing the ELBO $\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}}$ from Equation 4.9 with respect to the parameters of the model $\boldsymbol{\theta},\boldsymbol{\psi}$, the marginal likelihood $P_{\boldsymbol{\theta}}$ is approximately maximized resulting in a better generative model and the KL divergence between the approximated posterior and the true posterior is lowered.

To improve model performance, the ELBO is expanded with two additional terms which are included via multiplication to the likelihood from Equation 4.9 (Pastor-Serrano et al., 2023). The first is a spatial regularization term,

$$R(\Phi) = -\,w_{\text{REG}} \sum_{\mathbf{p}\in\Omega} \|\boldsymbol{\nabla}\Phi(\mathbf{p})\|_2, \tag{4.10}$$

where w_{REG} is a multiplication constant. This term penalizes large and unrealistic gradients in the deformation and encourages neighboring voxels to deform somewhat similarly.

The second is a segmentation regularization term using the DICE score is added, which is also multiplied by a constant w_{DICE}. This aims to improve the overlap between

the propagated and ground truth structures, and is written as

$$\mathrm{DICE}(\mathbf{s}_y^k, \hat{\mathbf{s}}_y^k) = 2\, w_{\mathrm{DICE}} \frac{\left|\mathbf{s}_y^k \cap \hat{\mathbf{s}}_y^k\right|}{\left|\mathbf{s}_y^k\right| + \left|\hat{\mathbf{s}}_y^k\right|}, \tag{4.11}$$

where k denotes the index of the structure present in the CT image, $k = 1, \ldots, K$, with K the total number of structures present, and $\mathbf{s}_y^k$ and $\hat{\mathbf{s}}_y^k$ are the k-th generated and ground truth structures respectively.

Including these two additional terms, and minimizing the negative ELBO instead, results in the following optimization problem

$$\min_{\boldsymbol{\theta},\boldsymbol{\psi}} \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}} \left[-w_{\mathrm{NCC}}\mathrm{NCC}(\mathbf{y}, \hat{\mathbf{y}}) - w_{\mathrm{DICE}} \frac{1}{K} \sum_{k=1}^{k} \mathrm{DICE}(\mathbf{s}_y^k, \hat{\mathbf{s}}_y^k) + w_{\mathrm{REG}} \sum_{\mathbf{p}\in\Omega} \|\boldsymbol{\nabla}\Phi(\mathbf{p})\|_2 \right] + w_{KL} D_{KL}(Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y}, \mathbf{s}_y, \mathbf{x}, \mathbf{s}_x) || P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)).$$

4.4 Dataset generation and training details

The dataset was acquired from the Holland Proton Therapy Center and came from 93 H&N patients with planning, repeat CT images and associated RT structures for each image. This resulted in 342 pCT - rCT pairs from which 10 %, corresponding to 9 patients, were set aside for final testing. The remaining part was divided into 5 % for validation and 95 % for training. The training dataset consisted of patients with a number of rCTs ranging from 1 to 6, with most patients having 3 (24 patients) and 4 (25 patients) rCTs taken. This creates a bias in the dataset for the anatomical changes present in patients that are more likely to be re-imaged. All the rCTs were rigidly registered to the pCTs using the Simple ITK library (Beare, Lowekamp, and Yaniv, 2018) with the resulting deformation vector fields used to register the RT masks. After this, all scans were interpolated to a $2 \times 2 \times 2$ mm grid and cropped around the center of mass of the present RT masks (the left and right parotid glands, the spinal cord and the constrictor muscle) into a shape of $96 \times 96 \times 64$ voxels. This resulted in volumes of $192 \times 192 \times 128$ mm^3 which were found to adequately cover the anatomical regions of interest.

The model was implemented in PyTorch (Paszke et al., 2017). The down-sampling path of the U-net (Encoder) and the Inference network were identical, and consisted of 4 blocks, where each block is composed of a 3D convolution layer, a Group Normalization layer, a rectified linear (ReLu) activation and a max pooling down-sampling operation. All convolution layers had a kernel of dimensions $3 \times 3 \times 3$. The convolution layer in the first block had 16 channels while the remaining blocks had 32. At the lowest level, a last convolution with 4 channels results in the encoded volume $\mathbf{r} \in \mathbb{R}^{4\times4\times4\times3}$. This volume is mapped to the means and variances via two different fully-connected layers. The up-sampling part of the U-net (Generator) concatenates the sampled latent variables to the volume $\mathbf{r}$ after a linear layer. Next, 7 blocks (with up-sampling as opposed to down-sampling max pooling operations) are applied, where for the first 5 the convolutional layer has 32 channels and for the last 2, the convolutional layer has 16 channels. This is followed by a last convolution with 3 channels. The model was trained using a batch size of 32, on a A40 NVIDIA GPU, for 1500 epochs with an early stopping patience of 300 epochs and the Adam optimizer with a learning rate of 1.0×10^{-4}.

The constants $w_{\mathrm{NCC}}, w_{\mathrm{DICE}}, w_{\mathrm{REG}}$ together with the constant w_{KL} that multiplied the D_{KL} loss term were considered as hyparparameters to be optimized. These hyperparameters were optimized on the validation set using a grid search method with

the validation loss defined as the sum of the NCC from equation 4.4 and DICE from Equation 4.11 with unity weights. Thus, for a given latent space dimension, ranges of allowed values were defined for each hyperparameter (w_{NCC} and w_{DICE} from 1000 to 5000 in steps of 1000, w_{REG} from 1.0×10^{-5} to 1.0×10^{-1} in multiples of 10 and w_{KL} from 1.0×10^{-3} to 1.0×10^{1} in multiples of 10). After each combination was tested, the model with the lowest validation loss was chosen. This resulted in the model with $w_{\text{NCC}} = 5000$, $w_{\text{DICE}} = 3000, w_{\text{REG}} = 1.0 \times 10^{-4}$ and $w_{\text{KL}} = 1$.

4.5 Results and discussion

This section presents and discusses the performance of the model in a series of tests. The section starts by presenting and discussing in Subsection 4.5.1 the performance of the model on the test set (a reconstruction accuracy test). Next, a baseline is set through a literature study for the expected anatomical changes in H&N patients in Subsection 4.5.2. The anatomical changes displayed by the training set are compared to the expectations set out by literature, in Subsection 4.5.3, in order to assess the degree to which the dataset used by this model is representative of the broader population. Given this framework, the generative performance of the model is presented and discussed in Subsection 4.5.4. To gain insight into the model, a latent space analysis is presented and discussed in Subsection 4.5.6. Lastly, a comparison to the recent diffusion model proposed by (Smolders et al., 2024) is given in Subsection 4.5.5.

4.5.1 Test set accuracy

The reconstruction accuracy of the model on the test set (composed only of the left and right parotids) was assessed. The accuracy was defined by two metrics, namely the normalized cross correlation (NCC) loss from Equation 4.4 and the DICE loss from Equation 4.11. Thus, each record in the test set (i.e., pair of pCT and rCT with associated masks) was used to generate through the inference network latent variables, which ultimately result in generated CTs and associated structures. The results were averaged over all records in the test set and the dimension of the latent space was varied between 2 and 256 in multiples of 2. The results of the two scores can be seen in Figure 4.3, which shows the mean of the individual scores and a band of one standard deviation around the mean. Both figures show considerable improvement in both metrics as the latent space is increased from 2 to 32, and thereafter a plateau occurring between 64 and 256. The same behavior is observed in both metrics (a rise in accuracy up to ≈ 32 latent variables and thereafter a plateau). It should be noted that the metrics are sensitive to different features (as the NCC metric is computed based on the HU values of the voxels while the DICE score is computed based on the overlap of the binary masks). Thus, the similar behavior that is observed is likely due to the chosen loss functions that include both the NCC and the DICE score. The model performs particularly well with regard to the DICE score, where it achieves a score of 0.82 with just 2 latent variables. The reconstruction accuracy values obtained were not directly comparable with the ones previously published in the work of (Pastor-Serrano et al., 2023), but exhibit the same behavior. The increased input size of this model ($96 \times 96 \times 64$ versus $64 \times 64 \times 48$), the more complex anatomical site (H&N versus prostate) and a different configuration of the layers in the Inference, Encoder and Generator networks likely explain the need for additional latent variables to achieve good performance.

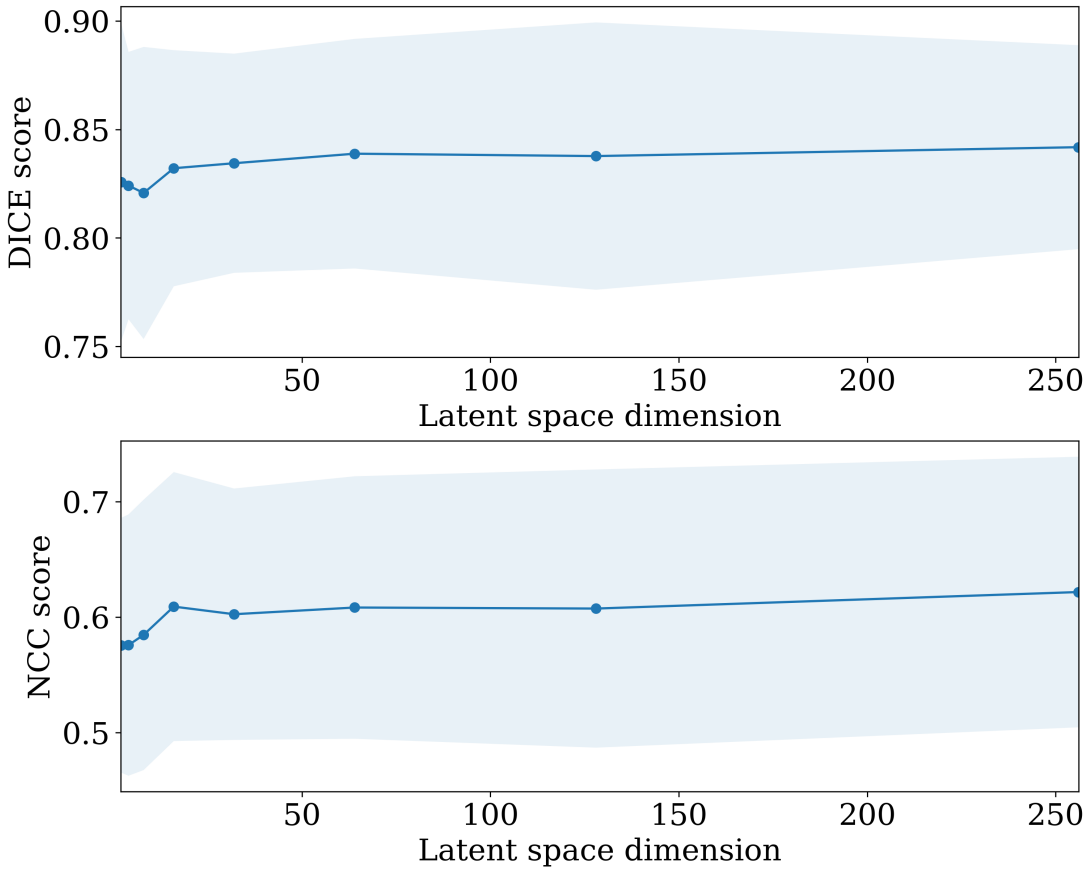


FIGURE 4.3: Reconstruction accuracy versus latent space dimensionality. The figure displays on the first row the DICE score and on the second row the NCC evaluated on the test set for models with the same hyperparameters and varying latent space dimensions.

4.5.2 Expected anatomical changes

This subsection details the anatomical changes in head and neck patients that literature studies report on. An overview of these changes can be seen in Table 4.1. This overview is used in Subsection 4.5.3 to assess the degree to which the changes observed in the training set, and therefore the changes that the DAM_{HN} learns to predict, correspond to the ones in the broader population.

The work of (Bhide et al., 2010) used repeat CT scans at weeks 2, 3, 4, and 5 during radiotherapy and compared the parotids and the target at succesive time points, i.e., pretreatment with week 2, week 2 with week 3, and so on. The greatest absolute and percent reduction in the volume of the parotid glands was 4200 mm^3 or 14.7 %, and occurred between week 0 and week 2. The absolute and percent reduction in the next two-week period was 4000 mm^3 or 16 %. The study found a significant medial shift of the parotid glands through the course of treatment, starting at week 2, with the highest mean movement of the COM being 2.3 mm at week 4. No significant movements of the COM in the anteroposterior and the inferosuperior directions were found.

In the work of (Vásquez Osorio et al., 2008) the impact of 46 Gy delivered to the tumor was assessed based on the planning and repeat CT images. They report that the parotids shrunk on average by 14 % and that the shrinkage occurred by keeping the regions nearby to bony anatomy as an anchor. Moreover, the parotids exhibited a tendency to move inward (right parotid leftward and left parotid rightward) with the

largest displacements being in the lateral and inferior regions. The region that moved the least was the medial region (partially adjacent to the bony structure). The study of (Barker et al., 2004) found a median medial shift of 3.1 mm for the center of mass of the parotid glands. They observed asymmetric shifts in parotid gland surfaces, with average displacements of 1 ± 3 mm and 3 ± 3 mm for the medial and lateral regions of the irradiated glands, respectively.

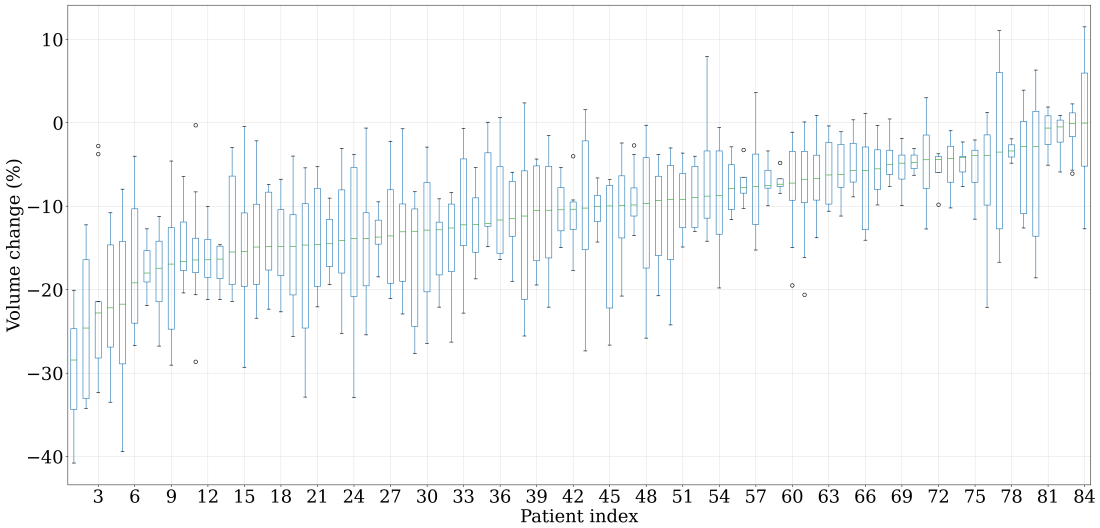
TABLE 4.1: Overview of documented quantitative and qualitative anatomical changes in the parotid glands. The table displays the study, the number of CTs used, the reported volumetric change (absolute, relative or both), the absolute shifts in the COM and its direction and qualitative notes on the reported changes.

Study	CT number	Volumetric loss	COM shift	Morphological alterations and notes
(Barker et al., 2004)	≥ 2	Median 190 mm^3 per day Range of 40 to 840 mm^3 per day	Median 3.1 mm Range 0 to 9.9 mm in medial direction	Shrinkage correlated with patient weight loss
(Vásquez Osorio et al., 2008)	2	Average 14 %	1 or 3 mm	Bony anatomy kept as anchor during shrinkage
(Bhide et al., 2010)	≥ 2	14 % or 4200 mm^3 between week 0 and 2 16 % or 4000 mm^3 between week 2 and 4 35 % over the course of chemoradiotherapy	2.3 mm by week 4 in the medial direction	COM shift insignificant in the anteroposterior and inferosuperior directions
(Santos et al., 2020)	2	Average 20.5 % or 6560 mm^3 between CTs	N.A.	Shape shift from convex to concave COM shift towards the medial and cranial directions

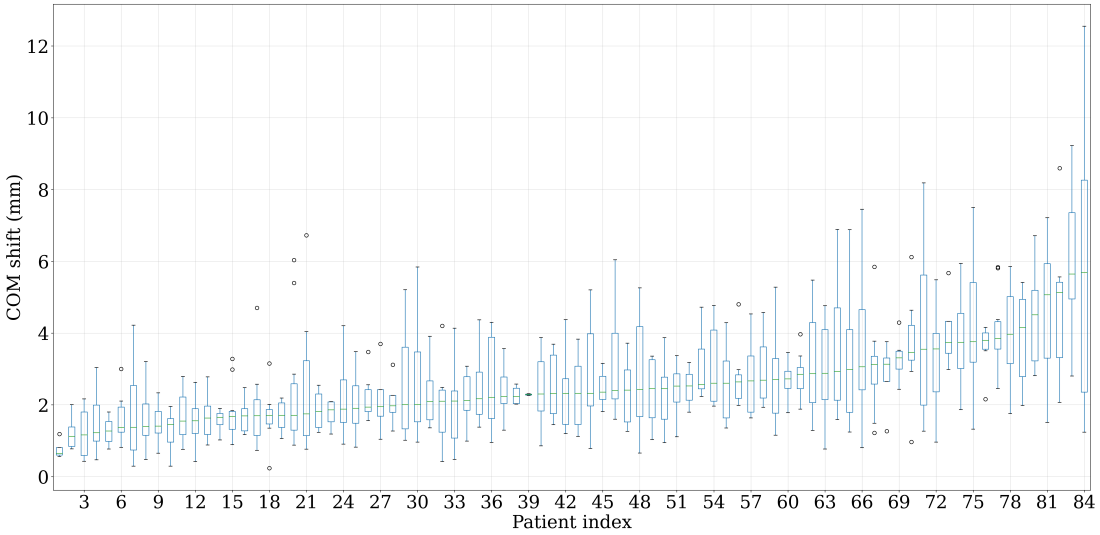
4.5.3 Training set anatomical changes

The generative performance of the model is tied to the data provided during training in the training set. Therefore, the anatomical changes in the training set and the literature reported changes from Table 4.1 were compared to assess the degree to which the training set is representative of the broader proton therapy head and neck patients population. The anatomical changes presented in Subsection 4.5.2 come from studies in which uni or bilateral photon-based radiotherapy (RT) or a combination of chemotherapy and RT was delivered. In contrast, the dataset of this work comes exclusively from proton therapy patients treated with mostly bilateral fields. The training set contained anonymized data and was composed of pairs of pCTs and consecutive rCTs (pCT-rCT_1, pCT-rCT_2, and so on). For each such pair and patient, the volume loss and COM shift in each parotid was computed and averaged over both parotid glands. Figure 4.4 displays, for each patient in the training set, boxplots of the distributions of percentage parotid glands volume changes and parotids center of mass shifts. Figure 4.4 shows that the median of the volumetric loss in the parotids is ≈ 11 % and the median of the COM shift is ≈ 3 mm. While the many patients have relatively unskewed volumetric change and COM shifts distributions, there are also patients (e.g., 3, 10, 60 and 72) that display skewed distributions with outliers. To facilitate comparison to previous publications, the data presented in Figure 4.4 is summarized in Table 4.2 where statistics on an individual parotid level are displayed. Specifically, the absolute volumes on the planning and repeat CT images, their difference (absolute and relative) and the COM shifts are characterized through their mean, standard deviation (SD), minimum, median and maximum.

The absolute volumes of the parotids on the pCT images are a mean of 35 878 mm^3 with a range of 16 984 to 83 520 mm3 for the left parotid and a mean of 35 447 mm^3 with a range of 11 344 to 87 352 mm3 for the right parotid. Both mean parotid volumes



(A) Patient specific box plot of the relative volume change distribution in both parotid glands.



(B) Patient specific box plot of the COM shift distribution in both parotid glands.

FIGURE 4.4: Training set characterization. The figures display median sorted boxplots (with the whiskers extending up to $1.5 \cdot \text{IQR}$, $\text{IQR} = Q_3 - Q_1$) with the x axis giving the patient identifying number and the y axis giving either the relative volumetric changes or the COM shifts distributions in both the parotid glands.

are roughly 23 % larger than the volumes reported by (Santos et al., 2020), namely 28 477 mm^3 for the left parotid and 29 274 mm^3 for the right parotid.

The differences between the parotid volumes in the training set are, a mean of -4307 mm^3 with a range of $-30\,456$ mm^3 to 4256 mm^3 corresponding to a mean of -12 % with a range of -41 % to 10 % for the left parotid and a mean of -3941 mm^3 with a range of $-29\,112$ mm^3 to 4584 mm^3 corresponding to a mean of -12 % with a range of -41 % to 10 % for the right parotid. This is slightly smaller but in line with previous studies, considering the averaging effect caused by the pCT-rCT pairings from the training set.

The COM shifts observed in the dataset are a median of 2 mm with a range of 0.2 mm to 13 mm for the left parotid and a median of 3 mm with a range of 0.4 mm to 12 mm for the right parotid. These values are in agreement with the median of 3.1 mm

TABLE 4.2: Training set statistics. The table displays for both parotid glands the mean, standard deviation, minimum, median and maximum of the volume on the planning and repeat CT images, the difference between these volumes (absolute and relative) and the center of mass shifts.

		Statistic				
Organ	Metric	Mean	SD	Min.	Median	Max.
Parotid L	Planning volume (mm^3)	35878	11290	16984	33280	83520
	Repeat volume (mm^3)	31571	10161	12976	29816	76632
	Difference (mm^3)	-4307	3880	-30456	-3548	4256
	Relative difference (%)	-12	9	-41	-11	10
	COM shift (mm)	3	2	0.2	2	13
Parotid R	Planning volume (mm^3)	35447	12568	11344	33024	87352
	Repeat volume (mm^3)	31507	11160	7496	29896	79136
	Difference (mm^3)	-3941	3955	-29112	-3320	4584
	Relative difference (%)	-11	8	-41	-10	10
	COM shift (mm)	3	2	0.4	3	12

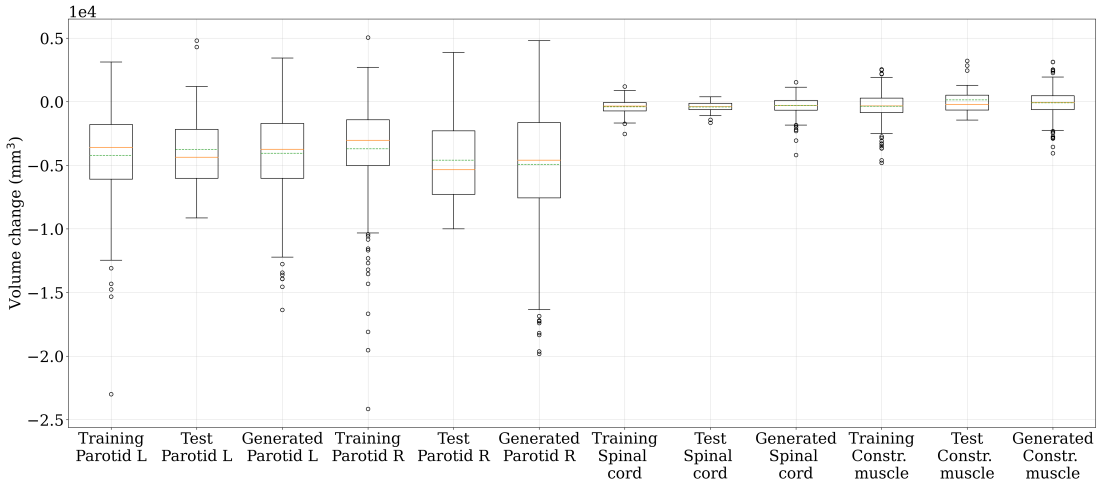
in a range of 0 mm to 9.9 mm reported by (Barker et al., 2004).

To conclude, the distributions from the training set are deemed in line with the expectations set out by previous studies. Differences between the data presented here and the one from previous studies, such as (Medbery et al., 2000) and (Santos et al., 2020) can be attributed to several factors. First, the pCT-rCT composition of the training set is bound to underestimate the changes when compared to studies based on only pCT-final CT pairs. Second, differences are expected due to the anonymization of the training set and the differences between the compared cohorts. Previous studies such as the ones of (Ericson, 1970; Santos et al., 2020; Vásquez Osorio et al., 2008) showed differences in parotid volumes depending on age, sex, weight, smoker status, planned doses, degree of parotid sparing and treatment modality, which are impossible to study in our current case. Third, a small effect could be expected due to inter-observer variability and systematic errors introduced by interpolating the original images on a new, coarser grid could also influence the observed absolute volumes.

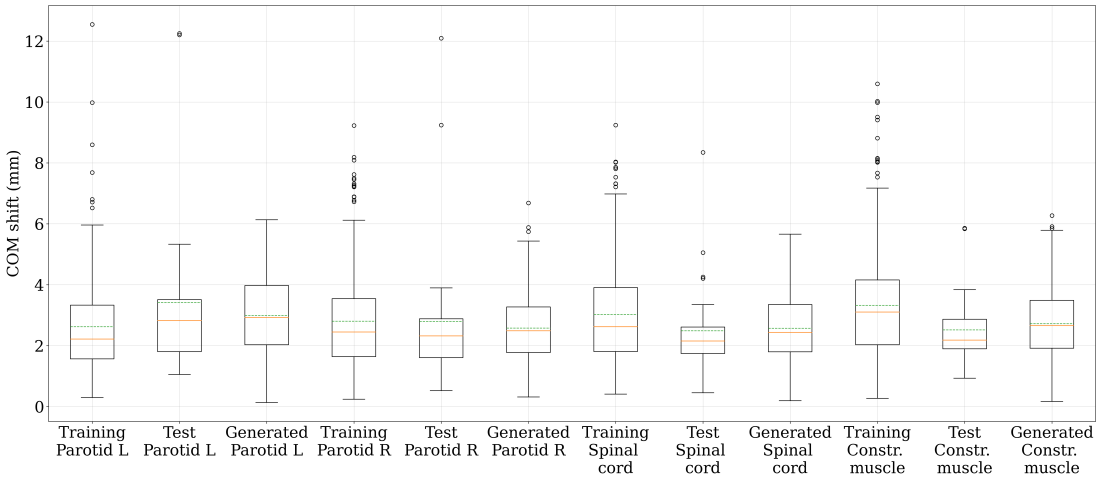
4.5.4 Generative performance

To assess the generative performance of the model, the test set, that contained 9 patients, was input into the final trained model and 100 samples were drawn for each record (pair of pCT-rCT) in the test set. Figure 4.5a displays for all present organs (left and right parotids, the spinal cord and the constrictor muscle) boxplots of the volume changes on the training, test and generated sets. Figure 4.5b displays for all present organs boxplots of the COM shifts on the training, test and generated sets.

In terms of volumetric change distributions, shown in Figure 4.5a, the figure shows that the parotid distributions on the training and test set are different. For example, the mean (indicated by the dotted green line) of the left parotid volume change distribution is below its median (indicated by the continuous orange line), while it is above it on the test set. A similar situation occurs for the right parotid. The same figure shows that the model generates volume change ranges that are broad enough to encompass the training and test sets, with means and medians in close agreement to the training and test set ones. The COM shift distributions, shown in Figure 4.5b, also display differences between the training and test sets. For example, the distribution of the



(A) Organ specific box plots of the absolute volume change distributions.



(B) Organ specific box plots of COM shift distributions.

FIGURE 4.5: Organ specific generative performance. The figures display boxplots (with the whiskers extending up to 1.5·IQR, IQR $= Q_3 - Q_1$) for the training, test and generated sets with the x axis showing the organ and the y axis giving either the absolute volume change or the COM shift distribution for the given organ. The median of the distributions is displayed with the continuous orange line while the mean is displayed with the dotted green line.

constrictor muscle COM shifts on the test set has a considerably smaller range of values, with smaller mean and median values. As was the case for Figure 4.5a, Figure 4.5b also shows that the model predicts distributions of COM shifts that are broad enough to encompass the test set ones, with means and medians in reasonable agreement. Some discrepancies can also be observed, for example in the difference between the median of the distribution of COM shifts of the constrictor muscle on the test and generated sets. Given the overall good agreement presented by both Figures 4.5b and 4.5b, it can be concluded that DAM_{HN} is capable of modelling volume and COM shift distributions present in the training and test set.

An illustration of the generative capabilities of the model is shown in Figure 4.6. The figure displays for 5 patients in the test set, in the first column the pCT, in the second column one of the rCTs and in the following 3 columns three patient specific generated CT images with corresponding contours (the left parotid colored in red and

the right parotid colored in orange, the spinal cord in green and the constrictor muscle in blue).

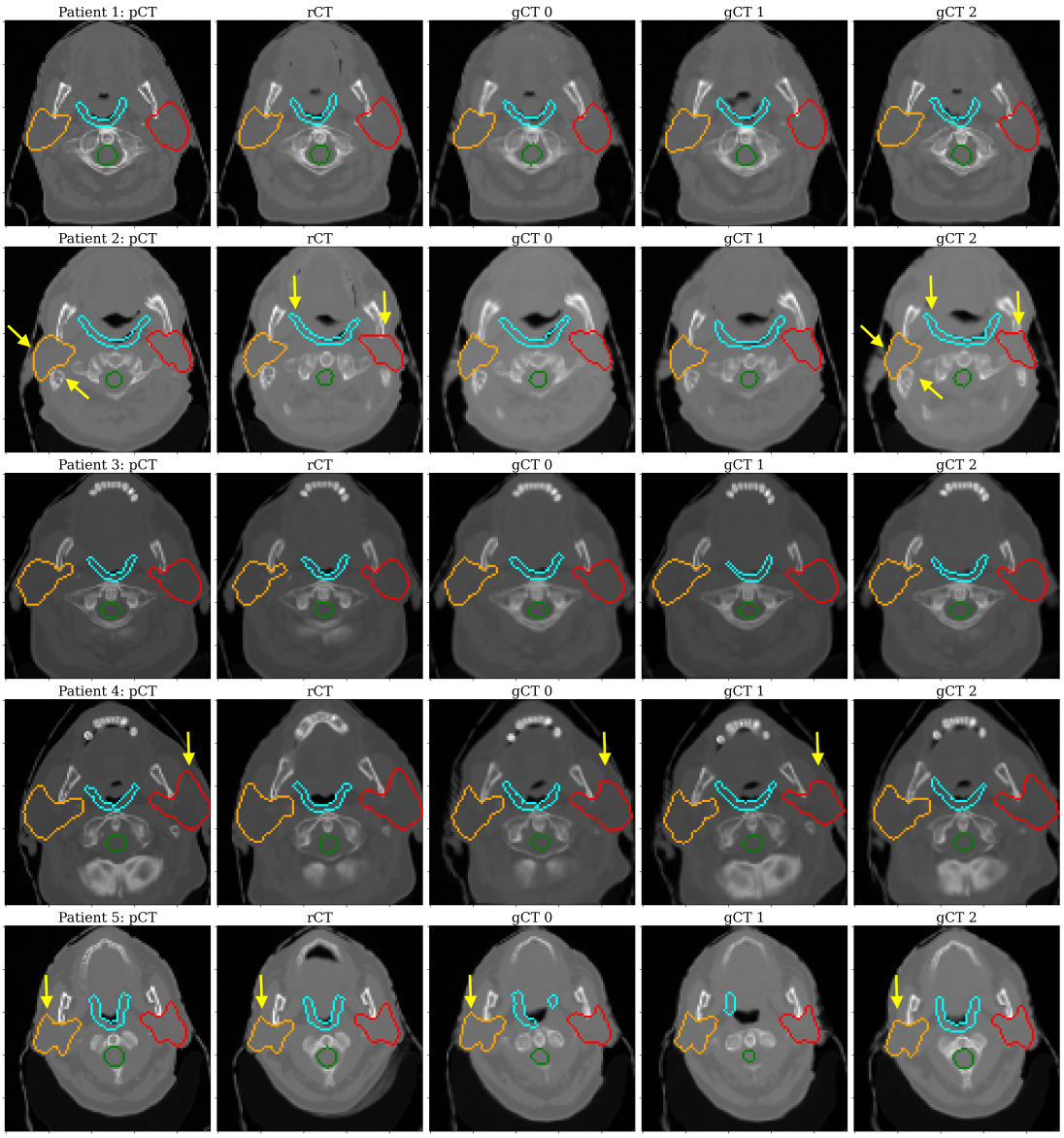


FIGURE 4.6: Example of generated images. The figure displays, for 5 randomly selected patients from the test set, in the first column the true pCT, in the second column one of the true rCTs and in the remaining columns generated CT images. Overlaid on all images are the left parotid (red), the right parotid (orange), the spinal cord (green) and the constrictor muscle (blue). Noteworthy anatomical changes are indicated with yellow arrows.

As already mentioned in Table 4.1, the flattening and medial movement of the parotids is expected. This feature is illustrated for the patient 4 through the yellow arrows in the planning and generated images shown in columns 3 to 5. Patient 2 displays shrinking in the right parotid (in orange) and flattening of the left parotid (in red) as illustrated by the yellow arrows. The model also appears to predict neck pose shifts, as illustrated by the changing air gap in the oral cavity of patient 1 in the second generated image or by the change in the shown dentition of patient 5 in the generated image 1. Weight loss, which is usually observed in radiotherapy patients, is prominent

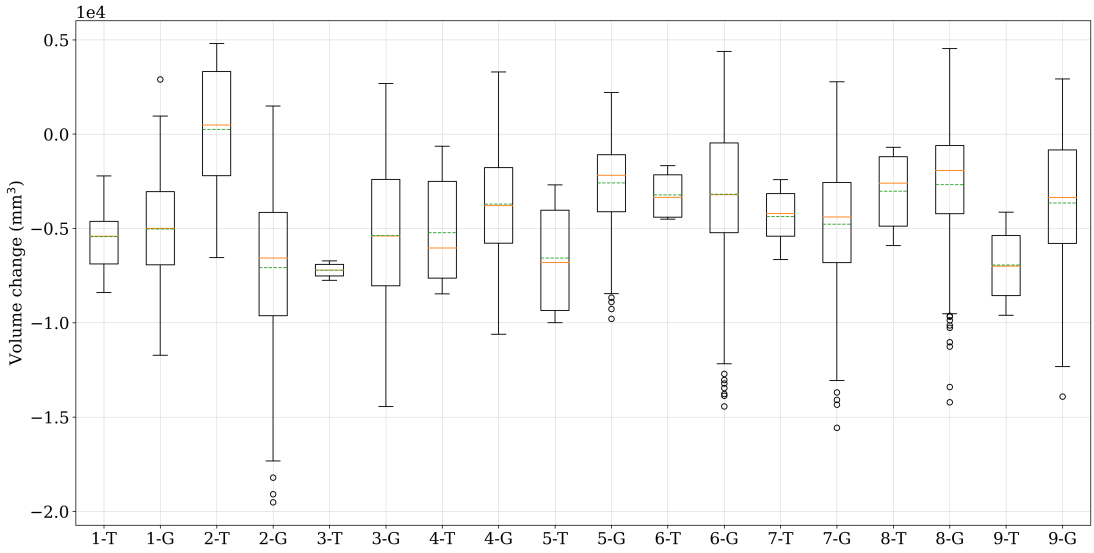
in the comparison between the pCT and the generated images for patient 4. Minimal overlap between the parotid glands and the mandible bone is visible for patient 1 on the pCT and the rCT. The generated images also display this feature, which illustrates the anatomical coherence of the generated anatomies. While it is difficult to definitively assert the feasibility of the generated image, the figure supports the conclusion that the model is capable of generating realistic anatomies that are coherent and involve posture shifts, shifting air gaps, weight loss and the typical expected anatomical changes in the parotid glands.

To further test the population based model, Figure 4.7 shows patient-specific boxplots of the anatomical changes in the parotid glands. Figure 4.7a displays for each patient in the test set, the true volumetric change (denoted by the patient number and -T) and the generated volumetric changes by drawing 100 samples (denoted by the patient number and -G). In terms of the volume change distributions illustrated in Figure 4.7a, the model largely predicts broad enough distributions that encompass the true ones. This is the case for patients 1, 3, 4, 6, 7, 8 and 9. Moreover, the means and medians are in close agreement for patients 1, 6, 7 and 8. Discrepancies in the means and medians can be observed for patients 2, 4 and 9. In terms of COM shift distributions, the model produces distributions with large enough ranges to encompass the test set ones, except for patient 5. The means and medians are in agreement for most patients, with the exception of 5, 6 and 9. The discepancies on a per-patient level could be explained by an insufficient number of recorded repeat CT images for those patients but also by the non-patient specific nature of the model. While the model attempts to provide patient specificity by allowing the parameters of the prior distribution to depend on the planning CT image and associated masks, the model optimizes the log likelihood of the full dataset, therefore resulting in a sample (or population) based model.

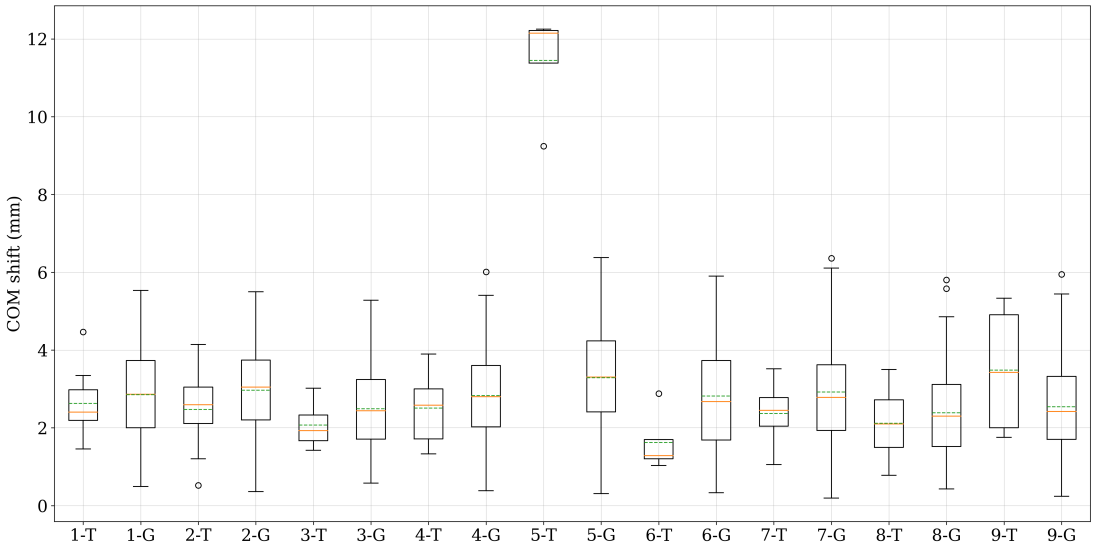
4.5.5 Comparison to DiffuseRT

The generative performance of DAM with respect to principal component analysis (PCA) based models has already been documented in the previous work of (Pastor-Serrano et al., 2023), where it was shown to outperform them. Thus, the generative performance of this model was compared with the recently published denoising diffusion probabilistic model (DDPM) of (Smolders et al., 2024). DDPM is also a generative deep learning model that approximates a data distribution, by inverting a gradual multi-step noise addition process. Similarly to the results shown by DDPM, Figure 4.8 displays for all organs, the true (training set) and generated volume change distributions (in Figures 4.8a, 4.8b, 4.8e and 4.8f) and COM shift distribution (in Figure 4.8c, 4.8d, 4.8g, 4.8h) together with a kernel density estimate for each. The kernel density estimate was computed using the Scikit library (Pedregosa et al., 2011) and a kernel bandwidth defined as one tenth of the range of values in the distribution. Both volume change and COM shift distributions that the DAM_{HN} training set exhibits are qualitatively different than the ones reported by DDPM, displaying less bimodality. This difference is likely attributable to the differences in the patient cohort and the specifics of treatment delivery (e.g., the chosen number and direction of beams). The kernel density estimates for the training and generated sets are generally in agreement, with disagreement occurring at the ends of the distributions, as is visible in Figures 4.8g and 4.8h.

DAM_{HN} and DDPM were also compared in terms of the Wasserstein distance (WD) between the true (training set) and generated anatomical changes distributions. The Wasserstein distance is a metric for probability distribution similarity, with a value of zero occurring when the distributions are the same and larger values indicating more different distributions. To compute it, the volume changes and COM shifts in the



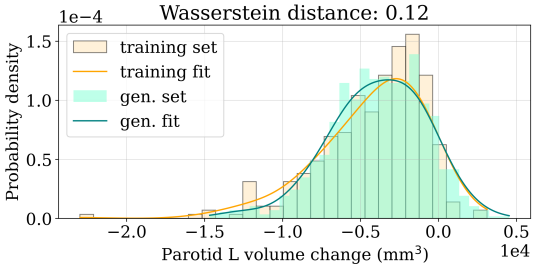
(A) Patient specific box plot of absolute volume change distributions in both parotid glands.



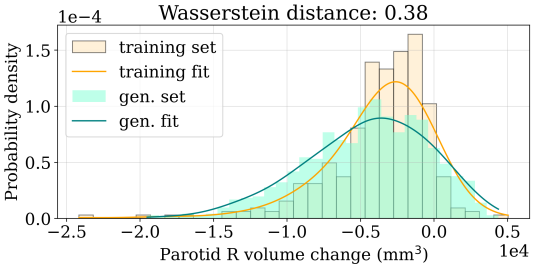
(B) Patient specific box plot of COM shift distributions in both parotid glands.

FIGURE 4.7: Patient specific generative performance. The figures display boxplots (with the whiskers extending up to $1.5 \cdot \text{IQR}$, $\text{IQR} = Q_3 - Q_1$) with the x axis showing the organ and the y axis giving either the absolute volumetric changes or the COM shifts for the given organ. The median of the distributions is plotted using the continuous orange line while the mean is plotted using the dotted green line.

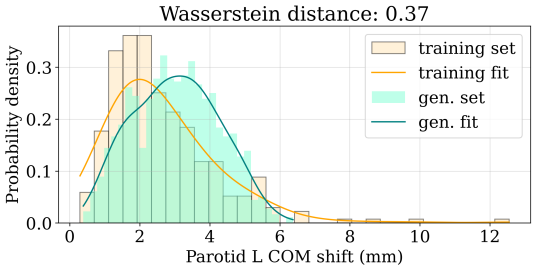
organs for both training and generated sets were normalized by the mean and standard deviation of the true (training set) values (to counter the scaling effect of the WD based on the range of the data) and thereafter input into the SciPy implementation (Virtanen et al., 2020). Table 4.3 shows the comparison between DDPM and DAM_{HN}. The qualitative agreement observed in Figure 4.8 is also illustrated by the low Wasserstein distances achieved by DAM_{HN}, which is comparable to the ones obtained by DDPM for all metrics.



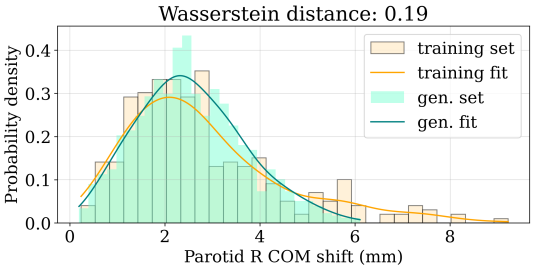
(A) Left parotid volume change distributions.



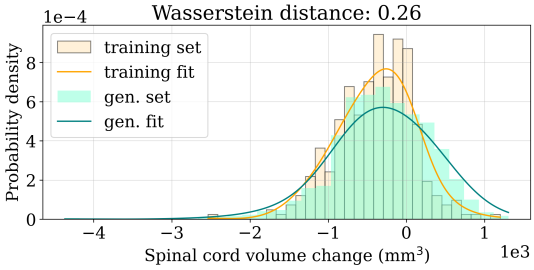
(B) Right parotid volume change distributions.



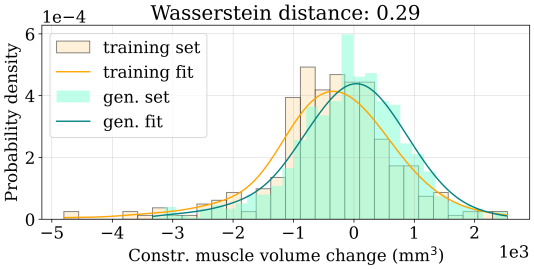
(C) Left parotid COM shift distributions.



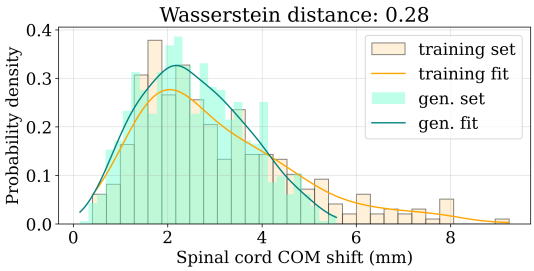
(D) Right parotid COM shift distributions.



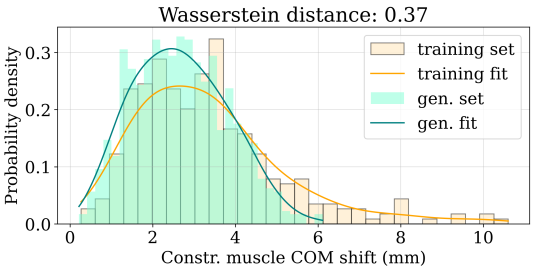
(E) Spinal cord volume change distributions.



(F) Constrictor muscle volume change distributions.



(G) Spinal cord COM shift distributions.



(H) Constrictor muscle COM shift distributions

FIGURE 4.8: Comparison of true and generated anatomical change distributions. The figures display for all organs (left and right parotid, constrictor muscle and spinal cord) the true and generated anatomical change (volume change and COM shift) distributions, their corresponding kernel density estimates and the WD between the true and generated distributions in the title.

TABLE 4.3: Wasserstein distance comparison between the best performing DDPM model of (Smolders et al., 2024) and DAM_{HN}. The table displays the Wasserstein distance between the true (training set) and generated volume loss and COM shift distributions in the left and right parotids.

Metric	Structure	Model	
		DDPM	DAM_{HN}
Δ Volume	Left parotid	0.60	0.12
	Right parotid	0.28	0.38
COM shift	Left parotid	0.31	0.37
	Right parotid	0.22	0.19

4.5.6 Latent space analysis

Given that DAM_{HN} encodes the information between the planning and repeat CT images into the latent space, the effect of varying individual latent variables while keeping the others fixed of the model on organ volume changes and COM shifts was investigated. Figure 4.9 illustrates the volume changes for each organ (left parotid, right parotid, spinal cord and the constrictor muscle) that occur when an individual latent variable is varied from -5σ to 5σ, while the others are kept fixed to 0. Similarly, Figure 4.10 displays the effect of varying individual latent variables on the COM shift.

Figure 4.9 shows consistently larger volumetric lossess in the parotid glands in comparison to spinal cord and constrictor muscle. This is expected, given that the spinal cord is smaller in volume than the parotid glands and is usually avoided during irradiation. Figure 4.9 also shows the relatively smooth latent space that the model learns and that the parotid glands volume changes are comparable, indicated the largely bilateral nature of the patient cohort. Variables that induce larger volumetric losses in one of the two parotids, could point to the presence of patients with unilateral fields, as non-irradiated parotids were shown to shrink less during treatment than radiated ones (Vásquez Osorio et al., 2008). Figure 4.10 shows that for both parotids, the COM deformations are roughly similar in absolute value. This is in line with the expectation, set by the work of (Vásquez Osorio et al., 2008), that both parotids move in the medial direction with similar amplitudes. Moreover, Figure 4.10 also shows that the learned latent space is smooth.

Volume and COM shifts are just one measure of latent space variations. Figure 4.11 shows, for a patient in the test set, a cut of the images produced when latent variables with numbers 1, 7, 20, 21 and 32 are varied. The particular latent variables were chosen due to the large changes they induce, as visible in Figures 4.9 and 4.10. The first column of Figure 4.11 displays the pCT, while the remaining columns display the image, the associated contours (as before the left parotid in red, the right parotid in orange, the spinal cord in green and the constrictor muscle in blue) and the overlaid deformation vector field that is created by the individual latent variables (with the value it was set to given in the title of the figure). As was already visible in Figures 4.9 and 4.10, latent variable 7 induces large changes in the right parotid for extreme values of the latent variable. This effect is also observed through the deformation vector field around this structure. Latent variable 21 displays a similar behavior, for both the left and right parotids. Figures 4.9 and 4.10 also show that latent variable 1 and 32 generate deformation fields in the oral cavity, perhaps pointing to shifting patient poses. A limitation of the framework, is that the latent variables are not encouraged to generate

non-correlated deformations and therefore, it is difficult to relate specific latent variables to specific induced anatomical changes.

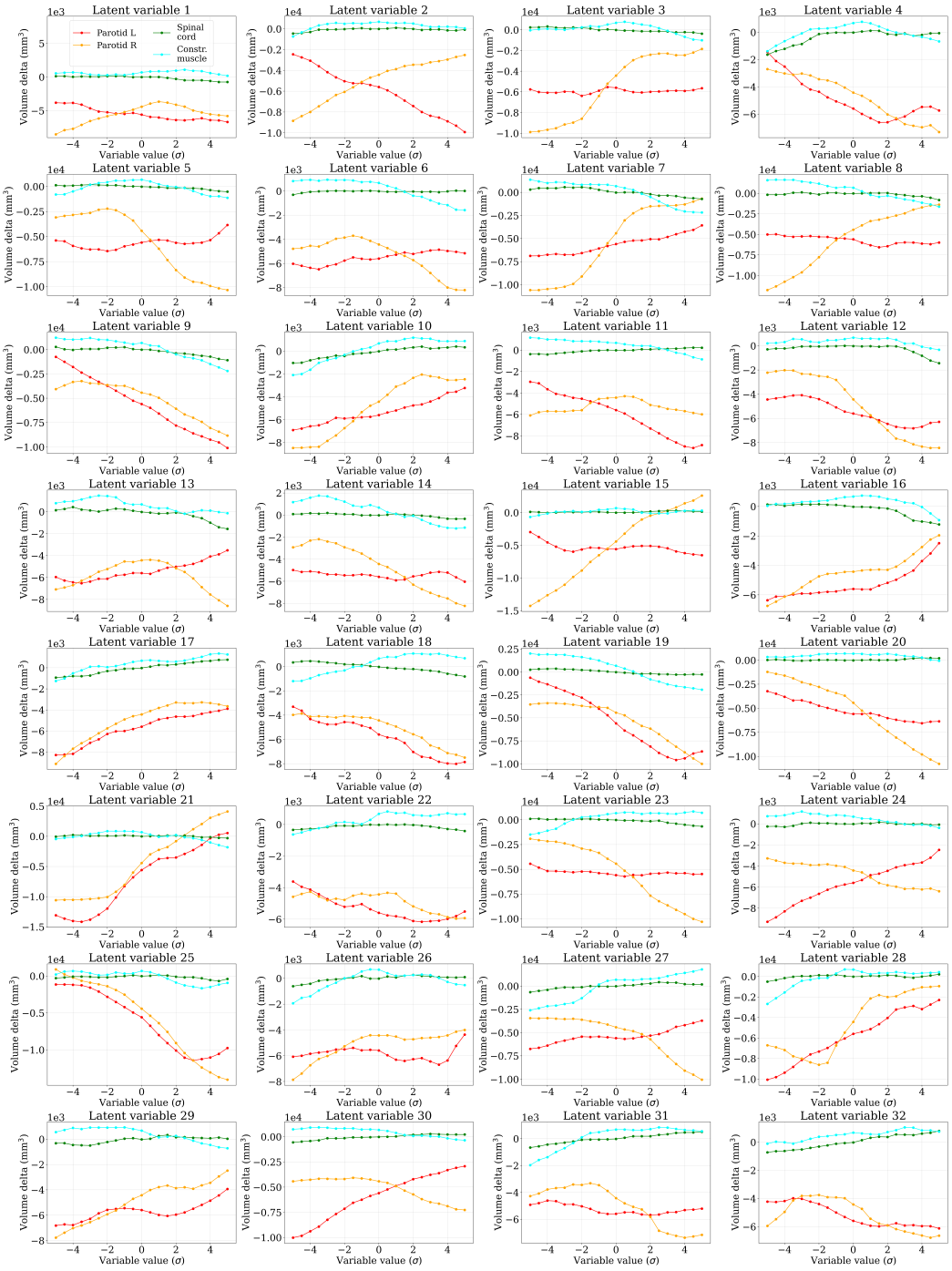


FIGURE 4.9: Latent space variations. The figure displays the organ (left parotid in red, right parotid in orange, spinal cord in green and constrictor muscle in blue) volume change that individual latent variables cause. The latent variables were varied from -5σ to 5σ while the remaining variables were set to 0.

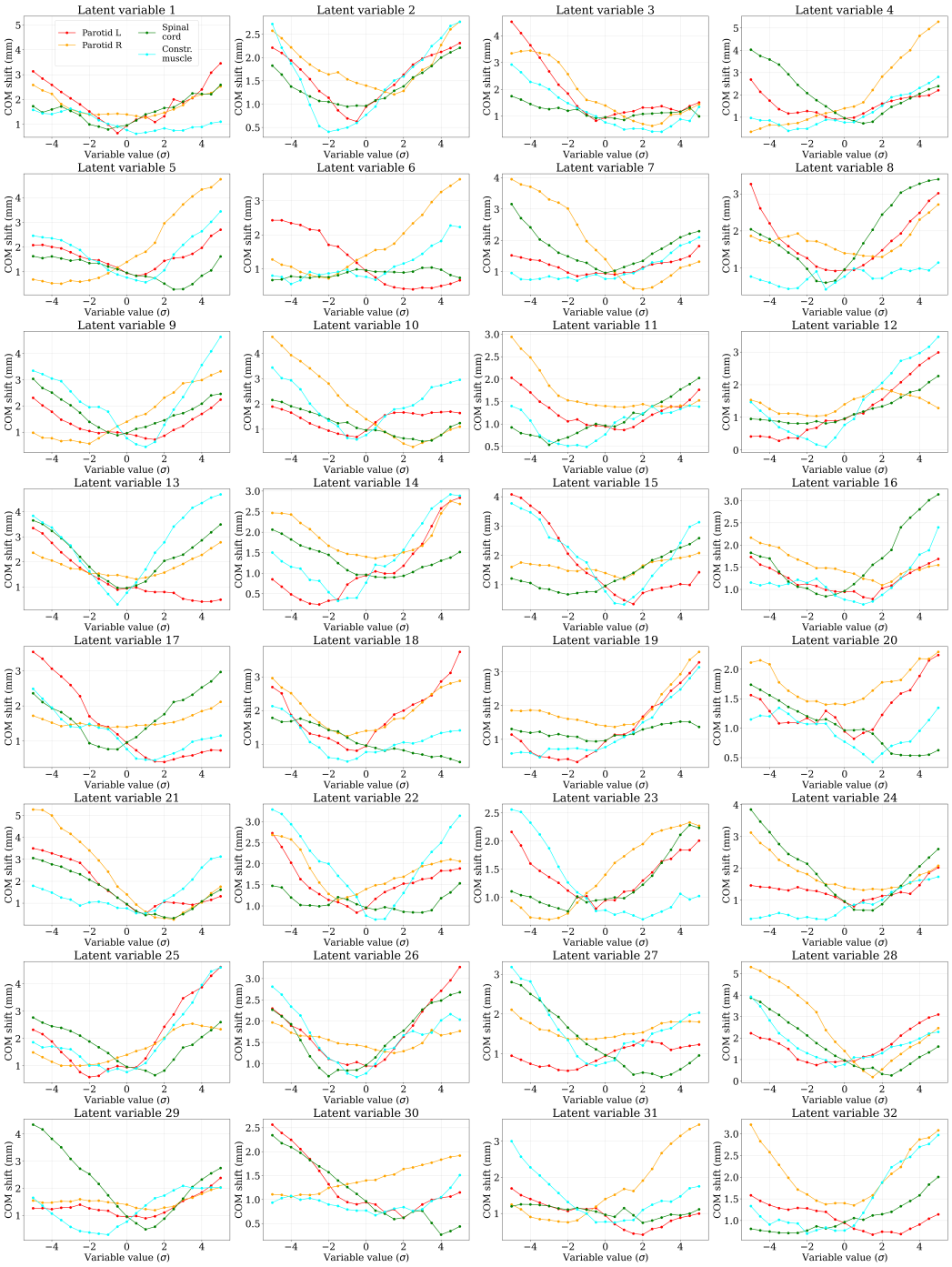


FIGURE 4.10: Latent space variations. The figure displays the organ (left parotid in red, right parotid in orange, spinal cord in green and constrictor muscle in blue) COM shift that individual latent variables cause. The latent variables were varied from -5σ to 5σ while the remaining variables were set to 0.

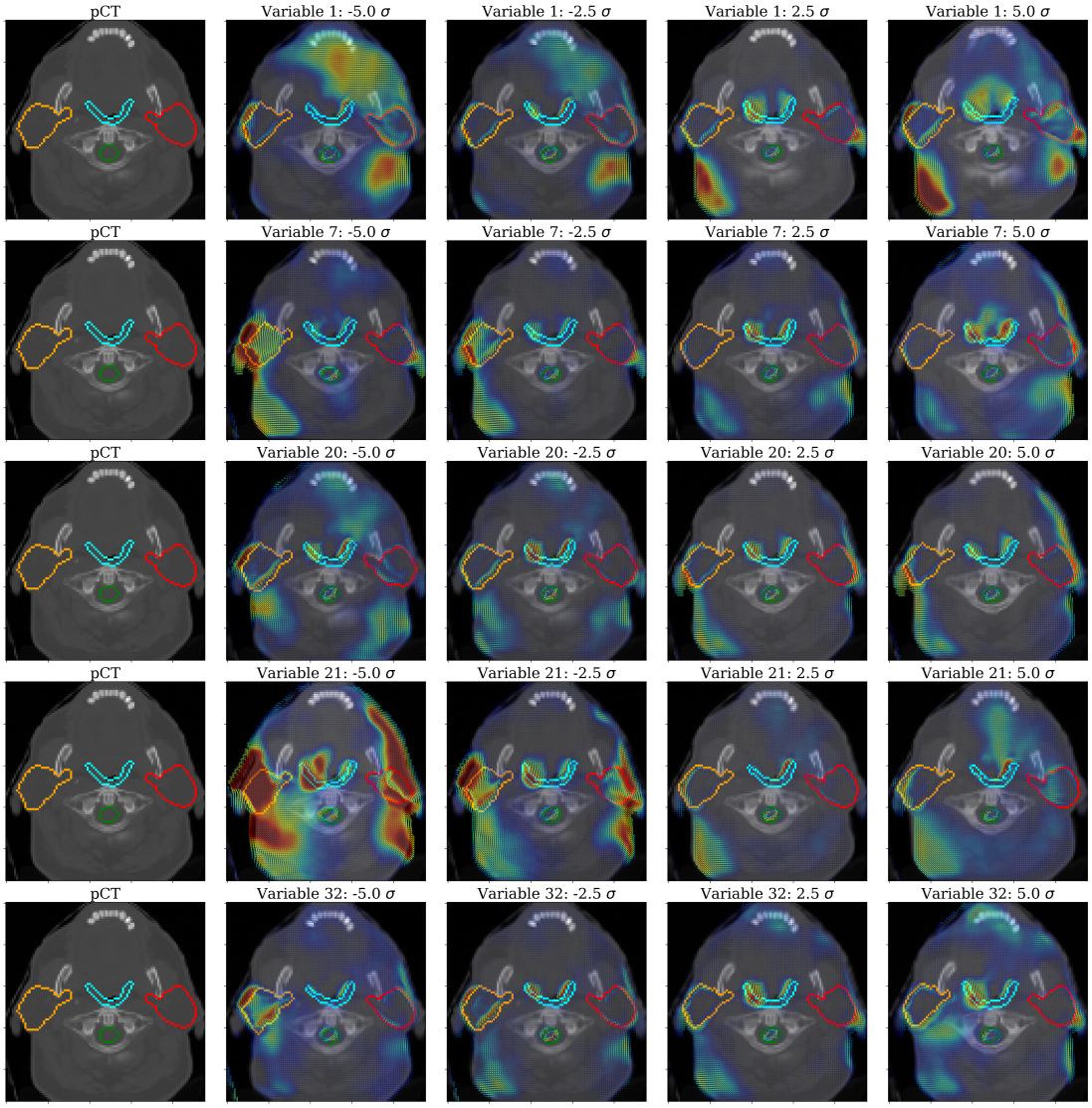


FIGURE 4.11: Latent space visualization. The figures display in the first column for a given patient, the pCT and associated organs (left parotid in red, right parotid in orange, spinal cord in green and constrictor muscle in blue). In the following columns the figure displays in the title the chosen latent variable number (and the value it was set to), the generated image and its organs. Overlaid is plotted the deformation vector field that the model learns, where the color represents the magnitude of the field.

4.6 Conclusion

This work presented a probabilistic deep learning model for generating future anatomical changes in H&N radiotherapy patients. The model was trained on a training set coming from 83 proton therapy head and neck patients and was assessed on test set coming from 9 patients. On the test set the model achieved a DICE score of 0.83 and an NCC score of 0.60 using 32 latent variables. The model produces volumetric changes and COM shift distributions that are broad enough to capture the real, observed ones, with the predicted means being close to the real ones. DAM_{HN} was compared to the state of the art denoising diffusion probabilistic model (DDPM) for H&N anatomical changes presented by (Smolders et al., 2024). For both parotid glands, DAM_{HN} achieved similar Wasserstein distances to the ones obtained by the DDPM model between the true and generated volume loss distributions (0.12 versus 0.60 and 0.38 versus 0.28) and between the COM shift distributions (0.37 versus 0.31 and 0.19 versus 0.22). The latent space analysis showed that the model learns a smooth latent space, that displays some correlation between the latent variables (which was not discouraged in the model framework). Although this work focused on data coming from a proton therapy patient cohort, the methodology is valid for a wider range of problems in adaptive radiotherapy, including adaptive photon radiotherapy.

There are several limitations of the current methodology and model framework and points for future improvement and studies. First, the dataset contained a different number of repeat CT images for each patient and is therefore biased towards patients with larger anatomical changes (as those patients are more likely to be re-imaged). This bias was not accounted for in this model and likely leads to the model predicting larger than observed anatomical changes for patients with small ones. However, given that a dataset with larger anatomical changes is more difficult to encode in the latent space, a dataset that contains repeat CT images from patients with small anatomical changes should not significantly decrease the overall population accuracy of DAM_{HN}. Second, a limitation of the model is that, despite allowing the parameters of the prior distribution to vary on an individual patient level, the model is intrinsically a population based one, as it optimizes the log likelihood of observing the full dataset. This, coupled to a limited number of repeat CTs in the dataset leads to degraded accuracy for some patients. Third, if the large number of structures present in the head and neck area would be included in the dataset, it is expected that the model would require a change in architecture (specifically an increase in the number and size of layers and latent space dimensionality) to correctly model those datasets. More generally, the necessary minimal architecture and the optimal weights of the different loss terms should be further investigated. Moreover, the inclusion of the regularization term that penalizes large gradients in the deformation could be detrimental for anatomical regions where such changes do occur (e.g., tongue position). Fourth, the comparison between DAM_{HN} and the model of (Smolders et al., 2024), is ultimately difficult due to the different datasets that the models were trained and evaluated on. Thus, both models should be trained and evaluated on the same sets that contain more structures than they presently do (e.g., additional useful structures could be the submandibular glands and the oral cavity). Fifth, the structure of the dataset could be changed from pCT-rCT_1, pCT-rCT_2, and so on to pCT-rCT_1, rCT_1-rCT_2 and so on. In doing so, a model that predicts changes on the time scales on which patients are re-imaged (weekly or daily depending on the workflow) could be obtained. Such a model would be applicable to an adaptive plan library approach or for plan quality assurance. Moreover, a time variable could be included in the architecture to encode information in addition to the repeat CT images. Sixth, as anatomy change predictions has applications in dose change

predictions, an analysis of the effect on dose characteristics (including a robustness analysis) of delivering treatment plans to the generated images is a natural next step for this model. Next to such a study, work on additional standards (beyond volume changes and COM shifts) for assessing the degree to which generated CT images are realistic should be established.

Overall, DAM_{HN} was capable of quickly generating hundreds of realistic images of inter-fractional anatomies. As already mentioned, such a model has a number of applications in the radiotherapy workflow, such as improving robust optimization, as a component in plan quality assurance in online adaptive proton therapy or in expanding the plan library approach.

Chapter 5

Conclusion

Adaptive proton therapy (APT) is proposed as a desirable solution to the over-reliance in intensity modulated proton therapy (IMPT) on the one robustly optimized irradiation plan generated prior to the start of treatment. One of the bottlenecks in the clinical implementation of online APT comes from the manual quality assurance (QA) procedures that are necessary for safe treatment delivery. Specifically, plan QA is the manual check by a medical physicist of the correct beam arrangement and parameters for each treatment plan. Patient-specific QA (PSQA) is the per-patient comparison between the dose generated by the treatment planning system (TPS) and the physical dose measurements performed in a water box.

This work presented two algorithms, namely Yet anOther Dose Algorithm (YODA) and Daily Anatomy Model for head and neck patients (DAM_{HN}), that can aid with the automation of PSQA and plan QA processes.

5.1 Outcomes

YODA is a fast, physics-based and adjoint-capable proton radiation transport algorithm. The physics-based approximations, as explained in Section 2.3, applied to the integro-differential Linear Boltzmann Equation, are the continuous slowing down approximation, energy loss straggling, the small angle Fokker-Planck approximation and the neglection of catastrophic inscatter integral. These approximations simplify the Linear Boltzman Equation into a partial differential equation (PDE). This PDE is separated into two PDEs by separating out the energy from the 6-dimensional proton phase-space density. One benefit of this approach is that one of the resulting PDEs (the one-dimensional Fokker-Planck equation) is efficiently solvable via numerical methods, as explained in Section 2.4, while the other (the Fermi-Eyges equation) has an analytical solution, as shown in Section 2.5 and Appendix B. This combination makes the algorithm fast. Another benefit is that the adjoint mathematical formalism can be applied to this system of PDEs. The adjoint formalism provides an estimate to the change in a metric (e.g., dose in a region of interest) as a function of a change in input parameters (e.g., the voxel HU value or the number of protons delivered to a spot). This approximation is advantageous when the number of input parameters that change is much larger than the number of metrics of interest, which is typically the case in proton therapy. The adjoint mathematical framework was developed in Section 2.7 for the case when the metric is the dose in a region of interest and the input parameters are the voxel HU values.

As a result of the physics-based approximations employed and the Gaussian beam splitting scheme developed in Section 3.3 to account for lateral heterogeneities, YODA computes doses with accuracy close to the golden standard Monte Carlo (MC) algorithm TOPAS in a fraction of the time. The dose computation capability was illustrated in Section 3.4 in a variety of set-ups such as different water boxes with varying slab inserts,

a head and neck CT scan, a prostate CT scan and a lung CT scan using beams with energies that encompass the clinical energy spectrum. To facilitate YODA's use in the clinic, a DICOM interface has been written. The interface takes clinical DICOM CT scans, RT structures and RT plans and uses them as inputs in YODA and thereafter compares the results with the RT doses predicted by the TPS. Moreover, a "commisioning" procedure was developed in Section 3.5 to match the doses produced by YODA to any clinically commissioned treatment planning system. This step is necessary for YODA to reproduce the actual optical characteristics of the in-room beams but also to account for raw data differences between the TPS dose engine and YODA. After "commisioning" YODA, the doses computed by the RayStation TPS were independently re-computed with YODA, thus illustrating YODA's independent dose calculation capabilities. YODA was also tested as a component in an adaptive workflow, namely as an adaptation trigger system, in Section 3.4. This scenario simulated an online workflow, where every day a decision must be made of whether or not to adapt the plan based on the dosimetric effect of delivering yesterday's plan to today's anatomy. To this end, two patients had robustly and non-robustly optimized plans generated in RayStation and exported alongside planning and repeat CT images to YODA. Next, the dose from the original plan to the target on the repeat CT image was computed in two ways: a simple and time-consuming re-computation of the treatment plan and the adjoint time-saving computation. The adjoint component was found to be accurate and fast (tens of seconds versus tens of minutes), thereby proving YODA's use in adaptation trigger systems.

DAM_{HN}, is a probabilistic deep learning algorithm capable of predicting, based on the patient planning CT image and associated contours, future possible anatomies observed throughout treatment. The probabilistic framework was embedded into a U-net generative framework that once trained, using the evidence lower bound, allowed to sample the latent space and generate deformation vector fields (DVFs), as explained in Section 4.3. In turn, the DVFs were used to generate possible patient CT images and associated RT structures. The performance of the algorithm was benchmarked in Section 4.5 in multiple tests such as its reconstruction accuracy, its generative capabilities and its latent space behavior and was found to perform well in all tests. DAM_{HN} has several possible applications within radiotherapy. First, the algorithm could be incorporated in the process of robust anatomical optimization in order to increase tumor coverage and decrease doses to surrounding organs at risk (Van de Water et al., 2018). Second, the algorithm could be included in the plan library approach (Oud et al., 2022). By predicting the future anatomy of the patient, and generating plans with varying degree of robustness, further NTCP reductions could be achieved. Lastly, the algorithm has applications in online adaptive proton therapy where it could aid in plan QA. Specifically, the algorithm can be used to generate future anatomies with truly optimal plans for each anatomy. Following this, on the given day when the patient is imaged and the adapted and refined plan is available, a quick plan check could be performed (in terms of coverage and dosimetric metrics and beam directions) between the adapted and refined plan and the truly optimal plan.

5.2 Recommendations and broader uses

YODA's improvement points are the accuracy (achieved through nuclear interactions modelling and improved splitting schemes), the speed (via a better multi-core implementation or a GPU one), the number of available metrics and input parameters that can be varied for the adjoint component and the lack of comprehensive multi-site studies on patient cohorts. Having these improvements in YODA would result in an accurate

physics-based TPS dose engine with accuracy close to MC, with the significant advantage of the adjoint component.

The simplest to tackle is the number of implemented metrics. Clinically relevant metrics such as minimum and maximum dose in a region, the tumor control probability or the normal tissue complication probability (NTCP) could easily be included. The effect of perturbations in additional input parameters should also be included. The easiest to compute (mathematically and computationally) in the adjoint mathematical formalism (as they result only in direct contributions) are variations in the spot MU values (which are related to the number of protons) and variations in the initial spot size. These quantities are easily extracted from log-files recorded during treatment delivery. Also recorded in the log-files are differences between the planned and actually delivered spot positions. To use these in the adjoint formalism it should be investigated how to convert the spot position changes into changes in the HU values along the beam path. This will allow YODA, once it is fast enough, to be used as a real-time dosimetric tool through the adjoint engine. Alternatively, log-files could be used in YODA's current implementation, by performing a-posteriori reconstructions of the actually delivered doses. By performing this type of QA, implicit machine QA would also be performed as failures to reach certain tolerances could imply failures in the beam delivery system.

YODA should model nuclear interactions, as they have a significant effect on the predicted doses, especially towards the high-side of the energy domain. This modelling should balance the speed and accuracy that YODA strives for. Simplistic implementations, such as the usual additional contribution to lateral scatter that classical pencil beam algorithms use (Soukup, Fippel, and Alber, 2005), could be sufficient for YODA's main purpose as a QA tool. Moreover, the effect of this additional parametrized scatter on the adjoint computations is bound to be minimal. If gamma index comparisons show that such a simplistic yet (likely) fast model is insufficient, more complex algorithms could be developed. One example would be to use a convolution based approach, similar to the collapsed cone method used in photon therapy (Ahnesjö, 1989). Another is to include the nuclear in-scattering integral in the Linear Boltzmann Equation but only from secondary protons (as they are the most significant particles that contribute to dose from nuclear interactions and are reasonably forward oriented (Paganetti, 2002)). Such an approach would likely increase the bandwidth of the system of linear equations solved in the Fokker-Planck equation and therefore would come with an increased computational burden. Another approach is given by the bipartition model, as explained by the work of (Luo, 1998). In this approach, the proton flux is split into a forward scattered component (the solution of which YODA already computes) and a large angle scattered component. Approximations could be employed for the large angle component to keep the computational expense to a minimum, as shown by the work of (Asadzadeh, Brahme, and Kempe, 2010).

YODA should also improve its accounting for lateral heterogeneities. As already discussed in Chapter 3, this is not trivial due to the complex interplay between the beam size, the CT grid size and the size and location of the lateral heterogeneities, and is a limitation that most pencil beam algorithms encounter. More lateral beamlets do not necessarily imply a more accurate result. A solution would be an adaptive beam splitting algorithm to balance speed and accuracy. Such an algorithm could start with a base level of splitting (e.g., 1+12) and upon detecting lateral heterogeneities, place lateral beamlets only in the relevant areas. As heterogeneities usually only last several voxels, upon reaching the end, the beamlets in that area could be re-consolidated into one beamlet (with potentially a non-Gaussian shape). The success of such an algorithm depends on the detection of lateral heterogeneities, on the re-consolidation method once the lateral heterogeneity has "ceased to exist" from a beam-eye point of view and on

the actual implementation of the algorithm. The detection of lateral heterogeneities could be a small cost as once the start and end positions of a spot in a treatment plan are computed, multiple ray traces would be started (again using a dense ring structure) in parallel. An alternative to the adaptive approach would be a fixed lateral grid in the beam-eye view, covering -3 to $+3$ of the initial spot size with a pre-defined (if needed non-constant) lateral resolution.

Despite YODA being aimed at proton therapy, its physics-based, PDE solving nature allows it to be applicable to more than proton therapy for humans. One application, that has already been tested, is YODA's use for small animal proton therapy treatment planning in the work of (Malimban et al., 2024), which found YODA to be a useful tool for the purpose. YODA uses physics-motivated approximations to the LBE to obtain the six-dimensional proton phase space density. Thus, YODA could be applied to the transport of other charged particles for radiotherapy purposes, as long as they are subject to the LBE equation and their scatter characteristics can be approximated by the same approximations that are applicable to protons. One example of this could be electron therapy. Broader in scope, are radiation transport problem in different environments from the radiotherapy one. One example of this could be space exploration, where the effect of cosmic and solar radiation on spacecraft materials is desired, or in battery applications, where the range of ions in different media is desired. Past radiation transport application, to which YODA could relatively easily be applied to, the solution to the Fokker-Planck equation (and the corresponding adjoint formalism) could make that component of YODA of use to different fields. Some example are the use of the Fokker-Planck equation in Biophysics to model the transport of molecules in cells, in population dynamic to study the growth and spread of populations, in statistical mechanics to model the time evolution of a probability density function in stochastic processes or in finance and economics where the Fokker-Planck equation is related to the price of financial options over time.

DAM_{HN} could also be improved in several ways. First, the algorithm should use more of the available masks in the dataset. Masks that identify the patient's skin, the submandibular glands, the oral cavity, the irradiation target and additional relevant organs could improve the algorithm's capability of generating realistic images, but also provide additional metrics (e.g., weight loss via the skin mask, tumor center of mass shifts) for further generative performance quantification. Second, the inclusion of additional masks would likely require a change in the number and size of layers in the network and in the required latent space dimensionality for appropriate modelling of a dataset with increased complexity. More generally, the necessary minimal architecture for a given dataset and the optimal weights of the different loss terms should be investigated. Moreover, the inclusion of the regularization term that penalizes large gradients in the DVFs could be detrimental for anatomical regions where such changes do occur (e.g., tongue position). Third, the input structure could be changed from the currently used pCT to multiple rCTs into a sequential pCT to rCT_1, rCT_1 to rCT_2 and so on. This would allow the algorithm to predict possible anatomies that occur over shorter time-frames, which would be useful in a week-to-week adaptive workflow (for example in anatomical robust optimization). Moreover, a time variable could be included in the architecture to encode information in addition to the repeat CT image. Fourth, as anatomy change predictions has applications in dose change predictions, an analysis of the effect on dose characteristics (including a robustness analysis) of delivering treatment plans to the generated images is a natural next step for this model. Next to such a study, work on additional standards (beyond volume changes and COM shifts) for assessing the feasibility of the generated anatomies should be investigated. Last, the dataset and test methodology should be opened to the broader scientific community, in

order to facilitate comparisons with other deep learning based algorithms that aim to model anatomical changes in head and neck patients.

Despite DAM_{HN} being applied to anatomical changes in head and neck proton therapy patients, the model is applicable to broader adaptive radiotherapy tasks. One example would be for predicting anatomical changes in patients that undergo photon based radiotherapy. This task would likely require minimal changes to the architecture and is an interesting avenue to explore. Another interesting study would be to use a combination of CT images and cone beam CT (CBCT) images. The frequent availability of CBCT images, could also be used for training an algorithm that creates CT images from CBCT ones. Additionally, image conversions such as magnetic resonance (MR) to CT could be investigated. Moreover, dose images could be included in the architecture, next to the CT images and structures, to generate possible future doses. Generic generative variational autoencoder tasks such as image generation, text generation and audio synthesis are also possible. The autoencoder architecture also has uses in dimensionality reduction, which can be used for visualization and clustering and anomaly detection, which can be done via detection in the latent space of outliers.

The presented models are solutions for the automation of patient-specific QA and anatomical change modelling in head and neck patients. Therefore, with sufficient improvements and testing in the clinical environment they could prove to be useful steps towards the clinical realization of automated QA in online adaptation in proton therapy. Through the automation of the QA processes, these algorithms will improve clinic efficiency, make proton therapy available to a wider group of patients and reduce toxicities associated with radiation therapy.

Appendix A

Proof of separability

Unlike in the homogeneous geometry considered by (Gebäck and Asadzadeh, 2012), in our more general, in-depth heterogeneous case it is not immediately straightforward that introducing the split from Equation 2.5 into Equation 2.4 yields the Fermi-Eyges Equation 2.7 and the 1D Fokker-Planck Equation 2.8. The difficulty is that the transport cross section $\Sigma_{tr}(z, E_a(z))$ depends on the depth-dependent mean energy $E_a(z)$. In the work of Gebäck and Asadzadeh (Gebäck and Asadzadeh, 2012), when also accounting for energy straggling this relation is defined implicitly via a line integral of the stopping power as $E_a(z) = E_0 - \int_0^z S(E_a(z'))dz'$. We define the average energy via

$$E_a(z) = \int_{E_{min}}^{E_{max}} \mathrm{d}E\, E\varphi_{FP} \Big/ \int_{E_{min}}^{E_{max}} \mathrm{d}E\, \varphi_{FP},$$

which means that $E_a(z)$ depends on the Fokker-Planck flux $\varphi_{FP}(z, E)$. Thus, the latter is present in the Υ term, namely

$$\underbrace{\frac{1}{\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}})} \Upsilon\Big(\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}}), \varphi_{FP}(z, E)\Big)}_{\lambda} + \underbrace{\frac{1}{\varphi_{FP}(z, E)} 1DFP(\varphi_{FP}(z, E))}_{-\lambda} = 0. \tag{A.1}$$

However, since the average energy $E_a(z)$ is defined as only depending on the energy integrated Fokker-Planck flux, the only independent variable $E_a(z)$ depends on is z. Consequently, even though $\varphi_{FP}(z, E)$ is present in the definition of $E_a(z)$, Υ does not have E dependence. The general solution of Equation A.1 is therefore to equate each component to a constant $\pm\lambda \in \mathbb{R}$, as shown in Equation A.1. Thus, the 1DFP component is written out as

$$\frac{\partial \varphi_{FP}}{\partial z} - \frac{\partial S(z, E)\varphi_{FP}}{\partial E} - \frac{1}{2}\frac{\partial^2 T(z, E)\varphi_{FP}}{\partial E^2} + (\Sigma_a(z, E) + \lambda)\varphi_{FP} = 0, \tag{A.2}$$

while the $\Upsilon\Big(\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}}), \varphi_{FP}(z, E)\Big)$ part reads

$$\frac{\partial \varphi_{FE}}{\partial z} + \Omega_x \frac{\partial \varphi_{FE}}{\partial x} + \Omega_y \frac{\partial \varphi_{FE}}{\partial y} - \Sigma_{tr}(z, E_a(z))\left(\frac{\partial^2 \varphi_{FE}}{\partial \Omega_x^2} + \frac{\partial^2 \varphi_{FE}}{\partial \Omega_y^2}\right) - \lambda\varphi_{FE} = 0. \tag{A.3}$$

The solution of Equation A.3 is found by Gebäck and Asadzadeh (Gebäck and Asadzadeh, 2012) to be the usual FE solution $\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}})$ given in Equation 2.24 with a multiplicative factor, namely

$$\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}}, \lambda \neq 0) = \varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}})e^{\lambda z}. \tag{A.4}$$

This solution also clearly follows from the structure of Equation A.3, containing the $\frac{\partial \varphi_{FP}}{\partial z}$ partial derivative and $-\lambda\varphi_{FE}$ terms, in addition to only other linear partial non-z derivatives. Due to the same structure, the solution of Equation A.2 can similarly be written as

$$\varphi_{FP}(z, E, \lambda \neq 0) = \varphi_{FP}(z, E)e^{-\lambda z}, \tag{A.5}$$

where $\varphi_{FP}(z, E)$ is the solution to usual FP Equation 2.9 without the additional λ term.

When multiplying $\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}}, \lambda \neq 0)$ and $\varphi_{FP}(z, E, \lambda \neq 0)$ it becomes clear that the value of λ does not affect the solution to Equation 2.4 and thus simply setting $\lambda = 0$ is a valid choice, yielding the usual FE Equation 2.7 and FP Equation 2.8, even in the in-depth heterogeneous setting with exact average energy calculation. Since the boundary conditions imply the existence of a unique solution, it is also guaranteed that the splitting proposed by Gebäck and Asadzadeh (Gebäck and Asadzadeh, 2012) in Equation 2.5 is exact in our more general, in-depth heterogeneous case too.

Appendix B

Solving the Fermi-Eyges equation

B.1 Preliminaries

The Fermi-Eyges equation is

$$\Upsilon(\phi_{FE}) = \frac{\partial \varphi_{FE}}{\partial z} + \Omega_x \frac{\partial \varphi_{FE}}{\partial x} + \Omega_y \frac{\partial \varphi_{FE}}{\partial y} - \overline{\Sigma_{tr}}(z) \left(\frac{\partial^2 \varphi_{FE}}{\partial \Omega_x^2} + \frac{\partial^2 \varphi_{FE}}{\partial \Omega_y^2} \right) = 0. \quad \text{(B.1)}$$

The x and y directions are separated by writing φ_{FE} as

$$\varphi_{FE}(\mathbf{r}, \hat{\mathbf{\Omega}}) = H(z, x, \Omega_x) H(z, y, \Omega_y),$$

which results in two separate PDE's for each direction, namely

$$\frac{\partial H(z, \xi, \omega)}{\partial z} + \omega \frac{\partial H(z, \xi, \omega)}{\partial \xi} - \overline{\Sigma_{tr}}(z) \frac{\partial^2 H(z, \xi, \omega)}{\partial \omega^2} = 0, \quad \text{(B.2)}$$

where ξ stands for either x or y and ω for either Ω_x or Ω_y. The associated boundary condition is a Gaussian at $z = 0$,

$$H(0, \xi, \omega) = C \exp\big(-(a_1 \xi^2 + a_2 \xi \omega + a_3 \omega^2)\big). \quad \text{(B.3)}$$

B.1.1 Fourier transform definitions

The transformed function H is given by

$$\tilde{H}(z, \alpha, \beta) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} \mathrm{d}\xi \mathrm{d}\omega \, H(z, \xi, \omega) \mathrm{e}^{-i(\alpha\xi + \beta\omega)}, \quad \text{(B.4)}$$

and the original by

$$H(z, \xi, \omega) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} \mathrm{d}\alpha \mathrm{d}\beta \, \tilde{H}(z, \alpha, \beta) \mathrm{e}^{i(\xi\alpha + \omega\beta)}. \quad \text{(B.5)}$$

B.2 The transformed PDE

The PDE from Equation B.2 is multiplied by $\frac{1}{2\pi} \mathrm{e}^{-i(\alpha\xi + \beta\omega)}$ and thereafter integrated over the whole domain. The first term results in

$$\frac{1}{2\pi} \iint \mathrm{d}\xi \mathrm{d}\omega \frac{\partial H}{\partial z} \mathrm{e}^{-i(\alpha\xi + \beta\omega)} = \frac{\partial}{\partial z} \frac{1}{2\pi} \iint \mathrm{d}\xi \mathrm{d}\omega H \mathrm{e}^{-i(\alpha\xi + \beta\omega)} = \frac{\partial \tilde{H}}{\partial z}.$$

The second term results in

$$\frac{1}{2\pi}\iint \mathrm{d}\xi\mathrm{d}\omega\,\omega\frac{\partial H}{\partial \xi}\mathrm{e}^{-i(\alpha\xi+\beta\omega)} = \frac{1}{2\pi}\int \mathrm{d}\omega\left[H\omega\mathrm{e}^{-i(\alpha\xi+\beta\omega)}\Big|_{\xi=-\infty}^{\xi=\infty} - \int \mathrm{d}\xi\, H\omega(-i\alpha)\mathrm{e}^{-i(\alpha\xi+\beta\omega)}\right]$$
$$= -\alpha\frac{\partial \tilde{H}}{\partial \beta}.$$

The last term results in

$$\frac{1}{2\pi}\iint \mathrm{d}\xi\mathrm{d}\omega\,\overline{\Sigma_{tr}}(z)\frac{\partial^2 H}{\partial \omega^2}\mathrm{e}^{-i(\alpha\xi+\beta\omega)} = -\overline{\Sigma_{tr}}(z)\beta^2\tilde{H}.$$

Thus the Fourier-transformed equation is

$$\frac{\partial \tilde{H}}{\partial z} - \alpha\frac{\partial \tilde{H}}{\partial \beta} + \overline{\Sigma_{tr}}(z)\beta^2\tilde{H} = 0,$$

which is the same as the one Eyges originally obtained. At this point Eyges introduces a change of variables

$$z'(z) = z$$
$$\epsilon(z,\alpha,\beta) = z + \frac{\beta}{\alpha}$$

Using this $\tilde{H}(z,\alpha,\beta)$ is transformed into $\tilde{H}(z',\epsilon)$. Via the chain rule,

$$\frac{\partial \tilde{H}}{\partial z} = \frac{\partial \tilde{H}}{\partial z'}\frac{\mathrm{d}z'}{\mathrm{d}z} + \frac{\partial \tilde{H}}{\partial \epsilon}\frac{\mathrm{d}\epsilon}{\mathrm{d}z} = \frac{\partial \tilde{H}}{\partial z'} + \frac{\partial \tilde{H}}{\partial \epsilon},$$
$$\alpha\frac{\partial \tilde{H}}{\partial \beta} = \alpha\left[\frac{\partial \tilde{H}}{\partial z'}\frac{\mathrm{d}z'}{\mathrm{d}\beta} + \frac{\partial \tilde{H}}{\partial \epsilon}\frac{\partial \epsilon}{\partial \beta}\right] = \frac{\partial \tilde{H}}{\partial \epsilon},$$

the equation becomes

$$\frac{\partial \tilde{H}}{\partial z'} = -\alpha^2(\epsilon - z')^2\overline{\Sigma_{tr}}(z')\tilde{H}.$$

B.3 The solution

The solution to the Fourier transformed PDE is

$$\tilde{H}(z',\epsilon) = \tilde{G}(\epsilon)\exp\left[-\alpha^2\int_{\kappa}^{z'}(\epsilon-\eta)^2\overline{\Sigma_{tr}}(\eta)d\eta\right]. \tag{B.6}$$

where $G(\epsilon)$ is the integration constant. To determine this constant, the initial condition is also Fourier transformed. To do that, the following relationship will be useful

$$\int_{-\infty}^{\infty}\mathrm{d}x\, e^{-(c_1x^2+c_2x)} = \sqrt{\frac{\pi}{c_1}}e^{\frac{c_2^2}{4c_1}}. \tag{B.7}$$

The Fourier transformed initial condition is written as

$$\begin{aligned}\tilde{H}(0,\alpha,\beta) &= \frac{C}{2\pi}\iint \mathrm{d}\xi\mathrm{d}\omega\, \mathrm{e}^{-(a_1\xi^2+a_2\xi\omega+a_3\omega^2)}\mathrm{e}^{-i(\alpha\xi+\beta\omega)} \\ &= \frac{C}{2\pi}\int \mathrm{d}\omega\, \mathrm{e}^{-(a_3\omega^2+i\beta\omega)}\int \mathrm{d}\xi\, \mathrm{e}^{-(a_1\xi^2+\xi(a_2\omega+i\alpha))}.\end{aligned}$$

The ξ integral is computed by making use of Equation B.7 to

$$\begin{aligned}\sqrt{\frac{\pi}{a_1}}\mathrm{e}^{\frac{(a_2\omega+i\alpha)^2}{4a_1}} &= \sqrt{\frac{\pi}{a_1}}\exp\left(\frac{a_2^2\omega^2-\alpha^2+2a_2\omega i\alpha}{4a_1}\right) \\ &= \sqrt{\frac{\pi}{a_1}}\exp\left(-\frac{\alpha^2}{4a_1}\right)\exp\left(\frac{a_2^2\omega^2}{4a_1}+\frac{a_2\omega i\alpha}{2a_1}\right),\end{aligned}$$

which when introduced into $\tilde{H}(0,\alpha,\beta)$ gives

$$\tilde{H}(0,\alpha,\beta) = \frac{C}{2\pi}\sqrt{\frac{\pi}{a_1}}\mathrm{e}^{-\frac{\alpha^2}{4a_1}}\int \mathrm{d}\omega\, \mathrm{e}^{-\left(\omega^2\frac{4a_1a_3-a_2^2}{4a_1}+\omega i\frac{2a_1\beta-a_2\alpha}{2a_1}\right)}.$$

The ω integral is also evaluated using Equation B.7 to

$$\sqrt{\frac{4a_1\pi}{4a_1a_3-a_2^2}}\exp\left[-\left(\frac{2a_1\beta-a_2\alpha}{2a_1}\right)^2\frac{a_1}{4a_1a_3-a_2^2}\right].$$

Introducing this back into $\tilde{H}(0,\alpha,\beta)$ results in

$$\begin{aligned}\tilde{H}(0,\alpha,\beta) &= \frac{C}{2\pi}\sqrt{\frac{4\pi^2}{4a_1a_3-a_2^2}}\exp-\left[\frac{4a_1^2\beta^2+a_2^2\alpha^2-4a_1a_2\alpha\beta}{4a_1(4a_1a_3-a_2^2)}+\frac{\alpha^2}{4a_1}\right] \\ &= \frac{C}{\sqrt{4a_1a_3-a_2^2}}\exp-\left(\frac{a_1\beta^2-a_2\alpha\beta+a_3\alpha^2}{4a_1a_3-a_2^2}\right),\end{aligned}$$

which can also be expressed into (z',ϵ) as

$$\tilde{H}(0,\alpha,\beta) = \tilde{H}\left(0,\frac{\beta}{\alpha}\right) = \frac{C}{\sqrt{4a_1a_3-a_2^2}}\exp\left[\frac{-\alpha^2}{4a_1a_3-a_2^2}\left(a_1\frac{\beta^2}{\alpha^2}-a_2\frac{\beta}{\alpha}+a_3\right)\right].$$

Making use of this into the solution allows finding $\tilde{G}\left(\frac{\beta}{\alpha}\right)$

$$\tilde{G}\left(\frac{\beta}{\alpha}\right) = \tilde{H}\left(0,\frac{\beta}{\alpha}\right)\exp\left[\alpha^2\int_{\kappa}^{0}\left(\frac{\beta}{\alpha}-\eta\right)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right],$$

which can also be written as

$$\tilde{G}\left(\frac{\beta}{\alpha}\right) = \tilde{H}\left(0,\frac{\beta}{\alpha}\right)\exp\left[-\alpha^2\int_{0}^{\kappa}\left(\frac{\beta}{\alpha}-\eta\right)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right].$$

Thus, $\tilde{G}(\epsilon)$ is equal to

$$\tilde{G}(\epsilon) = \tilde{G}\left(z + \frac{\beta}{\alpha}\right) = \tilde{H}\left(0, z + \frac{\beta}{\alpha}\right) \exp\left[-\alpha^2 \int_0^\kappa \left(\frac{\beta}{\alpha} + z - \eta\right)^2 \overline{\Sigma_{tr}}(\eta) d\eta\right].$$

Introducing this into the solution from Equation B.6 gives

$$\tilde{H}(z, \alpha, \beta) = \frac{C}{\sqrt{4a_1a_3 - a_2^2}} \exp - \frac{a_1(\beta + \alpha z)^2 - a_2\alpha(\beta + \alpha z) + a_3\alpha^2}{4a_1a_3 - a_2^2} \exp\left(-\alpha^2 \int_0^z \left(\frac{\beta}{\alpha} + z - \eta\right)^2 \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta\right).$$

This can be rewritten as

$$\tilde{H}(z, \alpha, \beta) = \frac{C}{\sqrt{4a_1a_3 - a_2^2}} \exp - \frac{a_1\beta^2 + \alpha\beta(2a_1z - a_2) + \alpha^2(a_1z^2 - a_2z + a_3)}{4a_1a_3 - a_2^2} \exp\left[-\beta^2 \int_0^z \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta - \alpha^2 \int_0^z (z - \eta)^2 \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta - 2\alpha\beta \int_0^z (z - \eta) \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta\right].$$

Let

$$\begin{aligned} D &= 4a_1a_3 - a_2^2, \\ A_0(z) &= \frac{a_1}{D} + \int_0^z \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta, \\ 2A_1(z) &= \frac{2a_1z - a_2}{D} + \int_0^z 2(z - \eta) \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta, \\ A_2(z) &= \frac{a_1z^2 - a_2z + a_3}{D} + \int_0^z (z - \eta)^2 \overline{\Sigma_{tr}}(\eta) \mathrm{d}\eta. \end{aligned}$$

Note that this is different from the coefficients given by (Gebäck and Asadzadeh, 2012), despite the same equation with the same boundary condition being solved. This solution has factors of 2 missing in the A_0 and A_2 coefficients, additional factors of 2 appearing in the A_1 coefficient, the roles a_1 and a_3 are swapped and a_2 appears with a - sign. Using the coefficients just defined, the solution becomes

$$\tilde{H}(z, \alpha, \beta) = \frac{C}{\sqrt{D}} \exp\big(-(A_0\beta^2 + 2A_1\alpha\beta + A_2\alpha^2)\big). \tag{B.8}$$

Transforming this back into the (ξ, ω) domain results in

$$\begin{aligned} H(z, \xi, \omega) &= \frac{1}{2\pi} \iint_{-\infty}^{\infty} \mathrm{d}\alpha \mathrm{d}\beta \, \frac{C}{\sqrt{D}} \mathrm{e}^{-(A_0\beta^2 + 2A_1\alpha\beta + A_2\alpha^2)} \mathrm{e}^{i(\alpha\xi + \beta\omega)} \\ &= \frac{1}{2\pi} \frac{C}{\sqrt{D}} \int \mathrm{d}\alpha \, \mathrm{e}^{-(A_2\alpha^2 - i\alpha\xi)} \int \mathrm{d}\beta \, \mathrm{e}^{-(A_0\beta^2 + \beta(2A_1\alpha - i\omega))}. \end{aligned} \tag{B.9}$$

The β integral is evaluated using Equation B.7 to

$$\sqrt{\frac{\pi}{A_0}}\exp\left[\frac{(4A_1^2\alpha^2-\omega^2-4A_1\alpha i\omega)}{4A_0}\right]=\sqrt{\frac{\pi}{A_0}}\exp-\frac{\omega^2}{4A_0}\exp\frac{A_1^2\alpha^2-A_1\alpha i\omega}{A_0}.$$

Using this results in

$$\frac{1}{2\pi}\frac{C}{\sqrt{D}}\sqrt{\frac{\pi}{A_0}}\mathrm{e}^{-\frac{\omega^2}{4A_0}}\int\mathrm{d}\alpha\,\mathrm{e}^{-\left[\alpha^2\left(A_2-\frac{A_1^2}{A_0}\right)+\alpha i\left(\frac{A_1\omega}{A_0}-\xi\right)\right]}.$$

The α integral is evaluated using Equation B.7 to

$$\sqrt{\frac{\pi A_0}{A_0A_2-A_1^2}}\exp\frac{-A_0}{4(A_0A_2-A_1^2)}\left(\frac{A_1^2\omega^2}{A_0^2}+\xi^2-2\xi\frac{A_1\omega}{A_0}\right).$$

Introducing this back into the final solution results in

$$H(z,\xi,\omega)=\frac{C}{2\sqrt{D}}\frac{1}{\sqrt{A_0A_2-A_1^2}}\exp-\frac{\left(A_0\xi^2-2A_1\xi\omega+A_2\omega^2\right)}{4(A_0A_2-A_1^2)}.$$

Introducing (in line with the work of (Gottschalk, 2012; Brahme, 1975))

$$B(z)=A_0A_2-A_1^2,$$

the solution is written as

$$H(z,\xi,\omega)=\frac{C}{\sqrt{D}}\frac{1}{\sqrt{4B(z)}}\exp\left(-\frac{A_0\xi^2-2A_1\xi\omega+A_2\omega^2}{4B(z)}\right). \tag{B.10}$$

As a check, this solution should reduce to the initial condition when $z=0$. In that case, the FE coefficients are equal to

$$A_0(0)=\frac{a_1}{D},\ 2A_1(0)=\frac{-a_2}{D},\ A_2(0)=\frac{a_3}{D},\ 4B(0)=\frac{1}{D},$$

which gives for the coefficients of the general solution

$$\frac{A_0(0)}{4B(0)}=a_1,\ \frac{-2A_1(0)}{4B(0)}=a_2,\ \frac{A_2(0)}{4B(0)}=a_3.$$

Thus, the general solution from Equation B.10 does indeed reduce to the imposed initial condition from Equation B.3. This is contrast to the solution obtained using the coefficients given by (Gebäck and Asadzadeh, 2012), which when evaluated at $z=0$ does not in fact reduce to the boundary condition.

B.4 Alternative with factor of 1/2

If $\overline{\Sigma_{tr}}$ from the FE Equation B.1 is in fact $\frac{\overline{\Sigma_{tr}}}{2}$ (which is the correct formulation according to the work of (Pomraning and Prinja, 1999; Pomraning, 1996)) the equations change from

$$\tilde{H}(z,\alpha,\beta)=\frac{C}{\sqrt{D}}\exp-\frac{a_1\beta^2+\alpha\beta(2a_1z-a_2)+\alpha^2(a_1z^2-a_2z+a_3)}{D}$$

$$\exp\left[-\beta^2\int_0^z\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta-\alpha^2\int_0^z(z-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta-2\alpha\beta\int_0^z(z-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right],$$

to

$$\begin{aligned}\tilde{H}(z,\alpha,\beta)=&\frac{C}{\sqrt{D}}\exp-\frac{1}{2D}\left[2a_1\beta^2+\alpha\beta(4a_1z-2a_2)+\alpha^2(2a_1z^2-2a_2z+2a_3)\right]\\&\exp-\frac{1}{2}\left[\beta^2\int_0^z\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta+\alpha^2\int_0^z(z-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta+2\alpha\beta\int_0^z(z-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right].\end{aligned}$$

Now let,

$$\begin{aligned}A_0(z)&=\frac{2a_1}{D}+\int_0^z\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta,\\2A_1(z)&=\frac{4a_1z-2a_2}{D}+2\int_0^z(z-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta,\\A_2(z)&=\frac{2a_1z^2-2a_2z+2a_3}{D}+\int_0^z(z-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta.\end{aligned}$$

In comparison to the work of (Gebäck and Asadzadeh, 2012), the factors of 2 that were missing in the previous section, now appear in the coefficients. The roles of a_1 and a_3 are still swapped and the sign of a_2 is still reversed. From this, it can be concluded that a typographical mistake was made, and that in the correct equation the $\overline{\Sigma_{tr}}$ term has a factor of 1/2 in front of it. When implemented, it is these coefficients that provide good agreement with MC simulations done in TOPAS. Then, the solution becomes

$$\tilde{H}(z,\alpha,\beta)=\frac{C}{\sqrt{D}}\exp-\frac{1}{2}\left[A_0\beta^2+2A_1\alpha\beta+A_2\alpha^2\right].$$

Fourier transforming this back gives,

$$\begin{aligned}H(z,\xi,\omega)&=\frac{C}{\sqrt{D}}\frac{1}{\sqrt{A_0A_2-A_1^2}}\exp-\frac{\left(A_0\xi^2-2A_1\xi\omega+A_2\omega^2\right)}{2(A_0A_2-A_1^2)}\\&=\frac{C}{\sqrt{D}}\frac{1}{\sqrt{B}}\exp-\frac{\left(A_0\xi^2-2A_1\xi\omega+A_2\omega^2\right)}{2B}.\end{aligned}$$

Integrating over ω gives

$$\frac{C}{\sqrt{D}}\sqrt{\frac{2\pi}{A_2}}\exp-\frac{\xi^2}{2A_2}.$$

Which gives the angular integral as

$$\frac{2\pi C^2}{D}\frac{1}{A_2}\exp-\frac{r^2}{2A_2}.$$

In the case of a normalized initial condition

$$\frac{2\pi C^2}{D}=\frac{1}{2\pi}.$$

And the angular integral is

$$\Psi_{FE} = \frac{1}{2\pi A_2} \exp -\frac{r^2}{2A_2}.$$

B.5 Normalized initial condition

If the boundary condition is a normalized two-dimensional Gaussian with zero means in ξ and ω, namely

$$H(0,\xi,\omega) = \frac{1}{2\pi\sigma_\xi\sigma_\omega\sqrt{1-\rho_\xi^2}} \exp \frac{-1}{2(1-\rho_\xi^2)} \left(\frac{\xi^2}{\sigma_\xi^2} - 2\rho_\xi \frac{\xi}{\sigma_\xi}\frac{\omega}{\sigma_\omega} + \frac{\omega^2}{\sigma_\omega^2} \right).$$

the coefficients $C, a_1, a_2,$ and a_3 can be identified as

$$\begin{aligned} C &= \frac{1}{2\pi\sigma_\xi\sigma_\omega\sqrt{1-\rho_\xi^2}}, \\ a_1 &= \frac{1}{2(1-\rho_\xi^2)\sigma_\xi^2},\ a_2 = \frac{-\rho_\xi}{(1-\rho_\xi^2)\sigma_\xi\sigma_\omega},\ a_3 = \frac{1}{2(1-\rho_\xi^2)\sigma_\omega^2} \\ D &= \frac{1}{(1-\rho_\xi^2)\sigma_\xi^2\sigma_\omega^2}. \end{aligned}$$

Thus, the solution can also be written as

$$H(z,\xi,\omega) = \frac{1}{2\pi}\frac{1}{\sqrt{4B(z)}} \exp\left(-\frac{A_0\xi^2 - 2A_1\xi\omega + A_2\omega^2}{4B(z)} \right).$$

B.6 Symmetric solution

If the same boundary condition is imposed for both x and y directions, the general FE flux is given by

$$\begin{aligned} \varphi_{FE}(\mathbf{r},\hat{\mathbf{\Omega}}) &= H(z,x,\Omega_x)H(z,y,\Omega_y) \\ &= \frac{C^2}{4B(z)D} \exp\left(-\frac{A_0(x^2+y^2) - 2A_1(x\Omega_x + y\Omega_y) + A_2(\Omega_x^2+\Omega_y^2)}{4B(z)} \right). \end{aligned}$$

B.7 Angular integral

To compute the dose in a certain region the angular integral of the FE flux must be computed. This is defined as Ψ_{FE} and is found to be

$$\begin{aligned} \Psi_{FE}(\mathbf{r}) &= \iint \mathrm{d}\hat{\mathbf{\Omega}}\varphi_{FE} \\ &= \int \mathrm{d}\Omega_x H(z,x,\Omega_x) \int \mathrm{d}\Omega_y H(z,y,\Omega_y). \end{aligned}$$

In ξ,ω notation, one of the integrals is equated to

$$\int \mathrm{d}\omega H(z,\xi,\omega) = \frac{C}{\sqrt{D}}\frac{1}{\sqrt{4B(z)}} \exp\frac{-A_0\xi^2}{4B} \int \mathrm{d}\omega \exp - \left(\frac{A_2}{4B}\omega^2 + \omega\frac{-2A_1\xi}{4B} \right).$$

Making use of Equation B.7 the ω integral is found to be

$$\sqrt{\frac{4\pi B}{A_2}} \exp \frac{A_1^2 \xi^2}{4A_2 B}.$$

Introducing this back

$$\begin{aligned}\int \mathrm{d}\omega H(z,\xi,\omega) &= \frac{C}{\sqrt{D}} \frac{1}{\sqrt{4B(z)}} \sqrt{\frac{4\pi B}{A_2}} \exp -\frac{\xi^2}{4B} \left(A_0 - \frac{A_1^2}{A_2}\right) \\ &= \frac{C}{\sqrt{D}} \sqrt{\frac{\pi}{A_2}} \exp -\frac{\xi^2}{4A_2}.\end{aligned}$$

Thus,

$$\Psi_{FE}(\mathbf{r}) = \frac{\pi C^2}{D A_2} \exp -\frac{(x^2+y^2)}{4A_2},$$

which becomes in the case of a normalized initial condition

$$\Psi_{FE}(\mathbf{r}) = \frac{1}{4\pi A_2} \exp -\frac{(x^2+y^2)}{4A_2}.$$

B.8 Iterative FE coefficients computation

In order to avoid expensive re-computations of the depth integrals, the coefficients at a position z_i can be written in terms of the coefficients at positions z_{i-1}. The first coefficient is written as

$$\begin{aligned}A_0(z_i) &= \frac{a_1}{D} + \int_0^{z_i} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta = \frac{a_1}{D} + \int_0^{z_{i-1}} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_{z_{i-1}}^{z_i} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta \\ &= A_0(z_{i-1}) + \frac{\overline{\Sigma_{tr}}(z_i) + \overline{\Sigma_{tr}}(z_{i-1})}{2}\Delta z,\end{aligned}$$

where the last integral was evaluated using the trapezoidal rule and $\Delta z = z_i - z_{i-1}$. The second coefficient is written as

$$\begin{aligned}2A_1(z_i) &= \frac{2a_1 z_i - a_2}{D} + \int_0^{z_i} 2(z_i - \eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta \\ &= \frac{2a_1}{D}(z_i - z_{i-1} + z_{i-1}) - \frac{a_2}{D} + \int_0^{z_{i-1}} 2(z_i - \eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_{z_{i-1}}^{z_i} 2(z_i - \eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta \\ &= \frac{2a_1}{D}\Delta z + \frac{2a_1 z_{i-1} - a_2}{D} + \int_0^{z_{i-1}} 2(z_i - \eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \Delta z^2 \overline{\Sigma_{tr}}(z_{i-1}),\end{aligned}$$

where the last integral was evaluated with the trapezoidal rule. The remaining integral can be written as

$$\int_0^{z_{i-1}} 2(z_i - z_{i-1} + z_{i-1} - \eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta.$$

Introducing this back gives

$$\begin{aligned}2A_1(z_i) &= \frac{2a_1}{D}\Delta z + \frac{2a_1 z_{i-1} - a_2}{D} + \int_0^{z_{i-1}} 2\Delta z\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_0^{z_{i-1}} 2(z_{i-1}-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \Delta z^2\overline{\Sigma_{tr}}(z_{i-1}) \\ &= 2A_1(z_{i-1}) + 2\Delta z\left[\frac{a_1}{D} + \int_0^{z_{i-1}} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right] + \Delta z^2\overline{\Sigma_{tr}}(z_{i-1}) \\ &= 2A_1(z_{i-1}) + 2\Delta z A_0(z_{i-1}) + \Delta z^2\overline{\Sigma_{tr}}(z_{i-1}).\end{aligned}$$

The third FE coefficient at a point z_i is equal to

$$A_2(z_i) = \frac{a_1 z_i^2 - a_2 z_i + a_3}{D} + \int_0^{z_i} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta.$$

The a_1, a_2, a_3 part in front of the integral is written as

$$\frac{a_1}{D}\Delta z^2 + \frac{a_1}{D}z_{i-1}^2 + \frac{a_1}{D}2\Delta z z_{i-1} - \frac{a_2}{D}\Delta z - \frac{a_2}{D}z_{i-1} + \frac{a_3}{D}.$$

The integral is split as before into

$$\int_0^{z_{i-1}} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_{z_{i-1}}^{z_i} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta.$$

The last integral is again evaluated using the trapezoidal rule to

$$\int_{z_{i-1}}^{z_i} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta = \frac{\Delta z^3}{2}\overline{\Sigma_{tr}}(z_{i-1}),$$

while the first integral is written as

$$\int_0^{z_{i-1}} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta = \int_0^{z_{i-1}} \left(\Delta z^2 + (z_{i-1}-\eta)^2 + 2\Delta z(z_{i-1}-\eta)\right)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta.$$

Introducing these results into the original $A_2(z_i)$ equation gives

$$\begin{aligned}A_2(z_i) = {} & \frac{a_1}{D}z_{i-1}^2 - \frac{a_2}{D}z_{i-1} + \frac{a_3}{D} + \int_0^{z_{i-1}} (z_{i-1}-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta \\ & + \Delta z^2\left[\frac{a_1}{D} + \int_0^{z_{i-1}} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right] \\ & + \Delta z\left[\frac{2a_1}{D}z_{i-1} - \frac{a_2}{D} + \int_0^{z_{i-1}} 2(z_{i-1}-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\right] \\ & + \frac{\Delta z^3}{2}\overline{\Sigma_{tr}}(z_{i-1}) \\ = {} & A_2(z_{i-1}) + \Delta z^2 A_0(z_{i-1}) + \Delta z 2A_1(z_{i-1}) + \frac{\Delta z^3}{2}\overline{\Sigma_{tr}}(z_{i-1}).\end{aligned}$$

Thus, the coefficients can be efficiently computed via

$$\begin{aligned}
A_0(z_i) &= A_0(z_{i-1}) + \frac{\overline{\Sigma_{tr}}(z_i) + \overline{\Sigma_{tr}}(z_{i-1})}{2}\Delta z,\\
2A_1(z_i) &= 2A_1(z_{i-1}) + 2\Delta z A_0(z_{i-1}) + \Delta z^2\overline{\Sigma_{tr}}(z_{i-1}),\\
A_2(z_i) &= A_2(z_{i-1}) + \Delta z^2 A_0(z_{i-1}) + \Delta z 2A_1(z_{i-1}) + \frac{\Delta z^3}{2}\overline{\Sigma_{tr}}(z_{i-1}).
\end{aligned}$$

Alternatively, in the case where an asymmetrical beam is implemented, only the z integrals should be expressed in terms of the previously computed ones. Let,

$$\begin{aligned}
I_0(z_i) &= \int_0^{z_i} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta,\\
I_1(z_i) &= \int_0^{z_i} 2(z_i-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta,\\
I_2(z_i) &= \int_0^{z_i} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta.
\end{aligned}$$

Then,

$$\begin{aligned}
I_0(z_i) &= I_0(z_{i-1}) + \int_{z_{i-1}}^{z_i} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta = I_0(z_{i-1}) + \frac{\overline{\Sigma_{tr}}(z_i) + \overline{\Sigma_{tr}}(z_{i-1})}{2}\Delta z,\\
I_1(z_i) &= \int_0^{z_{i-1}} 2(z_i-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_{z_{i-1}}^{z_i} 2(z_i-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\\
&= 2\Delta z\int_0^{z_{i-1}} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_0^{z_{i-1}} 2(z_{i-1}-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \Delta z^2\overline{\Sigma_{tr}}(z_{i-1})\\
&= 2\Delta z I_0(z_{i-1}) + I_1(z_{i-1}) + \Delta z^2\overline{\Sigma_{tr}}(z_{i-1}),\\
I_2(z_i) &= \int_0^{z_{i-1}} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_{z_{i-1}}^{z_i} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\\
&= \Delta z^2\int_0^{z_{i-1}} \overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \int_0^{z_{i-1}} (z_{i-1}-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta + \Delta z\int_0^{z_{i-1}} 2(z_{i-1}-\eta)\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\\
&\quad + \int_{z_{i-1}}^{z_i} (z_i-\eta)^2\overline{\Sigma_{tr}}(\eta)\mathrm{d}\eta\\
&= \Delta z^2 I_0(z_{i-1}) + \Delta z I_1(z_{i-1}) + I_2(z_{i-1}) + \frac{\Delta z^3}{2}\overline{\Sigma_{tr}}(z_{i-1}).
\end{aligned}$$

Appendix C

Sliding slab experiment results

This appendix presents the results of the sliding slab experiment, discussed in Chapter 3.

TABLE C.1: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
70	-1000	-4	1, 6, 6, 12	1, 1, 0	99.99
70	-1000	-4	1, 6, 6, 12	1, 1, 10	98.11
70	-1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.99
70	-1000	-4	1, 6, 6, 12, 12	1, 1, 10	96.86
70	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.99
70	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	97.19
70	-1000	-2	1, 6, 6, 12	1, 1, 0	99.99
70	-1000	-2	1, 6, 6, 12	1, 1, 10	93.80
70	-1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.99
70	-1000	-2	1, 6, 6, 12, 12	1, 1, 10	93.97
70	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.99
70	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	95.10
70	-1000	0	1, 6, 6, 12	1, 1, 0	99.99
70	-1000	0	1, 6, 6, 12	1, 1, 10	96.42
70	-1000	0	1, 6, 6, 12, 12	1, 1, 0	99.99
70	-1000	0	1, 6, 6, 12, 12	1, 1, 10	96.48
70	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.99
70	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	96.33
70	-1000	+2	1, 6, 6, 12	1, 1, 0	99.99
70	-1000	+2	1, 6, 6, 12	1, 1, 10	96.94
70	-1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.99
70	-1000	+2	1, 6, 6, 12, 12	1, 1, 10	96.65
70	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.99
70	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	97.78
70	-1000	+4	1, 6, 6, 12	1, 1, 0	100.00
70	-1000	+4	1, 6, 6, 12	1, 1, 10	99.69
70	-1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.99
70	-1000	+4	1, 6, 6, 12, 12	1, 1, 10	98.18
70	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	100.00
70	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	99.72

TABLE C.2: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
70	1000	-4	1, 6, 6, 12	1, 1, 0	100
70	1000	-4	1, 6, 6, 12	1, 1, 10	99.86
70	1000	-4	1, 6, 6, 12, 12	1, 1, 0	100
70	1000	-4	1, 6, 6, 12, 12	1, 1, 10	99.31
70	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	100
70	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	99.86
70	1000	-2	1, 6, 6, 12	1, 1, 0	99.99
70	1000	-2	1, 6, 6, 12	1, 1, 10	99.30
70	1000	-2	1, 6, 6, 12, 12	1, 1, 0	100
70	1000	-2	1, 6, 6, 12, 12	1, 1, 10	98.67
70	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	100
70	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	99.37
70	1000	0	1, 6, 6, 12	1, 1, 0	100
70	1000	0	1, 6, 6, 12	1, 1, 10	99.19
70	1000	0	1, 6, 6, 12, 12	1, 1, 0	100
70	1000	0	1, 6, 6, 12, 12	1, 1, 10	99.12
70	1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	100
70	1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	99.16
70	1000	+2	1, 6, 6, 12	1, 1, 0	99.99
70	1000	+2	1, 6, 6, 12	1, 1, 10	96.51
70	1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.99
70	1000	+2	1, 6, 6, 12, 12	1, 1, 10	97.09
70	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	100
70	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	97.28
70	1000	+4	1, 6, 6, 12	1, 1, 0	99.99
70	1000	+4	1, 6, 6, 12	1, 1, 10	98.79
70	1000	+4	1, 6, 6, 12, 12	1, 1, 0	100
70	1000	+4	1, 6, 6, 12, 12	1, 1, 10	97.51
70	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	100
70	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	97.87

TABLE C.3: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
160	-1000	-4	1, 6, 6, 12	1, 1, 0	99.98
160	-1000	-4	1, 6, 6, 12	1, 1, 10	97.86
160	-1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.95
160	-1000	-4	1, 6, 6, 12, 12	1, 1, 10	97.29
160	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.97
160	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	97.33
160	-1000	-2	1, 6, 6, 12	1, 1, 0	99.98
160	-1000	-2	1, 6, 6, 12	1, 1, 10	97.06
160	-1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.97
160	-1000	-2	1, 6, 6, 12, 12	1, 1, 10	96.53
160	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.98
160	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	96.97
160	-1000	0	1, 6, 6, 12	1, 1, 0	99.98
160	-1000	0	1, 6, 6, 12	1, 1, 10	97.24
160	-1000	0	1, 6, 6, 12, 12	1, 1, 0	99.97
160	-1000	0	1, 6, 6, 12, 12	1, 1, 10	96.96
160	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.97
160	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	96.97
160	-1000	+2	1, 6, 6, 12	1, 1, 0	99.98
160	-1000	+2	1, 6, 6, 12	1, 1, 10	97.48
160	-1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.98
160	-1000	+2	1, 6, 6, 12, 12	1, 1, 10	97.45
160	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.98
160	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	97.36
160	-1000	+4	1, 6, 6, 12	1, 1, 0	99.98
160	-1000	+4	1, 6, 6, 12	1, 1, 10	97.96
160	-1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.98
160	-1000	+4	1, 6, 6, 12, 12	1, 1, 10	97.60
160	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	99.98
160	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	97.60

TABLE C.4: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
160	1000	-4	1, 6, 6, 12	1, 1, 0	99.99
160	1000	-4	1, 6, 6, 12	1, 1, 10	98.26
160	1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.98
160	1000	-4	1, 6, 6, 12, 12	1, 1, 10	97.77
160	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.98
160	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	97.75
160	1000	-2	1, 6, 6, 12	1, 1, 0	99.98
160	1000	-2	1, 6, 6, 12	1, 1, 10	97.75
160	1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.98
160	1000	-2	1, 6, 6, 12, 12	1, 1, 10	97.66
160	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.98
160	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	97.59
160	1000	0	1, 6, 6, 12	1, 1, 0	99.99
160	1000	0	1, 6, 6, 12	1, 1, 10	98.17
160	1000	0	1, 6, 6, 12, 12	1, 1, 0	99.98
160	1000	0	1, 6, 6, 12, 12	1, 1, 10	97.91
160	1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.98
160	1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	97.88
160	1000	+2	1, 6, 6, 12	1, 1, 0	99.99
160	1000	+2	1, 6, 6, 12	1, 1, 10	99.20
160	1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.99
160	1000	+2	1, 6, 6, 12, 12	1, 1, 10	98.99
160	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.99
160	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	99
160	1000	+4	1, 6, 6, 12	1, 1, 0	99.99
160	1000	+4	1, 6, 6, 12	1, 1, 10	98.88
160	1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.99
160	1000	+4	1, 6, 6, 12, 12	1, 1, 10	98.75
160	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	99.99
160	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	98.78

TABLE C.5: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
190	-1000	-4	1, 6, 6, 12	1, 1, 0	99.92
190	-1000	-4	1, 6, 6, 12	1, 1, 10	95.76
190	-1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.86
190	-1000	-4	1, 6, 6, 12, 12	1, 1, 10	95.38
190	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.90
190	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	95.52
190	-1000	-2	1, 6, 6, 12	1, 1, 0	99.92
190	-1000	-2	1, 6, 6, 12	1, 1, 10	94.50
190	-1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.91
190	-1000	-2	1, 6, 6, 12, 12	1, 1, 10	93.96
190	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.92
190	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	94.36
190	-1000	0	1, 6, 6, 12	1, 1, 0	99.93
190	-1000	0	1, 6, 6, 12	1, 1, 10	95.27
190	-1000	0	1, 6, 6, 12, 12	1, 1, 0	99.93
190	-1000	0	1, 6, 6, 12, 12	1, 1, 10	95.29
190	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.93
190	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	95.31
190	-1000	+2	1, 6, 6, 12	1, 1, 0	99.93
190	-1000	+2	1, 6, 6, 12	1, 1, 10	95.62
190	-1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.93
190	-1000	+2	1, 6, 6, 12, 12	1, 1, 10	95.65
190	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.93
190	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	95.64
190	-1000	+4	1, 6, 6, 12	1, 1, 0	99.94
190	-1000	+4	1, 6, 6, 12	1, 1, 10	95.77
190	-1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.94
190	-1000	+4	1, 6, 6, 12, 12	1, 1, 10	95.72
190	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	99.94
190	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	95.77

TABLE C.6: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
190	1000	-4	1, 6, 6, 12	1, 1, 0	99.95
190	1000	-4	1, 6, 6, 12	1, 1, 10	95.92
190	1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.94
190	1000	-4	1, 6, 6, 12, 12	1, 1, 10	95.89
190	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.94
190	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	95.90
190	1000	-2	1, 6, 6, 12	1, 1, 0	99.94
190	1000	-2	1, 6, 6, 12	1, 1, 10	95.76
190	1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.94
190	1000	-2	1, 6, 6, 12, 12	1, 1, 10	95.87
190	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.94
190	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	95.81
190	1000	0	1, 6, 6, 12	1, 1, 0	99.95
190	1000	0	1, 6, 6, 12	1, 1, 10	96.15
190	1000	0	1, 6, 6, 12, 12	1, 1, 0	99.95
190	1000	0	1, 6, 6, 12, 12	1, 1, 10	96.12
190	1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.95
190	1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	96.10
190	1000	+2	1, 6, 6, 12	1, 1, 0	99.96
190	1000	+2	1, 6, 6, 12	1, 1, 10	97.30
190	1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.96
190	1000	+2	1, 6, 6, 12, 12	1, 1, 10	96.97
190	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.96
190	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	97.19
190	1000	+4	1, 6, 6, 12	1, 1, 0	99.96
190	1000	+4	1, 6, 6, 12	1, 1, 10	97.19
190	1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.95
190	1000	+4	1, 6, 6, 12, 12	1, 1, 10	97.04
190	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	99.96
190	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	97.18

TABLE C.7: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
230	-1000	-4	1, 6, 6, 12	1, 1, 0	99.42
230	-1000	-4	1, 6, 6, 12	1, 1, 10	96.88
230	-1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.13
230	-1000	-4	1, 6, 6, 12, 12	1, 1, 10	92.42
230	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.09
230	-1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	91.72
230	-1000	-2	1, 6, 6, 12	1, 1, 0	99.60
230	-1000	-2	1, 6, 6, 12	1, 1, 10	95.56
230	-1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.32
230	-1000	-2	1, 6, 6, 12, 12	1, 1, 10	94.40
230	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.09
230	-1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	90.98
230	-1000	0	1, 6, 6, 12	1, 1, 0	99.65
230	-1000	0	1, 6, 6, 12	1, 1, 10	96.62
230	-1000	0	1, 6, 6, 12, 12	1, 1, 0	99.39
230	-1000	0	1, 6, 6, 12, 12	1, 1, 10	96.35
230	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.15
230	-1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	92.85
230	-1000	+2	1, 6, 6, 12	1, 1, 0	99.65
230	-1000	+2	1, 6, 6, 12	1, 1, 10	96.89
230	-1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.40
230	-1000	+2	1, 6, 6, 12, 12	1, 1, 10	96.67
230	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.15
230	-1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	92.98
230	-1000	+4	1, 6, 6, 12	1, 1, 0	99.65
230	-1000	+4	1, 6, 6, 12	1, 1, 10	97.18
230	-1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.41
230	-1000	+4	1, 6, 6, 12, 12	1, 1, 10	96.73
230	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	99.16
230	-1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	93.05

TABLE C.8: Overview of the sliding slab experiment results. The table presents the beam energy, the slab HU value, the slab position with respect to the central beam axis, the used Gaussian splitting schemes, the gamma index settings and the gamma index pass rates.

Energy (MeV)	Slab value (HU)	Slab shift (mm)	Split scheme	Gamma settings (mm, %, %)	Gamma pass rate (%)
230	1000	-4	1, 6, 6, 12	1, 1, 0	99.65
230	1000	-4	1, 6, 6, 12	1, 1, 10	97.38
230	1000	-4	1, 6, 6, 12, 12	1, 1, 0	99.43
230	1000	-4	1, 6, 6, 12, 12	1, 1, 10	96.74
230	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 0	99.19
230	1000	-4	1, 6, 6, 12, 12, 24	1, 1, 10	92.98
230	1000	-2	1, 6, 6, 12	1, 1, 0	99.64
230	1000	-2	1, 6, 6, 12	1, 1, 10	97.18
230	1000	-2	1, 6, 6, 12, 12	1, 1, 0	99.41
230	1000	-2	1, 6, 6, 12, 12	1, 1, 10	96.65
230	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 0	99.17
230	1000	-2	1, 6, 6, 12, 12, 24	1, 1, 10	92.75
230	1000	0	1, 6, 6, 12	1, 1, 0	99.64
230	1000	0	1, 6, 6, 12	1, 1, 10	97.39
230	1000	0	1, 6, 6, 12, 12	1, 1, 0	99.40
230	1000	0	1, 6, 6, 12, 12	1, 1, 10	96.80
230	1000	0	1, 6, 6, 12, 12, 24	1, 1, 0	99.16
230	1000	0	1, 6, 6, 12, 12, 24	1, 1, 10	92.79
230	1000	+2	1, 6, 6, 12	1, 1, 0	99.61
230	1000	+2	1, 6, 6, 12	1, 1, 10	97.36
230	1000	+2	1, 6, 6, 12, 12	1, 1, 0	99.34
230	1000	+2	1, 6, 6, 12, 12	1, 1, 10	95.75
230	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 0	99.10
230	1000	+2	1, 6, 6, 12, 12, 24	1, 1, 10	91.95
230	1000	+4	1, 6, 6, 12	1, 1, 0	99.61
230	1000	+4	1, 6, 6, 12	1, 1, 10	97.56
230	1000	+4	1, 6, 6, 12, 12	1, 1, 0	99.32
230	1000	+4	1, 6, 6, 12, 12	1, 1, 10	95.65
230	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 0	99.08
230	1000	+4	1, 6, 6, 12, 12, 24	1, 1, 10	91.76

Bibliography

Ahnesjö, A. (1989). "Collapsed Cone Convolution of Radiant Energy for Photon Dose Calculation in Heterogeneous Media". In: *Medical Physics* 16.4, pp. 577–592. ISSN: 0094-2405. DOI: 10.1118/1.596360.

Albertini, Francesca et al. (Mar. 2020). "Online Daily Adaptive Proton Therapy". In: *The British Journal of Radiology* 93.1107, p. 20190594. ISSN: 0007-1285, 1748-880X. DOI: 10.1259/bjr.20190594. (Visited on 08/03/2020).

Anderson, E. et al. (1999). *LAPACK Users' Guide*. 3rd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics. ISBN: 0-89871-447-8 (paperback).

Asadzadeh, M., A. Brahme, and J. Kempe (Oct. 2010). "Ion Transport in Inhomogeneous Media Based on the Bipartition Model for Primary Ions". In: *Computers & Mathematics with Applications* 60.8, pp. 2445–2459. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2010.08.040. (Visited on 11/08/2024).

Barker, Jerry L. et al. (July 2004). "Quantification of Volumetric and Geometric Changes Occurring during Fractionated Radiotherapy for Head-and-Neck Cancer Using an Integrated CT/Linear Accelerator System". In: *International Journal of Radiation Oncology, Biology, Physics* 59.4, pp. 960–970. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2003.12.024. (Visited on 07/09/2024).

Barragán-Montero, Ana M. et al. (July 2023). "Dosimetrically Triggered Adaptive Radiotherapy for Head and Neck Cancer: Considerations for the Implementation of Clinical Protocols". In: *Journal of Applied Clinical Medical Physics* 24.11, e14095. ISSN: 1526-9914. DOI: 10.1002/acm2.14095. (Visited on 10/01/2024).

Beare, Richard, Bradley Lowekamp, and Ziv Yaniv (Sept. 2018). "Image Segmentation, Registration and Characterization in R with SimpleITK". In: *Journal of Statistical Software* 86, pp. 1–35. ISSN: 1548-7660. DOI: 10.18637/jss.v086.i08. (Visited on 12/24/2023).

Becquerel, Henri and Pierre Curie (1901). "Action physiologique des rayons du radium". In: *Compt. Rend. Acad. Sci* 132.1901, pp. 1289–1291.

Bhide, Shreerang A. et al. (Apr. 2010). "Weekly Volume and Dosimetric Changes during Chemoradiotherapy with Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Observational Study". In: *International Journal of Radiation Oncology, Biology, Physics* 76.5, pp. 1360–1368. ISSN: 1879-355X. DOI: 10.1016/j.ijrobp.2009.04.005.

Bodensteiner, Dayna (2018). "RayStation: External Beam Treatment Planning System". In: *Medical Dosimetry: Official Journal of the American Association of Medical Dosimetrists* 43.2, pp. 168–176. ISSN: 1873-4022. DOI: 10.1016/j.meddos.2018.02.013.

Börgers, Christoph (1999). "The Radiation Therapy Planning Problem". In: *Computational Radiology and Imaging*.

Ed. by Willard Miller, Christoph Börgers, and Frank Natterer. Vol. 110. New York, NY: Springer New York, pp. 1–16. ISBN: 978-1-4612-7189-5 978-1-4612-1550-9. DOI: 10.1007/978-1-4612-1550-9_1. (Visited on 03/17/2022).

Bortfeld, Thomas (1997). "An Analytical Approximation of the Bragg Curve for Therapeutic Proton Beams". In: *Medical Physics* 24.12, pp. 2024–2033. ISSN: 2473-4209. DOI: 10.1118/1.598116. (Visited on 04/21/2022).

Brahme, A. (1975). *Simple Relations for the Penetration of High Energy Electron Beams in Matter*. Tech. rep. Sweden, p. 47.

Brahme, A., J.-E. Roos, and I. Lax (Oct. 1982). "Solution of an Integral Equation Encountered in Rotation Therapy". In: *Physics in Medicine & Biology* 27.10, p. 1221. ISSN: 0031-9155. DOI: 10.1088/0031-9155/27/10/002. (Visited on 08/09/2024).

Burlacu, T., D. Lathouwers, and Z. Perkó (May 2023a). "OC-0117 A Deterministic Algorithm for Patient-Specific Quality Assurance in Online Adaptive Proton Therapy". In: *Radiotherapy and Oncology* 182, S78–S79. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/S0167-8140(23)08531-6. (Visited on 08/16/2024).

Burlacu, Tiberiu, Danny Lathouwers, and Zoltán Perkó (Jan. 2023b). "A Deterministic Adjoint-Based Semi-Analytical Algorithm for Fast Response Change Computations in Proton Therapy". In: *Journal of Computational and Theoretical Transport* 52.1, pp. 1–41. ISSN: 2332-4309. DOI: 10.1080/23324309.2023.2166077. (Visited on 08/30/2023).

— (July 2024). "Yet anOther Dose Algorithm (YODA) for Independent Computations of Dose and Dose Changes Due to Anatomical Changes". In: *Physics in Medicine & Biology* 69.16, p. 165003. ISSN: 0031-9155. DOI: 10.1088/1361-6560/ad6373. (Visited on 08/16/2024).

Burlacu, Tiberiu et al. (Mar. 2025). "A Deep Learning Model for Inter-Fraction Head and Neck Anatomical Changes in Proton Therapy". In: *Physics in Medicine & Biology* 70.6, p. 065011. ISSN: 0031-9155. DOI: 10.1088/1361-6560/adba39. (Visited on 05/07/2025).

Cancer Statistics (2024). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics. (Visited on 08/06/2024).

Chen, Allen M. et al. (2014). "Clinical Outcomes among Patients with Head and Neck Cancer Treated by Intensity-Modulated Radiotherapy with and without Adaptive Replanning". In: *Head & Neck* 36.11, pp. 1541–1546. ISSN: 1097-0347. DOI: 10.1002/hed.23477. (Visited on 08/17/2024).

Chen, Zhe et al. (Mar. 2023). "Proton versus Photon Radiation Therapy: A Clinical Review". In: *Frontiers in Oncology* 13, p. 1133909. ISSN: 2234-943X. DOI: 10.3389/fonc.2023.1133909. (Visited on 06/06/2024).

Courant, E. D. and H. S. Snyder (Apr. 2000). "Theory of the Alternating-Gradient Synchrotron". In: *Annals of Physics* 281.1, pp. 360–408. ISSN: 0003-4916. DOI: 10.1006/aphy.2000.6012. (Visited on 12/16/2024).

Coutard, Henri (July 1934). "PRINCIPLES OF X RAY THERAPY OF MALIGNANT DISEASES".

In: *The Lancet*. Originally Published as Volume 2, Issue 5784 224.5784, pp. 1–8. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(00)90085-0. (Visited on 09/24/2024).

Craft, D et al. (2014). *Supporting Material for: "Shared Data for IMRT Optimization Research: The CORT Dataset"*. DOI: 10.5524/100110. (Visited on 12/21/2023).

Cubillos-Mesías, Macarena et al. (Feb. 2019). "Including Anatomical Variations in Robust Optimization for Head and Neck Proton Therapy Can Reduce the Need of Adaptation". In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 131, pp. 127–134. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2018.12.008.

Curie, Marie (1921). *La Radiologie et La Guerre*. Paris: LIBRAIRIE FÉLIX ALCAN. (Visited on 09/24/2024).

Dai, Jian S. (Oct. 2015). "Euler–Rodrigues Formula Variations, Quaternion Conjugation and Intrinsic Connections". In: *Mechanism and Machine Theory* 92, pp. 144–152. ISSN: 0094-114X. DOI: 10.1016/j.mechmachtheory.2015.03.004. (Visited on 08/12/2024).

Dalca, Adrian V. et al. (Oct. 2019). "Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces". In: *Medical Image Analysis* 57, pp. 226–236. ISSN: 13618415. DOI: 10.1016/j.media.2019.07.006. arXiv: 1903.03545 [cs]. (Visited on 06/17/2024).

De Martino, Fortuna et al. (Jan. 2021). "Dose Calculation Algorithms for External Radiation Therapy: An Overview for Practitioners". In: *Applied Sciences* 11.15, p. 6806. ISSN: 2076-3417. DOI: 10.3390/app11156806. (Visited on 03/08/2024).

Deiter, Noelle et al. (Dec. 2020). "Evaluation of Replanning in Intensity-Modulated Proton Therapy for Oropharyngeal Cancer: Factors Influencing Plan Robustness". In: *Medical Dosimetry* 45.4, pp. 384–392. ISSN: 0958-3947, 1873-4022. DOI: 10.1016/j.meddos.2020.06.002. (Visited on 06/06/2024).

Duderstadt, J.J. and L.J. Hamilton (1991). *Nuclear Reactor Analysis*. Wiley. ISBN: 978-0-471-22363-4.

Elaimy, Ameer L. et al. (Aug. 2021). "History and Overview of Proton Therapy". In: *Proton Therapy - Current Status and Future Directions*. IntechOpen. ISBN: 978-1-83968-013-7. DOI: 10.5772/intechopen.95959. (Visited on 08/14/2024).

Ericson, S. (Oct. 1970). "The Normal Variation of the Parotid Size". In: *Acta Oto-Laryngologica*. DOI: 10.3109/00016487009181890. (Visited on 07/10/2024).

Eyges, Leonard (Nov. 1948). "Multiple Scattering with Energy Loss". In: *Physical Review* 74.10, pp. 1534–1535. DOI: 10.1103/PhysRev.74.1534. (Visited on 04/16/2021).

Feng, Mary et al. (Jan. 2018). "Individualized Adaptive Stereotactic Body Radiotherapy for Liver Tumors in Patients at High Risk for Liver Damage". In: *JAMA Oncology* 4.1, pp. 40–47. ISSN: 2374-2437. DOI: 10.1001/jamaoncol.2017.2303. (Visited on 08/17/2024).

Fiorentino, A et al. (Oct. 2012). "Parotid Gland Volumetric Changes during Intensity-Modulated Radiotherapy in Head and Neck Cancer". In: *The British Journal of Radiology* 85.1018, pp. 1415–1419. ISSN: 0007-1285. DOI: 10.1259/bjr/30678306. (Visited on 06/12/2024).

Frank, Steven J. and X. Ronald Zhu (2020). *Proton Therapy Indications, Techniques, and Outcomes*. ISBN: 978-0-323-73349-6.

Gebäck, Tobias and Mohammad Asadzadeh (Oct. 2012). "Analytical Solutions for the Pencil-Beam Equation with Energy Loss and Straggling".
In: *Transport Theory and Statistical Physics* 41.5-6, pp. 325–336. ISSN: 0041-1450. DOI: 10.1080/00411450.2012.671207. (Visited on 03/16/2022).

Gerbershagen, Alexander et al. (July 2017). "Simulations and Measurements of Proton Beam Energy Spectrum after Energy Degradation".
In: *Journal of Physics: Conference Series* 874, p. 012108.
DOI: 10.1088/1742-6596/874/1/012108.

Ghojogh, Benyamin et al. (May 2022).
Factor Analysis, Probabilistic Principal Component Analysis, Variational Inference, and Variational Autoencoder: Tutorial and Survey.
arXiv: 2101.00734 [cs, stat]. (Visited on 09/03/2024).

Gianfaldoni, Serena et al. (July 2017). "An Overview on Radiotherapy: From Its History to Its Current Applications in Dermatology".
In: *Open Access Macedonian Journal of Medical Sciences* 5.4, pp. 521–525.
ISSN: 1857-9655. DOI: 10.3889/oamjms.2017.122. (Visited on 09/24/2024).

Goldstein, Herbert, Charles Patton Jr. Poole, and John. Safko (2002).
Classical Mechanics. 3rd ed. Reading, Mass.: Addison-Wesley.
Chap. XVIII, 638 p. : illustrations ; 24 cm. ISBN: 0-201-65702-3 978-0-201-65702-9.

Gottschalk, Bernard (2012). *Techniques of Proton Radiotherapy: Transport Theory.*
DOI: 10.48550/ARXIV.1204.4470.

Green, Olga L., Lauren E. Henke, and Geoffrey D. Hugo (July 2019).
"Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy".
In: *Seminars in Radiation Oncology* 29.3, pp. 219–227. ISSN: 10534296.
DOI: 10.1016/j.semradonc.2019.02.004. (Visited on 08/15/2024).

Hajdu, Steven I. (2011a).
"A Note from History: Landmarks in History of Cancer, Part 1".
In: *Cancer* 117.5, pp. 1097–1102. ISSN: 1097-0142. DOI: 10.1002/cncr.25553.
(Visited on 08/06/2024).

— (2011b). "A Note from History: Landmarks in History of Cancer, Part 2".
In: *Cancer* 117.12, pp. 2811–2820. ISSN: 1097-0142. DOI: 10.1002/cncr.25825.
(Visited on 08/06/2024).

— (2012). "A Note from History: Landmarks in History of Cancer, Part 4".
In: *Cancer* 118.20, pp. 4914–4928. ISSN: 1097-0142. DOI: 10.1002/cncr.27509.
(Visited on 08/07/2024).

Hajdu, Steven I. and Manjunath Vadmal (2013).
"A Note from History: Landmarks in History of Cancer, Part 6".
In: *Cancer* 119.23, pp. 4058–4082. ISSN: 1097-0142. DOI: 10.1002/cncr.28319.
(Visited on 08/08/2024).

Hajdu, Steven I., Manjunath Vadmal, and Ping Tang (2015).
"A Note from History: Landmarks in History of Cancer, Part 7".
In: *Cancer* 121.15, pp. 2480–2513. ISSN: 1097-0142. DOI: 10.1002/cncr.29365.
(Visited on 08/08/2024).

Hanahan, Douglas and Robert A. Weinberg (Jan. 2000). "The Hallmarks of Cancer".
In: *Cell* 100.1, pp. 57–70. ISSN: 0092-8674, 1097-4172.
DOI: 10.1016/S0092-8674(00)81683-9. (Visited on 08/06/2024).

Hillewaert, Koen (2013). "Development of the Discontinuous Galerkin Method for High-Resolution, Large Scale CFD and Acoustics in Industrial Geometries".
PhD thesis. UCL - Université Catholique de Louvain. (Visited on 04/13/2022).

Huh, Hyun Do and Seonghoon Kim (Sept. 2020).
"History of Radiation Therapy Technology".

In: *Progress in Medical Physic* 31.3, pp. 124–134.
DOI: 10.14316/pmp.2020.31.3.124. (Visited on 08/08/2024).

Huh, Seung Jae, Won Park, and Do Ho Choi (Dec. 2019). "Recent Trends in Intensity-Modulated Radiation Therapy Use in Korea". In: *Radiation Oncology Journal* 37.4, pp. 249–253. ISSN: 2234-1900. DOI: 10.3857/roj.2019.00577. (Visited on 08/16/2024).

Jaderberg, Max et al. (Feb. 2016). *Spatial Transformer Networks*. DOI: 10.48550/arXiv.1506.02025. arXiv: 1506.02025 [cs]. (Visited on 06/17/2024).

Johnson, J. E. et al. (Feb. 2019). "Highly Efficient and Sensitive Patient-Specific Quality Assurance for Spot-Scanned Proton Therapy". In: *PLoS ONE* 14.2, e0212412. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0212412. (Visited on 02/19/2024).

Kang, Joanne H., Jan J. Wilkens, and Uwe Oelfke (May 2008). "Non-Uniform Depth Scanning for Proton Therapy Systems Employing Active Energy Variation". In: *Physics in Medicine and Biology* 53.9, N149–155. ISSN: 0031-9155. DOI: 10.1088/0031-9155/53/9/N01.

Kennedy, Christopher A and Mark H Carpenter (2016). *Diagonally Implicit Runge-Kutta Methods for Ordinary Differential Equations. A Review*. Tech. rep.

Kingma, Diederik P. and Max Welling (2019). "An Introduction to Variational Autoencoders". In: *Foundations and Trends® in Machine Learning* 12.4, pp. 307–392. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000056. arXiv: 1906.02691 [cs, stat]. (Visited on 09/02/2024).

Kooy, H M and C Grassberger (July 2015). "Intensity Modulated Proton Therapy". In: *The British Journal of Radiology* 88.1051, p. 20150195. ISSN: 0007-1285. DOI: 10.1259/bjr.20150195. (Visited on 09/25/2024).

Krebs, Julian et al. (Sept. 2019). "Learning a Probabilistic Model for Diffeomorphic Registration". In: *IEEE Transactions on Medical Imaging* 38.9, pp. 2165–2176. ISSN: 1558-254X. DOI: 10.1109/TMI.2019.2897112. (Visited on 06/10/2024).

Kwint, Margriet et al. (Dec. 2014). "Intra Thoracic Anatomical Changes in Lung Cancer Patients during the Course of Radiotherapy". In: *Radiotherapy and Oncology* 113.3, pp. 392–397. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2014.10.009. (Visited on 08/16/2024).

Lalee, Marucha, Jorge Nocedal, and Todd Plantenga (Aug. 1998). "On the Implementation of an Algorithm for Large-Scale Equality Constrained Optimization". In: *SIAM Journal on Optimization* 8.3, pp. 682–706. ISSN: 1052-6234. DOI: 10.1137/S1052623493262993. (Visited on 09/08/2023).

Li, Heng et al. (2013). "Use of Treatment Log Files in Spot Scanning Proton Therapy as Part of Patient-Specific Quality Assurance". In: *Medical Physics* 40.2, p. 021703. ISSN: 2473-4209. DOI: 10.1118/1.4773312. (Visited on 04/22/2022).

Liu, Wei et al. (2013). "Effectiveness of Robust Optimization in Intensity-Modulated Proton Therapy Planning for Head and Neck Cancers". In: *Medical Physics* 40.5, p. 051711. ISSN: 2473-4209. DOI: 10.1118/1.4801899. (Visited on 06/06/2024).

Loap MSc, Pierre, Renaud Huynh, and Youlia Kirova Md (June 2021). "The Centenary of the Fondation Curie (1921-2021)". In: *International Journal of Radiation Oncology*Biology*Physics* 110.2, pp. 331–336. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2020.12.013. (Visited on 09/24/2024).

Lomax, A J (Feb. 2008a). "Intensity Modulated Proton Therapy and Its Sensitivity to Treatment Uncertainties 1: The Potential Effects of Calculational Uncertainties". In: *Physics in Medicine and Biology* 53.4, pp. 1027–1042. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/53/4/014. (Visited on 04/20/2022).

— (Feb. 2008b). "Intensity Modulated Proton Therapy and Its Sensitivity to Treatment Uncertainties 2: The Potential Effects of Inter-Fraction and Inter-Field Motions". In: *Physics in Medicine and Biology* 53.4, pp. 1043–1056. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/53/4/015. (Visited on 04/20/2022).

Long, T. et al. (June 2012). "Sensitivity Analysis for Lexicographic Ordering in Radiation Therapy Treatment Planning". In: *Medical Physics* 39.6, pp. 3445–3455. ISSN: 0094-2405. DOI: 10.1118/1.4720218.

Luo, Zhengming (Sept. 1998). "An Overview of the Bipartition Model for Charged Particle Transport". In: *Radiation Physics and Chemistry* 53.3, pp. 305–327. ISSN: 0969-806X. DOI: 10.1016/S0969-806X(98)00112-1. (Visited on 11/08/2024).

Maeda, Yoshikazu et al. (2018). "Positioning Accuracy and Daily Dose Assessment for Prostate Cancer Treatment Using In-Room CT Image Guidance at a Proton Therapy Facility". In: *Medical Physics* 45.5, pp. 1832–1843. ISSN: 2473-4209. DOI: 10.1002/mp.12858. (Visited on 10/01/2024).

Malimban, Justin et al. (Apr. 2024). "A Deterministic Proton Dose Engine with Monte Carlo Accuracy for Preclinical Applications". In: *6th Conference on Small Animal Precision Image-guided Radiotherapy.*

MCsquare - Commissioning Procedure (2024). http://www.openmcsquare.org/. (Visited on 08/21/2024).

Medbery, Robbie et al. (Jan. 2000). "Variation in Parotid Gland Size, Configuration, and Anatomic Relations". In: *Radiotherapy and Oncology* 54.1, pp. 87–89. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/S0167-8140(99)00150-4. (Visited on 07/09/2024).

Meier, G. et al. (Mar. 2015). "Independent Dose Calculations for Commissioning, Quality Assurance and Dose Reconstruction of PBS Proton Therapy". In: *Physics in Medicine and Biology* 60.7, pp. 2819–2836. ISSN: 0031-9155. DOI: 10.1088/0031-9155/60/7/2819. (Visited on 04/11/2022).

Mohan, Radhe and David Grosshans (Jan. 2017). "Proton Therapy – Present and Future". In: *Advanced drug delivery reviews* 109, pp. 26–44. ISSN: 0169-409X. DOI: 10.1016/j.addr.2016.11.006. (Visited on 09/27/2024).

Moreno, Amy C. et al. (Jan. 2019). "Intensity Modulated Proton Therapy (IMPT) – The Future of IMRT for Head and Neck Cancer". In: *Oral oncology* 88, pp. 66–74. ISSN: 1368-8375. DOI: 10.1016/j.oraloncology.2018.11.015. (Visited on 08/17/2024).

National Institute of Standards and Technology (1993). *Penetration of Proton Beams through Water 1. Depth-Dose Distribution, Spectra and LET Distribution.* Tech. rep. NIST IR 5226. Gaithersburg, MD: National Institute of Standards and Technology, NIST IR 5226. DOI: 10.6028/NIST.IR.5226. (Visited on 09/09/2024).

Noshad, Houshyar and Seyyedeh Samira Bahador (Oct. 2012). "Investigation on Energy Straggling of Protons via Fokker–Planck Equation".

In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 288, pp. 89–93. ISSN: 0168-583X. DOI: 10.1016/j.nimb.2012.07.030. (Visited on 06/16/2022).

Nystrom, Hakan, Maria Fuglsang Jensen, and Petra Witt Nystrom (Mar. 2020). "Treatment Planning for Proton Therapy: What Is Needed in the next 10 Years?" In: *The British Journal of Radiology* 93.1107, p. 20190304. ISSN: 0007-1285. DOI: 10.1259/bjr.20190304. (Visited on 06/03/2024).

Oud, Michelle et al. (Nov. 2022). "An Online Adaptive Plan Library Approach for Intensity Modulated Proton Therapy for Head and Neck Cancer". In: *Radiotherapy and Oncology* 176, pp. 68–75. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2022.09.011. (Visited on 06/06/2024).

Oud, Michelle et al. (Mar. 2024). "A Fast and Robust Constraint-Based Online Re-Optimization Approach for Automated Online Adaptive Intensity Modulated Proton Therapy in Head and Neck Cancer". In: *Physics in Medicine & Biology* 69.7, p. 075007. ISSN: 0031-9155. DOI: 10.1088/1361-6560/ad2a98. (Visited on 10/25/2024).

Paganetti, H. (Mar. 2002). "Nuclear Interactions in Proton Therapy: Dose and Relative Biological Effect Distributions Originating from Primary and Secondary Particles". In: *Physics in Medicine and Biology* 47.5, pp. 747–764. ISSN: 0031-9155. DOI: 10.1088/0031-9155/47/5/305.

Paganetti, Harald, ed. (Nov. 2018). *Proton Therapy Physics*. 2nd ed. Boca Raton: CRC Press. ISBN: 978-1-315-22831-0. DOI: 10.1201/b22053.

Paganetti, Harald et al. (Nov. 2021). "Adaptive Proton Therapy". In: *Physics in Medicine & Biology* 66.22, 22TR01. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/1361-6560/ac344f. (Visited on 04/20/2022).

Pastor Serrano, O. (2023). "Artificial Intelligence in Radiotherapy: Probabilistic Deep Learning for Dose Prediction and Anatomy Modeling". In: DOI: 10.4233/uuid:c0c501d2-7c05-4e95-b8e8-d81aab627bb9. (Visited on 03/15/2024).

Pastor-Serrano, Oscar et al. (Apr. 2023). "A Probabilistic Deep Learning Model of Inter-Fraction Anatomical Variations in Radiotherapy". In: *Physics in Medicine and Biology* 68.8, p. 085018. ISSN: 1361-6560. DOI: 10.1088/1361-6560/acc71d.

Paszke, Adam et al. (Oct. 2017). "Automatic Differentiation in PyTorch". In: (visited on 12/03/2024).

Pedregosa, Fabian et al. (2011). "Scikit-Learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85, pp. 2825–2830. ISSN: 1533-7928. (Visited on 09/05/2024).

Perl, J. et al. (Nov. 2012). "TOPAS: An Innovative Proton Monte Carlo Platform for Research and Clinical Applications". In: *Medical Physics* 39.11, pp. 6818–6837. ISSN: 0094-2405. DOI: 10.1118/1.4758060.

Piessens, Robert et al. (1983). *QUADPACK: A Subroutine Package for Automatic Integration*. Springer.

Placidi, Lorenzo et al. (Mar. 2017). "Effect of Anatomic Changes on Pencil Beam Scanned Proton Dose Distributions for Cranial and Extracranial Tumors". In: *International Journal of Radiation Oncology, Biology, Physics* 97.3, pp. 616–623. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2016.11.013. (Visited on 10/01/2024).

Poenisch, Falk et al. (Jan. 2021). "CHAPTER 6 - Physics Quality Assurance". In: *Proton Therapy*. Ed. by STEVEN J. Frank and X. Ronald Zhu.

Philadelphia: Elsevier, 80–105.e2. ISBN: 978-0-323-73349-6.
DOI: 10.1016/B978-0-323-73349-6.00015-7. (Visited on 08/15/2024).
Pomraning, G. C. (Nov. 1996). "Higher Order Fokker-Planck Operators".
In: *Nuclear Science and Engineering* 124.3, pp. 390–397. ISSN: 0029-5639.
DOI: 10.13182/NSE96-A17918. (Visited on 06/02/2024).
Pomraning, G. C and A. K Prinja (May 1999).
"A Large Deflection Fermi–Eyges Pencil Beam Formula".
In: *Annals of Nuclear Energy* 26.7, pp. 595–609. ISSN: 0306-4549.
DOI: 10.1016/S0306-4549(98)00076-0. (Visited on 06/03/2024).
Praagman, Jaike et al. (Oct. 2022). *Kanker in Nederland*. Tech. rep. KNL, p. 85.
"Proceedings to the 61st Annual Conference of the Particle Therapy Cooperative Group" (Nov. 2023).
In: *International Journal of Particle Therapy* 10.2, pp. 118–396. ISSN: 2331-5180.
DOI: 10.14338/IJPT-23-PTCOG61-10.2. (Visited on 08/16/2024).
RayStation (Jan. 2019). *RayStation 9A, RayPhysics Manual*. (Visited on 08/12/2024).
Reed, Amy B. (Jan. 2011). "The History of Radiation Use in Medicine".
In: *Journal of Vascular Surgery* 53.1, 3S–5S. ISSN: 0741-5214, 1097-6809.
DOI: 10.1016/j.jvs.2010.07.024. (Visited on 09/24/2024).
Reiners, Keaton et al. (Jan. 2023). "CBCT-Based Dose Monitoring and Adaptive Planning Triggers in Head and Neck PBS Proton Therapy".
In: *Cancers* 15.15, p. 3881. ISSN: 2072-6694. DOI: 10.3390/cancers15153881. (Visited on 08/16/2024).
Rivière, Béatrice (Jan. 2008).
Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations.
Frontiers in Applied Mathematics.
Society for Industrial and Applied Mathematics. ISBN: 978-0-89871-656-6.
DOI: 10.1137/1.9780898717440. (Visited on 06/07/2021).
Röntgen, W. C. (1898). "Ueber Eine Neue Art von Strahlen".
In: *Annalen der Physik* 300.1, pp. 12–17. ISSN: 1521-3889.
DOI: 10.1002/andp.18983000103. (Visited on 08/07/2024).
Santos, Wellington Pereira dos et al. (2020).
"Morphology, Volume, and Density Characteristics of the Parotid Glands before and after Chemoradiation Therapy in Patients with Head and Neck Tumors".
In: *International Journal of Dentistry* 2020.1, p. 8176260. ISSN: 1687-8736.
DOI: 10.1155/2020/8176260. (Visited on 07/09/2024).
Scandurra, D. et al. (Jan. 2016).
"Assessing the Quality of Proton PBS Treatment Delivery Using Machine Log Files: Comprehensive Analysis of Clinical Treatments Delivered at PSI Gantry 2".
In: *Physics in Medicine and Biology* 61.3, pp. 1171–1181. ISSN: 0031-9155.
DOI: 10.1088/0031-9155/61/3/1171. (Visited on 05/31/2021).
Schneider, Wilfried, Thomas Bortfeld, and Wolfgang Schlegel (Feb. 2000).
"Correlation between CT Numbers and Tissue Parameters Needed for Monte Carlo Simulations of Clinical Dose Distributions".
In: *Physics in Medicine and Biology* 45.2, pp. 459–478. ISSN: 0031-9155, 1361-6560.
DOI: 10.1088/0031-9155/45/2/314. (Visited on 04/28/2022).
Smolders, Andreas Johan et al. (2024). "DiffuseRT: Predicting Likely Anatomical Deformations of Patients Undergoing Radiotherapy".
In: *Physics in Medicine & Biology*. ISSN: 0031-9155.
DOI: 10.1088/1361-6560/ad61b7. (Visited on 07/18/2024).
Sonke, Jan-Jakob and José Belderbos (Apr. 2010).
"Adaptive Radiotherapy for Lung Cancer".

In: *Seminars in Radiation Oncology*. Adaptive Radiotherapy 20.2, pp. 94–106. ISSN: 1053-4296. DOI: 10.1016/j.semradonc.2009.11.003. (Visited on 08/17/2024).

Soukup, Martin, Matthias Fippel, and Markus Alber (Oct. 2005). "A Pencil Beam Algorithm for Intensity Modulated Proton Therapy Derived from Monte Carlo Simulations". In: *Physics in Medicine and Biology* 50.21, pp. 5089–5104. ISSN: 0031-9155. DOI: 10.1088/0031-9155/50/21/010. (Visited on 06/20/2022).

Stammer, Pia et al. (Dec. 2024). *A Deterministic Dynamical Low-rank Approach for Charged Particle Transport*. DOI: 10.48550/arXiv.2412.09484. arXiv: 2412.09484 [math]. (Visited on 12/19/2024).

Taubmann, Oliver et al. (2018). "Computed Tomography". In: *Medical Imaging Systems: An Introductory Guide*. Ed. by Andreas Maier et al. Cham: Springer International Publishing, pp. 147–189. ISBN: 978-3-319-96520-8. DOI: 10.1007/978-3-319-96520-8_8. (Visited on 08/15/2024).

Taylor, A and M E B Powell (Mar. 2004). "Intensity-Modulated Radiotherapy—What Is It?" In: *Cancer Imaging* 4.2, pp. 68–73. ISSN: 1740-5025. DOI: 10.1102/1470-7330.2004.0003. (Visited on 09/25/2024).

Tian, Xiufang et al. (Jan. 2018). "The Evolution of Proton Beam Therapy: Current and Future Status". In: *Molecular and Clinical Oncology* 8.1, pp. 15–21. ISSN: 2049-9450. DOI: 10.3892/mco.2017.1499. (Visited on 08/09/2024).

Trnková, P. et al. (2016). "Factors Influencing the Performance of Patient Specific Quality Assurance for Pencil Beam Scanning IMPT Fields". In: *Medical Physics* 43.11, pp. 5998–6008. ISSN: 2473-4209. DOI: 10.1118/1.4964449. (Visited on 08/30/2023).

Unkelbach, Jan and Harald Paganetti (Apr. 2018). "Robust Proton Treatment Planning: Physical and Biological Optimization". In: *Seminars in Radiation Oncology* 28.2, pp. 88–96. ISSN: 10534296. DOI: 10.1016/j.semradonc.2017.11.005. (Visited on 04/21/2022).

van de Schoot, Agustinus J. A. J. et al. (July 2016). "Dosimetric Advantages of Proton Therapy Compared with Photon Therapy Using an Adaptive Strategy in Cervical Cancer". In: *Acta Oncologica* 55.7, pp. 892–899. ISSN: 0284-186X. DOI: 10.3109/0284186X.2016.1139179. (Visited on 06/06/2024).

Van de Water, Steven et al. (June 2015). "Shortening Delivery Times of Intensity Modulated Proton Therapy by Reducing Proton Energy Layers during Treatment Plan Optimization". In: *International Journal of Radiation Oncology, Biology, Physics* 92.2, pp. 460–468. ISSN: 1879-355X. DOI: 10.1016/j.ijrobp.2015.01.031.

Van de Water, Steven et al. (Jan. 2018). "Anatomical Robust Optimization to Account for Nasal Cavity Filling Variation during Intensity-Modulated Proton Therapy: A Comparison with Conventional and Adaptive Planning Strategies". In: *Physics in Medicine & Biology* 63.2, p. 025020. ISSN: 0031-9155. DOI: 10.1088/1361-6560/aa9c1c. (Visited on 06/15/2021).

van Kranen, Simon et al. (Apr. 2009). "Setup Uncertainties of Anatomical Sub-Regions in Head-and-Neck Cancer Patients After Offline CBCT Guidance". In: *International Journal of Radiation Oncology*Biology*Physics* 73.5, pp. 1566–1573. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2008.11.035. (Visited on 05/31/2021).

Vásquez Osorio, Eliana M. et al. (Mar. 2008). "Local Anatomic Changes in Parotid and Submandibular Glands During Radiotherapy for Oropharynx Cancer and Correlation With Dose, Studied in Detail With Nonrigid Registration". In: *International Journal of Radiation Oncology*Biology*Physics* 70.3, pp. 875–882. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2007.10.063. (Visited on 06/12/2024).

Virtanen, Pauli et al. (Mar. 2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17.3, pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. (Visited on 09/08/2023).

Visser, Sabine et al. (Dec. 2022). "Robustness Assessment of Clinical Adaptive Proton and Photon Radiotherapy for Oesophageal Cancer in the Model-Based Approach". In: *Radiotherapy and Oncology* 177, pp. 197–204. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2022.11.001. (Visited on 10/01/2024).

Way, Hansen (2015). "Eclipse Photon and Electron Algorithms Reference". In.

Webb, Steve and Philip M. Evans (Oct. 2006). "Innovative Techniques in Radiation Therapy: Editorial, Overview, and Crystal Ball Gaze to the Future". In: *Seminars in Radiation Oncology*. Innovative Technologies in Radiation Therapy 16.4, pp. 193–198. ISSN: 1053-4296. DOI: 10.1016/j.semradonc.2006.04.001. (Visited on 08/09/2024).

Williams, Evan James and William Lawrence Na8141 Bragg (Feb. 1932). "The Passage of α- and β- Particles through Matter and Born's Theory of Collisions". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 135.826, pp. 108–131. DOI: 10.1098/rspa.1932.0023. (Visited on 06/11/2021).

Wilson, Robert R. (Nov. 1946). "Radiological Use of Fast Protons". In: *Radiology* 47.5, pp. 487–491. ISSN: 0033-8419. DOI: 10.1148/47.5.487. (Visited on 08/08/2024).

Xiao, Mei et al. (2024). "Comparison of Cyclotron and Synchrotron in Particle Therapy". In: *Visualized Cancer Medicine* 5, p. 7. ISSN: 2740-4218. DOI: 10.1051/vcm/2024008. (Visited on 09/26/2024).

Yan, Di et al. (Jan. 1997). "Adaptive Radiation Therapy". In: *Physics in Medicine & Biology* 42.1, p. 123. ISSN: 0031-9155. DOI: 10.1088/0031-9155/42/1/008. (Visited on 08/15/2024).

Yang, Jinhe et al. (Sept. 2020). "An Improved Beam Splitting Method for Intensity Modulated Proton Therapy". In: *Physics in Medicine &amp$\mathsemicolon$ Biology* 65.18, p. 185015. ISSN: 0031-9155. DOI: 10.1088/1361-6560/ab9b55. (Visited on 06/20/2022).

Yorke, A. A. et al. (2019). *Pelvic Reference Data*. Imaging Data. The Cancer Imaging Archive. DOI: 10.7937/TCIA.2019.WOSKQ50O. (Visited on 03/08/2024).

Zheng-Ming, Luo and Anders Brahme (1993). "An Overview of the Transport Theory of Charged Particles". In: *Radiation Physics and Chemistry* 41.4, pp. 673–703. ISSN: 0969-806X. DOI: 10.1016/0969-806X(93)90318-O.

Acknowledgements

Limba Română pe pagina următoare

Pursuing this PhD has been an extraordinary journey, and I feel fortunate to have experienced it. It has been a time rich with personal and intellectual growth, presenting numerous opportunities and challenges, as well as moments of joy and disappointment. There are few experiences as rewarding as witnessing theoretical concepts come to life in my simulations, yet there are also few more self-doubting moments than spending months trying to understand why something fails to perform as anticipated and having to go back to the drawing board.

After such an extended period of intense effort, I imagined a great relief similar to the gray skies of the Netherlands parting, allowing a beam of sunlight to shine down, accompanied by the harmonious sounds of angels. I guess you cannot always get what you want. So, the predominant question on my mind now is: what comes next? However, before I contemplate my next steps, I would like to reflect on and acknowledge all the individuals who have supported me throughout this journey.

I would like to deeply thank both Danny and Zoltán who have been my supervisors in this project. Danny, thank you for your guidance during both my MSc and PhD projects. You helped me with being more independent in my approach, to trust myself when results do not intuitively seem correct and to be even more detail oriented. I have significantly improved my coding and scientific knowledge through the numerous discussions we have had. Additionally, thank you for your patience, for always having an open door and for your kindness (which you showed when you offered to take me to the hospital when I broke my ACL). Zoltán, given our respective nationalities, I think we did the best we could. Eastern European jokes aside, I also want to thank you for your guidance. You have taught me a lot about being thorough, putting in the hard work, thinking of the bigger picture and finding my place in it. Thank you too for always having an open door to me and for always being willing to get your hands dirty in finding solutions to issues. Frankly, I greatly enjoyed working with both of you. I particularly enjoyed when we went to the whiteboard and discussed physics, mathematics or programming and found solutions to the numerous problems that I encountered. I also enjoyed the several moments spent outside of work, during dinners at conferences or other events, where we discussed and laughed about wide-ranging issues. I think you make a great supervisory team, complement each other and I have been privileged to have both of you as my supervisors.

I would also like to thank my family. My fiancée Emma, my parents Cristina and Ginel, Emma's parents Lene and Per and brother Martin as well as my dog Otto (who despite not being able to read is still a good boy). Having a good foundation at home, allowed me to have the peace of mind that is needed to take tasks to completion in a PhD. Emma, thank you for being with me throughout my whole academic career and for providing not only a listening ear to my issues but also your unique perspectives. My life is greatly enriched by having you and Otto in it, and I look forward to continuing making an excellent team in all of our joint ventures. Dear parents, I clearly would not be where I am without having you. Thank you for providing me with a good education

with a focus on critical thinking, for always seeking solutions at the round table and for supporting me to leave Romania to establish myself in the Netherlands. Lene, Per and Martin, thank you for being another brick in the foundation I relied on. I always enjoyed visiting, discussing and spending time with you. You always provided independent and honest opinions and perspectives, which I have found greatly useful and challenged me to open up my horizons into topics that I would not naturally have approached.

Lastly, I would like to thank the people that I have bonded with in and outside the Medical Physics & Technology group. Marc and Celebrity who I joined in their cubicle roughly halfway in my PhD, and the "new wave" of PhDs, Mikolaj, Alessandro and Jelte. I greatly enjoyed all the moments we spent together: interesting discussions during walks, dinners at our places, various outings into the city and sports in different contexts. I think this was a great mix of like-minded yet different people, each with their own perspectives. I wish you all the best of luck, not only with the remaining parts of your PhDs, but also with what comes next for each of you. Thank you also to my old high-school friends, namely Tudor, Dan, Radu and the rest. Thank you for the all the fun we have had, and I am happy we stayed in touch despite the long distances involved.

Limba Română

Acest doctorat a fost o călătorie extraordinară și mă simt norocos că am trecut prin ea. A fost o perioadă plină de creștere personală și intelectuală, prezentând numeroase oportunități și provocări, precum și momente de bucurie și de dezamăgire. Există puține experiențe la fel de satisfăcătoare precum acelea de a vedea conceptele teoretice care prind viață în simulările mele. De asemenea, există puține momente de îndoială mai mare decât acelea în care pentru luni de zile ceva nu funcționează așa cum am anticipat și trebuie să mă întorc la început.

După o perioadă atât de lungă de efort intens, am crezut că cerul gri al Olandei se va despărți, permițând unui fascicul de lumină să strălucească în jos, însoțit de cântecele armonioase ale îngerilor. Bănuiesc că nu poți obține întotdeauna ceea ce îți dorești. Așadar, întrebarea predominantă în mintea mea acum este: ce urmează? Înainte de a-mi contempla următorii pași, aș dori să mulțumesc tuturor celor care m-au susținut pe parcursul acestei călătorii.

Aș dori să-i mulțumesc profund atât lui Danny, cât și lui Zoltán, cei care au fost supraveghetorii mei în acest proiect. Danny, îți mulțumesc pentru îndrumarea ta în timpul proiectelor mele de masterat și doctorat. M-ai ajutat să fiu mai independent în abordarea mea, să am încredere în mine atunci când intuitiv rezultatele nu par corecte și să fiu și mai atent la detalii. Mi-am îmbunătățit semnificativ programele și cunoștințele științifice prin numeroasele discuții pe care le-am avut. În plus, îți mulțumesc pentru răbdarea ta, pentru că ai mereu ușa deschisă și pentru amabilitatea ta (de care ai dat dovadă când te-ai oferit să mă duci la spital atunci când mi-am rupt ligamentul încrucișat). Zoltán, având în vedere naționalitățile noastre, cred că mai bine de atât nu se putea. Trecând peste glumele regionale, vreau și ție să îți mulțumesc îndrumarea ta. M-ai învățat multe despre cum să fiu minuțios, să lucrez din greu, să mă gândesc la imaginea de ansamblu și să-mi găsesc locul în ea. Îți mulțumesc și ție pentru că ai întotdeauna ușa deschisă pentru mine și pentru că ești întotdeauna dispus să găsim soluții la probleme. Sincer, mi-a plăcut foarte mult să lucrez cu amândoi. Mi-a plăcut în special când am mers la tablă și am discutat despre fizică, matematică sau programare și am găsit soluții la numeroasele probleme pe care le-am întâlnit. De asemenea, mi-au plăcut momentele petrecute în afara serviciului, în timpul cinelor la conferințe sau alte evenimente, în care discutam și râdeam despre varii subiecte. Cred că faceți o echipă de

supervizare grozavă, vă completați reciproc și mă simt norocos că am avut privilegiul să vă am pe amândoi ca supervizori.

De asemenea, aș dori să mulțumesc familiei mele. Logodnicei mele Emma, părinților mei Cristina și Ginel, părinților și fratelui Emmei Lene, Per și Martin și câinelui meu Otto (care, deși nu știe să citească, este un băiat bun). Având o fundație bună acasă, mi-a adus liniștea adesea necesară pentru a duce sarcinile din timpul doctoratului până la capăt. Emma, îți mulțumesc că ai fost alături de mine de-a lungul întregii mele cariere academice și că mi-ai oferit nu numai un umăr de sprijin pentru problemele mele, ci și perspectivele tale unice. Viața mea este grozav îmbogățită prin faptul că vă am pe tine și pe Otto în ea și sper că vom continua să facem o echipă excelentă în tot ceea vom întreprinde în viitor. Dragi părinți, este clar că nu aș fi unde sunt acum fără să vă am pe voi. Vă mulțumesc că mi-ați oferit o educație bună cu accent pe gandire liberă și critică, pentru toate discuțiile pe care le-am purtat fie in living room sau pe internet și pentru că m-ați ajutat cu dorința mea de a părăsi România pentru a mă stabili în Olanda. Lene, Per și Martin, vă mulțumesc că sunteți o altă cărămidă în fundația pe care m-am bazat. Întotdeauna mi-a plăcut să vă vizitez, să discut și să petrec timp cu voi. Ați oferit întotdeauna opinii și perspective independente și sincere, pe care le-am găsit foarte utile și care m-au provocat să-mi deschid orizonturile în feluri în care nu le-aș fi abordat.

În cele din urmă, aș dori să mulțumesc celor cu care am legat prietenii din interiorul și din afara grupei Medical Physics & Technology. Marc și Celebrity cărora li m-am alăturat în biroul lor la jumătatea doctoratului și „noul val" de doctoranzi, Mikolaj, Alessandro și Jelte. Mi-au plăcut foarte mult toate momentele petrecute împreună: discuții interesante în timpul plimbărilor, cine la noi acasă, diferite ieșiri în oraș și sport în diferite contexte. Cred că acesta a fost un amestec grozav de oameni asemănători, dar diferiți, fiecare cu perspective unice. Vă doresc mult succes, nu numai cu restul doctoratului, ci și cu ceea ce urmează pentru fiecare dintre voi. Mulțumesc și vechilor mei prieteni din liceu Tudor, Dan, Radu și restul. Vă mulțumesc pentru toate momentele pe care le-am avut și sper că vom continua să păstrăm legătura în ciuda distanțelor lungi dintre noi.

List of publications

Journal publications

1. Tiberiu Burlacu, Danny Lathouwers, and Zoltán Perkó (Jan. 2023b). "A Deterministic Adjoint-Based Semi-Analytical Algorithm for Fast Response Change Computations in Proton Therapy". In: *Journal of Computational and Theoretical Transport* 52.1, pp. 1–41. ISSN: 2332-4309. DOI: 10.1080/23324309.2023.2166077. (Visited on 08/30/2023)
2. Tiberiu Burlacu, Danny Lathouwers, and Zoltán Perkó (July 2024). "Yet anOther Dose Algorithm (YODA) for Independent Computations of Dose and Dose Changes Due to Anatomical Changes". In: *Physics in Medicine & Biology* 69.16, p. 165003. ISSN: 0031-9155. DOI: 10.1088/1361-6560/ad6373. (Visited on 08/16/2024)
3. Tiberiu Burlacu et al. (Mar. 2025). "A Deep Learning Model for Inter-Fraction Head and Neck Anatomical Changes in Proton Therapy". In: *Physics in Medicine & Biology* 70.6, p. 065011. ISSN: 0031-9155. DOI: 10.1088/1361-6560/adba39. (Visited on 05/07/2025).

Conference publications

1. Tiberiu Burlacu, Danny Lathouwers, Zoltán Perkó. "Performance assessment of a deterministic dose algorithm for quality assurance in online adaptive proton therapy in heterogeneous patient geometries" "Proceedings to the 61st Annual Conference of the Particle Therapy Cooperative Group" (Nov. 2023). In: *International Journal of Particle Therapy* 10.2, pp. 118–396. ISSN: 2331-5180. DOI: 10.14338/IJPT-23-PTCOG61-10.2. (Visited on 08/16/2024)
2. T. Burlacu, D. Lathouwers, and Z. Perkó (May 2023a). "OC-0117 A Deterministic Algorithm for Patient-Specific Quality Assurance in Online Adaptive Proton Therapy". In: *Radiotherapy and Oncology* 182, S78–S79. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/S0167-8140(23)08531-6. (Visited on 08/16/2024)
3. Justin Malimban et al. (Apr. 2024). "A Deterministic Proton Dose Engine with Monte Carlo Accuracy for Preclinical Applications". In: *6th Conference on Small Animal Precision Image-guided Radiotherapy*
4. Pia Stammer et al. (Dec. 2024). *A Deterministic Dynamical Low-rank Approach for Charged Particle Transport*. DOI: 10.48550/arXiv.2412.09484. arXiv: 2412.09484 [math]. (Visited on 12/19/2024)

About the author

Tiberiu Burlacu was born in 1995 in Dorohoi, Romania. He attended the "Colegiul Național Ștefan cel Mare" in Suceava, Romania from 5^{th} to 12^{th} grade. In 2014 he enrolled in the Bachelor of Physics at the Rijksuniversiteit Groningen, the Netherlands. There he specialized in Nano, Experimental and Theoretical Physics and graduated in 2017, after completing his thesis titled "Magnetic field compensation for the BaF eEDM experiment". During his Bachelors he gained an interest in particle physics and its application for medical purposes. Therefore, he moved to Technische Universiteit Delft to follow the Applied Physics Masters programme with a Research & Development track. At the TU Delft, he specialized in medical physics and received the Nuclear Science and Engineering annotation to his diploma. As part of his programme, he worked on a thesis titled "Gauging the effect of energy straggling on proton dose distributions" which constituted the base of the development of YODA. During this time, he also gained clinical exposure, by doing an internship at the Holland Proton Therapy Center. There he saw the largely manual quality assurance procedures that are being done. In 2020 he joined the Medical Physics & Technology group of the Radiation Science and Technology department of the TU Delft as a PhD candidate on software based methods for the automation of quality assurance in proton therapy. The developments from 4 years of research are presented in this thesis.