MCCXXII

UNIVERSITÀ DEGLI STUDI DI PADOVA
Dipartimento di Fisica e Astronomia 'G. Galilei'

PH.D. COURSE IN PHYSICS
XXXVI CYCLE

DEVELOPMENT OF COMPUTATIONAL TOOLS FOR AN IMPROVED RADIOTHERAPY EFFICIENCY

From treatment plan complexity to quantum-inspired optimization techniques

COORDINATOR:
Ch.mo Prof. Giulio Monaco
SUPERVISOR:
Ch.mo Prof. Simone Montangero
CO-SUPERVISOR:
Dr. Alessandro Scaggion, *Ph.D.*

PH.D. STUDENT:
Samuele Cavinato

Samuele Cavinato : *Development of computational tools for an improved radiotherapy efficiency* , From treatment plan complexity to quantum-inspired optimization techniques, © September 2023

A mio papà,
anima umile e generosa,
instancabile e valoroso guerriero
che, nel silenzio assordante
della sua malattia,
mi insegna ogni giorno
il valore incommensurabile
della vita.

ABSTRACT

Modern intensity-modulated radiotherapy (IMRT) techniques including fixed-gantry IMRT, VMAT and Helical Tomotherapy allow for the delivery of treatments with high levels of target dose conformity while sparing the sorrounding healthy tissues at the same time, ultimately reducing the risk for side effects. On the one hand, implementing IMRT treatments is a sophisticated process based upon the need of solving an optimization problem known as *inverse-planning* aimed at finding the optimal beam modulation for the achivement of the prescribed dose objectives. Typical inverse-planning problems involve the optimiztion of cost-functions of hundreds to thousands of variables. Therefore, the efficiency of this process is highly constrained by the capability of the used algorithms to explore the cost-function's landscape efficiently when looking for an optimal solution.

On the other hand, the use of intensity-modulated beams possiblity introduces factors that might affect the dosimetric and delivery accuracy of the resulting plan either because of inaccuracies in the dose calculation occuring in the treatment planning system (TPS) or errors during treatment delivery. In literature, the collection of such factors is referred to as plan *complexity*. Thereby, prior to treatment delivery to the patient, each plan is delivered onto a system of dose detectors to check for its dosimetric and delivery accuracy. However, this process which is known as *patient-specific quality assurance* (PSQA), highly impacts the overall efficiency of the radiotherapy workflow since it requires considerable time and the involvement of several resources.

It turns out that, with the aim of improving the overall efficiency of the radiotherapy workflow, the development and use of new tools are worth being investagated. The topic of this thesis falls within this context. On the one hand, new methods based on *complexity metrics* are proposed as possible support tools for the characterization of the dosimetric and delivery accuracy of radiotherapy treatment plans. A novel software package called `UCoMX` is presented, which is the first tool for the extraction of complexity metrics from IMRT/VMAT and Helical Tomotherpay plans made freely available. Morover, three original investigations aimed at validating the use of complexity indicators within the PSQA process are presented, showing that they would allow for a 50% reduction of the PSQA workload at the center where this research has been carried out. In the near future, the use of such tools could become effective alternative to standard measurement-based PSQA procedures.

On the other hand, on an independent research line, the use of quantum computers for the optimization of the fluence maps of fixed-gantry IMRT plans is proposed. Quantum computers are under intense development by several research groups and companies worldwide and might lead to a revolution in the field of computation. Therefore, their use in the inverse-planning process might lead to unprecedented benefits in terms of efficiency. In this work, two original investigations that explored the feasibility of using quantum computers in radiotheray are proposed through the definition of novel strategies for the reformulation of the inverse-planning problem in a form compatible with the features of quantum hardware.

The ultimate aim of the present doctoral project was to put the basis for the future adoption of the aforementioned tools in clinics, by investigating both their potentiality and limitations.

SOMMARIO

Le moderne tecniche di radioterapia ad intensità modulata (IMRT) tra cui la IMRT a campi fissi, la VMAT e la Tomoterapia Elicoidale, permettono di erogare trattamenti caratterizzati da un alto livello di conformità della dose al target e, contemporaneamente, di ottenere un notevole risparmio dei tessuti sani circostanti, riducendo complessivamente il rischio di danni collaterali. Da un lato, la realizzazione di un piano di trattamento IMRT è un processo articolato, alla cui base vi è la necessità di risolvere un problema di ottimizzazione noto come inverse-planning, che permetta di ottenere una modulazione della fluenza del fascio tale da garantire il raggiungimento degli obiettivi di dose prescritti. Tale problema di ottimizzazione è tipicamente descritti da una funzione costo definita su centinaia o migliaia di variabili. Pertanto, l'efficienza di tale processo è fortemente vincolata all'abilità degli algorimi di ottimizzazione utilizzati di esplorare in modo efficiente il panorama della funzione costo.

Dall'altro, l'utilizzo di fasci ad intensità modulata potenzialmente introduce nei piani di trattamento radioterapico alcuni fattori che potrebbero comprometterne la qualità dosimetrica e di erogazione a causa, per esempio, di inaccuratezze nel calcolo della dose effettuato dai sistemi di pianificazione (TPS) o di errori di erogazione da parte dell'unità di trattamento. In letteratura, l'insieme di tali fattori è noto come *complessità* di un piano di trattamento. Pertanto, in una fase precedente all'erogazione del piano di trattamento sul paziente, questo viene erogato su un sistema di rivelatori di dose con lo scopo di verificarne l'accuratezza dosimetrica e di erogazione. Tuttavia, tale processo, noto come controllo qualità paziente specifico (PSQA) ha un notevole impatto sull'efficienza complessiva del flusso di lavoro radioterapico poiché richiede tempo e notevole coinvolgimento di risorse.

Con lo scopo, quindi, di andare a migliorare l'efficienza complessiva del flusso di lavoro radioterapico, emerge l'esigenza di esplorare l'utilizzo di nuovi strumenti. In tal contesto si inserisce il presente lavoro di tesi. Da un lato, vengono proposti dei metodi basati sulle *metriche di complessità* quali potenziali strumenti di supporto alla caratterizzazione della qualità dosimetrica e di erogazione dei trattamenti radioterapici. Viene presentato `UCoMX`, il primo pacchetto software liberamente disponibile per l'estrazione di indicatori di complessità da piani IMRT a gantry fisso, VMAT e Tomoterapia Elicoidale, compatibile con la gran parte dei TPS e delle unità di trattamento utilizzati in clinica. Vengono, inoltre, proposti tre studi originali sull'impiego e la

validazione degli indicatori di complessità nell'ambito del processo di PSQA, mostrando che essi potrebbero permettere una riduzione del carico del lavoro stimata del 50% presso il centro in cui è stato svolto il presente lavoro. In un futuro, tali strumenti potrebbero configurarsi come validi sostituti dei sistemi basati su misura.

Dall'altro lato, su una linea di ricerca indipendente, si propone l'utilizzo dei computer quantistici per l'ottimizzazione delle mappe di fluenza dei piani IMRT a gantry fisso. I computer quantistici, ad oggi oggetto di intenso sviluppo da parte di gruppi di ricerca e industrie a livello globale, potrebbero portare a una rivoluzione nel campo computazionale in termini di potenza di calcolo. Pertanto, il loro utilizzo nel processo di inverse-planning potebbe introdurre dei benefici notevoli. Si propongono, quindi, due studi originali nei quali si è andati ad esplorare la fattibilità di utilizzo di tali calcolatori in ambito radioterapico mediante lo sviluppo di strategie originali per una riformulazione del problema di inverse-planning che sia compatibile con le caratteristiche dell'hardware quantistico.

Lo scopo finale del presente progetto di dottorato era, pertanto, quello di porre le basi per la futura adozione degli strumenti sopra menzionati in clinica, investigandone nel dettaglio sia le potenzialità e che gli attuali limiti.

AKNOWLEDGEMENTS

No man is an island entire of itself;
every man is a piece of the continent,
a part of the main.

— John Donne

It may seem superfluous, but it is also necessary to emphasise that this work would not have been possible without the active supervision and constant support of a number of people. First of all, I would like to thank those who, with their professional skills, were able to guide me day after day towards the achievement of my goals. Dr. Marta Paiusco and Dr. Alessandro Scaggion, who with care, patience, professionalism, and extreme competence introduced me to the world of Medical Physics and had the ability to indulge my inclinations, channel my ideas, and realise my ambitions. I would like to thank Prof. Simone Montangero, who from the very first moment had the ability to support and accompany me with intuition and competence, succeeding in indulging my enthusiasm and always granting me the utmost trust and freedom in managing my work. Furthermore, I would like to thank all the staff of the UOC Fisica Sanitaria of the Istituto Oncologico Veneto, which in these three years has become my second home, an extended family made up of splendid and brilliant souls, always open to discussion and willing to help. A home where I have found true and sincere friends.

However, at the completion of this journey, I have to say a special thanks to those who have been the driving force behind my going all these years. I thank my mom and dad who passed on to me the love of culture and knowledge and pushed and supported me to chase my dreams. I thank E. and G, my lifelong friends, the famous 'few but good' that everyone would like and who, in spite of everything and in spite of everyone, have always been there. Finally, I thank Silvia, my wife, irreplaceable traveling companion. With her I have shared day after day the tears and smiles of an academic journey that, especially in these last years, has been intertwined with a sometimes impervious life path, during which I have often stumbled and fallen, but which in the end, thanks to her, has always seen me get back up.

Today, exactly nine years after the day I first set foot inside the Physics Department of the University of Padua, one of the richest and most important chapters of my life is about to end. A chapter made up of study, encounters, friendships, loves, detachments, departures, travels, research and experiences, during which I had the opportunity

to enjoy both the beauty and the toil of life from countless perspectives. Therefore, as I write these words, I can only look with humble gratitude at the road traveled so far and give thanks, perhaps God or who knows, for the gift received.

RINGRAZIAMENTI

Nessun uomo è un'isola, completo in se stesso;
Ogni uomo è un pezzo del continente,
Una parte del tutto.

— John Donne

Apparirà forse superfluo, ma è altresì doveroso, sottolineare che questo lavoro non sarebbe stato possibile senza la supervisione attiva e il supporto costante di alcune persone. Dapprima, ringrazio coloro che, con le loro competenze professionali, hanno saputo guidarmi giorno dopo giorno verso il raggiungimento degli obiettivi preposti. La Dott.ssa Marta Paiusco e il Dott. Alessandro Scaggion, che con cura, pazienza, professionalità ed estrema competenza mi hanno introdotto al mondo della fisica medica e hanno avuto l'abilità di assecondare le mie inclinazione, incanalare le mie idee e concretizzare le mie ambizioni. Il Prof. Simone Montangero, che fin dal primo momento ha avuto la capacità di affiancarmi e accompagnarmi con intuito e competenza, riuscendo ad assecondare il mio entusiasmo e accordandomi sempre la massima fiducia e libertà nella gestione del mio lavoro. Ringrazio, inoltre, tutto il personale della UOC di Fisica Sanitaria dell'Istituto Oncologico Veneto, che in questi tre anni è diventata la mia seconda casa, una famiglia allargata fatta di anime splendide e brillanti, sempre aperte al confronto e disponibili all'aiuto. Una casa in cui ho trovato amici veri e sinceri.

Tuttavia, al termine di questo percorso, non possono fare a meno di ringraziare esplicitamente coloro che sono stati motore del mio andare in tutti questi anni. Ringrazio mia mamma e mio papà che mi hanno trasmesso l'amore per la cultura e la conoscenza e mi ha hanno spinto e sostenuto a rincorrere i miei sogni. Ringrazio E. e G, i miei amici di una vita, i famosi 'pochi ma buoni' che tutti vorrebbero e che, nonostante tutto e nonostante tutti, ci sono sempre stati. Infine, ringrazio Silvia, mia moglie, insostituibile compagna di viaggio. Con lei ho condiviso giorno dopo giorno lacrime e sorrisi di un percorso accademico che, specialmente in questi ultimi anni, si è intrecciato con un percorso di vita talvolta impervio, durante il quale sono spesso inciampato e caduto, ma che alla fine, grazie a lei, mi ha sempre visto rialzarmi.

Oggi, esattamente a distanza di nove anni dal giorno in cui misi piede per la prima volta all'interno del Dipartimento di Fisica dell'Università di Padova, si sta per concludere uno dei capitoli più ricchi ed importanti della mia vita. Un capitolo fatto di studio, incontri, amicizie, amori, distacchi, partenze, viaggi, ricerca ed esperienze, durante il quale ho avuto l'opportunità godere sia della bellezza che della fatica della vita da innumerevoli prospettive. Pertanto, mentre scrivo questo parole, non posso che guardare con umile gratudine alla strada percorsa e ringraziare, forse Dio o chissachì, per il dono ricevuto.

CONTENTS

APPENDIX

LIST OF FIGURES

LIST OF TABLES

PREFACE

I have carried out my PhD research activity between the Quantum group of the Department of Physics and Astronomy of the University of Padova, and the Department of Medical Physics of the Veneto Insititute of Oncology. During my doctoral career, I pursued two different directions of research. First, I analysed and developed complexity indicators for characterizing the dosimetric accuracy of radiotherapy treatment plans. Second, I developed and implemented strategies for optimizing classical non-polynomial functions of continous variables on quantum hardware, with a particular focus on the cost-functions typically used for the inverse-planning optimization of treatment plans in radiotherapy. With the aim of reflecting the heterogeneity of my doctoral career and for the sake of clarity, I have divided the present thesis into two parts: *The complexity of radiotherapy treatment plans* and *Towards the use of quantum computers in radiotherapy*. Although it is noteworthy that the two topics comfortably align with the main focus of my doctoral project, namely developing computational tools to improve the efficiency of the radiotherapy workflow, making them converge was far beyond the scope of this thesis.

Most of the contents presented in this thesis are original research investigations taken from the papers that I have contributed to as the primary author (either those published on peer-review journals or the ones that are currently under review or being written). Only minor modifications have been performed to make their reciprocal connections explicit to the reader. Additionally, two introductory chapters have been added to convey the needed theoretical prerequisites to allow the reader to get the details of the investigations and developments herein presented. Finally, three additional scientific papers that I have contributed to as a co-author during my doctoral career have been reported in the Appendix.

PREFAZIONE

Ho svolto la mia attività di ricerca di dottorato tra il gruppo Quantum del Dipartimento di Fisica e Astronomia dell'Università di Padova e la UOC di Fisica Sanitaria dell'Istituto Oncologico Veneto. Durante il mio percorso di dottorato, ho perseguito due diverse direzioni di ricerca. In primo luogo, ho analizzato e sviluppato indicatori di complessità per caratterizzare l'accuratezza dosimetrica e di erogazione dei piani di trattamento radioterapico. In secondo luogo, ho sviluppato e implementato strategie per l'ottimizzazione di funzioni di variabili continue su hardware quantistico, con particolare attenzione alle funzioni costo tipicamente utilizzate durante il processo di pianificazione inversa per l'ottimizzazione dei piani di trattamento in radioterapia. Con l'obiettivo di riflettere l'eterogeneità del mio percorso di dottorato e per motivi di chiarezza, ho diviso la presente tesi in due parti: *La complessità dei piani di trattamento radioterapico* e *Verso l'uso dei computer quantistici in radioterapia*. Sebbene sia degno di nota il fatto che i due argomenti si allineino comodamente con l'obiettivo principale del mio progetto di dottorato, ovvero lo sviluppo di strumenti computazionali per migliorare l'efficienza del flusso di lavoro in radioterapia, farli convergere era ben oltre lo scopo del progetto qui presentato.

La maggior parte dei contenuti presentati in questa tesi sono ricerche originali tratte dai lavori a cui ho contribuito come autore principale (sia quelli pubblicati su riviste peer-review, sia quelli attualmente in fase di revisione o di scrittura). Sono state apportate solo piccole modifiche per rendere esplicito al lettore il loro reciproco collegamento. Inoltre, sono stati aggiunti due capitoli introduttivi, uno nella prima e uno nella seconda parte, per trasferire al lettore i necessari prerequisiti teorici per la comprensione dei dettagli delle indagini e degli sviluppi qui presentati. Infine, in Appendice sono riportati altri tre lavori scientifici a cui ho contribuito come coautore durante il mio percorso di dottorato.

MOTIVATION AND OUTLINE

Cancer is the leading cause of death worldwide, a burden that affects approximately 19 millions people each year and causes approximately 10 millions deaths[1]. According to the World Health Organization (WHO), between 30% and 50% of cancers can be prevented by avoiding risk factors and implementing existing evidence-based prevention strategies. But where prevention fails, treatment is needed. In this context, *radiotherapy* treatments play a crucial role in several cases. Radiotherapy is a compound word that connects the two terms *radiation* and *therapy*. It is a medical technique that exploits different types of ionizing radiation to kill tumor cells in patients affected by cancer. It is well documented that approximately 50% of such patients require at least one course of radiotherapy during their disease history globally, a value that rises to 87% for breast cancer. The typical radiotherapy workflow is a complex process which involves the participation of several professionals. Among tyem, the Medical Physics Expert (MPE) is essential to achieve proper management and practical implementation of the treatment. In particular, according to both International and Italian regulations, it is among the responsibility of the MPE to ensure the use of sufficient and appropriate technological tools for calculating, monitoring and evaluating the dosimetric accuracy of the treatment plans.

Among the different steps composing the radiotherapy workflow where the MPE is actively involved, two of them are central in this thesis: the optimization of radiotherapy treatment plans, which occurs using optimization algorithms implemented in the treatment planning systems (TPSs) used for the creation of radiotherapy treatment plans, and the patient-specific pre-treatment verification of the dosimetric and delivery accuracy of the resulting plans, which usually occurs prior to the actual plan delivery to the the patient. Without these two fundamental processes, external beam intensity-modulated radiotherapy treatments could not be delivered with the needed precision, accuracy and safety. In general, several human and technological resources are involved in these processes and a significant amount of time is required for their execution, leading to constraints on the number of patients that can receive radiotherapy each day. Therefore, any procedure or strategy aimed at improving the overall radiotherapy workflow efficiency should consider the two aforementioned steps.

One the one hand, the inefficient nature of the well established measurement-based patient-specific quality assurance (PSQA) can be

---

[1] Last available update: GLOBOCAN 2020

attributed to the requirement of administering every treatment plan on a dose detector system before it can be administered to patients. This procedure is crucial to ensure that the calculated dose distribution from the TPS aligns with the measured distribution within some clinical tolerance and action limits. This measurement-based procedure is the universally accepted standard, endorsed worldwide by professional organizations such as ESTRO and ASTRO. However, it invariably diminishes the machine time available for treating patients. In addition, when the dosimetric accuracy of a treatment plan does not fulfill the tolerance and action thresholds, it compels the need to return to the planning process. This not only degrades the efficiency of the process but also may cause delays in the treatment of the patient.

To reduce the typical workload caused by the measurement-based PSQA process, research teams worldwide have attempted to develop alternative not-measurement-based PSQA methodologies to evaluate the dosimetric and delivery accuracy of plans. Among the different approaches proposed, some of them rely on the so-called complexity metrics, mathematical indicators devised to identify factors of radiotherapy treatment plans that may increase the risk of poor dosimetric and delivery accuracy. In this context, several groups have applied artificial intelligence tools to create machine-learning and deep-learning models capable of predicting the dosimetric and delivery accuracy of treatment plans and provide support to the planner during this process. However, the lack of a shared consensus and evidence on which are the most effective complexity indicators to consider as well as the lack of commercial platforms for their computation is somehow slowing down their systematic adoption across institutions worldwide. The first part of this thesis, *The complexity of radiotherapy treatment plans*, focuses on this topic.

Prior to treatment verification, however, it comes the creation of the radiotherapy plan. In modern intensity-modulated radiotherapy, a treatment plan is the result of an optimization process called inverse-planning where hundreds to thousands of variables associated with the intensity of the radiation beams need to be optimized. It turns out that, the efficiency of this process is intimately related to the ability of the adopted optimization algorithms to explore the landscape of the associated cost-function. In clinic, this task is routinely tackled through optimizers implemented in the commercial TPSs available. However, the introduction of a novel computational paradigm, namely quantum computing, might give an unprecedented boost to the efficiency of this process in the future. Although large-scale quantum computers that can be used like classical calculators are absent, these novel machines are undergoing intense development by tenths of research groups and companies worldwide. Although different approaches to realisation are still under investigation, the basic working principle of using quantum bits (qubits) to process information has been well established and accepted across the expert community. Qubits behaviour is ruled by the laws of quantum mechanics, and this is what should provide quantum computers with unprecedented computational power compared to classical calculators. In fact, exploiting quantum features such as the superposition principle and the entanglement effect they will be endowed with inborn parallelization capabilities that even the most powerful classical supercomputers on Earth cannot match. This revolution will affect several sectors, including the simulation of many-body quantum systems, material science, chemistry, cryptography, and optimization.

It turns out that, exploiting the parallelization potential of future quantum computers in radiotherapy might improve the efficiency of the inverse-planning optimization process, possibly leading to better plans in the same time or allowing to implement more sophisticated functions capable to account for additional factors. Therefore, this might also impact the overall quality of the opimization process. However, there are two critical aspects that need to be addressed to achive this aim: establishing efficient strategies to map continuous variables onto the discrete system of qubits building up the quantum processors, and finding a suitable and effective representation of the initial cost function in terms of a quantum Hamiltonian whose ground-state corresponds to the minimum-cost configuration of the cost-function describing the inverse-planning optimization problem. In the literature, different strategies to solve classical combinatorial optimization problems on qubits-based hardware (e.g. quantum annealers) have been proposed, such as Quadratic Unconstrained Binary Optimization (QUBO) or Polynomial Unconstrained Binary Optimization (PUBO). However, such strategies are generally very expensive in terms of the number of qubits required, especially if continuous variables problems are considered. In fact, they are naturally designed for problems involving discrete-valued variables (e.g. binary) for whom a one-to-one (or one-to-few) mapping with the qubits in the quantum processor can be established quite straightforwardly. Therefore, new more efficient strategies are needed for more general continuous variables problems, especially with the aim of using them on near-term quantum devices that are expected to be equipped with limited numbers of qubits. These aspects lay at the core of the investigation presented in the second part of the thesis, *Towards the use of quantum computers in radiotherapy*.

The work proceeds as follows.

CHAPTER 1 conveys some basics elements of modern radiotherapy that are necessary to understand the research contents presented later on in the work. In particular, the following topics are discussed: intensity-modulated radiotherapy techniques, the inverse-planning optimization problem, the comparison between dose distributions, the measurement-based patient specific quality assurance and complexity metrics.

CHAPTER 2 provides a detailed and comprehensive overview of `UCoMX` (Universal Complexity Metrics Extractor), the software package developed within this PhD project for the extraction of complexity metrics from radiotherapy treatment plans. The package implements most of the complexity indicators proposed in the current literature and can handle DICOM RT-Plan files from several TPSs and treatment units. In this chapter, the complexity indicators implemented in the package are explained thoroughly.

CHAPTER 3 presents two original investigations on the complexity of Helical Tomotherapy treatment plans. First, a validation of the effectivness of the original complexity metrics developed within this doctoral project is proposed. In particular, their relation with PSQA results and plan efficiency is investigated. Second, the training and validation of machine-learning models for the prediction of the PSQA results is presented.

CHAPTER 4 presents an analysis of the complexity of VMAT plans where the relation between two complexity limitation strategies available in a commercial TPS and complexity metrics has been investigated. Furthermore, results on the definition of operational limits on some complexity metrics as decision-support tools for the reduction of PSQA workload is discussed.

CHAPTER 5 provides a theoretical introduction to many-body quantum systems and Tensor Network Methods. The latter are among the most powerful alternatives available to date for the simulation of quantum many-body systems and allow, among other things, to emulate the behaviour of quantum hardware thus configuring themselves as valuable test benches.

CHAPTER 6 presents the first application of Tensor Network Methods to a fixed-gantry IMRT fluence optimization problem. In particular, the formulation of the beamlets intensity inverse-planning optimization problem in terms of a qubits Hamiltonian is discussed. First, test on toy-models are presented. Second, a more realistic application to a simplified version of a prostate cancer treatment is proposed.

CHAPTER 7 presents a novel approach developed within this doctoral project for the optimization classical non-polynomial functions of continuous variables on quantum hardware (e.g. quantum computers). The core of this approach consists of a novel efficient strategy for the discretization of continuous variables that exploits a small

number of qubits. The validation of the developed strategy on different non-polynomial, non-convex optimization problems is presented, namely: unconstrained optimization and constrained optimization of the Gramacy-Lee function, and a fixed-gantry IMRT fluence optimization problem.

MOTIVAZIONE E STRUTTURA DELLA TESI

Il cancro è la principale causa di morte nel mondo, un fardello che colpisce ogni anno circa 19 milioni di persone e causa circa 10 milioni di morti[2]. Secondo l'Organizzazione Mondiale della Sanità (OMS), tra il 30% e il 50% dei tumori possono essere prevenuti evitando i fattori di rischio ad essi associati e attuando le strategie di prevenzione esistenti basate sull'evidenza. Ma quando la prevenzione fallisce, è necessario procedere al trattamento. In questo contesto, la *radioterapia* svolge un ruolo cruciale in molti casi. 'Radioterapia' è una parola composta che collega i due termini *radiazioni* e *terapia*. È una tecnica medica che sfrutta diversi tipi di radiazioni ionizzanti per uccidere le cellule tumorali nei pazienti affetti da cancro. È ben documentato che circa il 50% di questi pazienti richiede almeno un ciclo di radioterapia nel corso della loro storia di malattia, un valore che sale a 87% per il cancro al seno. Il tipico flusso di lavoro della radioterapia è un processo complesso che prevede la partecipazione di diversi professionisti. Tra queste, lo specialista in fisica medica è essenziale per ottenere una corretta gestione e attuazione pratica del trattamento. In particolare, secondo le normative internazionali e italiane, è tra le responsabilità dello specialista in fisica medica garantire l'utilizzo di strumenti tecnologici sufficienti e appropriati per il calcolo, il monitoraggio e la valutazione dell'accuratezza dosimetrica dei piani di trattamento.

Tra le diverse fasi che compongono il flusso di lavoro della radioterapia in cui lo specialista in fisica medica è attivamente coinvolto, due sono centrali in questa tesi: l'ottimizzazione dei piani di trattamento radioterapico, che avviene utilizzando un algoritmo di ottimizzazione implementato nel sistema di pianificazione del trattamento (TPS) adottato per la creazione dei piani di trattamento radioterapico, e la verifica pre-trattamento paziente-specifica dell'accuratezza dosimetrica e di erogazione dei piani risultanti, che di solito avviene prima dell'effettiva erogazione del piano al paziente. Senza questi due processi fondamentali, i trattamenti di radioterapia a fasci esterni erogati mediante tecniche ad intensità modulata non potrebbero essere effettuati con la precisione, l'accuratezza e la sicurezza necessarie. In generale, in questi processi sono coinvolte diverse risorse umane e tecnologiche e per la loro esecuzione è necessaria una quantità significativa di tempo, con conseguenti limitazioni del numero di pazienti che possono essere trattati ogni giorno. Pertanto, qualsiasi procedura o strategia volta a migliorare l'efficienza complessiva del flusso di

---

[2] Ultimo aggiornamento disponibile: GLOBOCAN 2020

lavoro della radioterapia dovrebbe prendere in considerazione anche le due fasi sopra descritte.

Da un lato, la natura inefficiente della verifica pretrattamento paziente-specifica (PSQA) basata su misure può essere attribuita alla necessità di somministrare ogni piano di trattamento su un sistema di rivelatori di dose prima di poterlo somministrare al paziente. Questa procedura è fondamentale per garantire che la distribuzione di dose calcolata dal TPS si allinei con quella misurata entro i cosidetti limiti di tolleranza e azione clinica. Questa procedura basata sulla misura è lo standard universalmente accettato, approvato e sostenuto in tutto il mondo da organizzazioni professionali come ESTRO e ASTRO. Tuttavia, riduce inevitabilmente il tempo macchina disponibile per il trattamento dei pazienti. Inoltre, quando l'accuratezza dosimetrica di un piano di trattamento non soddisfa le soglie di tolleranza e di azione, è necessario tornare al processo di pianificazione. Ciò non solo riduce l'efficienza del processo, ma può anche causare ritardi nel trattamento del paziente.

Per ridurre il carico di lavoro tipico del processo di PSQA basato sulle misure, gruppi di ricerca in tutto il mondo hanno cercato di sviluppare metodologie alternative di PSQA non basate sulle misure per valutare l'accuratezza dosimetrica e di erogazione dei piani. Tra i diversi approcci proposti, alcuni si basano sulle cosiddette metriche di complessità, indicatori matematici ideati per identificare i fattori dei piani di trattamento radioterapico che possono aumentare il rischio di scarsa accuratezza dosimetrica e di erogazione. In questo contesto, diversi gruppi hanno applicato strumenti di intelligenza artificiale per creare modelli di apprendimento automatico (machine-learning e deep-learning) in grado di prevedere l'accuratezza dosimetrica e di erogazione dei piani di trattamento e di fornire supporto al pianificatore durante questo processo. Tuttavia, la mancanza di un consenso condiviso e di prove su quali siano gli indici di complessità più efficaci da considerare, così come la mancanza di piattaforme commerciali per il loro calcolo, sta in qualche modo rallentando la loro adozione sistematica nelle istituzioni di tutto il mondo. La prima parte di questa tesi, *La complessità dei piani di trattamento radioterapico*, si concentra su questo argomento.

In una fase precedente alla verifica del trattamento, tuttavia, avviene la creazione del piano di radioterapia. Nella moderna radioterapia basata su tecniche a intensità modulata, un piano di trattamento è il risultato di un processo di ottimizzazione chiamato pianificazione inversa, in cui si vanno ad ottimizzare centinaia o migliaia di variabili associate all'intensità dei fasci radianti. L'efficienza di questo processo è intimamente legata alla capacità degli algoritmi di ottimizzazione adottati di esplorare efficientemente il panorama della funzione-costo associata al problema di pianificazione inversa. In clinica, questo compito viene affrontato nella routine quotidiana attraverso gli algoritmi di ottimizzazione implementati nei TPS commerciali. Tuttavia, l'introduzione di un nuovo paradigma computazionale, ossia l'informatica quantistica, potrebbe dare in futuro un impulso senza precedenti all'efficienza di questo processo. Sebbene non esistano computer quantistici su larga scala che possano essere utilizzati analogamente ai calcolatori classici, queste nuove macchine sono in fase di intenso sviluppo da parte di decine di gruppi di ricerca e aziende in tutto il mondo. Sebbene i diversi approcci alla realizzazione siano ancora in fase di studio, è condiviso all'interno della comunità degli esperti il principio di funzionamento di base per elaborare le informazioni che prevede l'impiego dei bit quantistici (qubit). Il comportamento dei qubit è regolato dalle leggi della meccanica quantistica e questo dovrebbe fornire ai computer quantistici una potenza di calcolo senza precedenti rispetto ai calcolatori classici. Infatti, sfruttando caratteristiche quantistiche come il principio di sovrapposizione e l'effetto entanglement, saranno dotati di capacità di parallelizzazione innate che nemmeno i più potenti supercomputer classici potranno eguagliare. Questa rivoluzione interesserà diversi settori, tra cui la simulazione di sistemi quantistici a molti corpi, la scienza dei materiali, la chimica, la crittografia e l'ottimizzazione.

Ne consegue che lo sfruttamento del potenziale di parallelizzazione dei futuri computer quantistici in radioterapia potrebbe migliorare l'efficienza del processo di ottimizzazione della pianificazione inversa, possibilmente portando a piani migliori nello stesso tempo o permettendo di implementare funzioni più sofisticate in grado di tenere conto di fattori aggiuntivi. Tutto ciò potrebbe avere un impatto sulla qualità complessiva del processo di ottimizzazione. Tuttavia, ci sono almeno due aspetti critici che devono essere affrontati per raggiungere questo obiettivo: stabilire strategie efficienti per mappare le variabili continue sul sistema discreto di qubit che costituiscono i processori quantistici e trovare una rappresentazione adeguata ed efficace della funzione costo iniziale in termini di una Hamiltoniana quantistica il cui ground-state corrisponda alla configurazione a costo minimo del problema di ottimizzazione in pianificazione inversa. In letteratura sono state proposte diverse strategie per risolvere problemi di ottimizzazione combinatoria classica su hardware basato su qubit (ad esempio, annealer quantistici), come la Quadratic Unconstrained Binary Optimization (QUBO) o la Polynomial Unconstrained Binary Optimization (PUBO). Tuttavia, tali strategie sono generalmente molto costose in termini di numero di qubit necessari, soprattutto se si considerano problemi a variabili continue. Infatti, esse sono naturalmente concepite per problemi che coinvolgono variabili a valore discreto (ad esempio binarie) per le quali è possibile stabilire in modo abbastanza semplice una mappatura uno-a-uno (o uno-a-poco) con i qubit del processore quantistico. Pertanto, sono necessarie nuove strategie più efficienti per problemi più generali di variabili continue, soprattutto con l'obiettivo di utilizzarle su dispositivi quantistici a breve termine, i quali si prevede saranno dotati di un numero limitato di qubit. Questi aspetti sono al centro dell'indagine presentata nella seconda parte della tesi, *Verso l'uso dei computer quantistici in radioterapia*.

La presente tesi si articola come di seguito riportato.

Il CAPITOLO 1 introduce ad alcuni elementi di base della moderna radioterapia, necessari per comprendere i contenuti della ricerca presentati più avanti nel lavoro. In particolare, vengono trattati i seguenti argomenti: le tecniche di erogazione ad intensità modulata, il problema di 'ottimizzazione in pianificazione inversa, il confronto tra distribuzioni di dose, il controllo qualità paziente-specifico pre-trattameto basato su misure e le metriche di complessità.

Il CAPITOLO 2 fornisce una panoramica dettagliata e completa di `UCoMX` (Universal Complexity Metrics Extractor), il software sviluppato nell'ambito di questo progetto di dottorato per l'estrazione di metriche di complessità dai piani di trattamento radioterapico. Il pacchetto implementa la maggior parte degli indicatori di complessità proposti nella letteratura corrente e può gestire file DICOM RT-Plan provenienti da diversi TPS e unità di trattamento. In questo capitolo, gli indicatori di complessità implementati nel pacchetto sono spiegati in modo approfondito.

Il CAPITOLO 3 presenta due indagini originali sulla complessità dei piani di trattamento realizzati mediante Tomoterapia Elicoidale. In primo luogo, viene proposta una validazione dell'efficacia delle metriche di complessità originali sviluppate nell'ambito di questo progetto di dottorato. In particolare, si va a studiare la loro relazione con i risultati di PSQA e di efficienza del piano. In secondo luogo, vengono presentati l'addestramento e la validazione di modelli di apprendimento automatico per la previsione dei risultati di PSQA.

Il CAPITOLO 4 presenta un'analisi della complessità dei piani VMAT in cui è stata studiata la relazione tra due strategie di limitazione della complessità disponibili in un TPS commerciale e le metriche di complessità. Inoltre, vengono discussi i risultati relativi alla definizione di limiti operativi su alcune metriche di complessità come strumenti di supporto decisionale per la riduzione del carico di lavoro associato al PSQA.

Il CAPITOLO 5 fornisce un'introduzione teorica ai sistemi quantistici a molti corpi e ai metodi di rete tensioriale. Questi algoritmi sono tra le più potenti alternative ad oggi disponibili per la simulazione di sistemi quantistici a molti corpi e permettono, tra le altre cose, di emulare il funzionamento dell'hardware quantistico configurandosi, quindi, come dei preziosi banchi prova.

Il CAPITOLO 6 presenta la prima applicazione dei metodi di rete tensoriale ad un problema di ottimizzazione della fluenza di piani IMRT a gantry fisso. In particolare, viene discussa la formulazione del problema di ottimizzazione dell'intensità dei beamlet in termini di una Hamiltoniana descrivente un sistema di qubit interagenti. In primo luogo, vengono presentati dei test su modelli giocattolo. Successivamente, viene proposta un'applicazione più realistica a una versione semplificata di un trattamento per cancro alla prostata.

Il CAPITOLO 7 presenta un nuovo approccio sviluppato nell'ambito di questo progetto di dottorato per l'ottimizzazione di funzioni di variabili continue su hardware quantistico (ad esempio, computer quantistici). Il cuore di questo approccio consiste in una nuova strategia efficiente per la discretizzazione di variabili continue che sfrutti un ridotto numero di qubit. Viene presentata la validazione di tale strategia su tre diversi problemi di ottimizzazione non-convessi e non-polinomiali, in particolare: ottimizzazione non vincolata e ottimizzazione vincolata della funzione di Gramacy-Lee e un problema non polinomiale di ottimizzazione della fluenza in piani IMRT a gantry fisso.

PUBLICATIONS

This is the full list of publications that I have contributed to during my doctoral career either as the first author or contributing author. They are reported in chronological order starting from the most recent one.

[1] Samuele Cavinato et al. "Prediction models as decision-support tools for virtual patient-specific quality assurance of helical tomotherapy plans." In: *Physics and Imaging in Radiation Oncology* 26 (2023), p. 100435. ISSN: 2405-6316. DOI: `https://doi.org/10.1016/j.phro.2023.100435`.

[2] Alessandro Scaggion et al. "Updating a clinical Knowledge-Based Planning prediction model for prostate radiotherapy." In: *Physica Medica* 107 (2023), p. 102542. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2023.102542`.

[3] Samuele Cavinato et al. "Quantitative assessment of helical tomotherapy plans complexity." In: *Journal of Applied Clinical Medical Physics* 24.1 (2023), e13781. DOI: `https://doi.org/10.1002/acm2.13781`.

[4] Pedrame Bargassa et al. "Quantum algorithm for the classification of supersymmetric top quark events." In: *Phys. Rev. D* 104 (9 2021), p. 096004. DOI: `10.1103/PhysRevD.104.096004`.

[5] Marco Fusella et al. "Analysis of clinical patient-specific pretreatment quality assurance with the new helical tomotherapy platform, following the AAPM TG-218 report." In: *Radiation Oncology* 16 (2021), p. 226. DOI: `10.1186/s13014-021-01952-w`.

[6] Samuele Cavinato et al. "Optimizing radiotherapy plans for cancer treatment with Tensor Networks." In: *Physics in Medicine & Biology* 66.12 (2021), p. 125015. DOI: `10.1088/1361-6560/ac01f2`.

Part I

THE COMPLEXITY OF RADIOTHERAPY TREATMENT PLANS

1

ELEMENTS OF MODERN RADIOTHERAPY

Radiotherapy boasts a long-standing story that has lasted since the very first observation of X-ray by Wilhelm Conrad Roentgen in 1896. However, despite the impressive quality, accuracy and precision of modern radiotherapy treatments, the pace of evolution of this cancer treatment modality is far from slowing down. On the one hand, this is due to the technological developments of recent years, including automation and artificial intelligence [1, 2], the increasing availability of clinical facilities for particle therapy [3], the evolution of medical imaging techniques with the development of novel MRI-linacs [4], and the spreading relevance of the so-called FLASH effect that has opened the possibility of fast and ultra-high dose rate treatments in the future [5]. On the other hand, evidence of biological benefits arising from the combination of radiotherapy with other cancer treatment modalities such as chemotherapy or immunotherapy [6–8] are pushing the boundaries of its applicability way further.

The content of this chapter will focus on the so-called external beam radiotherapy (EBT), which constitutes approximately 80% of radiotherapy applications globally [9]. More precisely, treatment with X-ray photons will be considered only, thus not covering those performed using electrons, protons or light ions. The chapter is structured as follows. In Sec. 1.1, the working principles of three intensity-modulated radiotherapy (IMRT) delivery methods, namely fixed-gantry IMRT, Volumetric Modulated Arc Therapy (VMAT) and Helical Tomotherapy (HT) will be presented. In Sec. 1.2, the process of verification of the dosimetric and delivery accuracy of radiotherapy treatments, also known as patient-specific quality assurance (PSQA) will be discussed. Finally, Sec. 1.3 will offer an introduction to the complexity of radiotherapy treatment, including a discussion on the so-called *complexity metrics*, briefly retracing their history and describing how they are related to the PSQA process and can contribute to improving both the efficiency and effectiveness of the radiotherapy workflow. At the end of this chapter, the reader will own all the necessary pieces of knowledge to access and understand the details of the research work presented in this thesis.

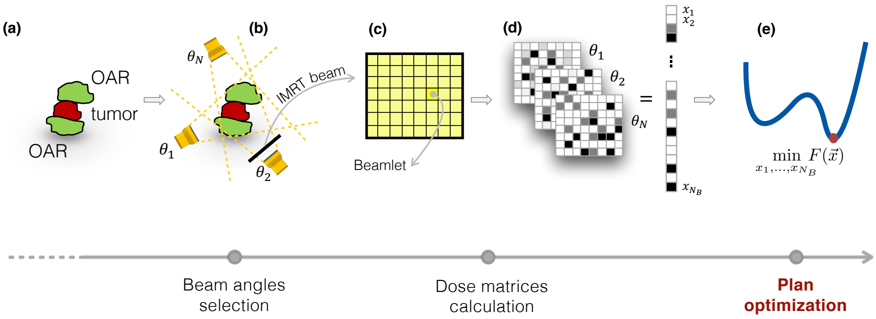


Figure 1.1: Schematics of the main steps involed in the realization of an IMRT treatment. FROM CAVINATO ET AL (2021) [15].

1.1 INTENSITY MODULATED RADIOTHERAPY TECHNIQUES

Based on the data contained in IAEA DIRAC database, roughly 80% of the radiotherapy facility installed globally is devoted to EBT [9]. In contrast with other delivery methods such as brachytherapy [10], EBT radiation is delivered from outside the patient's body. In the specific case of X-ray photons, this is done using a linear accelerator (LINAC). The delivery of these types of treatments is a sophisticated process where several healthy organs are generally involved in addition to the target tumor, and an optimal radiotherapy treatment should always account for delivering a sufficient amount of radiation dose to the target while avoiding damages to the sorrounding healthy tissues.

Today's standard-of-care delivery methods are represented by the so-called Intensity Modulated Radiotherapy (IMRT) techniques. Compared to older and less advanced 2D or 3D conformal radiotherapy treatments (3DCRT) which are not covered in the present introduction, IMRT allows reaching higher tumor dose conformity and better sparing of the surrounding healthy tissues at the same time, with increased local control and reduced local re-occurrence, toxicities and possible complications of treatments [11–13]. The two peculiar features of these types of treatments are the following:

1. The intensity (or *fluence*) of the radiation beams is not spatially uniform;
2. The final treatment results from an optimization process known as *inverse-planning* [14].

Figure 1.1 shows a schematics of the main steps involved in the realization of a radiotherapy treatment using IMRT techniques. The key fact in IMRT is that each radiation beam gets subdivided into a grid of hundreds of smaller beams called *beamlets* whose intensity can be tuned independently from each other allowing for non-uniform beam intensity profiles. The optimal set of beamlet intensity levels is the result of the associated inverse-planning problem, which is usually described by an optimization problem with following general formulation:

$$\vec{w}_{opt} = argmin_{\vec{w}} F(\vec{w}|\{A_1, \ldots, A_n\, C_1, \ldots, C_k\, \alpha_1, \ldots, \alpha_k\}) \tag{1.1}$$

where F is the optimization function describing the inverse-planning problem and $\vec{w}$ is the set of beamlet intensity levels, namely the optimization variables. The A_is are the so-called *influence matrices*, which contain the information on the dose delivered to a given structure (either tumor or surrounding organ) by each radiation beam per unitary values of the beamlet intensity levels. The entries of each A_i depend on several factors including the energy of the radiation beams, the anatomy and geometry of the patient and the geometry of the treatment, namely the arrangement of the radiation beams around the patient. The latter depends on the details of the adopted IMRT delivery technique. The C_is are the so-called *constraints*, namely the requests or limits in dose both for the tumor and the surrounding organs, while the α_ks are the associated priorities that represent the relative importance of each constraint during the optimization process. The result of such an optimization problem is here represented by $\vec{w}_{opt}$ which is the optimal set of beamlet intensity levels that best satisfies the constraints for a particular treatment. In Ch. 6 the mathematical formulation of the inverse planning problem in Eq. (1.1) will be discussed in more detail.

During the delivery of treatment, the local intensity modulation is obtained through a very peculiar mechanical collimator called multi-leaf collimator (MLC) consisting of pairs of highly absorbing tungsten leaves that can block the incident radiation and move against each other. Based on the specific IMRT delivery technique used, the construction details of the MLC may vary as it will be discussed later in this work [16]. IMRT delivery techniques can be divided into two main categories: fixed gantry and moving gantry. Fixed-gantry IMRT was the first proposed IMRT technique, where radiation is delivered only when the gantry is stationary while the MLC is allowed to move. In particular, the MLC can be operated in two modes: dynamic (or sliding window) and step-and-shoot (or stop-and-shoot). The former involves the continuous movement of MLC leaves from a specific field opening. The desired fluence map is achieved by either narrowing or widening the aperture or varying the leaf speed while the beam is on [17]. The fluence generated by a pair of opposing leaves A and B at point x is proportional to the time interval between their passage through x. Differently, in step-and-shoot mode, the field is divided into several sub-fields (or segments). The beam is switched off while the MLC moves from one segment to the next. In general, sliding-windows IMRT allows for producing more conformal dose distributions compared to step-and-shoot, but it results in less efficient delivery and increased leakage and total body dose [16].

In 2007, Karl Otto proposed a new delivery technique, which has become widely adopted globally across institutions: the Volumetric Modulated Arc Therapy (VMAT) [18]. This technique has evolved from the fixed-gantry IMRT method. Like other 'arc' therapies, such as tomotherapy (described in the following), VMAT treatments involve the patient being treated from a full 360° beam angle. During the treatment, the radiation source (i.e. gantry) moves continuously. Moreover, it allows for the simultaneous variation of three parameters, namely the dose rate, gantry rotation speed, and opening shapes of the treatment aperture through the movement of MLC leaves. In comparison to standard fixed-gantry IMRT, it enables enhanced delivery efficiency by reducing the delivery time and lowering the total number of MUs administered.

On the other side of the coin, Helical Tomotherapy (HT) is another advanced arc therapy technique available which differs consistently from VMAT [19]. During an HT treatment, beam fluence modulation is achieved through a pneumatically powered binary MLC that acts on a fan beam that rotates around the patient, while the couch is translated through the gantry at a constant speed. As for VMAT, thousands of beamlets are typically involved in this process, a fact that makes this technique very suitable to deliver a highly conformal dose to the PTV and optimal OARs sparing at the same time, making it a valid competitor of VMAT [20–22]. However, despite HT was developed before VMAT, it has had a less fortunate history. In fact, HT requires the installation of dedicated equipment which are not the same used for fixed-gantry IMRT, while the treatment units already available for the latter could be adapted to VMAT, too. Also thanks to this fact, the VMAT became predominant. This fact has also influenced the development of complexity indicators for HT, as it will be discussed in Sec. 1.3.

1.2 PATIENT-SPECIFIC QUALITY ASSURANCE (PSQA)

As discussed in Sec. 1.1, the dose distributions obtained using IMRT delivery techniques, either fixed and moving gantry, are generally more heterogeneous compared to those of 3D conformal radiotherapy plans, making use of more complex fields geometry and higher degrees of modulation. If on the one hand, this allows achieving more conformal treatments and improved patient outcome, procedures to assess the accuracy of calculated dose distributions and to detect clinically relevant errors that may occur during the radiation delivery constitute a necessary part of the IMRT process. To this purpose, different professional organizations around the globe (American Association of Physics in Medicine (AAPM), American Society for Radiation Oncology (ASTRO) , European Society for Radiation Oncology (ESTRO), etc.) have recommended the introduction of so-called *patient-specific quality assurance* (PSQA) programs as a needed and fundamental part of the IMRT process, also through the publication of reference guidelines for their implementation [23].

As discussed in Sec. 1.1, each IMRT plan is the result of an inverse-planning optimization process which is carried out for specifically for each patient. Therefore, also the process of quality assurance of such plans becomes patient-specific, and needs to be carried out for each treatment plan independently. Currently, PSQA programs are routinely performed at all clinics worldwide as a fundamental step of the radiotherapy workflow. This verification process is necessary to get patient-specific information on the dosimetric and delivery accuracy of each treatment plan and allows for the safe delivery of the treatment to the patient. However, it is a time-consuming and resource-intensive process. For this reason, several clinics worldwide have been trying to overcome this limitation by proposing alternative PSQA approaches to those currently adopted. This will be discussed in more detail later in this work and it is one of the main topics of the research activity presented in this thesis.

In general, two main sources of errors are aimed to be detected through this process: planning errors and delivery errors [23]. The former have to do with the TPS and includes innacuracies in the modeling of MLC leaf ends, tongue-and-groove effects, leaf/collimators transmission, collimators/MLC penumbra, and output factors for small field sizes. Conversely, delivery errors have to do with the inaccuracy of the treatment unit in delivering the treatment created with the TPS due to the finite precision of both mechanical systems and feed-back controls systems, and may include MLC leaf position errors, MLC leaf speed or acceleration errors, gantry rotational stability or beam stability. The ability to detect one or more sources of errors strictly depends on the particular PSQA methods adopted. Several approaches for PSQA have been proposed, including measurement-based, independent re-calculation of the dose, analysis of the treatment machine's log files, and, more recently virtual PSQA approaches. A brief introduction to some of these PSQA methodologies will be provided later in this chapter for what is ancillary to this work. However, before moving to that topic an introduction to another important concept is provided: the comparison of dose distributions, which is at the base of most of the commonly used PSQA approaches.

1.2.1 *Comparing dose distributions: the γ-index passing rate*

Broadly speaking and neglecting the specific implementation details of each PSQA methodology, most of the PSQA procedures adopted in clinics nowadays foresee the comparison between two 'doses': the one calculated using the treatment planning system and the one obtained using an external tool, either a phantom detector or an indepen-

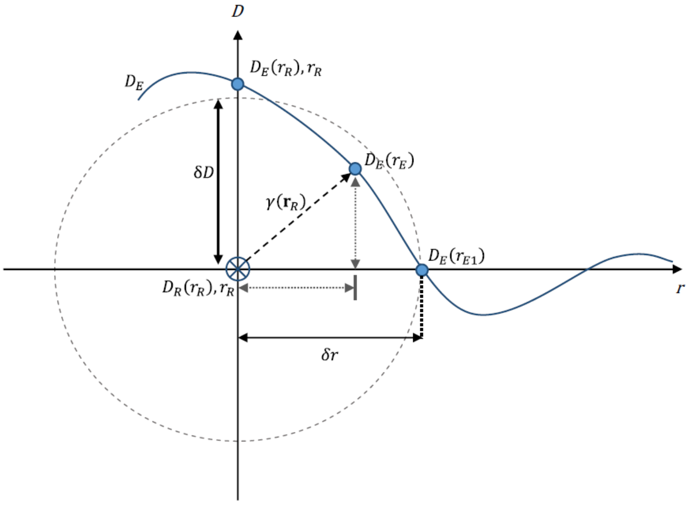


Figure 1.2: Geometrical representation of the *gamma*- index. ADAPTED FROM HUSSEIN ET AL (2017) [24].

dent calculation algorithm. Based on the methodology adopted, such comparison can involve either point-dose or 1D, 2D or 3D dose distributions, and focus on performing either dose difference or distance-to-agreement tests [25–27].

Each of the two approaches has its own advantages and disadvantages [23]. However, within the scope of this work, this section focuses on the most commonly adopted measure of agreement between dose distributions which tries to gather the two types of tests into a unique one: the γ-passing rate which was introduced by Low and colleagues [28, 29] and whose use is suggested by the report of the AAPM Task Group 218 [23].

Let's consider two dose distributions and call them the *evaluated* (E) and *reference* (R) dose distributions, respectively. In most clinical applications, the first one corresponds to that calculated with the TPS used the create the treatment plan, while the second one that obtained using the external PSQA tool. However, this is not a strict rule and different conventions might be adopted across institutions [REFs from TG218]. Then, based on the two aforementioned distributions, the following quantity is defined:

$$\Gamma(\vec{r}_E, \vec{r}_R) = \sqrt{\left(\frac{\delta(\vec{r}_E, \vec{r}_R)}{\Delta D}\right)^2 + \left(\frac{r(\vec{r}_E, \vec{r}_R)}{\Delta d}\right)^2} \tag{1.2}$$

with $\delta(\vec{r}_E, \vec{r}_R)$ the dose difference between an evaluated and a reference point, and $r(\vec{r}_E, \vec{r}_R)$ their distance. The term ΔD denotes the so-called dose difference (DD) criterion which is usually expressed as a percentage of a dose value chosen as reference. In particular, if a *global* normalization is used, the reference dose value is the same for all pairs of points. Typical values used in this case are the maximum dose value in one of the two distributions, the prescription dose, or the value of the evaluated dose distribution located in a specific point of the reference one. Conversely, when *local* normalization is used, for each pair of points the dose value of the point of the evaluated dose distribution is used. The term Δd indicates the so-called distance-to-agreement (DTA) criterion and it is usually expressed in millimeters.

The γ-index is defined for each point of the reference dose distribution R as the minimum displacement computed considering all the points of the evaluated dose distribution E in terms of $\Gamma(\vec{r}_E, \vec{r}_R)$, namely:

$$\gamma(\vec{r}_R) = \min\{\Gamma(\vec{r}_E, \vec{r}_R)\}\forall\{\vec{r}_E\} \tag{1.3}$$

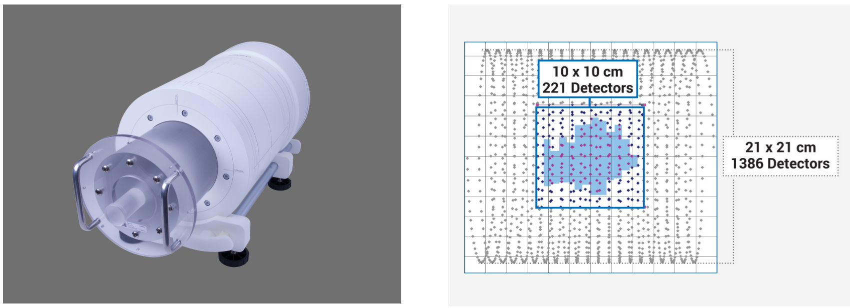
Points with $\gamma(\vec{r}_R) \in [0, 1]$ satisfy both the DD and DTA criteria. Conversely, $\gamma(\vec{r}_R) > 1$ indicates a failure to fulfill either one or the other criterion.

The γ-passing rate (PR$_\gamma$%) is the percentage of points passing the γ-test. A geometrical interpretation of the γ-index formulation can be provided as it shown in Fig. 1.2. in fact, Eq. (1.2) forms an ellipse whose x and y semi-axes coincide with the DTA and DD criteria. Points falling inside the ellipse pass the γ-test, while those falling outside fail it. In general, in the computation of the PR$_\gamma$% points with a dose smaller than a user-selected value, TH, are not included in the analysis. This helps the user to focus on clinically relevant doses. Typical settings for γ-index computation are 3% or 2% DD combined either with 2 mm or 3 mm DTA and $TH = 5 - 10\%$. Both local and global normalization are used. However, other choices might be found in the literature, which might also depend on the specific detector used in case of measurement-based PSQA.

Following the most recent clinical guidelines contained in the AAPM TG-218 report [23], *action* and *tolerance* limits can be defined on the number of points passing the γ test as criteria to evaluate the accuracy of delivery of radiotherapy treatment plans. Based on the definition provided in [23], action limits are defined as 'the amount the quality measures are allowed to deviate without risking harm to the patient' or, similarly 'limit values for whom clinical action is required'. On the other side, tolerance limits are defined as 'the boundaries within a process is considered to be operating normally, that is, subject to only random errors'. In the same document, universal action and tolerance limits are suggested on the PR$_\gamma\%(3\%, 2mm)$ with global normalization as 90% and 95%, respectively. However, center-specific action and tolerance limits can be defined using statistical process control techniques [30] as it is also discussed in App.A.

1.2.2 *Measurement-based PSQA: the* `ArcCheck`® *phantom*

Among all the alternatives available for pre-treatment PSQA of IMRT plans, those based on measurements are the current gold standard



(a)

(b)

Figure 1.3: *(a)* The `ArcCheck`® phantom with the `CavityPlug`™ homogeneous PMMA cylinder. *(b)* Helical grid detector arrangement of the `ArcCheck`® phantom. ADAPTED FROM SUNNUCLEAR.COM.

adopted at all institutions worldwide. The most common measurement-based methods are three: true composite (TC), perpendicular field-by-field (PFF), and perpendicular composite (PC). However, based on the tools used for the analysis presented later on in this thesis, the present introduction focuses on TC, only.

Each one of the aforementioned approaches has its own peculiarity; however, three steps are foreseen for all of them. First, the plan's dose distribution is recalculated onto the detector geometry whose 3D images (i.e. CT scan) are available within the TPS. Second, the plan is delivered to the detector using the same treatment machine that will be used also for treating the patient. The radiation beams are delivered to the detector using the same treatment parameters as for the patients, including MLC and JAW positions, Monitor Units (MUs) and gantry angles, and the radiation dose delivered to the detector is collected. Third, the measured dose distribution is compared to the calculated one as described in Sec. 1.2.1.

The main advantage related to the use of TC methods is that they simulate the delivery of the treatment to patients. This is usually done by means of phantoms equipped with matrices of radiation detectors. In particular, they allow measuring inaccuracies of the gantry, collimators, couch angles, and MLC leaf positions with gantry angle (the so-called gravity effects). Additionally, the dose distribution obtained is very close to the one that would be delivered to the patient, making the comparison simpler and more significant from a clinical point of view [23].

At the institution where the research project was carried out, the PSQA TC measurement method adopted is based on the `ArcCheck`® detector [31]. This phantom detector consists of 1386 n-Si diodes with dimensions 0.8×0.8 *mm* arranged in a helical shape at 3 *cm* depth along the long-axis of a cylindrical phantom made of PMMA. The diodes are placed at 1 *cm* center-to-center distance and measure both exit and entrance doses during the delivery of treatment. The cylinder is 21 *cm* long and has a diameter of 21 *cm*. The cylinder has a central cavity that can be filled using the PMMA `CavityPlug`™. In Fig. 1.3 a picture of the `ArcCheck`® phantom and a schematic of the arrangement of the detectors is shown.

1.3 COMPLEXITY METRICS

In Sec. 1.1 three modern IMRT delivery techniques have been described. If the use of these techniques allows, on the one hand, improved target volume conformity, maximum normal tissue sparing, dose escalation with potential patient benefits due to a reduction of acute and late toxicities, on the other hand, there are some side effects to consider. For instance, the high degree of dynamic modulation of machine parameters allowed places high demands on both the treatment machines and the TPSs, a fact that may affect the overall dosimetric and delivery accuracy of the resulting treatments [32].

In the literature, the collection of factors that may contribute to this phenomenon is commonly referred to as plan complexity [33, 34], an idea based on the concept that treatment plans that are *too complex* would result in significant discrepancies between the delivered and calculated dose distributions. This could be due to either inaccuracies in dose calculation by the TPS or errors in treatment delivery. In this context, several mathematical indicators defined from machine parameters or radiotherapy plan properties have been proposed to quantify such factors and link them to the results of the pre-treatment PSQA process. These are known as *complexity metrics*.

1.3.1 *Complexity metrics for fixed-gantry IMRT and VMAT*

Historically, the first papers proposing the use of complexity metri+cs were published in the early 2000s [35–39]. The metrics presented in those works were derived from the 2D fluence map of IMRT treatment plans and, therefore, they were later categorized among the so-called *fluence metrics* by several authors [33, 34, 40]. Thanks to their TPS- and machine-agnostic definition, they can be generalized to VMAT, too. However, they are affected by a lack of sensitivity to fluence-map degeneracy, namely, they cannot recognize whether a fluence map has been generated by a single large beam or by a combination of small beams [41]. Additionally, they do no provide any information on TPS or machine-related characteristics. A list of the main fluence metrics proposed in the literature is reported in Tab. 1.1.

Table 1.1: List of the main fluence metrics proposed in the literature. Adapted from Antoine et al (2019) [34] and Chiavassa et al (2019) [33]

Name	Description	Main reference
Fluence map complexity -*FMC*	The FMC is a normalized root sum over the local differences between bixel values and their two neighbours.	Llacer et al [38]
Modulation index - *MI*	It quantifies the variations of photon fluence along one direction between neighboring pixels in the fluence map including a threshold defined as a fraction of the standard deviation in the beam	Webb [36]
2D modulation index - 2*D MI*	Generalization of the MI which considers changes along x, y and diagonal directions	Nicolini et al [42]
Maximum intensity ration - *MIR*	It considers the maximum intensity allowed for each beamlet during the optimization process	Coselmon et al [39]
Plan intensity map variation - *PIMV*	It considers the intensity difference between neighbour beamlets along the x, y and diagonal directions.	
Fractal dimension analysis	Analysis of three fractal dimensions: variation, power spectrum and variogram methods.	Nauta et al [43]
Angular Second Moment - *ASM*	Metric based on the GLCM that indicates a measure of the homogeneity of a fluence map	Park et al [44]

Continued on next page

Table 1.1: List of the main fluence metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33] (Continued)

Name	**Description**	**Main reference**
Inverse Difference Moment - *IDM*	Metric based on the GLCM that indicates a measure of the local homogeneity of a fluence map	
Contrast - *CTR*	Metric based on the GLCM that indicates a measure of the local variation in a fluence map	
Variance - *VAR*	Metric based on the GLCM that indicates a measure of the inhomogeneity of a fluence map	
Correlation - *COR*	Metric based on the GLCM that indicates a measure of the linear dependency of gray levels in a fluence map	
Entropy - *S*	Metrics based on the GLCM that indicates a measure of randomness of a fluence map	

With the advent of the VMAT technique, several groups around the world began to develop new complexity metrics, each focusing on different aspects of treatment plans, also in order to overcome the limitations of existing fluence metrics. The proposed metrics can be divided into two additional categories, as is generally accepted [33, 40]:

- Deliverability: metrics based on the idea that the ability of treatment machines to deliver treatments as planned is affected by variations in machine mechanical and dosimetric parameters [35, 41, 42, 45–56]. A list of the main deliverability metrics is reported in Tab. 1.2.
- Accuracy: metrics aimed at quantifying challenging MLC configurations that may compromise dose calculation accuracy due to machine modelling and algorithm inaccuracies in the TPS [37,

40, 41, 43, 57–63]. A list of key accuracy metrics is provided in Tab. 1.3.

For the sake of completeness, it should be noted that some metrics fall between the two categories mentioned above because of their definition. It is for this reason that other authors have preferred to merge the two groups into a single one, called the degree of freedom variation [34]. Overall, more than fifty complexity metrics for IMRT/VMAT have been proposed in the literature including fluence, accuracy, and deliverability metrics.

Table 1.2: List of the main deliverability metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33]

Name	**Description**	**Main reference**
Monitor Unit - *MU*	Total number of monitor units of the plan	Du et al [41], Mohan et al [35], Masi et al [47]
Monitor unit per cGy - *MUcGy*	Total number of monitor units of the plan normalized by the fraction dose	
Plan normalized monitor units - *PMU*	Total number of monitor units of the plan normalized to a single fraction of 2 *Gy*	
Plan averaged beam irregularity -*PI*	Field aperture irregularity per CP, by considering the non-circularity of the aperture area	Du et al [41]
Plan averaged beam modulation - *PM*	Field aperture area per CP by considering the union area of all aperture areas of a beam	
Aperture area variability - *AAV*	Field area variation from a maximum area	McNiven et al [48]
Leaf sequence variability - *LSV*	Field irregularity comparing adjacent leaf positions	

Continued on next page

Table 1.2: List of the main deliverability metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33] (Continued)

Name	**Description**	**Main reference**
Modulation complexity score - *MCS*	Product between the *AAV* and *LSV*	
Dose rate variation - *DRV*	Variation of the norminal dose rate	Nicolini et al [42]
Gantry speed variation - *GSV*	Variation of the nominal gantry speed	
Gantry angle per MU - deg/MU	Average distance traveled by the gantry per MU delivered	Miura et al [49]
Leaf travel per MU - mm/MU	Average distance travel by the MLC leaves per MU delivered	
Monitor units per control point - MU/CP	Average number of MU delivered at each CP	Shen et al [50]
MU < 3 - $\%MU/CP < 3$	Proportion of CPs with MU < 3	
S_{l-h} and $A_{l,h}$	Proportion of leaf speeds and leaf accelerations from a given range, respectively	Park et al [52]
Modulation index for speed of MLC - *MIs*	MLC leaves speed modulation between different control points	Park et al [51]
Modulation index for speed and acceleration of MLC - *MIa*	MLC leaves speed and acceleration modulation between different control points	
Total modulation index - *MIt*	Modulation of speed and acceleration of the MLC, gantry acceleration and dose rate variation	

Continued on next page

Table 1.2: List of the main deliverability metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33] (Continued)

Name	Description	Main reference
Modulation complexity score for VMAT plans - *MCSv*	Adaptation of the MCS designed for IMRT to VMAT	Masi et al [47]
Leaf travel - *LT*	Entire distance traveled by each leaf over the VMAT arc averaged over all in-field moving leaves	
Leaf travel modulation complexity score - *LTMCS*	Combination of LT and MCSv	
Modulation index for station parameter optimized radiation therapy - MI_{SPORT}	Modulation of MLC around a given CP considering a certain range of adjacent CPs, weighted by the corresponding MU per gantry angle	Li and Xing [53]
Delivery complexity - *DC*	Sensitive to the Monitor Units and the total number of segments of an IMRT plan	Anker et al [54]
Adapted modulation complexity score - *oMCS*	Combination of aperture area variability and sectored leaf sequence variability, which considers a specific organ located in the field, per segment for an IMRT beam	Sumida et al [55]

Continued on next page

Table 1.2: List of the main deliverability metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33] (Continued)

Name	Description	Main reference
MLC leaf velocity - MLC_{velo}	Ratio between the distance travelled by an active MLC leaf between two consecutive CPs and the time between two consecutive CPs	Agnew et al [56]

Table 1.3: List of the main accuracy metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33].

Name	Description	Main reference
Average leaf gap -ALG	Average distance between opposing leaf pairs	Nauta et al [43]
Mean field area -MFA	Segment areas without making a distinction between single and split fields into a given segment and between line-like and circular-like fields	Kairn et al [57], Crowe et al [40]
Small aperture Score - $SAS(x)$	Proportion of apertures smaller than x	
Closed leaf score - CLS	Proportion of closed leaf within jaw field	
Cross-axis score - CAS	Proportion of leaf pairs with leaf crossing central axis	
Mean asymmetry distance - MAD	Distance between the leaf pair center and the central axis	
Segment area per control point - SA/CP		

Continued on next page

Table 1.3: List of the main accuracy metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33]. (Continued)

Name	Description	Main reference
Modulation degree - *MD*	Inverse of the sum over all segments of the segment area multiplied by the segment MU	Heijmen et al [58]
Plan averaged beam area -*PA*	MLC aperture area weighted by the MU	Du et al [41]
Segment area/Perimeter - *SA*/*CP* or Circumference/area - *C*/*A*		Carlsson et al [59], Godsted et al [60]
Edge metric - *EM*	Ratio between MLC perimeter and area with different weight on cross-axis and cross-leaf perimeter	Younge et al [61]
Edge area metric - *EAM*	Relative amount of edge region for the MLC aperture	Godstedt et al [60]
Converted aperture metrics *CAM*	Field aperture per CP by considering the distance between the MLC leaves in both X and Y directions	
Leaf offset impact on calculation - *LOIC*	PTV mean dose with respect to a change in the leaf offset parameter of the machine model	Mathot et al [62]
Average leaf pair opening - *ALPO*	Ratio between the sum of the aperture area and the sum of the fractional MU during which a leaf pair is open.	Zygmanskin et al [37]

Continued on next page

Table 1.3: List of the main accuracy metrics proposed in the literature. ADAPTED FROM ANTOINE ET AL (2019) [34] AND CHIAVASSA ET AL (2019) [33]. (Continued)

Name	**Description**	**Main reference**
Tongue-and-groove - *TG*	It is a measure of the MLC aperture irregularity	Yao and Farr [63]
Equivalent Field Size - *EFS*	Side of a square field equivalent to a rectangular one	Wolfs et al [64]

1.3.2 *Complexity metrics for HT*

Although HT was introduced into the clinic before the advent of VMAT, the development of complexity metrics for this delivery technique followed a much slower evolution. There are at least two main reasons behind this fact: first, the adoption of HT units in clinics worldwide was less fortunate than that of standard C-arm Linacs used for fixed-gantry IMRT and VMAT, with several clinics that found it more convenient to adapt the existing units for fixed-gantry IMRT to VMAT instead of installing a new HT platform. Second, the extension of the metrics designed for fixed-gantry IMRT to HT is not straightforward because of their different delivery modalities and characteristic of the MLC (see Sec. 1.1).

Most of the studies present in the literature focused their efforts on determining the role of typical HT delivery parameters such as the modulation factor (MF), the maximum and mean leaf opening time (maxLOT and mLOT, respectively), and the pitch as possible indicators of complexity [65, 66]. Other authors have investigated the effect of individual leaf latencies on the dosimetric accuracy of plans with a high proportion of small LOTs or LOTs close to the projection time [67, 68]. In general, the effort of all those papers was aimed at defining clinical practices for the planner to get treatment plans with both high efficiency and delivery accuracy without affecting the clinical quality.

In 2020, Santos and colleagues [69] proposed a set of complexity metrics specifically designed for HT. The authors proved that those metrics provide a quantitative description of the differences among three categories of treatments that were known in advance to exhibit qualitative differences in the plans' sinograms (i.e., stereotactic treatments of the brain, prostate plans, and head and neck plans). Some of the indexes showed a mild correlation with PSQA results of stereotactic brain plans. The work done by Santos and colleagues provided a first significant step forward in the characterization of the complexity of HT plans and it was one of the starting points of the research work presented in this thesis for what HT is concerned.

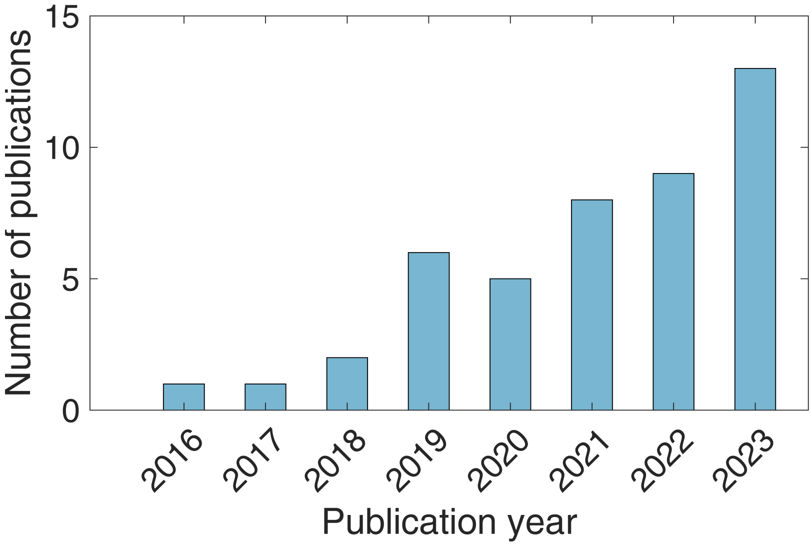


Figure 1.4: Number of publications on the use of Arfiticial Intelligence tools for PSQA since 2016.

1.3.3 *Complexity metrics and PSQA*

The development of complexity indicators is closely linked to enhancing the efficiency, accuracy and safety of the PSQA process. On the one hand, the currently implemented measurement-based PSQA verification has been identified as a bottleneck in the radiotherapy workflow. This is due to it being a resource-intensive and time-consuming process, as discussed in the previous pages. Therefore, the implementation of robust and reliable complexity indicators that can identify factors that may affect the accuracy of radiotherapy plans would improve this process. For instance, establishing operational boundaries on such metrics, to be utilised as "traffic lights," could enable the early detection of plans that may result in potential PSQA failures. This approach would offer a dual advantage: firstly, more attention can be given during the planning stage to those plans expected to be more crucial; secondly, the measurement-based PSQA can be avoided for plans falling within a safe zone determined by previously authenticated complexity indicators.

On the other hand, it is important to consider that the commonly used discrepancy measures the PSQA relies on (i.e. DD, DTA, $PR_\gamma\%$) are subject to various limitations. Therefore, it is also imperative to investigate the use of more sensitive and precise tools capable of provide more accurate results in terms of error detection to guarantee the highest safety for the patient being treated.

In the last years, this research area has gained increasing interest and attention thanks to the constantly increasing availability of artificial intelligence tools. Several groups around the globe has applied well-grounded machine- and deep-learning strategies to create models capable of combining the information provided by different types of input information, including complexity metrics, to predict the results of the PSQA process or to enhance the capability of detecting several types of dosimetric and delivery errors. All those approaches go under the name of *virtual* PSQA (vPSQA).

Based on a bibliographic research performed on PubMed, more than 40 papers have been published on this topic since 2016. Figure 1.4 shows the temporal trend of the number of publications per years on this topic. In general, approximately 80% of the investigations proposed in the literature focused on classification [70–84] or regression [85–97] tasks as well as combinations of the two approaches [98–103] aimed at predicting the PSQA results either in terms of $PR_\gamma\%$ or statistics of the gamma distributions (i.e. mean γ), or at detecting and identifying several type of errors (e.g.random errors, mechanical/positioning errors or MUs errors). This was done using well-established machine and deep-learning algorithms including Logistics Regression, Poisson regression, Support Vector Machines (SVMs), tree-based learners (i.e. AdaBoost, Random Forest, XGBoost, etc.), Feed-Forward Neural Networks (FFNNs) and Convolutional Neural Networks (CNNs). The remaining papers focused on different tasks, including quality assurance-based optimization as well as the generation of synthetic gamma maps or measured dose maps [104–109] also thanks to the most recents development coming from the field of artificial intelligence such as Variational AutoEncoders (VAEs), Generative Adversarial Networks (GANs) and Transformers.

More than half of the examined paper used either complexity metrics or a combination of the latter with other types of variables, such as DVH's metrics, radiomics features, dosiomics features or deep-features automatically extracted using deep-neural network based models. Most of the other studies focused on input variable different from complexity metrics as well as on different input information such as fluence maps, gamma maps or dose-difference maps.

The interest in this research area is rapidly increasing over the years and this topic is expected to become central in clinical radiotherapy in the near future. However, all the developments and investigation coming from literature have to do with the IMRT/VMAT technique. Later in this work, in Ch. 3 the first and only example of the realization of a vPSQA model for HT will be presented.

2

THE UCOMX SOFTWARE PACKAGE

In Ch. 1, an introduction to complexity metrics was provided and their role in the PSQA process was delineated. Establishing a robust set of complexity indicators would possibly allow for efficient assessment of plan delivery and dosimetric accuracy, ultimately driving standardization of the optimization approaches and reducing the workload associated with pre-treatment PSQA. This calls for the necessity of shared and valiated tools devoted to the evaluation of treatment plan complexity.

THIS CHAPTER presents `UCoMX (Universal Complexity Metrics Extractor)` a free software package named that has been developed within this PhD project. In a context where the interest in plan complexity is constantly increasing, it is crucial to lay the foundations for shared tools for its evaluation. As reported in Tab. 2.1 `UCoMX` is platform-agnostic software that has been validated on most of the common commercial combinations of TPSs, treatment units, and MLCs available on the market. It includes metrics either for fixed-gantry IMRT, VMAT, and HT, a fact which makes it a universal tool.

Table 2.1: List of the combinations of treatment unit, TPS and MLC used to test the functioning of `UCoMX`.

Availability	Model	MLC	Number of leaves	TPS
In-house	TrueBeam	HD120	120	Eclipse, RayStation
	Unique	Millenium 120	120	Eclipse, RayStation
	Halcyon-Ethos	Dual-layer stacked staggered	114 (56-58)	Eclipse, Ethos
	RadixAct	Binary	64	RayStation, Eclipse

Continued on next page

Table 2.1: List of the combinations of treatment unit, TPS and MLC used to test the functioning of `UCoMX`. (Continued)

Availability	**Model**	**MLC**	**Number of leaves**	**TPS**
Other	TrueBeam	Millenium 120	120	Eclipse
	Axesse	Agility	160	RayStation
	Sinergy	MLCi2	80	RayStation, Monaco, Pinnacle
	MRIdian	Dual-layer stacked staggered	138 (60-70)	ViewRay

Two extraction engines build up the core of `UCoMX`: `VCoMX (VMAT Complexity Metrics Extractor)`, designed for fixed-gantry IMRT and VMAT plans, and `TCoMX (Tomotherapy Complexity Metrics Extractor)`, suitable for HT plans. They are described in details in the following sections. The platform was developed on `MATLAB`® `R2020b (The MathWorks Inc, Natick, MA, USA)` and can be downloaded from `10.5281/zenodo.8276838`. Two different versions are available:

- **`MATLAB`®`-based version`**, which requires a valid `MATLAB`® licence installed.
- **`Stand-alone compiled version`**, which works without the need of `MATLAB`®.

Both the two version are equipped with a Graphic User Interface (GUI) which allows for a simplified experience for the user. However, the `MATLAB`®-based version's GUI might not be compatible with releases older than R2020a. Dedicated text files are made available to the user to set the input and overcome this compatibility issue. The functioning of `UCoMX` is documented in the related user manual which can be found in App. **??**. A brief overview is provided in Sec. 2.3. In Sec. 2.1 and 2.2 the IMRT/VMAT and HT complexity metrics are described.

2.1 VCOMX

The VMAT/IMRT complexity metrics implemented in the VCoMX extraction engine were selected from those proposed in the literature and well summarized in Refs. [33] and [34] as well as in Tab. 1.1, 1.2, 1.3 of Ch. 1. For consistency with the existing literature, the metrics were subdivided into categories according to what was proposed by Crowe [40] and, later, by Chiavassa and colleagues [33] as it was already discussed in Sec. 1.3. A few additional metrics developed during this project have also been included. In total, 43 complexity metrics are currently implemented in the `VCoMX` engine.

In total, 26/40 metrics were taken from either Ref.[33] or Ref.[34]. The remaining 16 were taken from the other relevant literature presented in Ch. 1. Fifteen metrics contained either in [33] or [34] were excluded due to their strong similarity to other ones. However, in a future version of the package they will be included for sake of completeness. The detailed description of the implemented metrics is provided in Sec. 2.1.4. However, before moving to that, in Sec. 2.1.1, 2.1.2 and 2.1.3, a description of the general approach used for their computation is provided.

2.1.1 *Control points, control arcs and segments*

In the DICOM RT-Plan files of VMAT treatments, each delivery arc is divided into several sectors called control arcs (CAs), each of which has extremals called control points (CPs). For each CP, the MLC leaves positions, jaws positions, and other relevant pieces of information are stored. The total number of CPs in a complete arc depends on various factors, such as the arc length, the use of the 'dual-arc' option available on some TPSs, and the angular distance between adjacent CPs which also varies depending on the specific TPS used. In the `VCoMX` implementation, all the values are initially mapped from the extremals of the control arcs to their centres to account for the continuous VMAT delivery across CPs Figure 2.1 provides a visual representation of both CPs and CAs. Regarding the structure of DICOM RT-Plan files, fixed-gantry IMRT (either step-and-shoot or sliding window) is a special case of VMAT, where 'arcs' become 'fields', and CPs become 'segments'. For the rest of this document, we will use the terminology that is based on CAs/CPs. following replacements:

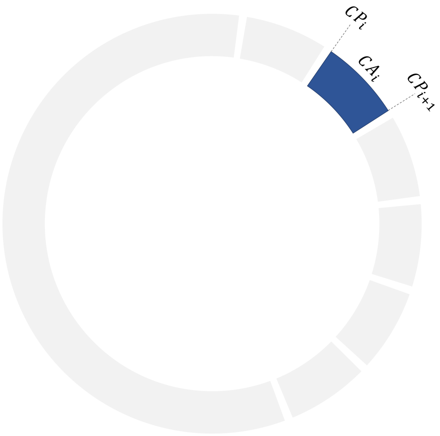


Figure 2.1: Schematic representation of CAs and CPs in VMAT treatments.

- arcs → fields
- control arcs → field's segments

In the following, the nomenclature based on arcs and CAs will be used.

2.1.2 *Active control arcs and active leaves pairs*

During VMAT treatment delivery, there may be portions of arcs where radiation is interrupted. To address this, beam-on CAs are defined as those where a non-zero number of Monitor Units (MUs) are delivered. The current version of `VCoMX` only considers active arcs, while inactive ones are ignored. Furthermore, the active leaf pairs are defined at each CA based on two criteria. (1) Facing leave pairs must form a gap greater than the minimum possible gap set on the TPS, which is derived directly from the DICOM RT-Plan file; (2) active leaves tip must not be covered by the jaws. Both two criteria must be fulfilled for a leaf pair to be considered in computing the metrics.

2.1.3 *Plan value and notation used*

Metrics are computed at each active CA of each arc. In the current version, the plan value (i.e. the final value of the metric reported in the output table) is generally calculated as the average over all arcs and CAs. More precisely, given a metric m_{ij} computed at each CA j and arc i, the plan value is computed as:

$$m = \frac{1}{\sum_i NCA_i} \sum_{i=1}^{N_A} \sum_{j=1}^{NCA_i} m_{ij} \tag{2.1}$$

where N_A and NCA_i denote the number of arcs and corresponding active CAs. Additionally, for some metrics a weight corresponding to the relative number of monitor units delivered might be also included as:

$$m = \frac{1}{\sum_i NCA_i} \sum_{i=1}^{N_A} \sum_{j=1}^{NCA_i} m_{ij} \frac{MU_{ij}}{MU_{tot}} \tag{2.2}$$

where MU_{ij} is the number of monitor unit delivered during the $j-th$ control-arc of the $i-th$ arc, and MU_{tot} is the total number of monitor unit delivered during the plan. The latter is particularly useful in the case of geometrical metrics describing the shape, dimensions and irregularity of the beam eye view (BEV).

Consistently with this description, the notation used in the following pages is the following:

- The index i runs from 1 to the number of plan arcs N_A;

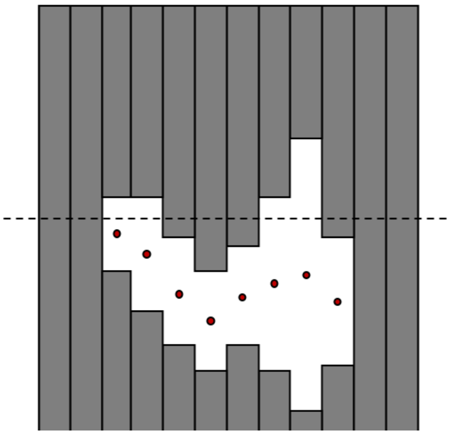


Figure 2.2: Schematic example of the asymmetry distance. The red points indicate the centroid of the leaf gaps. The MAD for each segment is computed as the average of distance of the red points from the central axis of the MLC (dotted line).

- the index j runs over the active CAs of a given arc i, namely from 1 to NCA_i;
- the index l runs over the leaves of each MLC bank, namely from 1 to $NL_{ij}/2$;
- $x_{ijl}^{A(B)}$ and $y_{ijl}^{A(B)}$ indicate the positions of the leaves with respect to the center of the beam in the cross-leaf and along-leaf direction, respectively.
- In some case, the index $k \in [1; NCP]$ will be used to indicated CPs. NCP is the number of CPs. In particular, CA j is found between CP k and $k+1$.

2.1.4 *Complexity metrics: accuracy*

`Leaf gap (LG).` This metric measures the distance between each pair of opposing at each arc and CA. It is defined as:

$$LG_{ijl} = (y_{ijl}^{A} - y_{ijl}^{B}) \tag{2.3}$$

The plan value is called *average leaf gap* (ALG).

`Small aperture score (SASn).` This metric was first introduced by Crowe and colleagues [40]. It measures the fracion of leaf gaps smaller than a given threshold, and it is computed on the whole plan as:

$$SAS_n = \frac{1}{\sum_{ij} NL_{ij}/2} \sum_{i=1}^{N_A} \sum_{j=1}^{NCA_i} \sum_{l=1}^{NL_{ij}/2} [LG_{ij} < n] \tag{2.4}$$

where $\left[LG_{ij} < n\right]$ is 1 if $LG_{ij} < n$, 0 otherwise.

`Asymmetry distance (MAD)`. This metric measures the average distance between the centroid of the BEV and the central axis of the MLC. It is measured in mm and it is calculated as follows:

$$AD_{ij} = \sum_{l=1}^{N_{ij}/2} \frac{\left(|y_{ijl}^{B}| + |y_{ijl}^{A}|\right)}{2} \tag{2.5}$$

The plan values is the so-called mean asymmetry distance (MAD). A schematic representation is shown in Fig. 2.2

`Perimeter (P).` The perimeter of the BEV is measured in mm and it is computed as follows:

$$P_{ij} = 2(x_{ij} + y_{ij}) \tag{2.6}$$

The plan value is computed using Eq. (2.2).

`Equivalente field size (EFS).` The EFS is implemented based on the definition provided in [64], and it is computed as follows:

$$EFS_{ij} = \frac{2L_{ij}W_{ij}}{L_{ij} + W_{ij}} \tag{2.7}$$

with L_{ij} and W_{ij} shown in Fig. 2.3a. The plan value is compute using Eq. (2.2).

`Percentage of small fields (psmall).` This metrics counts the percentage of fields with $EFS_{ij} < 30\ mm$ and it is defined as follows:

$$psmall_{ij} = \left[EFS_{ij} < 30\ mm\right] \tag{2.8}$$

where the term in the square brackets is equal to 1 if $EFS_{ij} < 30\ mm$, 0 otherwise.

`Edge metric (EM)` This metric was first proposed by Younge and colleagues [61] and quantifies the 'amount of edge' in a VMAT/IMRT plan. It is measured in mm^{-1} and it is computed as follows:

$$EM_{ij} = \frac{C_1 Px_{ij} + C_2 Py_{ij}}{BA_{ij}} \tag{2.9}$$

where Px_{ij} and Py_{ij} denote the length of the perimeter of defined by the MLC ends and sides, respectively, at the $j-th$ CA of the $i-th$ arc. The two constants C_1 and C_2 are scaling factors to adjust the relative importance of Px_{ij} and Py_{ij}, independently. In the current implementation, $C_1 = 0$ and $C_2 = 1$ consistently to what used in the original paper. In a future version of the code, the user will be able to set these two parameters arbitrarily. A schematic representation of Px_{ij} and Py_{ij} is shown in Fig. 2.3b.

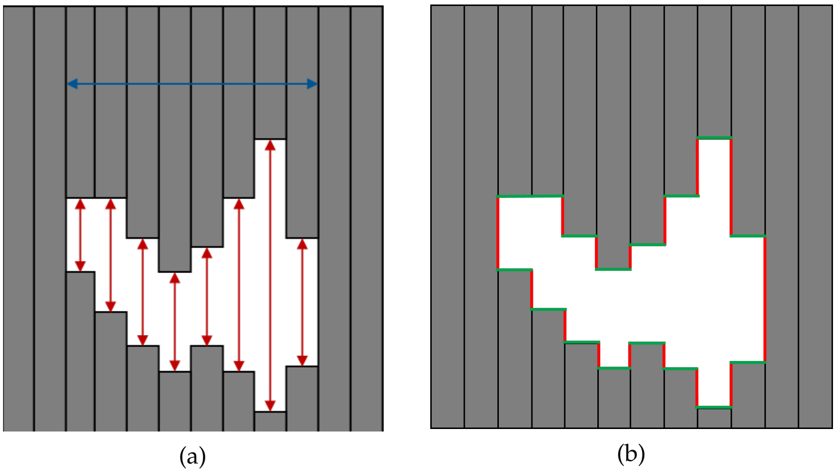


Figure 2.3: *(a)* Schematic representation of L_{ij} (blue arrow) and W_{ij} (average of the red arrows) in the computation of the EFS_{ij}. *(b)* Schematic representation of Px_{ij} (green) Py_{ij} (red) in the definition of EM.

2.1.5 *Complexity metrics: deliverability metrics*

`Monitor units (MUs).` This metric indicates the number of monitor units delivered during a CA. In this case, the plan value is computed as the sum over all arcs and CAs, namely:

$$MUs = \sum_{i=1}^{N_A} \sum_{j=1}^{NCA_i} MUs_{ij} \tag{2.10}$$

`Monitor units per unitary prescribed dose (MUcGy).` This metric measures the ration between MUs and the prescribed dose per fraction. It is measured as MUs/cGy, and it is computed as follows:

$$MUcGy = \frac{MUs}{DF} \tag{2.11}$$

`Monitor units per control-arc (MUCA).` This metric measures the average number of MUs delivered per each CA. It is computed as follows:

$$MUCA_i = \frac{MUs_i}{NCA_i} \tag{2.12}$$

`Monitor units per unitary gantry angle (MUdeg).` This metric measures the average number of MUs delivered per unitary gantry rotation. It is measured in $MU/°$, and it is computed as follows:

$$MUdeg_{ij} = \frac{MUs_{ij}}{CAL_{ij}} \tag{2.13}$$

where CAL_{ij} is the angular length (i.e. in °) of CA i of arc j.

`Leaf travel (LT).` This metric was introduced by Masi and colleagues [47] and the total distance travelled by the MLC leaves during a CA. It is measured in mm, and it is computed as follows:

$$LT_{ij} = \sum_{l=1}^{NL_{ij}/2} |\Delta y_{ijl}^{A}| + |\Delta y_{ijl}^{B}| \tag{2.14}$$

with:

$$\Delta y_{ijl}^{A(B)} = y_{ikl}^{A} - y_{i(k+1)l}^{A} \tag{2.15}$$

which correspond to the distance travelled by leaf l between CP k and $k+1$.

`Leaf travel per monitor unit (LTMU).` This metric measures the distance travelled by the MLC leaves per delivered MU. It is measured in mm/MU, and it is computed as follows:

$$LTMU_{ij} = \frac{LT_{ij}}{MUs_{ij}} \tag{2.16}$$

The plan value is computed using Eq. (2.2).

`Leaf travel per leaf and monitor unit (LTNLMU).` This metric measures the distance traveled by the MLC leaves per delivered MU and active leaf. It is measured in mm/MU, and it is computed as follows:

$$LTNLMU_{ij} = \frac{LT_{ij}}{(NL_{ij}MUs_{ij})} \tag{2.17}$$

`Leaf travel per leaf and unitary CA length (LNA).` This metric measures the average leaf travel per single leaf and unitary CA length. It is measured in $mm/°$, and it is computed as follows:

$$LNA_{ij} = \frac{LT_{ij}}{NL_{ij}CAL_{ij}} \tag{2.18}$$

`Delivery Time (dt).` The total delivery time is not stored in the DICOM RT-Plan file and it is handled by the machine during treatment. Therefore, its computation is based on a strategy aimed at mimicking what the machine does. In particular, for each CA the following time intervals are computed:

- $dt_{ij}^{(MAX,MLC)}$: estimated time interval for the MLC moving at its maximum speed;

- $dt_{ij}^{(MAX,GANTRY)}$: estimated time interval for the gantry moving at its maximum speed;
- $dt_{ij}^{(MAX,JAWS)}$: estimated time interval for the jaws moving at their maximum speed;
- $dt_{ij}^{(MAX,DR)}$: estimated time interval for the machine at its maximum allowed dose rate.

During delivery, the machine adapts the speeds of its mechanical axis and the dose rate to the element leading the longest time. Therefore, the delivery time for a given CA is computed as:

$$dt_{ij} = max\left(dt_{ij}^{(MAX,MLC)}, dt_{ij}^{(MAX,GANTRY)}, dt_{ij}^{(MAX,JAWS)}, dt_{ij}^{(MAX,DR)}\right) \tag{2.19}$$

The total delivery time is thus computed as:

$$dt = \sum_{i=1}^{N_A} \sum_{j=1}^{NCA_i} dt_{ij} \tag{2.20}$$

`Dose rate (DR).` This metric measures the average dose rate during the treatment. It is measured in MU/min, and it is computed as follows:

$$DR_{ij} = \frac{MU_{ij}}{dt_{ij}} \tag{2.21}$$

`Mean dose rate variation (mDRV).` This metric measures the average variation of dose rate per degree of gantry rotation. It is measured in $MU/(min\ ^\circ)$, and it is computed as follows:

$$mDRV_{ij} = \left| \frac{DR_{ij+1} - DR_{ij}}{(CAL_{ij+1} + CAL_{ij}/2)} \right| \tag{2.22}$$

`Gantry speed (GS).` This metric measures the average speed of rotation of the gantry. It is measured in $^\circ/s$, and it is computed as follows:

$$GS_{ij} = \frac{CAL_{ij}}{dt_{ij}} \tag{2.23}$$

`Mean gantry speed variation (mGSV).` This metric measures the average variation of the gantry speed per degree of gantry rotation. It is measured in $1/s$, and it is computed as follows:

$$mGSV_{ij} = \left| \frac{GS_{ij+1} - GS_{ij}}{(CAL_{ij+1} + CAL_{ij}/2)} \right| \tag{2.24}$$

`Leaf speed (LS).` This metric measures the average translation speed of the leaves of the MLC. It is measured in mm/s, and it is computed as follows:

$$LS_{ij} = \frac{1}{NL_{ij}} \sum_{l=1}^{NL_{ij}} \frac{\Delta y_{ijl}}{dt_{ij}} \tag{2.25}$$

where Δy_{ijl} is computed according to Eq. (2.15).

`Aperture area variability (AAV).` This metric is computed according to the definition provided by Masi and colleagues [47]. It is calculated 'as the area defined by apertures of opposing leaves in the single control point normalized to the maximum area in the arc', namely:

$$AAV_{ik} = \frac{\sum_{l}^{NL_{ik}/2} (y^B_{ikl} - y^A_{ikl}) \cdot w_l}{\sum_{l=1}^{NL_{ij}/2} \left[max(y^B_l) - max(y^A_l) \right] \cdot w_l} \tag{2.26}$$

where $max(y_l^{A(B)})$ denotes the position of the $l-th$ leaf of bank $A(B)$, and w_l its width.

Based on the approach proposed in Ref. [47], the AAV in Eq. (2.26) is computed at each CP and then mapped to the CAs as follows:

$$AAV_{ij} = \left[\frac{AAV_{ik} + AAV_{ik+1}}{2} \right] \tag{2.27}$$

The plan value is computed using Eq. (2.2).

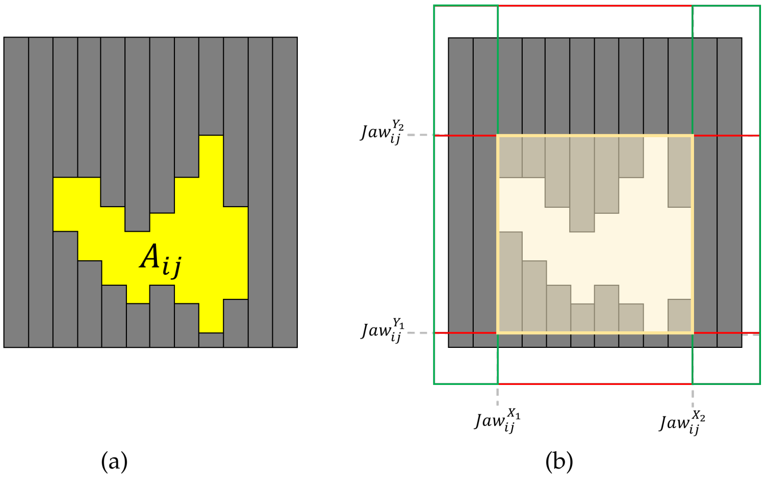
`Leaf sequence variability (LSV).` This metric is computed according to the definition provided by Masi and colleagues [47]. It is calculated 'for each control point considering in each bank the differences in position between adjacent MLC leaves' as follows:

$$LSV_{ik} = \prod_{\alpha=\{A,B\}} \left(\frac{\sum_{l}^{NK_{ik}/2-1} y_{max,l} - |y_{kl} - y_{kl+1}|}{(NL_{ik}/2 - 1) \cdot y_{max,l}} \right)_{\alpha} \tag{2.28}$$

where:

$$(y_{max,l})_{A(B)} = \left[max(y_l^k) - min(y_l^k) \right]_{A(B)} \tag{2.29}$$

indicates the possible maximum change in leaf position within CP k.



(a) (b)

Figure 2.4: *(a)* Schematic representation of the area of the BEV A_{ij}. *(b)* Schematic representation of jaw area JA_{ij}.

Based on the approach proposed in Ref. [47], the LSV in Eq. (2.28) is computed at each CP and then mapped to the CAs as follows:

$$LSV_{ij} = \left[\frac{LSV_{ik} + LSV_{ik+1}}{2}\right] \tag{2.30}$$

The plan value is computed using Eq. (2.2).

Modulation complexity score for VMAT (MCSv). This metric was first introduced by McNiven and colleagues [48] for the IMRT technique and then adapted to the VMAT technique by Masi and colleagues [47]. It combines the AAV and the LSV in a single numerical score and it is computed as follows:

$$MCSv_{ij} = LSV_{ij} \cdot AAV_{ij} \tag{2.31}$$

Plan area (PA). This metrics evaluated the area of the BEV. It is measured in mm^2, and it is computed as follows:

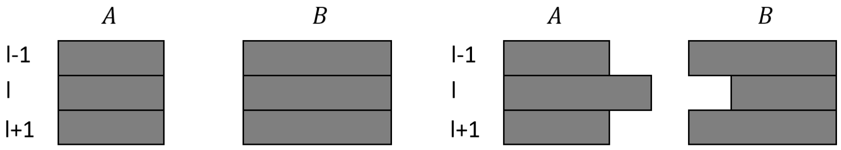
$$A_{ij} = \sum_{l=1}^{N_{ij}/2} LG_{ijl} \cdot w_l \tag{2.32}$$

The plan value is computed according to Eq. (2.2), and it is called the plan area (PA). A schematic of A_{ij} is shown in Fig. 2.4a.

Jaw area (JA). This metric measures the average area defined by the Jaws. It is measured in mm^2, and it is computed as follows:

$$JA_{ij} = \left(\left|Jaw_{ij}^{X_1} - Jaw_{ij}^{X_2}\right| \times \left|Jaw_{ij}^{Y_1} - Jaw_{ij}^{Y_2}\right|\right) \tag{2.33}$$

where $Jaw_{ij}^{X(Y)}$ are shown in Fig. 2.4b. The plan values is computed using Eq. (2.2).



(a) (b)

Figure 2.5: *(a)* Example of leaf pairs with zero toungue-and-groove. *(b)* Example of leaf pairs with non-zero tongue-and-groove.

`BEV-Jaw area ratio (BJAR).` This is a novel metric which measures the ratio between the area of the BEV and that of the JAW. It is computed as follows:

$$BJAR_{\mathbf{ij}} = \frac{A_{ij}}{JA_{ij}} \tag{2.34}$$

`Plan irregularity (PI).` This metric was proposed by Du and collegues [41], and it is computed as follows:

$$BI_{ij} = \frac{PI_{ij}^2}{4\pi A_{ij}} \tag{2.35}$$

The plan values is computed using Eq. (2.2), and it is called *plan irregularity* (PI) according to the original paper.

`Plan modulation (PM).` This metric was proposed by Du and collegues [41], and it is computed as follows:

$$BM_{ij} = 1 - \frac{\sum_{j=1}^{NCA_i} MU_{ij} A_{ij}}{MU_i A_i^{tot}} \tag{2.36}$$

where A_i^{tot} is the area of the union of the BEVs of al CAs of beam i. Following the original paper, the plan value is called *plan modulation* (PM), and it is computed using Eq. (2.2). The values of PM range from 0 to 1, with the two values corresponding to the lowest and highest level of modulation, respectively.

`Tongue-and-groove (TG).` This metric was first proposed by Yao and collegues [63]. For each leaf in the two banks compute its contribution to the tongue-and-groove effect as:

$$TG\alpha_{ij}(l-1,l) = \begin{cases} 0 & y_{ijl-1} - y_{ijl} < 0 \\ LG_{ijl} & y_{ijl-1} - y_{ijl} > LG_{ijl} \\ y_{ijl-1} - y_l & otherc wise \end{cases} \tag{2.37}$$

with $\alpha = \{A, B\}$.

The TG effect for each leaf pair is thus computed as follows:

$$TG_{ijl} = \frac{1}{2}\sum_{\alpha}\left[TG\alpha_{ij}(l-1,l) + TG\alpha_{ij}(l+1,l)\right] \tag{2.38}$$

The value of the tongue-and-groove is thus computed as follows:

$$TG_{ij} = \sum_{l=1}^{NL_{ij}/2} TG_{ijl} \tag{2.39}$$

The plan values is computed using Eq. (2.2). In Fig. 2.5a and 2.5b, schematics of leaf pairs with zero and non-zero TG effect are shown, respectively.

`Modulation degree (MD).` The use of this metric was proposed in Ref. [58], and it is computed as follows:

$$MD = \frac{MUs}{\sum_{ij} \frac{A_{ij} \cdot MU_{ij}}{A_{tot,i}}} \tag{2.40}$$

`Modulation index (MI).` The modulation indexes were first introduced by Park and collegues [51]. We refer to the original paper for all the mathematical details concerning its computation. In VCoMX, the metrics has been implemented as described by the authors.

2.2 TCOMX

The `TCoMX` extraction engine implements 51 metrics for the HT technique. In total, 27/51 were taken either from Ref. [69], or from other relevant literature [65–67], and includes also typical parameters reported by the TPSs [110]. The remaining 24 metrics were developed as part of this thesis project. Metrics are organised in categories and subcategories based on the features of the HT plans described. The three categories were identified as delivery, leaf open time (LOT) statistics, and sinogram, and they are described in Sec. 2.2.2, 2.2.3 and 2.2.4.

2.2.1 *Plan sinogram and open leaves.*

In the DICOM RT-Plan files of HT plans, each complete rotation of the gantry is divided into 51 parts called projections. This information is stored in the sinogram, a matrix containing the fractional LOT values for each leaf at each projection. TCoMX calculates the complexity metrics using these data. To obtain the absolute LOTs, the sinogram is multiplied by the projection time, which is saved as a separate field in the DICOM RT-Plan file. Each LOT value correspond to the time a certain leaf remains open during a given projection. Therefore, the *open leaves* at a given projection are all those with $LOT > 0\ ms$.

2.2.2 *Complexity metrics: delivery parameters*

`Modulation Factor (MF).` It is a dimensionless quantity computed as:

$$MF = \frac{mLOT}{maxLOT} \tag{2.41}$$

with $mLOT$ the mean LOT and $maxLOT$ the maximum LOT. These two quantities are explicitely defined in Sec. 2.2.3.

`Number of projection per rotation (`$N_{proj,rot}$`).` Each rotation of the gantry is subdivided into $N_{proj,rot}$ discrete steps called *projections*. To date, this parameter is fixed by construction to 51 for all the existing machine. However, it was left as a free parameter for practicality.

`Number of projections (`N_{proj}`).` It is the total number of projections in the treatment and corresponds to the number of rows of the sinogram stored in the RT-PLAN files minus one.

`Number of rotations (`N_{rot}`).` It quantifies the number of gantry rotations during the treatment. It is defined as follows:

$$N_{rot} = \frac{N_{proj}}{N_{proj,rot}} \tag{2.42}$$

`Projection Time (PT).` The duration of each projection is refered to as *projection time*. It is measured in seconds (s).

`Gantry Period (GP).` It quantifies the duration of each rotation and it is related to PT through the following relation:

$$GP = N_{proj,rot} \times PT \tag{2.43}$$

and it is measured in seconds (s).

`Treatment Time (TT).` It quantifies the duration of the treatment. It is related to the gantry period by the following relation:

$$TT = GP \times N_{rot} \tag{2.44}$$

It is measured in seconds (s).

`Field Width (FW).` Is is defined as the distance between the Y jaws. It is measured in millimeters (mm).

`Pitch` It is a dimensionless quantity defined as follows:

$$pitch = \frac{FW}{\Delta y} \tag{2.45}$$

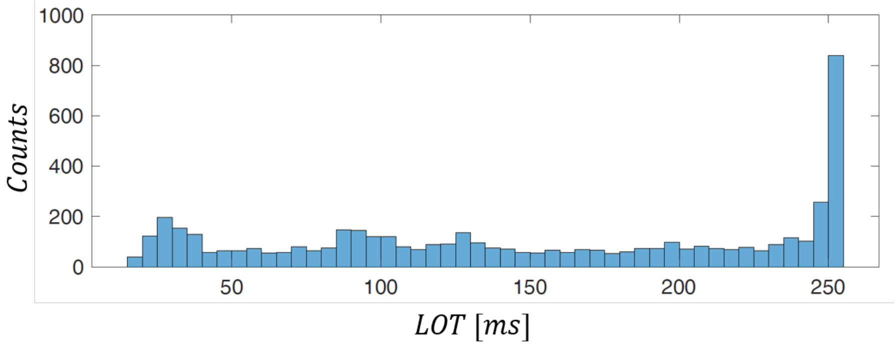


Figure 2.6: Example of LOT distribution.

where Δy is the translation of the couch after each full gantry rotation.

`Couch Translation (CT).` It is defined as follows:

$$CT = N_{rot} \times FW \times Pitch = N_{rot} \times \Delta y \tag{2.46}$$

It is measured in millimiters (mm):

`Couch Speed (CS).` It is defined as follows:

$$CS = \frac{CT}{TT} \tag{2.47}$$

It is measured in millimiters per second (mm/s).

`Target Length (TL).` It is defined as follows:

$$TL = CT - FW \tag{2.48}$$

It is measured in millimiters (mm). A detailed explanation of this relation is provided in Ref. [110].

`Treatment time over fraction dose (TTDF).` It is defined as follows:

$$TTDF = \frac{TT}{DF} \tag{2.49}$$

where DF is the dose per fraction expressed in cGy. It is measured in s/cGy.

2.2.3 *Complexity metrics: absolute and relative LOT statistics*

`LOT mean (mLOT)` It is computed as the arithmetic average of the non-zero entries of the sinogram multiplied by the projection time. The unit of measure used is the millisecond (ms).

`LOT standard deviation (sdLOT)` It is computed as the sample standard deviation of the non-zero entries of the sinogram multiplied by the projection time. The unit of measur used is the millisecond (ms).

`LOT median (mdLOT)` It is computed as the median of the non-zero entries of the sinogram multiplied by the projection time. The unit of measure used is the millisecond (ms).

`LOT mode (moLOT)` It is computed as the mode of the non-zero entries of the sinogram multiplied by the projection time. The unit of measure used is the millisecond (ms).

`LOT maximum (maxLOT)` It is computed as the maximum of the non-zero entries of the sinogram multiplied by the projection time. The unit of measure used is the millisecond (ms).

`LOT minimum (minLOT)` It is computed as the minimum of the non-zero entries of the sinogram multiplied by the projection time. The unit of measure used is the millisecond (ms).

`LOT kurtosis (kLOT)` It is computed as the kurtosis of the non-zero entries of the sinogram.

`LOT skewness (sLOT)` It is computed as the skewness of the non-zero entries of the sinogram.

Cumulative LOT Number Score ($\mathbf{CLNS}_n$) It is a dimensionless metrics defined as follows:

$$CLNS_n = \frac{1}{\sum_{ij} O_{ij}} \sum_{i=1}^{N_{proj}} \sum_{j=1}^{N_{leaves}} O^*_{ij}(n) \qquad n \in [0; PT] \tag{2.50}$$

with:

$$\begin{cases} O^*_{ij}(n) = 1 & if\ S_{ij} \times PT < n \\ O^*_{ij}(n) = 0 & otherwise \end{cases} \tag{2.51}$$

It counts the fractional number of LOTs smaller than n.

Cumulative LOT Number Score at Projection Time ($\mathbf{CLNS}_{pt,n}$) It is dimensionless metrics defined as follows:

$$CLNS_{pt,n} = \frac{1}{\sum_{ij} O_{ij}} \sum_{i=1}^{N_{proj}} \sum_{j=1}^{N_{leaves}} O^*_{ij}(n) \qquad n \in [0, PT] \tag{2.52}$$

with $O^*_{ij}(n)$ defined as in Eq. (2.51). It counts the fractional number of LOTs which are closer than n ms to the projection time.

Relative LOT statistics are computed in the same way absolute ones, considering the fractional leaf open times (FLOT), namely the LOT divided by the projection time. The metrics implemented in `TCoMX` corresponding to this category are: mFLOT, sdFLOT, mdFLOT, moFLOT, maxFLOT, maxFLOT, minFLOT, CFNS$_n$.

2.2.4 *Complexity metrics: sinogram geometry and modulation*

`Leaves with n Open Nearest Neighbors (LnNS).` It was introduced by Santos et al. in [69] and it is a dimensionless metric defined as follows:

$$LnNS = \frac{1}{N_{proj}} \sum_{i=1}^{N_{proj}} \left[\frac{N_O(n)}{N_O} \right]_i \times 100\% \tag{2.53}$$

where $N_O(n)$ indicates the number of open leaves with n open neighbors and N_O the number of open leaves. The sub-index i indicates that the ratio has to be evaluated at each projection. The plan value is obtained as the average over all the projections. The index n can take three different values: 0, 1 and 2.

`Treatment Area (TA).` It is defined as follows:

$$TA = \frac{1}{N_{proj}} \sum_{i=1}^{N_{proj}} (|R_i - L_i| + 1) \tag{2.54}$$

where R_i and L_i are the right-most and left-most open leaves at projection i, respectively.The plan value is obtained by averaging over all the projections. Since the LOTs values are not considered in this definition, TA represents the average cumulative open area. It is measured in number-of-leaves.

`Number of Connected Components (nCC).` It is a dimensionless metric which counts the number of independent groups of connected leaves inside the treatment area. It is computed at each projection and then averaged over all the projections to get the plan value. In Fig. 2.7 an example of two projections with different numbers of connected components is shown.

`Length of the Connected Components (lengthCC).` It is defined as follows:

$$lengthCC = \frac{1}{\sum_{i=1}^{N_{proj}} nCC_i} \sum_{i=1}^{N_{proj}} \sum_{k=1}^{nCC_i} |L_k - R_k| + 1 \tag{2.55}$$

where L_k and R_k are the positions of the leftmost and rightmost open leaves of the $k-th$ connected component at the $i-th$ projection, respectively. It is measured in number-of-leaves. The plan value measures the average length of the connected components.

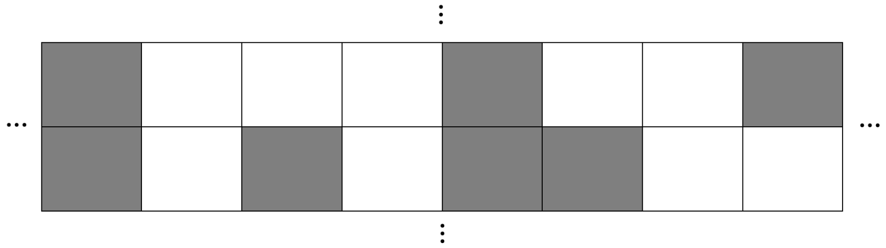


Figure 2.7: Schematic representation of two projections with different numbers of connected components. White and gray squares represent open and closed leaves, respectively. The upper projection has 2 connected components, the lower one has 3.

In Fig. 2.7, four examples of connected components (white blocks) with different lengths are shown. The two connected components of the upper projection have length 3 and 2, respectively. The three connected components of the lower projection have length 1, 1 and 2, respectively. The plan value is obtained by averaging over the total number of connected components of the sinogram.

Fraction of Discontinuous Projections (fDISC). It is a dimensionless metric defined as follows:

$$fDISC = \frac{1}{N_{proj}} \sum_{i=1}^{N_{proj}} [nCC_i > 1] \tag{2.56}$$

where the term $[nCC_i > 1]$ is 0 if $nCC_i = 0$ or 1, 1 otherwise. It counts the fraction of projections with two or more connected components, namely all the projections having at least one closed leaf within the treatment area.

Closed Leaf Score (CLS). It was introduced in [69] and it is a dimensionless defined as follows:

$$CLS = \frac{1}{N_{proj}} \sum_{i=1}^{N_{proj}} \left[\frac{N_{leaves} - \sum_j^{N_{leaves}} O_{ij}}{N_{leaves}} \right] \tag{2.57}$$

It can take values in $[0, 1]$, being 1 when all leaves remain closed.

Closed Leaf Score within the treatment area (CLS$_{in}$). It was introduced in Ref. [69] with this definition:

$$CLS_{in} = \frac{1}{N_{proj}} \sum_{i=1}^{N_{proj}} \left[\frac{TA_i - \sum_j^{N_{leaves}} O_{ij}}{N_{leaves}} \right] \tag{2.58}$$

A more general form has been devised and three variants have been introduced:

1. $CLS_{in,area}$: the number of closed leaves within the treatment area at projecton i is normalized by TA_i instead of N_{leaves};
2. $CLS_{in,disc}$: it is computed by considering the discontinuous projections $(nCC_i > 1)$ only, namely the first summation in Eq. (2.58) runs from 1 to the total number of discontinuous projections and is then divided by the number of discontinuous projections instead of by N_{proj} ;
3. $CLS_{in,area,disc}$: it corresponds to the combination of the $CLS_{in,area}$ with the $CLS_{in,disc}$;

The three quantities are in general correlated with each other. However, the different definitions should help to characterize the geometry of the leaf openings in a more intuitive way.

`Centroid`. It is defined as follows:

$$centroid = \frac{1}{N_{proj}} \sum_{i=1}^{N_{proj}} \frac{1}{\sum_{j=1}^{N_{leaves}} O_{ij}} \left[\sum_{j=1}^{N_{leaves}} O_{ij} P_j \right] \tag{2.59}$$

The centroid is measured in number-of-leaves. The plan value is obtained by averaging over N_{proj}. The centroid of a plan represents the mean of the average positions of the open leaves at each projection.

`Leaf Open Time Variability (LOTV)`. It was introduced in Ref. [69] and it is a dimensionless metric defined as follows:

$$LOTV = \frac{1}{N_{leaves}} \sum_{j=1}^{N_{leaves}} \left[\frac{\sum_{i=1}^{N_{proj}-1} max(S_j) - |S_{ij} - S_{i+1,j}|}{(N_{proj} - 1) \times max(S_j)} \right] \tag{2.60}$$

where S_j marks the j-th column (leaf) of the sinogram. It takes values in $[0, 1]$, being 1 when all the leaves have the same opening time at each projection. For the leaves which do not open during the treatment it is set to 1 by definition.

`Extended Leaf Open Time Variability (ELOTV`$_{\Delta p}$`)`. It is a dimensionless metric defined as follows:

$$ELOTV_{\Delta p} = \frac{1}{N_{leaves}} \sum_{j=1}^{N_{leaves}} \left[\frac{\sum_{i=1}^{N_{proj}-\Delta p} |S_{ij} - S_{i+\Delta p,j}|}{(N_{proj} - \Delta p) \times max(S_j)} \right] \tag{2.61}$$

where Δp is the projection step and S_j the j-th column (leaf) of the sinogram. It takes values in $[0, 1]$, being 0 when all the leaves have the same opening time at each projection. The $ELOTV_{\Delta p}$ is first evaluated for each leaf and then averaged over all the leaves to obtain the plan value. For the leaves which do not open during the treatment it is set to 0 by definition. It is worth noticing that the $ELOTV_{\Delta p}$ includes the $LOTV$ as a special case. In particular:

$$LOTV = 1 - ELOTV_1 \tag{2.62}$$

Compared to the *LOTV* it shows two main differences:

- It is positively correlated with the inter-projection (F)LOT variability;
- It allows the comparison of projections which lie at arbitrary distances.

`Plan Sinogram Time Variation (PSTV).` It was introduced by Santos et al. in [69] and it is a dimensionless metrics defined as follows:

$$PSTV = \frac{1}{N_{proj}-1} \sum_{i=1}^{N_{proj}-1} \sum_{j=1}^{N_{leaves}-1} |S_{i+1,j} - S_{i,j}| + |S_{i,j+1} - S_{i,j}| \quad (2.63)$$

which is the arithmetic average over all the projections of the sum of FLOTs difference between pairs of adjacent leaves and projections. Higher values of PSTV value correspond to a higher variability of the (F)LOT across the leaves and the projections.

`Extended Plan Sinogram Time Variation (EPSTV`$_{\Delta p,\Delta l}$`)` .It is a dimensionless metric defined as follows:

$$EPSTV_{\Delta p,\Delta l} = \frac{1}{N_{proj}-\Delta p} \sum_{i=1}^{N_{proj}-\Delta p} \sum_{j=1}^{N_{leaves}-\Delta l} |S_{i+\Delta p,j} - S_{i,j}| + |S_{i,j+\Delta l} - S_{i,j}| \quad (2.64)$$

The $EPSTV_{\Delta p,\Delta l}$ contains the $PSTV$ defined in Eq. (2.63) as a special case, namely:

$$PSTV = EPSTV_{1,1} \quad (2.65)$$

This new formulation allows the comparison of projections/leaves which lie at arbitrary distances. Furthermore, there are two special cases included in it, namely:

- $EPSTV_{0,\Delta l}$: time variation along the leaves direction;
- $EPSTV_{\Delta p,0}$: time variation along the projection direction.

In general, a higher $EPSTV_{\Delta p,\Delta l}$ corresponds to higher inter-leaf and/or inter-projection (F)LOT variability.

`Modulation Index (MI).` The *Modulation Index* was introduced by Park and collegues [51] for VMAT plans. It was adapted to HT plans by Santos and colleques [69]. It is a dimensionless metric defined as follows:

$$MI = \int_0^{2\sigma} Z(f)df \quad (2.66)$$

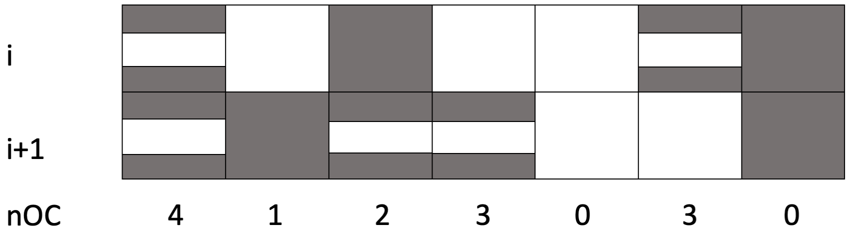


Figure 2.8: Schematic representation of possible combinations of the LOTs leading to different numbers of openings and closures. Rows represent two adjacent projections, while columns show seven possible leaf movements. White and gray spaces represent open and closed leaves, respectively. Each LOT is centered about the middle point of the corresponding projection. The number under each leaf represents the cumulative number of openings and closures for each leaf over the two projections.

where Z(f) is defined as:

$$Z(f) = \frac{Z_x(f) + Z_y(f) + Z_{xy}(f) + Z_{yx}(f)}{4} \tag{2.67}$$

with each $Z_k(f)$ defined as:

$$Z_k(f) = \frac{1}{N_{proj}} N_k(f; \Delta t_k > f\sigma) \tag{2.68}$$

where $k = \{x, y, xy, yx\}$ represents four different directions of the sinogram S (leaves, projections, diagonal and anti-diagonal), Δt_k represents the time variation between adjacent elements in the k direction, f a fraction of the $sdFLOT$ and $N_k(f, \Delta t_k > f\sigma)$ the number of FLOTs exeeding this fraction. Higher values of MI should correspond to a higher plan modulation.

`Number of Openings and Closures (nOC).` It is a dimensionless metric and it is computed by counting the number of times each leaf opens and closes during the treatment. The plan value is obtained by averaging over the leaves and normalizing by the total number of projections. The number of openings and closures is computed considering that each (F)LOT is centered with respect to the projection time [67]. In Fig. 2.8, a schematic of the different conditions that might be encountered during the treatment is shown. This metric is associated with the mechanical stress of the MLC during the treatment. Moreover, this metric is related to the CLS (average fraction of closed leaves per projectoin) by the following approximate linear relationship:

$$CLS \approx 1 - 0.5 \times nOC \tag{2.69}$$

This is due to the fact that fully closed ($S_{ij} = 0$) have $nOC_{ij} = 0$ and fully open ($S_{ij} = 1$) appear only in a negligible amount. Therefore, each $0 < S_{ij} < 1$ correspond to $nOC_{ij} = 2$.

`Mean Sinogram Asymmetry (mSA)` It is defined as follows:

$$MSA = \frac{\sum_{j=1}^{N_{leaves}} P_j \times LPS_j}{\sum_{j=1}^{N_{leaves}} LPS_j} \tag{2.70}$$

where LPS_j is called *Leaf Projected Sinogram* and is defined as:

$$LPS_j = \sum_{i=1}^{N_{proj}} \frac{S_{ij}}{N_{proj}} \tag{2.71}$$

The MSA represents the weighted average displacement of the (F)LOTs from the vertical axis passing through the center of the sinogram and it is measured in number-of-leaves. It allows to highlight left-right asymmetries of the sinogram.

`Sinogram intensity (SI).` Starting from the LPS defined in Eq. (2.71), it is possible to compute another metric related to it, namely the *Sinogram Intensity*. In particular, the mean sinogram intensity (mSI), is computed as the average of the LPS values over the MLC leaves. Leaves with $LPS = 0$ are included in the computation and this fact distinguishes mSI from mLOT (mFLOT). In addition to mSI, also its standard deviation (sdSI) and the median (mdSI) are computed.

2.3 METRICS EXTRACTION WORKFLOW

Figure 2.9, depicts a schematic of the extraction process with the `UCoMX`. The system consists of two layers: the GUI interface and the back-end libraries. The user has access only to the first layer, with the second one automatically managed by the package. `UCoMX` extracts complexity metrics from the DICOM RT-Plan files of radiotherapy treatment plans.

To begin using `UCoMX`, the user should select the desired extraction engine from the main panel. After this selection, a new panel will appear where the user has to type the input and output folders. The input folder should contain the DICOM RT-Plan files for analysis (even categorised into subfolders), and the output folder will store the results. As an alternative, the user can place an .xlsx file in the input folder that contains the path and file name of the plans to analyse. If the GUI is not available, all the necessary information should be provided through a dedicated text file called CONFIG.in. For VMAT plans, machine-related parameters which are not stored in the DICOM RT-Plan file, e.g. maximum gantry speed,maximum MLC leaf speed, can be provided by using a dedicated file. This file must be

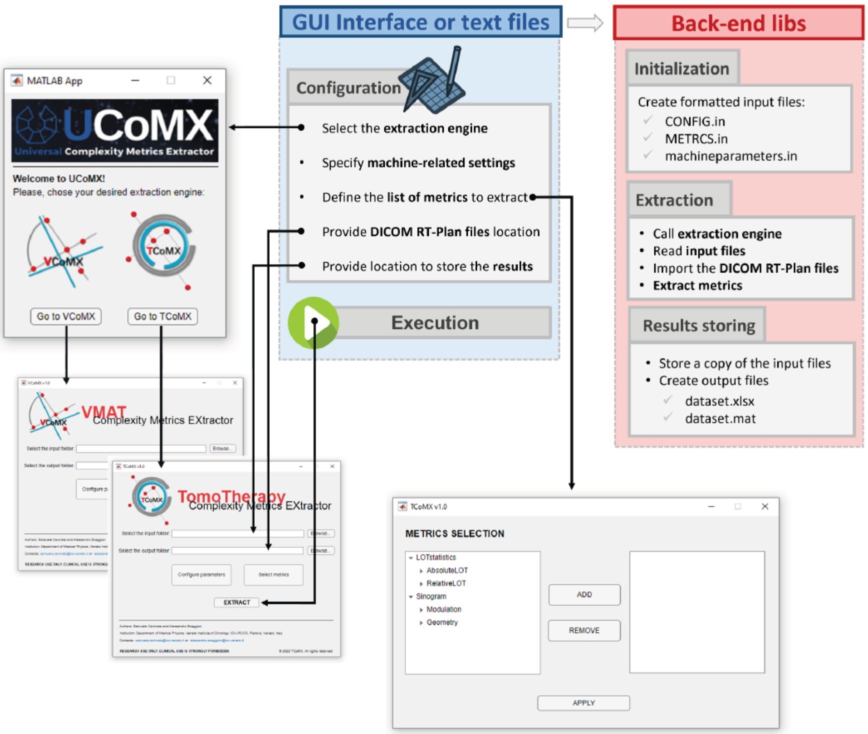


Figure 2.9: Main panel of `UCoMX`. From the dedicated buttons, the user can select the extraction engine to use.

stored in a subfolder prepared specifically for this purpose within the UCoMX main folder. The extraction engine requires these parameters to estimate the delivery time and accurately derive all the dynamical quantities (e.g. speed or acceleration). Default parameters are used if this file is not provided. The 'Select metric' button enables the user to access a panel dedicated to creating the list of metrics to extract. If the GUI is not available, the user must manually fill in a file called METRICS.in with the list of metrics he wishes to extract.. UCoMX automatically stores the following output files upon completion of each execution:

- A datasheet (dataset.xlsx) which summarises the information about the analysed DICOM RT-Plan files and reports the values of the extracted complexity metrics. This file is additionally saved in .mat format.
- A copy of the input files used (CONFIG.in and METRICS.in);
- A log file that summarises the main details of the execution;

These output files are automatically stored in a subfolder within the designated output folder. This subfolder's name is determined by the session's date and time to avoid overwriting older files in the event of multiple executions.

Table 2.2: List of the treatment units and TPSs in the reference dataset.

Structure	**Treatment unit**	**TPS**
Prostate, Head/Neck, MultiTarget, C-Shape	TrueBeam STX	Eclipse v16.1
	Unique	Eclipse v16.1
	Ethos	Eclipse v16.1
	RadixAct	RayStation v1.0a
	RadixAct	Precision v1.0.02
C-shape	Unique	RayStation v1.0a
	Ethos	RayStation v1.0a

2.4 REFERENCE DATASET

The `UCoMX` package is provided with a reference dataset of 58 treatment plans created using three different TPSs, and for four different treatment units, resulting in five and two combinations for fixed-ganty IMRT/VMAT and HT, respectively, as shown in Table 2.2. In total, 28 VMAT, 14 IMRT and 16 HT plans are included in the dataset. For VMAT and Helical Tomotherapy, two versions of each plan have been created by setting two different levels of modulation during treatment optimization. For convenience, they will be referred to as *low modulation (LM)* and *high modulation (HM)*. For VMAT plans, this was done by using two different combinations of MU limits and ASC on Eclipse, while enabling the 'double-arc' option on RayStation. For Helical Tomotherapy plans, two different values of modulation factor (MF) were used. The treatment plans were created following AAPM TG-119 IMRT commissioning test [111], which include four cases:

- **`MultiTarget:`** it consists of three cylindrical targets that are stacked along the axis of rotation, each one with diameter 4 *cm* and length 4 *cm*.
- **`Mock prostate:`** it consists of an ellipsoidal CTV with RL, AP and SI measuring 4.0 *cm*, 2.6 *cm* and 6, 5 *cm*, respectively. The PTV is also included and it is obtained by expanding the CTV 0.6 *cm* all around. Two OARs are delineated, namely bladder and rectum.
- **`Mock Head/Neck:`** the PTV is composed of all anterior volume from the base of the skull to the upper neck. Posterior neck nodes are included, too. A gap of 0.6 *cm* exist between the PTV and the skin. Cord and parotids are the delinated OARs.

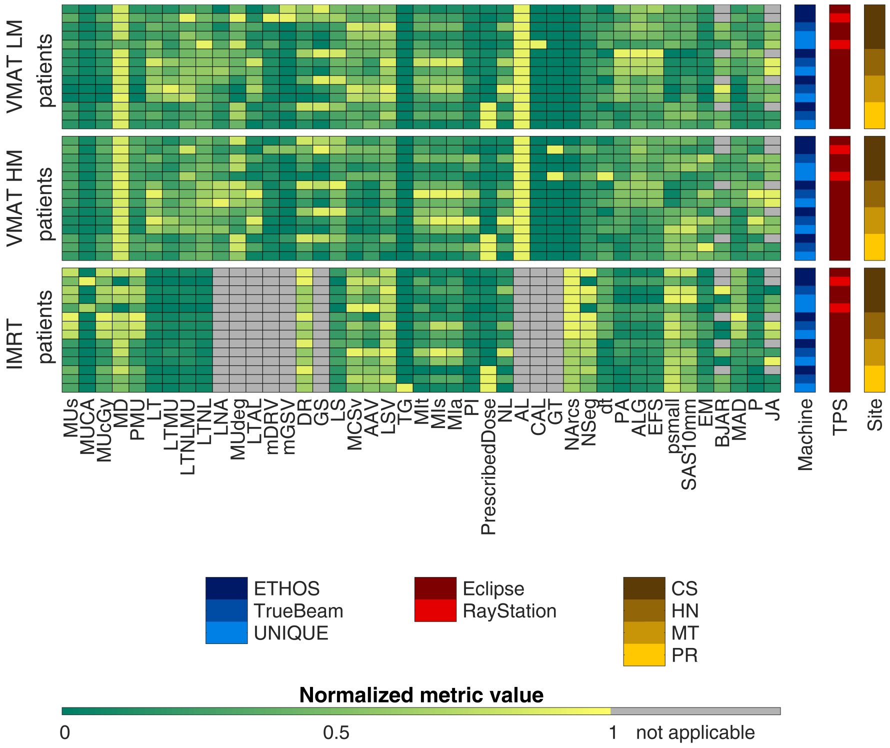


Figure 2.10: Heatmaps showing the values of all the metrics extracted from the 28 VMAT and 14 IMRT plans in the reference dataset. Each metric was normalised to take values between 0 and 1.

- **`C-shape`:** it is a C-shaped target of length 8 *cm* sorrounds a central cylindrical avoidance structure of 1 *cm* radius and 10 *cm* long. The inner and outer arcs of the PTV measure 1.5 and 3.7 *cm*, respectively. A gap of 1.5 *cm* separate the PTV and the avoidance structure.

More detail on the phantom used can be found in Ref. [111].

Figure 2.10 summarises the results of the metrics extraction from the 28 VMAT and the 14 IMRT plans. In total, 43 metrics were extracted. A subgroup of those metrics are not applicable to some plans (grey squares in the heat map). On the one hand, the bev-jaw area ratio (BJAR) and jaw area (JA) are not applicable to ETHOS due to the absence of the jaws on that machine. On the other hand, metrics such as the LNA, MUdeg, LTAL mDRV, mGSV, GS, AL, CAL and GT are not applicable to IMRT since they describe dynamical quantities That are not involved in IMRT treatment delivery, such as gantry speed/acceleration or dose rate variations. In general, the different modulation levels for the three groups of plans (VMAT LM, VMAT HM and IMRT) lead to different values of the complexity indicators. On the other hand, also the treatment machine, the TPS and the site have different effects on the metric values.

Figure 2.11 shows the same result for the 16 helical tomotherapy plans in the dataset. In total, 69 metrics were extracted. Also in this

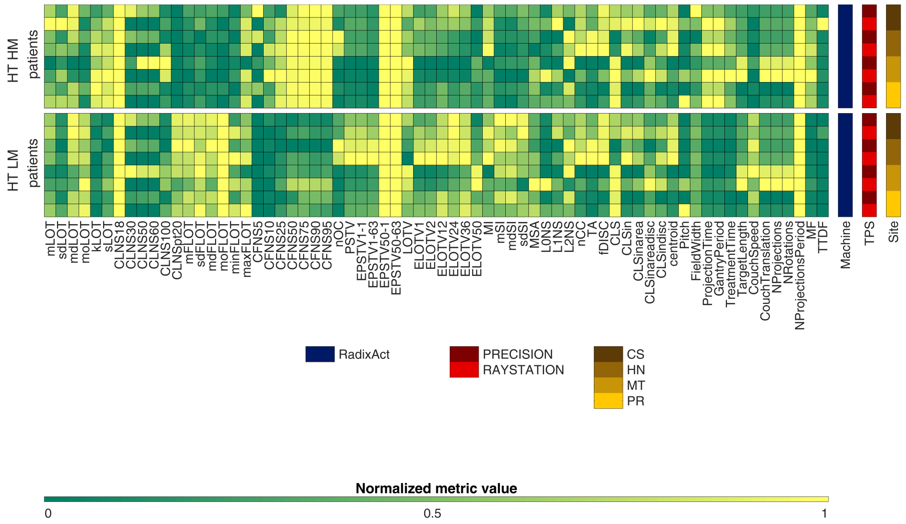


Figure 2.11: Heatmaps showing the values of all the metrics extracted from the 16 helical tomotherapy plans in the reference dataset. The upper and lower plots correspond to high and low modulation plans, respectively. Each metric was normalised to take values between 0 and 1.

case, the different levels of modulation (HT LM vs HT LM) to different values for the complexity metrics on the same plan as we can observe from the different patterns arising in the two maps.

2.4.1 *Impact and significance of the* `UCoMX` *package*

In this work, we presented `UCoMX`, a free software package for the extraction of complexity metrics from radiotherapy treatment plans. Thanks to the extensive set of metrics included, the different versions available and the ease of use, it is meant to meet the necessity of any type of user. Additionally, thanks to its compatibility with most of the commercially available TPSs and treatment units, it possesses all the characteristics to become a standard reference tool to be used across institutions for the evaluation of treatment plan complexity. Its availability should overcome the current difficulties and limitations in comparing results between different institutions. Typically, results published in the literature have been obtained using in-house packages that are not made available, and whose implementation details might differ substantially. Indeed, there are choices that can be made when implementing the complexity metrics that could impact the final result, a fact that affects their significance and transferability between centres. On this concern, `UCoMX` is provided with an extensive documentation where the mathematical details concerning the implementation of each complexity metric are accurately described. Furthermore, the provided reference dataset serves as a benchmark for comparisons with other in-house routines.

In a context where plan complexity is considered as one the building blocks of plan quality [112], it turns out that less than 25% of centres worldwide make actual use of complexity metrics [113]. As a matter of fact, the are two main reasons behind this fact. First, a shared consensus on which are the most significant indicators to consider still lack. Second, commercial TPSs do not offer any tool to perform the extraction of complexity metrics. Therefore, the availability of a common tool to be used for metrics extraction was a first needed step to increase the number of people that can access this field as well as to move the first steps towards a standardisation of the extraction approach. This was the rationale behind the development of `UCoMX`.

`UCoMX` is undergoing continuous and intense development from different sides. First, thanks to the active and new collaborations between our group and different institutions in Italy, the package is going to be tested on new combinations of treatment units and TPSs, extending the number of users that can use it. Second, new metrics are about to be implemented. In particular, among all the indicators currently included in the package, radiomics features are not present, despite having been used in different studies applied to fluence maps, gamma maps or dose difference maps [87, 99, 114]. However, all such features were proposed before 2015, while the field of radiomics has moved several steps forward in recent years, including the definition of standardised features and softwares for their computation [115, 116]. Therefore, with the aim of making UCoMX a modern and standardised tool, its integration with S-IBEX is under development. S-IBEX is a free and open-source package created at our institution and is one of the most standardised tools for the extraction of radiomics features according to the IBSI standard [115–117]. Third, new tools for visualisation and comparisons plans are under development, including the possibility to extract complexity metrics from machine log-files and to compare the latter with those extracted from the RT-plans.

3

THE COMPLEXITY OF HELICAL TOMOTHERAPY PLANS

Helical tomotherapy (HT) is one of the most advanced technologies currently available for external beam radiotherapy treatments. In Ch. 1, we learned that due to its unique delivery method, thousands of beamlets are typically involved in the process, a fact that makes this technique very suitable for achieving a highly conformal dose to the PTV and optimal OARs sparing at the same time. However, this might introduce factors that could impact the overall delivery and dosimetric accuracy of the resulting plans.

THIS CHAPTER outlines the advancements and accomplishments made in this PhD project regarding HT complexity. The material presented is mostly taken from Refs. [102, 118].

3.1 QUANTITATIVE ASSESSMENT OF HT PLANS COMPLEXITY

It was discussed in Ch. 1 that, there are works published in the literature that investigated the role of typical HT delivery parameters such as the modulation factor (MF), the maximum and mean LOT as well as individual leaf latencies [66–68, 119] as possible indicators of complexity. More precisely, most of the published works proposed action limits or operative rules to support the planner in the realization plans that should fulfill the tolerance and action limits on the dosimetric and delivery accuracy [23]. Despite the importance of having practical indications that could drive the planning process towards plans satisfy the tolerance and action limits on the delivery and dosimetrics accuracy, the strategies that have been proposed so far do not allow an in-depth and accurate characterization of HT plan complexity. Finding a set of robust complexity indicators would facilitate the plan comparison and would have the potential to drive a better standardization of the optimization process, so as to reduce the inter- and intra-planner variability and share effective planning strategies across the community.

As presented in Ch. 2, significant steps have been moved towards the definition of new complexity metrics for HT in this work, with the proposal of several indicators capable of quantifying different factors, including the statistics of the leaf open time (LOT) distribution, the geometry of the binary MLC apertures and the modulation of the LOTs over space and time. In this section, a study to validate the effectiveness of the newly developed complexity indicators is proposed.

3.1.1 *Dataset*

Between June 2018 and April 2021, a total of 464 treatments was planned with the Precision TPS v1.0.02 (Accuray, Inc., Madison, Wisconsin, USA) using the GPU-based optimizer VOLO coupled with the Collapsed Cone Convolution Superposition dose calculation algorithm [110, 120]. The dataset used for this analysis is an extension of the one used in Ref. [30] for the analysis presented in App. **??**. The treatment plans included in the dataset were related to different sites and diseases and were grouped into the following categories: brain, head-and-neck (HN), thorax, abdomen, pelvis, prostate, and others. The dose per fraction (D/fr) ranged from 1.5 to 5 Gy. For all the plans in the dataset, dedicated PSQA sessions were performed, and the delivery accuracy of each plan was accessed in terms of two different scores: the $PR_\gamma(3\%G, 2\ mm)$ and the $PR_\gamma\%(2\%L, 2\ mm)$. The former is typically used in the clinical routine, while the second allows for increased sensitivity in the detection of delivery errors [121, 122]. The two scores were computed by setting $TH = 10\%$ [123].

The PSQA measurements were performed with the `ArcCheck`® detector array without using the PMMA `CavityPlug`™. Prior to each PSQA measurement, each plan was recomputed on the homogeneous synthetic `ArcCheck`® CT (density of 1.1836 g/cm^3) with a grid size of 1.87 mm. Acquisition, analysis, and calculation of the PR_γ%s were performed with the software `SNC Patient`™ version 6.7.

3.1.2 *Metrics Extraction*

A set of 65 metrics was extracted from the dataset using the `TCoMX` extraction engine of the `UCoMX` package [124], which was presented in Ch. 2. Within the scope of the analysis presented herein, the extracted metrics were subdivided into two groups: the existing ones reported in the literature, and a set of new ones chosen among those developed as part of this thesis work. In the following, the two groups will be referred to as *Old* and *New*, respectively. As presented in Ch. 2, some metrics are endowed with free-parameters that can be tuned by the user. Therefore, different extraction parameters were set on the same metric, and this resulted in a total amount of 29 and 36 metrics for the *Old* and *New* groups, respectively.

A comprehensive list of the extracted metrics and corresponding parameters is reported in Tab. 3.1. In the case of the $CLNS_n$, $n = \{20, 30, 50, 100\}$ were chosen, whereas for the $CFNS_{nx100}$ $n = \{0.05,$

$0.10, 0.5, 0.75, 0.9\}$ were set. All the possible variants of the $LnNS$ and CLS were computed, whereas a choice was made for the $ELOTV_{\Delta p}$ and $EPSTV_{\Delta p,\Delta_l}$. The former was computed for $\Delta p \in [1, 5]$ to show the behavior one should expect when the projection step is increased. The case $\Delta p = 1$ was chosen to show the relation in Eq. (2.62) explicitly. The $EPSTV_{\Delta p,\Delta l}$ was computed for three different combinations of $(\Delta p, \Delta l)$, namely, $(1, 1), (1, 0), (0, 1)$. The first pair of parameters was chosen to show the relation in Eq. (2.65) explicitly, whereas the other two in order to consider only the projection and leaf directions, respectively.

Table 3.1: List of the metrics extracted in the work.

Category	**Group** → **Subcategory** ↓	**Old** (29)	**New** (36)
TPS	Delivery	Pitch, FW, PT, GP, TT, TL, CS, CT, N_{proj}, N_{rot}, MF, TTDF	
LOT statistics	Absolute LOT	mLOT, sdLOT, mdLOT, moLOT, minLOT, maxLOT, $CLNS_{100}$, $CLSN_{50}$, $CLSN_{30}$, $CLNS_{pt,20}$	kLOT, sLOT, $CLSN_{20}$
	Relative LOT		mFLOT, sdFLOT, moFLOT, mdFLOT, minFLOT, maxFLOT, $CFNS_{5}$, $CFNS_{10}$, $CFNS_{50}$, $CFNS_{75}$, $CFNS_{90}$

Continued on next page

Table 3.1: List of the metrics extracted in the work. (Continued)

Category	**Group →** **Subcategory ↓**	**Old** (29)	**New** (36)
Sinogram	Geometry	L0NS, L1NS, CLS, CLS_{in}	L2NS, nCC, lengthCC, TA, fDISC, $CLS_{in,area}$, $CLS_{in,disc}$, $CLS_{in,area,disc}$, centroid
	Modulation	PSTV, LOTV, MI	nOC, $EPTSV_{1,1}$, $EPSTV_{0,1}$, $EPSTV_{1,0}$, $ELOTV_1$, $ELOTV_2$, $ELOTV_3$, $ELOTV_4$, $ELOTV_5$, mSI, mdSI, sdSI, MSA

3.1.3 *Correlations between metrics*

To determine, the reciprocal relation between the different aspects described by the available metrics, a correlation analysis was performed. This allows highlighting possible intrinsic dependencies between different properties of HT plans, and to investigate the mutual relation between extracted metrics. The analysis was performed on the entire dataset with no distinction among groups of plans. Both the inter- and intra-subcategory correlations were analyzed. Spearman's correlation coefficient was chosen as the indicator of correlation, and the significance level was set to 0.05. As generally accepted, the correlation values were classified into five classes: very weak ($0 \leq |r| < 0.2$), weak ($0.2 \leq |r| < 0.4$), moderate ($0.4 \leq |r| < 0.6$), strong($0.6 \leq |r| < 0.8$), and very strong ($|r| \geq 0.8$).

In Fig. 3.1, the full correlation map between the metrics is shown. All the five subcategories introduced before are considered, and the metrics in each subcategory are sorted as in Tab. 3.1. Only strong or very strong correlations are purposely shown. The two clusters of metrics composed of Delivery, Absolute LOT, and Relative LOT on one side, and Geometry and Modulation on the other have shown to be not (very) strongly correlated each other. As expected from Eq. (2.69), a very strong negative correlation between nOC and CLS ($r = 1$) was found, which confirms the inverse linear dependence between the two quantities. Strong correlations were found between MF and

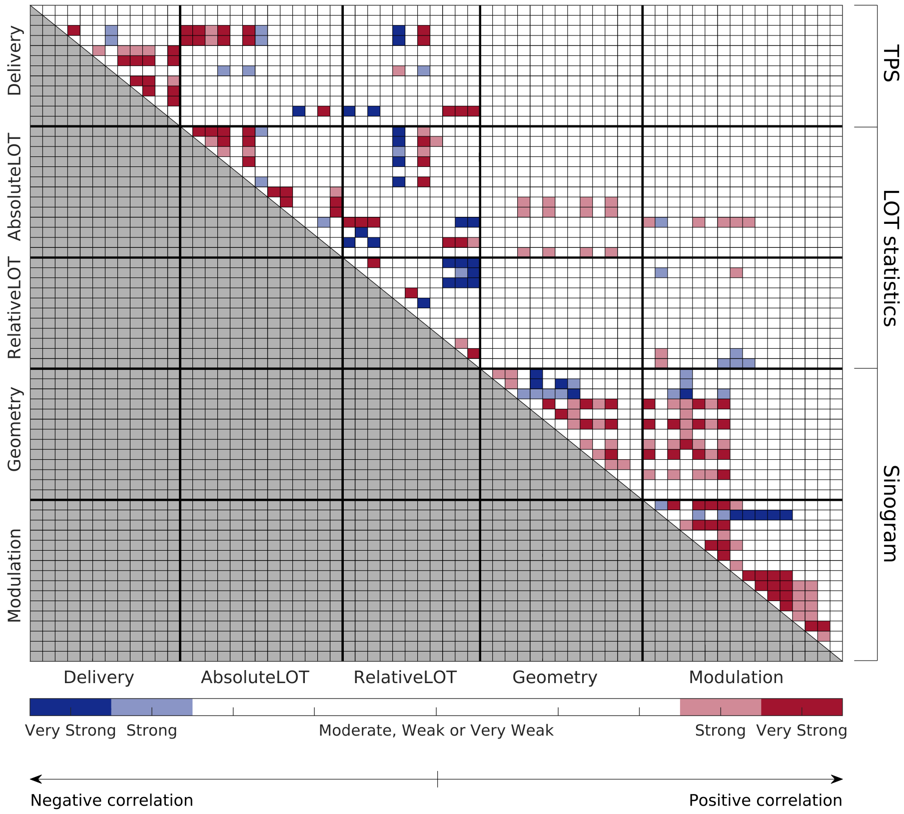


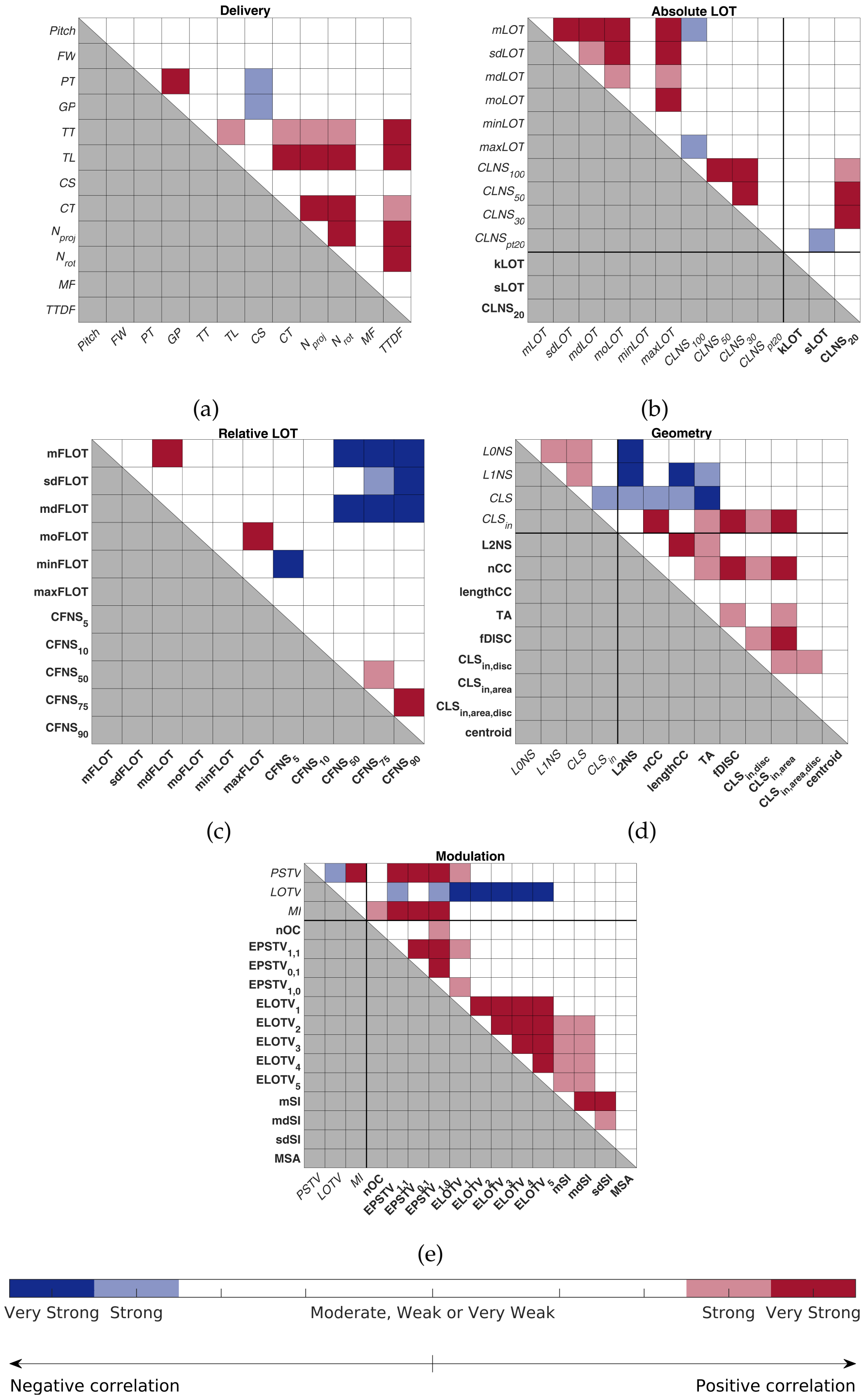
Figure 3.1: Complete inter-subcategory Spearman's correlation map. Only strong and very strong correlations are reported. The black thicker solid lines mark the separation between the subcategories.

$CLNS_{pt,20}$ $(r = -0.81)$, between $(MI, EPSTVs)$ and the CLS_{in} $(r > 0.74)$, and between the $L1NS$ and CLS $(r = 0.76)$.

In Fig. 3.2, the intra-subcategory correlations are shown. Among all, it is interesting to observe that the analysis of the geometrical properties of the sinogram shows that TA is at least strongly correlated with most of the metrics describing the discontinuity of the projections: the different variants of the CLS $(r = -0.98)$, $nCC(r = 0.68)$, $fDISC$ $(r = 0.65)$, and $L1NS$ $(r = 0.66)$. No(very) strong correlations were found instead between the centroid and the other geometrical metrics. Three groups of strongly correlated metrics arose in the subcategory Modulation: the $EPSTVs$, the $ELOTVs$, and the SIs. A strong positive correlation was found between MI and nOC $(r = 0.63)$. No strong correlations were found between the MSA and any other metric.

3.1.4 *Complexity metrics and plan efficiency*

Following the analysis by Boyd and colleagues [65], the relation between delivery efficiency and complexity was investigated. Two independent sub-samples of plans were extracted from the dataset based



(a) (b)

(c) (d)

(e)

Figure 3.2: Significant Spearman's correlations between the metrics in each of the five subcategories considered. Metrics belonging to the *New* and *Old* group are marked using the bold and italic font, respectively. The black thicker solid lines mark the separation between the two groups.

on the $CFNS_{75}$ value. Plans with $CFNS_{75} < 25^{th}$ percentile represent the most efficient plans, while plans with $CFNS_{75} > 75^{th}$ percentile

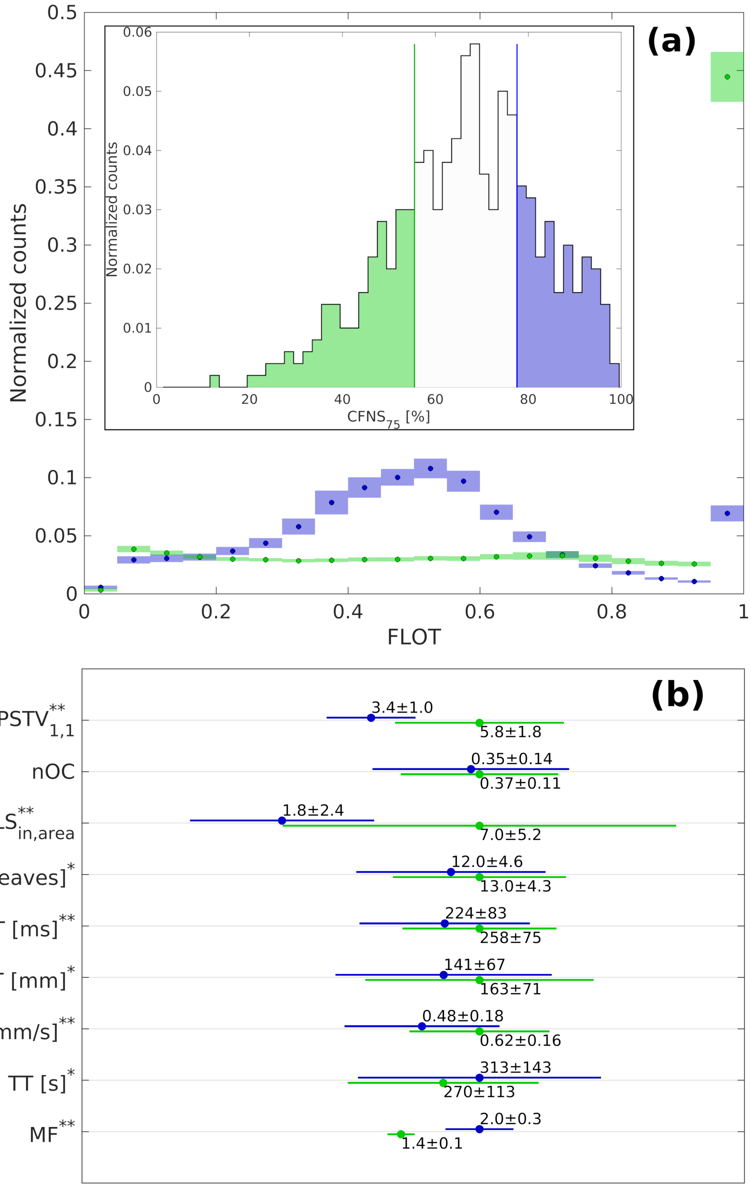


Figure 3.3: *(a) [Inner panel]* Distribution of the $CFNS_{75}$over the whole dataset of 464 plans. Data corresponding to the most (least) efficient plans are highlighted in green (blue). Vertical lines represent the 25th and 75th percentile. *[Outer panel]* Average FLOT histogram for the two sub-samples. Points mark the mean number of (normalized) counts for each bin and the colored boxes represent the 95% Gaussian CI around it. *(b)*. Mean values and standard deviations of the nine metrics selected. Green (blue) lines and markers refer to the most (least) efficient plans. The symbol * (**) near the metrics' names indicates statistically significant comparisons according to a Student'st-test at 0.05 (0.01) significance level

represent the least efficient plans. The two groups are compared using a Student's t-test.

Table 3.2: List of the metrics with the largest Spearman's correlation coefficients with the PSQA results.

Group	$\boldsymbol{PR_\gamma\%(2\%L, 2mm)}$		$\boldsymbol{PR_\gamma\%(3\%G, 2mm)}$	
	Metric	ρ	**Metric**	ρ
Brain	**MSA**	0.63	$\mathbf{CLS_{in,disc}}$	0.46
HN			**fDISC**	0.34
Thorax	LoNS	-0.36	LoNS	-0.25
Abdomen	mLOT	-0.49	**nCC**	0.47
Pelvis			TTDF	0.25
Prostate			minLOT	0.33
Others	FW	0.55	$\mathbf{CLS_{in,area,disc}}$	0.48
All	**sdSI**	0.21	**TA**	0.33

In Fig. 3.3, the average FLOT distributions obtained for the two sub-samples corresponding to the least and most efficient plans, respectively, are shown. The mean values of nine representative metrics selected from the different subcategories are reported. The least efficient plans showed significantly higher values of MF and TT and this resulted in significantly smaller values of $EPSTV_{1,1}$, $CLS_{in,area}$, TA, $mLOT$, TL,and CS. No statistical differences arose for the number of openings and closures nOC.

3.1.5 *Relation between the metrics and the PSQA results*

The relation between the metrics and the results of PSQA analysis was investigated through a correlation analysis. This analysis was performed for each of the seven categories of plans independently. Additionally, the whole dataset was considered too. The metric-PR Spearman's correlation matrix, $\rho_{metric,PR_\gamma}\%$, was computed for each of the two considered $PR_\gamma\%$. Only the metrics that turned out to be significantly correlated with the $PR_\gamma\%$ $\left(p_{value}\left(\rho_{metric,PR_\gamma}\%\right) < 0.05\right)$ were kept. Furthermore, if two metrics were (very) strongly correlated to each other, the one with the smallest $\left|\rho_{metric,PR_\gamma}\%\right|$ was removed. The number of metrics selected from each of the two groups after this procedure was collected and compared.

Figure 3.4 shows the numbers of metrics selected from the two groups of metrics (*Old* and *New*) and from each subcategory. At least one significant correlation was found with $PR_\gamma\%(3\%G, 2\ mm)$ for each group of plans. No significant correlations were found between the metrics and the $PR_\gamma\%(2\%L, 2\ mm)$ for HN, pelvis, and prostate. Metrics belonging to Geometry appear in 11 out of the 13 cases where significant correlations were found, 8 for the $PR_\gamma\%(3\%G, 2\ mm)$ and

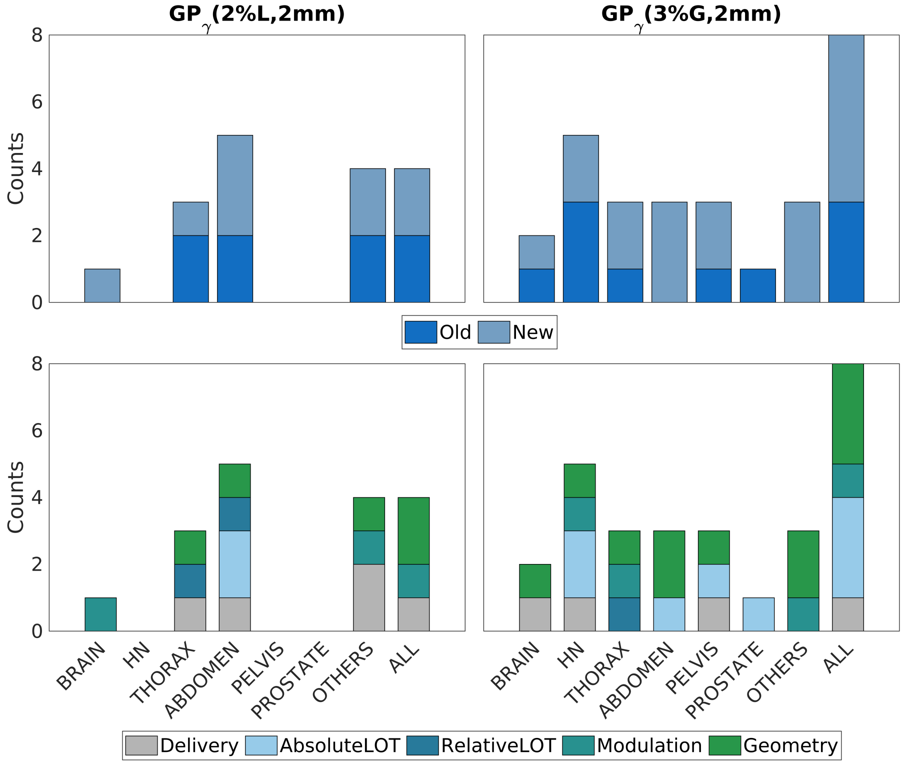


Figure 3.4: Barplots showing the final number of metrics (y-axis) correlated to the patient-specific quality assurance (PSQA) results after the selection procedure for each of the groups of plans considered (x-axis). *(Upper panels)* Separation between Old and New metrics is highlighted.*(Lower panels)* Separation between the subcategories is highlighted.

the remaining 5 for the $PR_\gamma\%(2\%L, 2\ mm)$. Geometry is followed by Delivery, Modulation, Absolute LOT, and Relative LOT in descending order. In Tab. 3.2, the metrics with the highest correlation with each of the two $PR_\gamma\%$ for each group of plans and for all the plans are reported. Seven out of thirteen metrics belong to the group *New*. In general, the correlation values span between weak and moderate, with one case of strong correlation between the MSA and the $PR_\gamma\%(2\%L, 2\ mm)$ for the brain.

3.1.6 *Impact and significance of the results*

The analysis of Spearman's correlation among the metrics confirmed some of the results already reported in the literature, in particular the very strong correlations among the $CLNS_n$, the strong negative one between MF and the $CLNS_{pt,20}$, the one between MI and $PSTV$ and the CLS_{in}, and the one between the $L1NS$ and CLS [68, 69, 119]. However, the introduction of a set of newly developed complexity metrics allowed focusing on additional aspects of the HT treatment plans. In particular, the introduction of the Relative LOT statistics, and more precisely the $CFNS_{75}$, allowed for the indenfication of two sub-samples of plans characterized by different characteristics LOT distribution's shape and treatment's efficiency. These results closely match the idea proposed by Boyd et al. [65]: the least efficient plans (those with $CFNS_{75} > 75^{th}$ percentile) exhibit an average bimodal LOT distribution with a first peak around $FLOT = 0.5$ and a second one around $FLOT = maxFLOT$. On the other hand, the most efficient plans (those with $CFNS_{75} < 25^{th}$ percentile) show an overall flat trend characterized by a marked peak in the last bin.

The modulation of delivery of the most efficient plans, namely those with small values of $CFNS_{75}$, is achieved mainly by acting on the geometry of the beam by splitting the field at each projection into two or more disjoint sub-fields. This results in a higher number of in-field closed leaves (higher $CLS_{in,area}$), as well as a generally higher number of leaves involved (higher TA). Additionally, they showed faster treatments (smaller TT and higher CS), despite that each projection was found to be less disjoint for the least efficient plans, namely those with high values of $CFNS_{75}$, so that the modulation seems to be mainly achieved by acting on the LOTs values (i.e., higher MF). The two groups of plans do not exhibit significant differences in the FW, making all the previous conclusions independent from it. Furthermore, no significant differences were observed in the PSQA results, suggesting that future implementation of the indicators proposed in this work in the optimization process might help to get efficient plans without affecting deliverability.

In general, the metrics in the *New* group have been shown to be at least as correlated to the PSQA $PR_{\gamma}\%$ results as the already existing ones. At least one of them was selected when there were significant, apart in one case ($PR_{\gamma}\%(3\%G, 2\ mm)$ for prostate). More precisely, the newly introduced metrics have shown to be generally more correlated with the ($PR_{\gamma}\%(3\%G, 2mm)$ where 19 out of the 29 metrics kept after the selection procedure belongs to this group. On the other hand, no net difference arose between the *New* and *Old* metrics for the $PR_{\gamma}\%(2\%L, 2\ mm)$ where the metrics kept after the selection procedure amount to 8 and 9, respectively.

In three cases, $PR_{\gamma}\%(2\%L, 2\ mm)$ for the brain, $PR_{\gamma}\%(3\%G, 2\ mm)$ for the abdomen and others), only metrics pertaining to the *New* group are selected as influential. Although they might be correlated with metrics belonging to the group *Old*, they completely substitute them when considered in the selection process. This suggests a higher predictive power. It is worth observing that all the significant correlations arising for "others" need to be taken with care due to the heterogeneity of anatomical districts included in it. No significant correlations were found between the complexity metrics and the $PR_{\gamma}\%(2\%L, 2\ mm)$ for prostate and HN cases, similarly to the results obtained in Ref. [69] for the 3D $PR_{\gamma}\%(2\%G, 2\ mm)$. On the other hand, the present study provides the first evidence of a set of complexity metrics specifically designed for HT obtained by combining existing metrics and the set of new metrics proposed herein, which show significant correlations with PSQA results for all the groups of plans considered when the $PR_{\gamma}\%(3\%G, 2\ mm)$ is used. Among all the considered metrics, geometrical ones are clearly the most related to the PSQA results. The correlation analysis with the $PR_{\gamma}\%(3\%G, 2\ mm)$ showed that they play a prominent role for the majority of the groups, with the most correlated metric belonging to this subcategory for five out of the seven groups of plans (brain, HN, thorax, abdomen, others) as well as for the complete set.

All the geometric complexity indicators give different measures of the discontinuity of the binary MLC openings. However, differently from what might have been hypothesized, the correlations are positive in all cases but one (thorax), suggesting that a higher geometrical complexity of the MLC openings might lead to higher plan deliverability. These results go in the same direction as the ones found in Ref. [69], where strong positive correlations of the CLS and the $L1NS$ with the 3D $PR_{\gamma}\%(2\%G, 2\ mm)$ were observed for stereotactic brain plans. However, its interpretation can now be completely reversed. As a matter of facts, based on the inverse linear relation between CLS and nOC, a larger number of closed leaves implies a smaller number of opening/closing leaves in motion. Following existing literature [67], this fosters the idea that increased deliverability might be related to a reduced contribution of leaf latency. In fact, despite the impact of small individual leaf errors is not significant [125, 126] their composition over the whole treatment might result in significant deviation in the dose deposition [127, 128]. Relying on the collected data, it is not possible to disentangle the two competing effects, therefore specifically designed experiments will be undertaken in the future.

The reported correlations between complexity metrics and PSQA results are strongly related to the clinical practice followed to obtain both the plans and the PSQA results (e.g., Tomotherapy unit, TPS, details of optimization approach, PSQA phantom, and its setup). Therefore, the relevance of the proposed metrics should be grounded and validated through large comprehensive comparisons across different centers. Furthermore, although the new metrics proposed and results obtained provide new intuition concerning possible new strategies to orient the planning process, to date these parameters are not directly accessible by the planner on the TPS during the planning process.

To summarize, the validation of the complexity indicators proposed herein has provided four major pieces of evidence. First, the new set of HT complexity metrics is complementary to the existing one and allows for a more in-depth characterization of HT plans complexity. Second, the geometrical properties, in particular the discontinuity of the beam projections (as measured by $CLS_{in,area,disc}$, $CLS_{in,area,disc}$,$fDISC$, and nCC), are the most related to PSQA results, at least in the considered database. Third, the intrinsic relation observed between the beam spatial discontinuity and the number of leaf opening/closing, that is, $CLS \approx nOC$, allowed for a reformulation of the conclusions reported by Santos and colleagues [69] and to enforced the role of leaf latency on plan deliverability. Fourth, FLOTs seem to discriminate among plans having different levels of delivery efficiency, extending what was qualitatively described by Boyd and colleagues [65].

3.2 DECISION-SUPPORT TOOLS FOR PSQA OF HT PLANS

The availability of robust and reliable complexity indicators to predict the dosimetric and delivery accuracy of treatment plans would impact the radiotherapy workflow greatly. Indeed, as discussed in Ch. 1, the conventional measurement-based PSQA approach highly impacts the radiotherapy workload. This asks for alternative quality assurance tools and protocols. Decision-support models based on complexity metrics may be suitable candidates for implementing virtual PSQA (vPSQA) programs to support planners in this process. However, the development of such tools for HT is still in the embryonic stage.

Following the validation of the complexity indicators for HT proposed in Sec. 3.1, in this section proposes an analysis aimed at combining complexity metrics, delivery parameters and radiomics features, to create to create prediction models as decision-support tools for PSQA of HT plans. These models were repeatedly trained and cross-validated on a large clinical dataset of HT plans generated by two treatment planning systems (TPSs). Ultimately, this research aimed to provide guidance on the potential use of such models in clinics as part of implementing a vPSQA program, and to quantify the PSQA workload reduction that could be expected from their use.

3.2.1 *Dataset*

The considered dataset contained 881 clinical HT treatments that were delivered between March 2019 and September 2022 using the RadixAct (Accuray, Inc., Madison, Wisconsin, USA) treatment unit available at our institution. In total, 512 treatments were planned using Precision v1.0.02 TPS (Accuray, Inc., Madison, Wisconsin, USA) equipped with the GPU-based optimizer VoLO™, while the remaining 369 were created with RayStation v8a and v9a (RaySearch Laboratories, Stockholm, Sweden) (Fig. 1a). Collapsed Cone Convolution Superposition dose calculation algorithm [110, 120] was used on both TPSs. The two TPSs will be referred as TPS 1 and TPS 2, respectively. The dataset is an extension of the one already described in Sec. 3.1 and contained treatment plans for several sites including brain, head and neck, thorax,

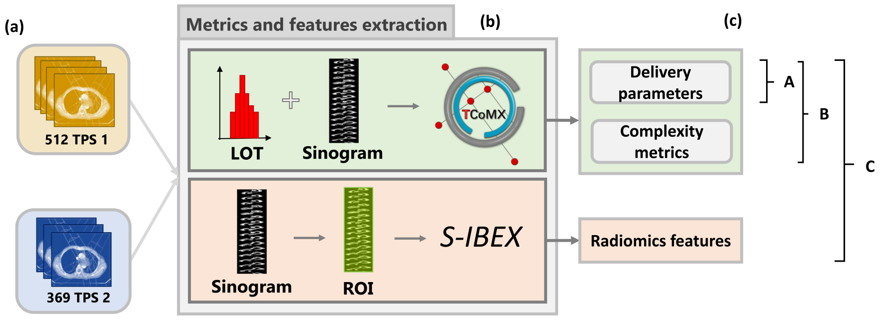


Figure 3.5: Schematic summary of the workflow of the metrics extraction and the creation of the groups of variables. (a). Sixty-five indicators including 12 delivery parameters and 53 complexity metrics were computed using the dedicated library TCoMX. Additionally, 174 radiomics features were extracted from the plans' sinograms using SIBEX (b). Three groups of variables were created (c).

abdomen, pelvis and prostate. The dose per fraction (D/fr) ranged from 1.5 Gy to 8 Gy.

Dedicated PSQA verifications were performed using the `ArcCheck`® detector array (Sun Nuclear Corporation, Melbourne, FL, USA) without using the PMMA `CavityPlug`™. The plans were recomputed on the homogeneous synthetic `ArcCheck`® CT. Acquisition, analysis and calculation of the $(PR_{\gamma}\%)$ were performed with `SNC Patient`™ version 6.7. In this work, $PR_{\gamma}\%(3\%G, 2\ mm)$ at 3% used in our clinical routine was considered [23], with $TH = 10\%$ set for its computation [123]. Universal action level $PR_{\gamma}\%(3\%G, 2mm) > 90\%)$ and tolerance level $PR_{\gamma}\%(3\%G, 2\ mm) > 95\%)$ are used in our clinical practice.

In total, 239 variables were extracted including 12 delivery parameters (DP), 53 complexity metrics (CM) and 174 radiomics features (RF). Three groups of variables called A (DP), B (DP + CM) and C (DP+CM+RF) were created and used as the input for training three XGB regression models for each TPS (Fig. 3.5 b-c). When group B and C were created, only variables that were not strongly correlated ($r_{Pearson} < 0.90$) with those already contained in the previous groups (A and B, respectively) were added to limit the amount of redundant information. Delivery parameters and complexity metrics were extracted using the `TCoMX` extraction engine of the `UCoMX` package described in Ch. 2. The indicators were chosen among those proposed in the literature [69, 118] to describe many aspects of HT treatments delivery, field geometry and modulation, and properties of the leaf open time (LOT) distribution. Radiomics features were extracted from the sinograms of the plans to quantify their textural, intensity and morphological properties (Fig. 3.5b) using the software SIBEX [116] which was chosen for its compliance with the IBSI standard [117] and for the extensive set of implemented features [115]. Each sinogram was exported as a single-slice CT image to DICOM format. A region-

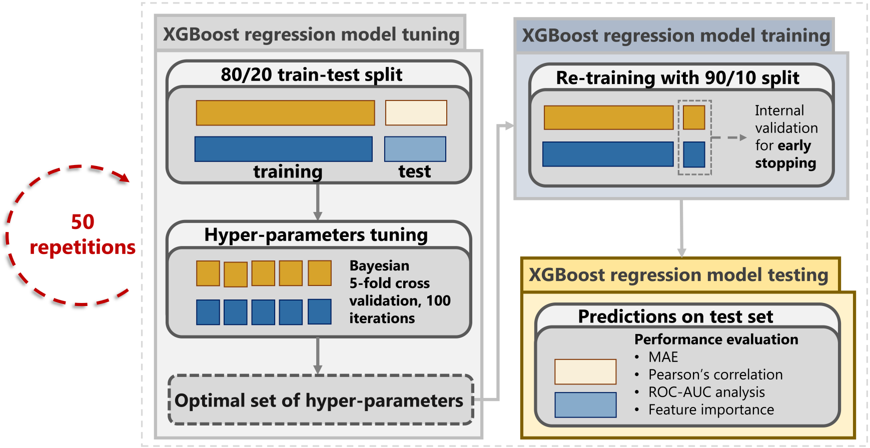


Figure 3.6: Schematic of the workflow adopted to train the XGBoost regression models.

of-interest (ROI) was associated to each sinogram image as the binary mask corresponding to its active area (i.e. non-zero LOT values). The radiomics features were extracted from the resulting ROI.

3.2.2 *Training and testing XGboost regression models for vPSQA*

Table 3.3: List of the XGB hyper-parameters tuned through the Bayesian cross-validation procedure, and corresponding search-space ranges. The names reported in this table match those used in the code.

Hyper-parameter name	**Search space**
col_sample_bytree	Real, [0.1; 1.0]
reg_alpha	Integer, [0; 5]
reg_lambda	Integer, [0; 5]
learning_rate	Real, [0.0001; 1.0]
max_depth	Real, [0.0; 20]
gamma	Real, [0.0; 1.0]
min_child_weight	Integer, [1; 10]
subsample	Real, [100; 2000]
n_estimators	Integer, [100; 2000]
scale_pos_weight	Real, [0.1;1.0]

The prediction capability of the variables in the three groups was assessed. XGBoost [129] (XGB) regression models were trained to predict the $PR_{\gamma}\%$ at 3%. XGB is a gradient-boosted decision tree machine learning library and it was chosen for its proven effectiveness [90, 101], flexibility, easiness of implementation, in-born regularization and easy interpretability of the results through the feature importance evaluation. Fig. 3.6 sketches the strategy adopted for training and testing the models. For each TPS, the set of plans was split randomly into two disjoint subsets (80% training, 20% test). The test set did not enter any stage of the training process to avoid any form of training-test leakage that could have altered the reliability and robustness of the results.

Five-fold Bayesian cross-validation consisting of 100 iterations was applied to tune some hyper-parameters of the model. Bayesian cross-validation exploits Gaussian Process [130] to explore a user-defined set of hyper-parameters and return the optimal ones. It is an alternative to brute-force grid-search approaches or randomized search ones which has the advantage of exploiting the information at each iteration to orient the next one. The list of XGB hyper-parameters considered and the corresponding search spaces is summarized in Tab. 3.3. More details can be found in Ref. [131]. Using the best set of hyper-parameters, the model was re-trained: 90% of the training data was used to train the final model, while the remaining 10% was used as an internal-validation set to determine an early-stopping point for the training. At the end of each training iteration, the performance of the model was calculated on the internal validation set. In case the latter did not improve for 10 iterations, the training was stopped. This strategy may help preventing the overfitting. Once trained, the model was applied to perform predictions on the test set. For each group of variables, (A, B and C), this procedure was repeated 50 times by changing the training-test split at each repetition. For the sake of comparison, the same set of 50 seeds was used for each group to split the dataset.

The performance of the trained model on the test set was evaluated using two metrics: the mean absolute error (MAE) and the Pearson's correlation coefficient (PCC), both calculated between the true and the predicted $PR_{\gamma}\%(3\%G, 2mm)$. The global performance on the test set for each of the three groups of variables was computed as the average MAE and PCC over the 50 repetitions. The performance of the different groups was compared using a Wilcoxon signed-rank test with a significance level of 0.05. A distinguishing feature of XGB is the built-in feature importance estimation. In this work, the 'Gain' was used to measure such importance. This metric quantifies the improvement in regression accuracy induced by a feature to the branches of the tree it is present. A larger Gain of a specific variable indicates a greater importance to make a decision. For each group of variables, a global measure of the importance of the different predictors was obtained by averaging the Gain over the 50 repetitions. Variables were ranked in descending order using those values.

A crucial step of the PSQA process is to determine whether the plan owns the required deliverability criteria. According to reference

AAPM guidelines adopted in our clinical practice [15], the clinical tolerance level on the $PR_{\gamma}\%(3\%, 2mm)$ was set to 95%. Plans were subdivided into two disjoint classes, with the positive one corresponding to plans having measured $PR_{\gamma}\%(3\%, 2mm) \leq 95\%$, that are called 'non deliverable' plans in this work. The scores returned by the trained models were used to predict whether a plan was 'deliverable' or 'non-deliverable'. This procedure was applied for each group of variables at the end of each repetition. The classification performance was evaluated through ROC analysis. The latter provides a curve of sensitivity (or true positive rate, TPR) and specificity (or true negative rate, TNR) for all possible cutoff values on the $PR_{\gamma}\%(3\%G, 2mm)$ predicted by the model, and the corresponding area under the curve (AUC) which provides a cutoff-independent measure of classification performance. In this work, sensitivity is the ability of the model to correctly label non-deliverable plans as failing; conversely, specificity is the ability to label a deliverable plan as passing. Additionally, to quantify the expected workload reduction associated with the use of the model in our clinical workflow, for each repetition the highest value of specificity corresponding to 100% sensitivity was collected. This corresponds to a condition where all non-deliverable plans and a fraction of deliverable plans are correctly labelled. In a clinical use of the model, this fraction of plans would not need the measurement-based PSQA. The average ROC curve, AUC and fraction of deliverable plans correctly detected at 100% sensitivity over the 50 repetitions were computed.

3.2.3 *Performance of the models*

Fig. 3.7 shows that for both TPSs improvements (i.e. decreasing MAE and increasing PCC) were observed when CM and RF were included in the set of input predictors. Significant improvements were observed in all cases but one (i.e. TPS 1 MAE between groups B and C). The improvement was generally larger for TPS 2 which also exhibited better MAE and PCC compared to TPS 1: 1.61 ± 0.12 and 0.33 ± 0.09 for TPS 1 versus 1.05 ± 0.12 and 0.46 ± 0.11 for TPS 2. For both TPSs, the best average performance was reached for group C, despite the comparison between B and C was statistically significant for TPS 2 only.

Table 3.4: List of the five most important predictors for the two TPSs and the three groups of variables. 'Gain' was used as the importance score to sort the variables and the results here reported refer to the average over the 50 repetitions.

Group →	**A**	**B**	**C**
TPS ↓			

Continued on next page

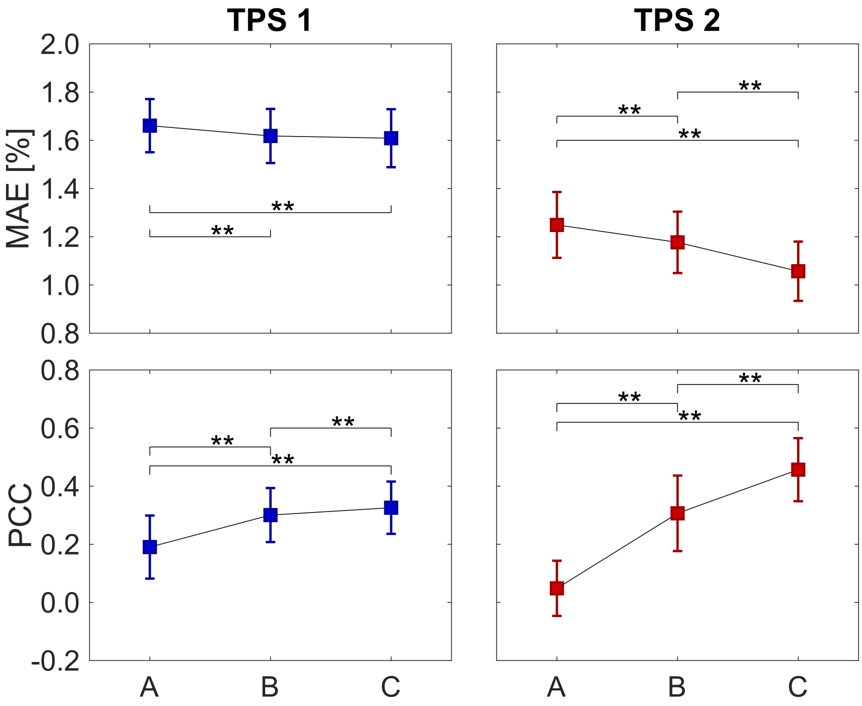


Figure 3.7: Trend of the MAE (upper panel) and PCC (lower panel) on the test set for TPS 1 (left) and TPS 2 (right) for the three groups of variables. Square markers indicate the average over the 50 repetitions, while errorbars represent the standard deviations of the corresponding samples. The '**' mark groups with performance that resulted statistically different to a Wilcoxon signed-rank test at 0.05 significance level. Abbreviations: DP, Delivery Parameters; CM, Complexity Metrics; RF, Radiomics Features; MAE, Mean Absolute Error; PCC, Pearson's Correlation Coefficient

Table 3.4: List of the five most important predictors for the two TPSs and the three groups of variables. 'Gain' was used as the importance score to sort the variables and the results here reported refer to the average over the 50 repetitions. (Continued)

Group →	**A**	**B**	**C**
TPS 1	Number of projections	CLS	LIF LocalIntensity Peak
	Target length	sdSI	sdSI
	Treatment Time	MSA	ID Intensity-based Energy
	Projection Time	L0NS	MSA
	Couch Speed	Centroid	Centroid

Continued on next page

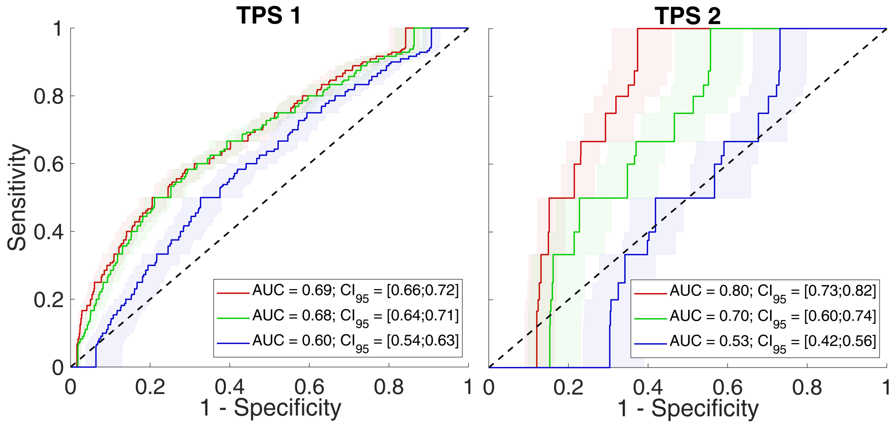


Figure 3.8: ROC curves for the two TPSs on the test data. Results for the models trained on A (blue), B (green) and C (red) are reported. Each line represents the average ROC curve over the 50 repetitions and the shaded areas the corresponding 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3.4: List of the five most important predictors for the two TPSs and the three groups of variables. 'Gain' was used as the importance score to sort the variables and the results here reported refer to the average over the 50 repetitions. (Continued)

Group →	**A**	**B**	**C**
TPS 2	Treatment Time	minLOT	GLDZM High Grey Level Zone Emphasis
	Pitch	moFLOT	minLOT
	Number of projection	sdSI	GLDZM Grey Level Variance
	Projection Time	sdFLOT	NGLDM Low Dependence High Grey Level Emphasis
	Couch Speed	Pitch	NID Complexity

From the results summarized in Table 3.4, it can be observed that the behaviors of the two TPSs were quite similar for group A, while they differed in the other two cases. In particular, for group B, three geometrical and two modulation metrics were the most important ones for TPS 1, while three LOT/FLOT statistical metrics, one modulation metric and one delivery parameter were the most relevant for TPS 2.

The average ROC curves and AUC values reported in Fig. 3.8 shows that Group A was modestly predictive for TPS 1, while it was completely ineffective for TPS 2. An evident increase in the AUC was observed whenever a group of variables was added to the input predictors for TPS 2. Conversely, for TPS 1 the results obtained for groups B and C were very close and better than those obtained for group A. On average, the use of model C would have allowed the detection of around 16% and 63% of the deliverable plans for TPS 1 and TPS 2 with 100% sensitivity, respectively. Considering that the two datasets included 445 and 349 deliverable plans for TPS 1 and TPS 2, respectively, this corresponds to 291 plans (i.e. $0.16 \times 445 + 0.63 \times 349$). In a real clinical scenario, this would have allowed for approximately 35% (291/881) PSQA workload reduction.

3.2.4 *Significance and impact of the results*

The analysis presented in this section was the first attempt to develop models to support the PSQA of HT plans. Three models trained on different groups of predictors were used to perform PSQA predictions. Using the best model (model C) approximately 16% and 63% of the deliverable plans for TPS 1 and TPS 2, respectively, could have been detected with 100% sensitivity, reducing the whole PSQA workload of about 35% . A fully comprehensive and unbiased comparison of the effectiveness of different groups of indicators that included all the state-of-art complexity metrics reported in HT literature [69, 118] and radiomics features extracted from the sinograms [115–117] was proposed.

As the second key novelty, a dataset including plans created using alternatively one of the two TPSs available for HT planning was considered [132, 133]. The results showed that the PSQA of the plans obtained with TPS 2 can be more successfully predicted than those obtained with TPS 1, despite providing an extensive paired comparison was beyond the aim of this analysis. We know from clinical practice that, plans created with TPS 2 have lower inter-planner variability because of the limited number of planners who use that TPS (two experienced planners) compared to those using TPS 1 (six planners with different experience levels). It follows that the $PR_{\gamma}\%(3\%G, 2mm)$ of the plans created using TPS 1 is a process which is intrinsically noisier because of the wide variety of planning strategies and planners' experiences. This might explain the higher discriminative power observed in the ROC analysis for group A on TPS 1. In fact, the planning approach - which is the same for both TPSs to the extent the two optimizers allow - aims at maximizing the delivery efficiency while keeping a satisfactory clinical quality. To this aim, the planner can act on delivery parameters only, namely trying to minimize the Modulation Factor on TPS 1 and the delivery time on TPS 2. Thereby, personal habits and experience might lead to a large variability of planning choices [134].

The choice of the TPS is related to the everyday management of our department (human resources, satellite department, distributed computing possibilities). Given the higher predictive ability of the models on TPS 2, a possibility could be increasing the fraction of plans created using the latter to increase the expected workload reduction associated with the possible use of the model. However, further investigations should focus on the paired comparison of the two TPSs on the same set of plans to get a more comprehensive picture [135]. The effectiveness of vPSQA models may be affected by several factors related to the clinical practice adopted, including the details of the planning stage. Therefore, monitoring the stability of the planning process could be an important element to consider for reducing the impact of many confounding factors, such as the experience of the operator, that can affect or mask the actual predictive power of the proposed predictors. In this context, extending the use of the statistical process control techniques proposed in the AAPM-218 Task Group [23] also to set center-specific action and tolerance limits on the complexity indicators might be a possible solution to harmonize the planning process. Therefore, the availability of standardized and integrated tools to compute all those indicators during the planning may allow for the online monitoring of this process [101].

The results reported herein are extremely similar to the recent studies which proposed predictive models for vPSQA of IMRT and VMAT plans delivered with C-arm linacs. Relevant works found in the literature reported on models that could predict the $PR_\gamma\%$ rate with MAEs between 0.7% and 4.2% and AUC scores between 0.7 and 0.9 [93, 101, 136]. The variability of the results is due to several factors including the size of the datasets, the models' details and the particular $PR_\gamma\%$ considered, as well as factors related to the clinical practice adopted at each institution. In this context, a number of works have investigated the use of radiomics features to characterize the properties of IMRT/VMAT fluence maps, gamma maps and dose difference maps for vPSQA and error-detection purposes [71, 74, 99, 137]. However, a part from the different treatment technology, the additional value of our paper is the use of the open-source IBSI-compliant software SIBEX which makes the conclusion of this work more generalizable and may facilitate its implementation.

Additionally, the effectiveness of radiomics features to detect sinogram properties that can predict PSQA performance opens to future investigations of automatic features extractions strategies, such as Convolutional Neural Networks (CNNs) [87, 137, 138]. The validity of the results reported is limited to the equipment and the specific measurement protocol adopted in our center for the PSQA, including the specific detector used. It could be extended through a multi-institution comparison but this was far beyond the scope of this analysis [135]. A second partial limitation of the proposed analysis is that each pathology was not considered separately. The choice of using the whole set of plans to train and validate the XGB models was mainly driven by the amount of data available, since the number of plans for each pathology was too small to create pathology-specific models. This direction will be explored by adding new plans to the dataset. However, this did not affect the comparison of the models which was performed on the same sets of input data. In conclusion, the results presented fill the gap towards the formulation of a model for vPSQA in HT, foster the use of complexity metrics and radiomics features and pave the way to clinical applications in the near future.

4

THE COMPLEXITY OF VMAT PLANS

Radiation therapy's ever-increasing drive for more conformal and less harmful treatment plans has driven treatment technologies to deliver plans of ever-increasing complexity. In Ch. 3, we have seen examples of this behaviour in the context of HT plans. THIS CHAPTER presentes original results for the VMAT technique.

VMAT is the standard of care treatment technique performed on C-arm linacs for most diseases and treatment sites. The success of VMAT is due to its ability to delivery highly spatially modulated dose distribution efficiently through the simultaneous variation of machine's gantry speed and dose rate, and MLC conformation. However, higher rates of variation delivery parameters as well as using complex MLC configurations might lead to an increased plan complexity [112] which can ultimately impact the both the dose calculation accuracy and treatment delivery, and increase the total treatment delivery time. Since treatment plans with similar dose distributions may exhibit significant differences in plan complexity, the planning effort be aimed at obtaining a clinically optimal dose distribution which allows for accurate and robust radiation treatments, while minimizing the overall plan complexity [32, 112, 135].

It was discussed in Ch. 1 that an increasing number of studies have proposed various strategies to predict the phantom-based PSQA outcomes of VMAT plans based on complexity metrics [71, 85, 87, 90, 101, 136]. Alongside the efforts devoted to these predictive tools, the possibility to actively reduce, control, and quantify plan complexity is only marginally present in the commercial TPSs and sparsely implemented in the clinics worldwide [113]. The first explicit tool devoted to limit VMAT plan complexity appeared only a few years ago in Eclipse v15.5 (Varian Medical System, 2017), the so-called Aperture Shape Controller (ASC). This tool controls the local curvature and regularity of the beam eye views' (BEVs) shape [32, 139–141] and helps in reducing the amount of challenging MLC configurations that may affect the dose calculation accuracy. Additionally, the ASC is complemented by the possibility of limiting the total number of deliverable Monitor Units (MUs). When properly tuned, these tools have been shown to increase the delivery accuracy of the treatments without significantly compromising their clinical quality [32, 139, 142–144].

On the other hand, the possible knowledge gained by using TPS-specific tools is difficult to transfer to the rest of the community using different planning systems. In fact, there is still limited and not exhaustive literature evidence on the relationship between the aforementioned complexity limitation strategies and the existing complexity metrics [32, 139, 142]. Furthermore, there is no shared consensus yet on which of the proposed complexity metrics are the most appropriate and effective to accomplish this objectives. The aim of this work is, first, to provide a detailed investigation of how different combinations of the two aforementioned tools affect the clinical quality and deliverability of VMAT treatment plans. Then, the definition of operational limits on the complexity indicators to be used to identify deliverable plans prior to the measurement-based PSQA is proposed.

4.1 PLANS WITH DIFFERENT LEVELS OF COMPLEXITY

4.1.1 *Plan complexity limitation strategies*

When optimizing VMAT plans on the Eclipse TPS (Varian Medical Systems, Palo Alto, CA), plan complexity can be directly limited by acting on two knobs: the Aperture Shape Controller (ASC) value and the maximum number of deliverable Monitor Units. The ASC is a component of the VMAT leaf sequencer available in the PO algorithm from version 15.5. By increasing the penalty of the ASC, apertures with minimal local curvature (associated with the positions of adjacent leaf tips modulating the same spatially continuous target projection) are favored. The user can choose from five discrete ASC levels to vary the value of a multiplicative penalization term in the optimization cost function (Varian Medical System, 2017). The ASC values allowed are “Very Low”, “Low”, “Moderate”, “High” and ““Very High”. The maximum total number of deliverable Monitor Units can be set via the MU objective to a specific maximum MU number with an objective weight that can vary between 0 and 100.

4.1.2 *Patient sample and baseline plans*

Thirty patients treated in the previous twelve months were retrospectively retrieved from our clinical database:

- Ten patients treated for stage III or IV oropharyngeal or nasopharyngeal primary tumors. Radiotherapy doses to high- and intermediate-risk planning target volumes were 69.96 Gy and 59.4 Gy delivered in 33 fractions (2.12 and 1.80 Gy/fx). Contouring was based on various international guidelines adopted at our institution [145–148]

- Ten patients with a histologically confirmed diagnosis of stage III lung cancer (either squamous cell carcinoma or adenocarcinoma) were treated with radical EBRT to the tumor and involved lymph nodes. The delivered dose was 60 Gy in 30 fractions. The contouring practice followed the ESTRO ACROP guidelines for NSCLC [149].
- Ten patients treated for low-risk prostate cancer with a moderately hypofractionated regimen 60 Gy in 20 fractions following recent evidence [150–152]. CTV was the prostate gland only, delineated according to the NCCN v.3.2016 guidelines and the PTV was obtained with a 7 mm isotropic margin, except in the posterior direction where the margin was reduced to 5mm. All plans were optimized according to the prostate radical treatment protocol of our department, which is based on the ASTRO/ASCO/AUA guidelines [153].

A set of baseline plans were generated without any specific strategy to limit plan complexity and are used as a reference in all the comparisons. They were planned using Eclipse TPS v15.6.1 for a TrueBeamSTx with high-definition MLC (Varian Medical Systems, Palo Alto, CA, USA) and 6 MV energy. Optimization was performed with the Photon Optimizer v.15.5 and calculated with AcurosXB v.15.5 dose engine (Varian Medical Systems, Palo Alto, CA, USA). Prostate plans were planned with two full arcs with approximately orthogonal collimator rotations: one close to 5°, the other close to or exactly at 90°. Head and Neck plans were obtained with two complete arcs with complementary collimator angles, approximately $\pm$ 10°. Occasionally, a third arc with 90° collimator rotations was added to meet minimal clinical requirements. Lung plans were generated using 2 or 3 arcs with complementary or orthogonal collimator arrangements, depending on the PTV extension and laterality.

4.1.3 *Combining the ASC penalty and MU/cGy limit*

Each baseline plan was re-optimized several times using different complexity limitation strategies using the tools presented in Sec. 4.1.1. Each optimization was performed without any human intervention and by using exactly the same geometry and optimization objectives as for the corresponding baseline plan. Furthermore, in order to avoid any possible bias resulting from human intervention during the optimization process, the baseline plans were re-optimized without any interaction from the planner. This procedure resulted in six (seven) treatment plans for each HN and Lung (Prostate) patient, for a total amount of 190 treatment plans. A summary of the strategies adopted to limit complexity is shown in Tab. 4.1.

Table 4.1: Summary of the ASC and MU/cGy values associated with the different complexity limitation strategies adopted to realize the plans in the dataset used for this study.

MU/cGy $\rightarrow$ **/ ASC** $\downarrow$	≤ 3	≤ 4	≤ 5
Very High	VH_MU3	VH_MU4	VH_MU5 (Prostate)
Moderate	Mod_MU3	Mod_MU4	
Off	Off_MU3	Off_MU4 (HN,Lung)	Off_MU5 (Prostate)

4.2 INVESTIGATING QUALITY, DELIVERABILITY AND COMPLEXITY

4.2.1 *Assessment of the clinical quality*

The clinical quality of the treatment plans was assessed via the Plan Quality Metric (PQM) score, in order to limit the subjectivity of judgment and to facilitate the comparison between plans. The PQM was first introduced by Nelms and colleagues [121] and it is implemented in the PlanIQ software (v2.1.1, Sun Nuclear Corp., Melbourne, FL). The PQM is a user-defined metric that summarizes the judgment of quality into a single number and attempts to mimic the judgment of a clinical team. It consists of a list of submetrics (e.g. DVH metrics), each of which has an associated numerical scoring function designed to model as accurately as possible the clinicians' judgment criteria. The final PQM score is the sum of the scores obtained for each sub-metric and measures the extent to which the plan conforms to the list of defined goals. The percentage PQM, $PQM\%$, is the ratio of the achieved score to the maximum achievable score and provides a relative measure of the soundness of the plan. In this study, a dedicated PQM algorithm was used to assess the plans for each of the three sites considered. In order to assess whether plans were obtained with different trade-offs between target coverage and OARs sparing, the $PQM\%$ was also split into two complementary measures: the $PQM_{target}\%$ which combines together all of the sub-metrics related to target coverage, homogeneity, and conformance; and the $PQM_{OARs}\%$ which represents the sum of the sub-metrics related to OARs sparing. The description of the three PQM algorithms can be found in the Supplementary Material.

Fig.4.1 shows the clinical quality as assessed via the $PQM\%$ score. No statistically significant differences were observed, on average all plans performed well compared to the relative reference. This conclu-

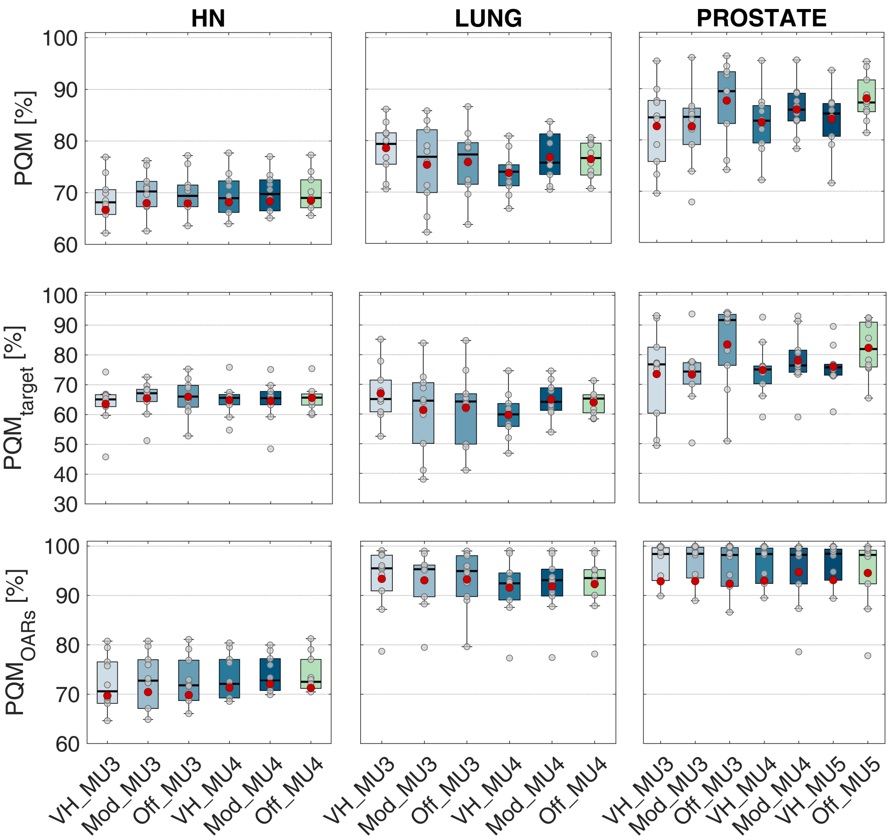


Figure 4.1: Comparison of the PQM% score of the plans at different complexity levels. (Upper panels) Overall PQM% results. (Middle panels) Target PQM% score. (Lower panels). OARs PQM% score. Each plan value is represented using gray circles. The red points represent the samples' mean and the horizontal black lines are the median values. The distance between the top and the bottom of each box is the inter-quartile range. The vertical lines extend between the minimum and maximum values that are not outliers.

sion can be also extended to the *PQM*% scores obtained for the target and OARs separately. In general, a greater variability was observed for the target scores among the different combinations of ASC and MUs, especially for the prostate. The only slight trend is related to OAR sparing in the HN plans, which seem to increase with increasing plan complexity.

4.2.2 *Assessment of the delivery accuracy*

Dedicated patient-specific quality assurance (PSQA) sessions were performed using the `ArcCheck`®detector array (Sun Nuclear Corporation, Melbourne, FL, USA) without the PMMA `CavityPlug`™. The reference 3D dose distribution was computed with AcurosXB v.15.5 on a virtual homogeneous phantom (1.1836 g/cm3 density). Dose comparisons were performed using the software `SNC Patient`™ version 6.7.2, com-

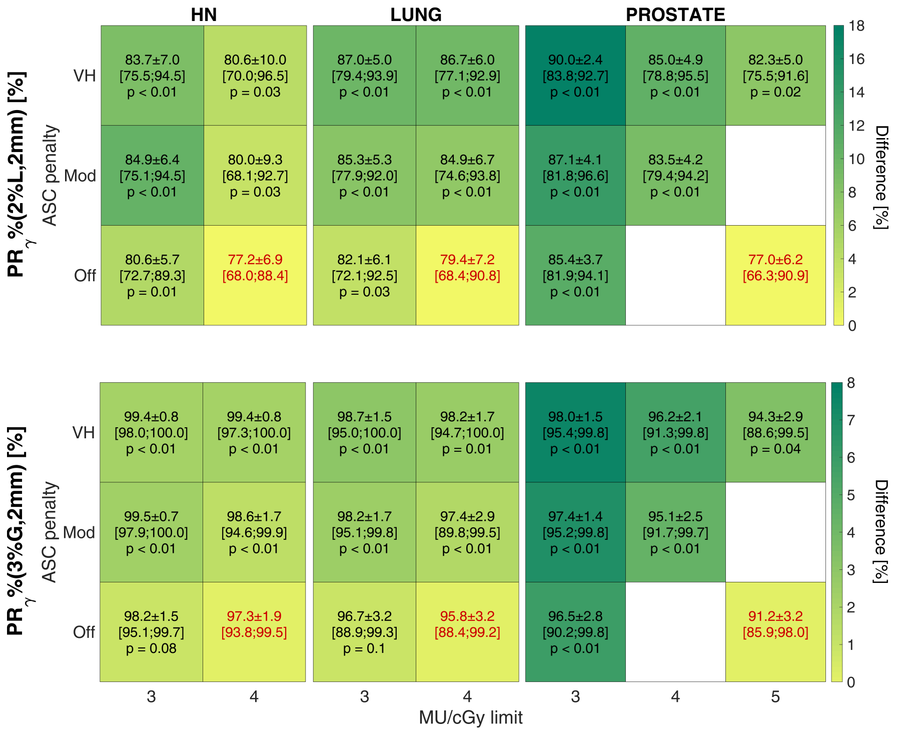


Figure 4.2: Heat maps showing the absolute differences in $PR_\gamma\%(2\%L, 2\ mm)$ and $PR_\gamma\%(3\%G, 2\ mm)$ (bottom) for the ASC-MU combinations used. Inside each box, the mean ± standard deviation, range and t-test p-value is reported.

puting the $PR_\gamma(3\%G, 2\ mm)$ and $PR_\gamma(2\%L, 2\ mm)$. A 10% threshold was set. The plans with $PR_\gamma(3\%G, 2\ mm) > 95\%$ were considered as deliverable according to the AAPM TG-218 [23].

Figure 4.2 shows the $PR_\gamma(2\%L, 2\ mm)$ and $PR_\gamma(3\%G, 2\ mm)$ for the entire database, also showing the p-values obtained from the two-tailed paired Student's t-test against the reference plan (no complexity reduction). Reducing the maximum deliverable MUs seems to have a more significant impact on the PR_γ scores of the prostate plans, while the effect of the ASC seems to be the most prominent for the Lung and the HN plans. This is true for both the $PR_\gamma(2\%L, 2\ mm)$ and $PR_\gamma(3\%G, 2\ mm)$. Prostate plans show the best PSQA results when the highest ASC level (Very High) and the lowest MUs limit $(MU/cGy \leq 3)$ are combined, while using the highest level of ASC does not seem to improve the deliverability of HN plans. Among the three considered anatomical districts, the prostate plans exhibit the greatest average improvement in both $PR_{gamma}(2\%L, 2\ mm)$ and $PR_\gamma(3\%G, 2\ mm)$, namely 18% and 8% with respect to the baseline. Improvement of 12% (10%) and 5% (5%) were observed in the HN (Lung) plans for the two scores, respectively. Regardless of the site, a MU/cGy 3 limit grants a 100% of deliverable plans $(PR(3\%G, 2\ mm) > 95\%)$

when coupled with an ASC level higher or equal to Moderate. This rule can be relaxed for HN: $MU/cGy \simeq 3$ and ASC set to Off, or $MU/cGy \simeq 4$ and ASC set to Very High grant 100% deliverability.

4.2.3 *Assessment of plan complexity through complexity indicators*

Forty-two different indicators of complexity were extracted from each treatment plan by choosing among those reported presented in Ch. 2 and implemente in the `UCoMX` package.

Figure 4.3 represents using spider plots the average value of a set of eleven representative complexity metrics for the different complexity limitation strategies. The three anatomical districts show overall comparable trends. A reduction of the limit on the MUs is generally associated with an increase of the area, average leaf gap (ALG) and dose rate (DR), and with a reduction of the SAS10mm. In the case of prostate plans, it is also associated with a net reduction of the mean gantry speed variation (mGSV) and leaf travel (LT). For a given limit on the MUs, plans with higher ASC levels are generally characterized by larger area and smaller LT, SAS10mm and EM.

In Figure 4.4 the Spearman's correlation maps between the complexity metrics and the $PR_\gamma\%$ results is shown. In general, higher correlation values are observed for the $PR_\gamma\%(3\%G, 2mm)$, with ALG and the EM showing the highest correlation values ($|\rho| > 0.7$). Additionally, stronger correlations arose when all the sites are considered together.

4.3 USING COMPLEXITY METRICS AS DECISION-SUPPORT TOOLS FOR PSQA

4.3.1 *Definition of operational limits on the complexity metrics*

A procedure to set operational limits on the complexity metrics was implemented. The aim of such limits is to help the planner in the identification of deliverable plans prior to PSQA. The adopted procedure was based on a ROC-type analysis with the positive class represented by non-deliverable plans (i.e. $PR_\gamma(3\%G, 2mm) < 95\%$, according to the guidelines adopted at our institution [23]). For each complexity metric, the the maximum or minimum threshold value (depending on the metric) corresponding to a 100% sensitivity (perfect classification of non-deliverable plans) was computed. This value that we call the operational limit splits the corresponding metric's space into two disjoint regions, one of them containing only plans that pass the PSQA and the second one that may contain either plans passing or failing the PSQA. The former is called the operational region. A compound operational region can be defined by combining operational limits defined on more than one metric.

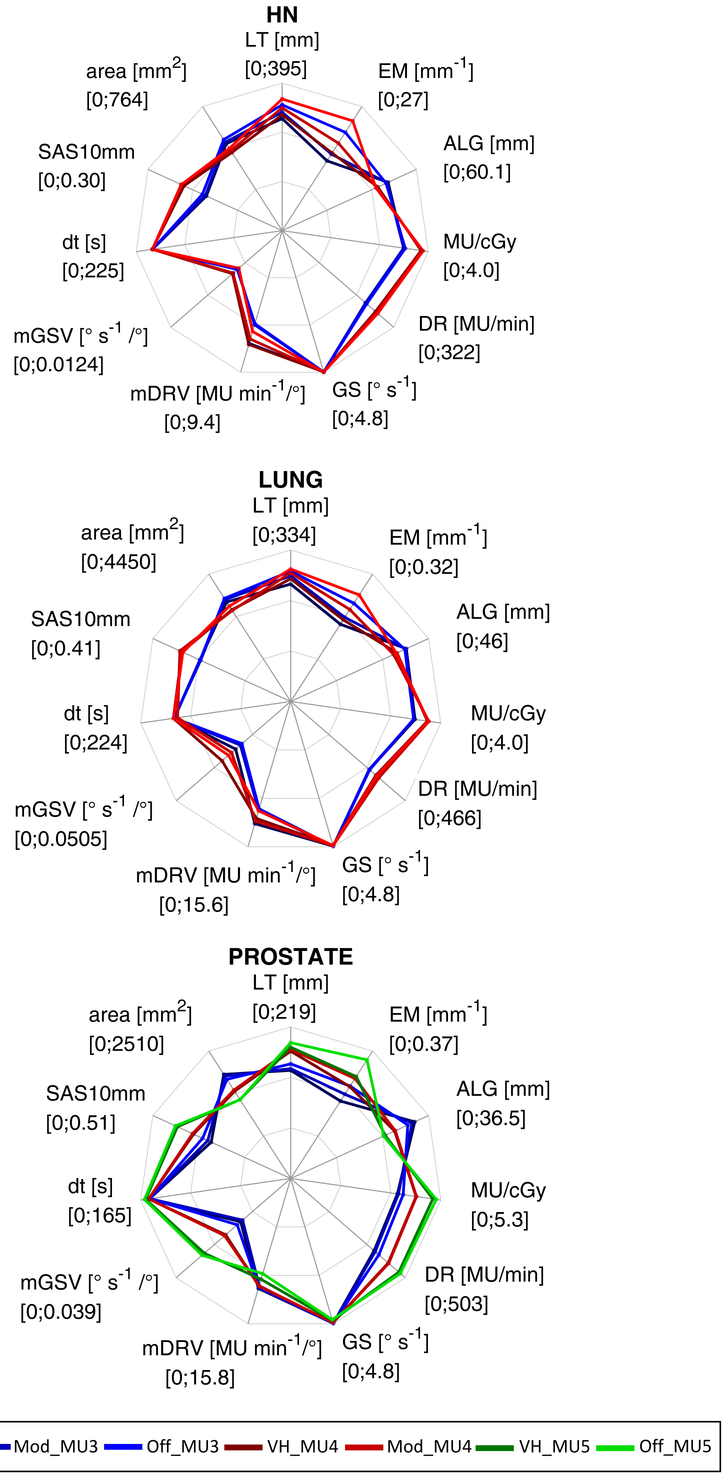


Figure 4.3: Spider plots showing the trend of eleven representative complexity metrics for different values of ASC level and MUs limit for the three anatomical districts considered.

The operational region is thus built to allow identifying in advance the deliverable plans with the highest possible specificity corresponding to a 100% sensitivity and can be adopted as a virtual PSQA tool

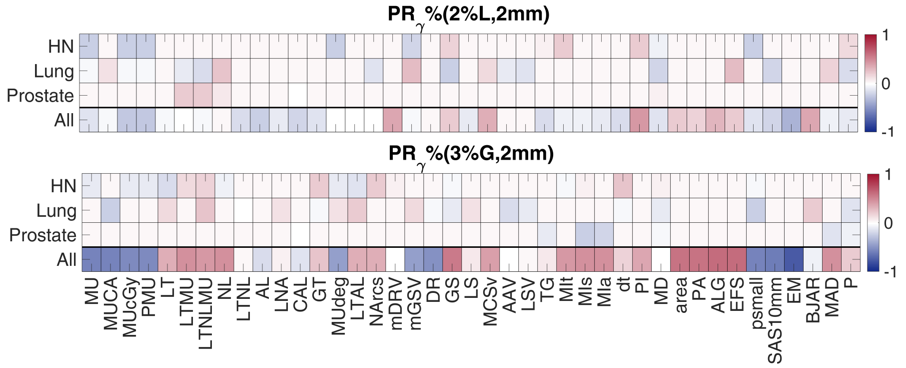


Figure 4.4: Spearman's correlation maps showing the values of Pearson's Correlation Coefficient between the metrics and the $PR_\gamma\%(2\%L, 2mm)$ (upper panel) and the $PR_\gamma\%(3\%G, 2mm)$ (lower panel)

that can reduce the typical workload associated with the measurement-based process process. The strategy adopted in this work is the same used in Ch. 3 when a threshold on the XGBoost model prediction was set. In fact, the specificity, also known as true negative rate, correspond to the number of negative (deliverable) plans that can be correctly identified normalize by the total number of negative plans. Therefore, upon a strict clinical validation, for each plan that will fall in the operational region the PSQA could be avoided because its delivery accuracy is granted. In this work, the operational limits were computed by considering all the plans in the dataset together without making any distinction among the three anatomical regions. However, their effect on the different treatment sites as well as their relation to the different complexity limitation strategies adopted were investigated.

Table 4.2: Operational limits and specificity values for the four best performing metrics taken as the arithmetic average over the bootsrap samples. Values in the square brackets correspond to the 95% bootstrap confidence bounds.

Metric	**ALG**	**EM**	**SAS10mm**	**MCSv**
Operational limit	33.1 mm [30.7; 33.8]	0.160 mm^{-1} [0.155; 0.189]	16.4 % [15.4; 17.0]	0.285 [0.245; 0.310]
Specificity	46.5 % [34.7; 55.1]	47.7 % [37.9; 66.2]	33.5 % [24.0; 40.3]	15.5 % [6.5; 24.5]

Tab. 4.2 reports the operational limits obtained for the four best performing metrics (highest specificity at 100 % sensitivity). As it was previously explained, the operational limits represent the threshold above or below which (depending on the metric) no plans fail the QA. Consistently to what was pointed out in when the results on the correlation between the metrics and the $PR_\gamma\%$ where shown, the

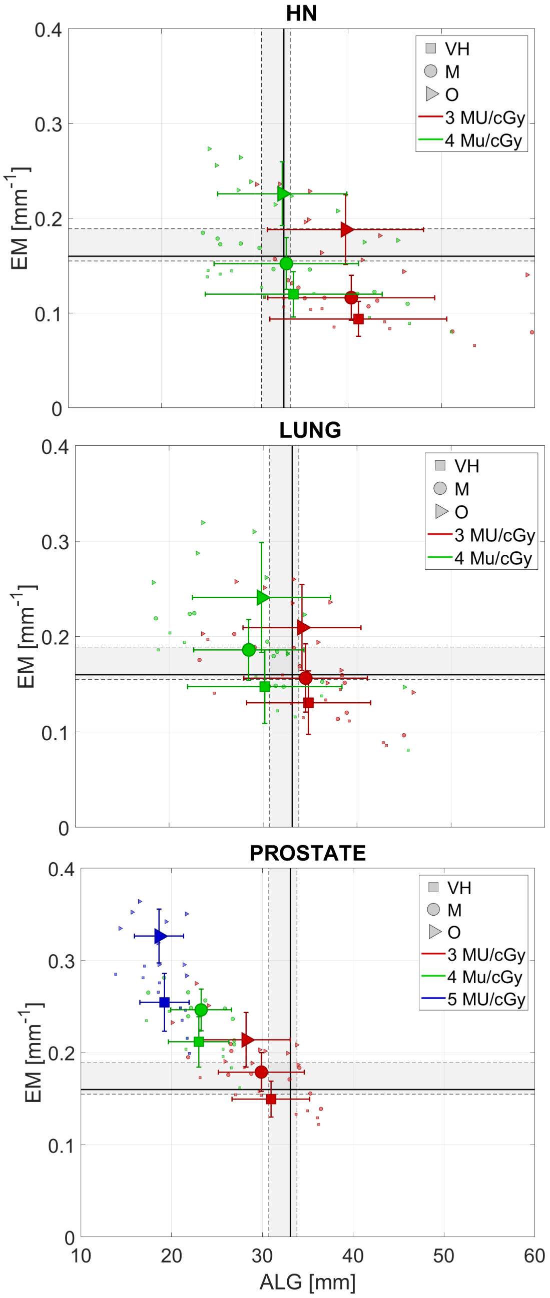


Figure 4.5: Edge Metric vs Average Leaf Gap]. The biggest markers represent the average value for each group of plans and the error bars the corresponding standard deviations. The smallest markers represent the data points for each group of plans. Solid black lines represent the operational limits on the two metrics while the shaded area indicates the corresponding 95% confidence interval computed using the bootstrap approach. The operational region is the one corresponding to the bottom-right corner.

ALG and the EM are the two best performing metrics with specificity values of of 46.5% and 47.7% respectively, that are significantly higher than the ones obtained for the SAS10*mm* and the MCSv. Table. 4.3

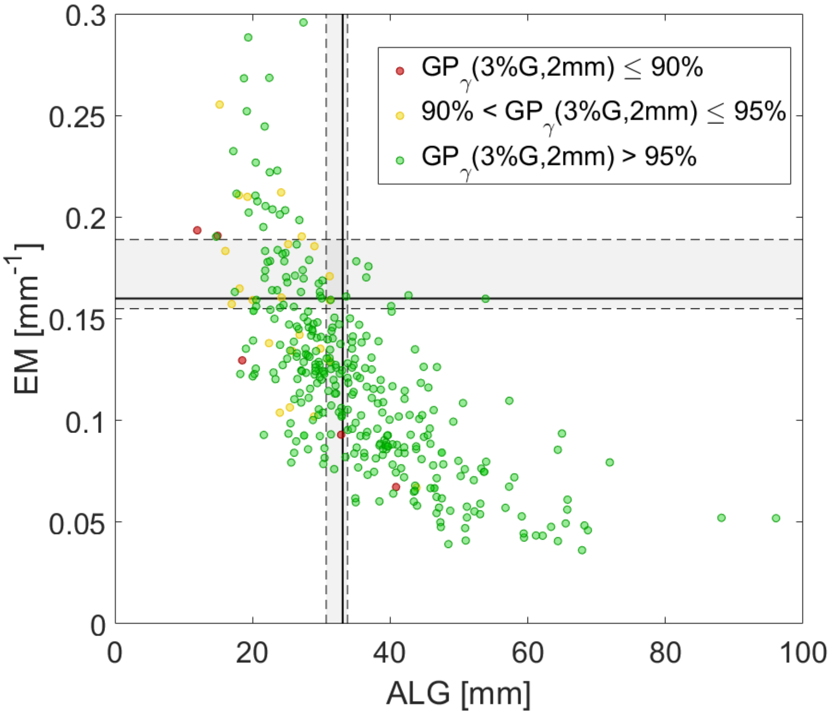


Figure 4.6: EM vs ALG on the test set of 350 plans used for the clinical validation. Vertical and horizontal solid lines indicated the (average) operational limits found on the other dataset, while shaded region the corresponding 95 % confidence intervals.

shows the specificity values for each site separately with respect to the average operative limits reported in Tab. 4.2.

Table 4.3: Specificity of the operational limits on the four best performing metrics for each site.

Metric	ALG	EM	SAS10mm	MCSv
HN	61.1 %	64.3 %	66.1 %	10.2 %
Lung	43.3 %	41.7 %	11.7 %	5.0 %
Prostate	13.0 %	14.5 %	7.2 %	20.3%

Figure 4.5 shows the behaviour of the EM and ALG for the different plan complexity limitation strategies adopted and with respect to the operational limits found. In general, for a fixed ASC values, a lower limit on the MU/cGy increases the probability to fulfill the operational limit on the ALG. Conversely, for a fixed MU/cGy limit, plans with a higher ASC penalty are more prone to satisfy the limit on EM.

4.3.2 *Clinical validation of the operational limits*

The clinical effectiveness of an operation limits defined according to the procedure presented in Sec. 4.3.1 was investigated. In particular, the EM and ALG were considered because of their significantly higher specificity values on the whole dataset compared to the other metrics. A dataset containing 350 VMAT plans created between October 2019 and February 2022 was retrospectively collected. This new dataset did not include the plans used for the determination of the operational limits. The plans in this new dataset were planned using the typical complexity limitation strategies adopted in the clinical practice and included a large variety of treatment sites. In total, 322 plans had $PR_{\gamma}\%(3\%G, 2mm) > 95\%$.

Figure 4.6 shows the distribution of the plans on the EM-ALG plane. The first interesting thing to notice is that only two positive (non-deliverale) plans fall in the operative region (bottom-right corner). In particular, the ALG allows for a higher sensitivity (approximately 100%) compared to the EM, and resulted in a specificity close to 40%.

4.3.3 *Impact and significance of the results*

The study explored the impact of applying different combinations of ASC levels and limits on the maximum deliverable MUs for treatment plans' clinical quality, deliverability and complexity. The findings demonstrate that these complexity limitations strategies do not significantly influence the PQM% evaluation of the resulting plans' clinical quality. Additionally, no significant differences in dose conformity, dose uniformity, and OARs sparing were identified. These findings align with the outcomes of a previous study conducted by this research group [32], which examined the impact of three distinct complexity limitation strategies utilising the same tools as in the present study. Additionally, they concur with Binny and colleagues' report [139] stating that there was no decrease in the optimized plans' dose conformality when using the ASC.

The PSQA results analysis illustrated that utilizing the two complexity limitation tactics present in the Eclipse TPS can notably enhance the plans' deliverability. The restriction on the MU/cGy and the ASC penalty mutually enhance the deliverability of the plans, whereby the former primarily impacts prostate plans, and the latter exhibits greater efficacy on HN and Lung plans. The data presented in this study suggest that, to ensure maximum delivery success rate ($PR_{\gamma}(3\%G, 2mm) > 95\%$), it is optimal to restrict MU/cGY to < 3 and set ASC to Moderate, regardless of the treatment site location. Nevertheless, this standard can be relaxed in HN plans where $MU/cGy < 3$ with ASC Off or $MU/cGy < 4$ with ASC Very High can guarantee the same level of deliverability

Regarding treatment complexity, the integration of distinct strategies to limit complexity, as proposed in this study, effectively reduced treatment duration and curtailed gantry speed and dose rate fluctuations. These findings advance and corroborate earlier research in the field [135]. Additionally, these strategies yielded more regular BEVs shapes (smaller EM), decreased leaf travel, and expanded treatment areas. The thorough and systematic analysis of over forty complexity metrics, across three anatomical districts, and seven ASC pentaly-MU/cGy limit combinations revealed a group of complexity indicators that exhibit the strongest correlations with PSQA results. Operative limits for the extracted complexity metrics were established as criteria for identifying deliverable plans, in accordance with adopted clinical practice and reference guidelines [23]. Specifically, effective operative limits were found for EM ($< 0.160mm^{-1}$) and ALG ($> 33.1mm$), enabling forecasting of the PSQA outcome for approximately 40% of external clinical dataset plans. If the thresholds defined by this procedure demonstrate similar performance in a prospective trial, it will significantly enhance the efficiency of radiotherapy workflow at this institution in the future. Additionally, the study explicitly showed the correlation between certain complexity indicators and the different ASC penalty/MU/cGY limit combinations. This could potentially extend these complexity limitation strategies to institutions that use different TPSs. In fact, it is important to note that although the values acquired from this study have a close relationship with the specific equipment and procedure used and should be validated with different settings, the approach proposed in this study can be effortlessly extended to other institutions, without depending on any particular equipment or machine/TPS.

Part II

TOWARDS THE USE OF QUANTUM COMPUTERS IN RADIOTHERAPY

5

SIMULATING MANY-BODY QUANTUM SYSTEMS WITH TENSOR NETWORK METHODS

The Universe is home to billions of constituents living at different scales. However, these constituents do not exist in isolation; they interact constantly, forming systems characterized by behaviour markedly distinct from the sum of their individual elements. Within this interaction, constituents exchange different kind of information in the form of energy or matter, and this is exactly what makes our universe looking and existing the way it does. In Physics, we call a many-body system a physics systems that contain numerous degrees of freedom. It is intriguing to witness their occurrences at every level, making their study and comprehension highly significant in various Physics areas and leading to implications in diverse fields including, the investigation of stars and planets, the animal and human social behaviour, ecosystems, bacterial colonies and many more.

Among the several many-body systems studied in contemporary Physics, there are the so-called quantum many-body systems. These are physical systems whose degrees of freedom behave according to the rules of quantum mechanics, resulting in behaviour and ensemble phenomena that are profoundly different from those observed in classical systems, and sometimes surprising. This branch of research is of great importance because of its implications in areas such as materials sciences, quantum chemistry, criptography, information theory and, not least, computations, thanks to the the development of quantum computers.

THIS CHAPTER provides some essentials for the physical description of quantum many-body systems. Additionally, it presents Tensor Network Methods (TNMs), a very successful numerical approache for the simulation of such systems on classical computers. The contents of this chapter have been selected to provide the reader with all the needed ingredients to understand the developments and results presented in this second part of the thesis.

5.1 MANY-BODY QUANTUM SYSTEMS

5.1.1 *Quantum lattice systems*

Many-body quantum systems are physical ensembles of several quantum degrees of freedom. A particular class is represented by quantum lattice systems where each degree of freedom i occupies a vertex V_i of a graph $\mathcal{G} = \{V, E\}$, with the interaction between sites represented by the links E. A local Hilbert space $\mathcal{H}_i$ is associated with each lattice site i which represents the natural container where each single-particle wave function $|\psi\rangle_i$ lives. The latter is defined as follows:

$$|\psi\rangle_i = \sum_{k=1}^{d} \alpha_k \, |\alpha\rangle_k \tag{5.1}$$

where $\{|\alpha\rangle\}$ a set of d basis vector used to represent the wave function $|\psi\rangle$, and $\{|\alpha_k|^2\}$ measure the probability of finding the system in the configuration $|\alpha\rangle_k$, and are defined such that $\sum_k |\alpha_k|^2 = 1$. $\mathcal{H}_i$ is also called the *local* Hilbert space, and its dimension d is the *local dimension*.

The description of the physics of a quantum lattice of N sites as a whole passes through the definition of the *global* Hilbert space, which is given by the tensor product of the single-particle ones according to the postulates of quantum mechanics, namely:

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \ldots \otimes \mathcal{H}_N \tag{5.2}$$

which has dimension d^N, assuming each local Hilbert space has the same dimension d. $\mathcal{H}$ is the space where the many-body wave function naturally lives. The latter is given by:

$$|\psi\rangle = \sum_{\vec{\alpha}} \psi_{\alpha_1,\ldots,\alpha_N} \, |\alpha_1 \ldots \alpha_N\rangle \tag{5.3}$$

where the set $\{|\alpha_1, \ldots, \alpha_N\rangle\}$ form an orthonormal d^N-dimensional basis of $\mathcal{H}$ and $\mathcal{P}_{\alpha_1,\ldots,\alpha_N} = |\psi_{\alpha_1,\ldots,\alpha_N}|^2$ returns the probability of finding the system in the configuration $|\alpha_1 \ldots \alpha_N\rangle$.

The physics of the resulting quantum lattice is described by the system's Hamiltonian H. In general, the latter contains both single-particle and $n-$body interaction terms. As an example, let's consider the case of a $1D$ Ising model in transverse field. The Hamiltonian associated with site i, including also the interactions with neighbor sites, is the following:

$$H_i = -\lambda \sigma_i^z + \frac{1}{2} \sum_{j \in \langle i,j \rangle} J_{ij} \sigma_i^x \sigma_j^x \tag{5.4}$$

where $\sigma_i^{x(z)}$ are the Pauli's matrices, λ is the magnetic field and $\langle i, j\rangle$ means that the two-body interaction is limited to nearest-neighbor sites only. For a system of N sites, the formal representation of the nearest-neighbor interaction between sites i and $i+1$ is given by:

$$\sigma_i^x \sigma_{i+1}^k = \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \ldots \otimes \mathbb{I}_{i-1} \otimes \sigma_i^x \otimes \sigma_{i+1}^x \otimes \mathbb{I}_{i+2} \otimes \ldots \otimes \mathbb{I}_N \quad (5.5)$$

which is a square matrix of dimension 2^N, with $\mathbb{I}$ the identity matrix of dimension 2×2. The formulation in Eq. (5.5) works also for the local term at site i by performing the two substitutions $\sigma_{i+1}^x \rightarrow \mathbb{I}_{i+1}$ and $\sigma_i^x \rightarrow \sigma_i^z$.

Simulating many-body quantum lattice systems efficiently is not possible using classical calculators. In fact, $|\psi\rangle$ defined in Eq. (5.3) is a $N-$rank tensor whose dimension scales exponentially with the number of sites in the system. Therefore, the simulation efficiency is dramatically impacted as it demands conducting operations involving large vectors and matrices. This is one of the main reasons behind the need for quantum computers in this field.

Several approaches have been proposed over the past decades to represent $N-$body wave functions efficiently by *compressing* or statistically describe the information in them contained, which is also the basic idea behind Tensor Network Methods (TNMs). However, before moving to a description of TNMs, in the next section, we provide an introduction to another fundamental concept at the base of TNMs: the entanglement of quantum systems.

5.1.2 *Entanglement in quantum systems*

The definition provided in Eq. (5.3) can be extended to a generic many-body state and we call *vector states* or *pure states* all those states for whom a representation in terms of *ket* $|\psi\rangle$ is possible. Given a generic many-body state, we will call it a *separable* state if it can be rewritten as the tesor product of single-particle state, namely:

$$|\psi\rangle_{MB,separable} = |\psi\rangle_1 \otimes \ldots \otimes |\psi\rangle_N \quad (5.6)$$

Otherwise, the state is *entangled*.

From a more general point of view, given a generic quantum state ψ, we can defined its associated *density matrix* as follows:

$$\rho = |\psi\rangle \langle\psi| \quad (5.7)$$

which has the following three fundamental properties:

1. $\rho = \rho^\dagger$
2. $\mathrm{Tr}(\rho) = 1$
3. $\mathrm{Tr}(\rho^2) \leq 1$

It can be demonstrated that the equality hold for pure states only [154].

Let's now consider a quantum system of two particles A and B that we assume to be in a pure state $|\psi\rangle_{AB}$ described by the density matrix $\rho_{AB} = |\psi\rangle_{AB} \langle\psi|_{AB}$. We can introduce the *reduced density matrix* for one of the two bipartitions, A for example, as follows:

$$\rho_A = Tr_B(\rho_{AB}) = \sum_i \langle B_i| \rho_{AB} |B_i\rangle \tag{5.8}$$

which correspond to the so-called *partial trace* computed with respect to particle B. Now, the connection with systems' entanglement arises naturally. In fact, in case the two particle are in an entangled state, we will have that $Tr[\rho_A^2] < 1$.

A quantitative measure of entanglement can be introduced, the *Von Neumann entropy* or *entanglement entropy* which is defined for a generic state ρ as follows:

$$S(\rho) = -Tr(\rho log(\rho)) = -\sum_m \lambda_m log(\lambda_m) \tag{5.9}$$

with λ_m the m^{th} eigenvalue of ρ.

From this description, we get two important pieces of information:

1. Defining entanglement of a systems requires considering bipartition of the systems into two sub-systems A and B.

2. $S(\rho) > 0$ for an entangled state, while it is zero for a separable state.

A paradigmatic example of an entangled state is the following Bell's state of two qubits:

$$|\psi\rangle_{Bell} = \frac{1}{\sqrt{2}} (|10\rangle + |01\rangle) \tag{5.10}$$

which has the followig density matrix:

$$\rho_{Bell} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.11}$$

This density matrix fulfills all the three aformentioned properties and, in particular, has $\mathrm{Tr}(\rho_{Bell}^2) = 1$ as expected being it a pure state. If instead we consider the reduced density matrix for the second qubits obtained by tracing out the first one, it reads:

$$\rho_2 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.12}$$

which describes a classical mixture of states (the anti-diagonal terms are null, namely there is no coherent superposition). In particular,

$\mathrm{Tr}(\rho_2^2) = 0.5 < 1$ and $S(\rho) = \log 2 > 0$, which prove that $|\psi\rangle_{Bell}$ describes a pair of (maximally) entangled qubits.

The description of the entanglement presented in the previous lines can be further extended by considering the *Schmidt decomposition* of a state. In fact, given a many-body state $|\psi\rangle$ describing a state of N particles, we can always rewrite it in a matricial form L with respect to a given bipartition $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ of the system (for more details see Sec. 5.2.2). In particular, the Schmidt decomposition consists in performing the Singular Value Decomposition of the resulting matrix, namely:

$$L = S \times V \times D \tag{5.13}$$

where $S(D)$ is a unitary matrix whose columns (rows) are the so-called left (right) singular vectors $\{|\alpha\rangle_A\}$ ($\{|\alpha\rangle_B\}$) which form an orthonormal basis for the space $\mathcal{H}_A$ ($\mathcal{H}_B$), and V is a diagonal matrix whose entries are the *singular values* $\Lambda_\alpha \geq 0$. The Schmidt decomposition of the state is thus is written as:

$$|\psi\rangle = \sum_\alpha \Lambda_\alpha |\alpha\rangle_A |\alpha\rangle_B \tag{5.14}$$

where the index α runs over all the non-zero coefficients Λ_α. The number of non-zero Schmidt coefficients is called the *Schmidt rank* and the set of non-zero Λ_α is the *entanglement spectrum*. In particular, the set of $\{\Lambda_\alpha\}^2$ corresponds to the eigenvalues of the reduced density matrix of the bipartition (see Eq. (5.8)).

Equation (5.14) is the starting point for any Tensor Network Methods representations of a many-body quantum state. In particular, a cutoff χ is usually introduced such that:

$$|\psi\rangle = \frac{1}{Z} \sum_\alpha^\chi \Lambda_\alpha |\alpha\rangle_A |\alpha\rangle_B \tag{5.15}$$

with $Z = \sqrt{\sum_{\alpha=1}^{\chi} \Lambda_\chi^2}$a normalization constant. Usually, the cutoff χ is called the *bond-dimension* and its meaning will become clearer in the following Section. The quantum state representation in Eq. (5.15) has been proven to be the best possible lower-rank approximation in term of the Frobenius norm $\|.\|_F$ [155], namely:

$$\|M - M_\chi\|_F^2 = \sum_{\alpha=\chi+1}^{S_R} \Lambda_\alpha^2 \tag{5.16}$$

with S_R being the Schmidt rank. The introduction of such a lower-rank representation based on the Schmidt decomposition allows to compress the representation of a quantum state and to interpolate between a fully-efficient mean-field approximation and the complete (but inefficient) representation. The following section will present the fundamentals of TNMs, demonstrating how they integrate the aforemetioned concepts to provide an efficient representation of many-body quantum states.

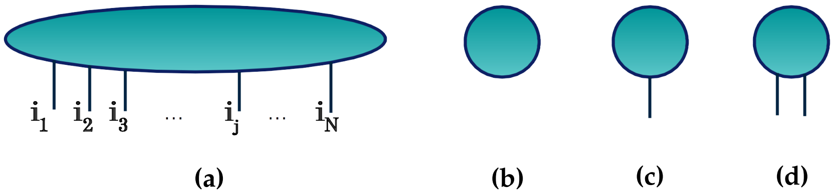


Figure 5.1: **(a)** Generic tensor representation; **(b)** representation of a scalar; **(c)** representation of a vector; **(d)** representation of a matrix.

5.2 TENSOR NETWORK METHODS

5.2.1 *Tensor definition and representation*

Tensor calculus is the natural extension of linear algebra where *vectors* and *matrices* are replace by *tensors*. If we consider a finite set of vector spaces $\{\mathcal{H}_1, \ldots, \mathcal{H}_\mathcal{N}\}$ over a common field $\mathcal{F}$, a new vector space $\mathcal{H}$ can be formed as the result of the tensor product operation, namely $\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_\mathcal{N}$. A tensor is the generic element of $\mathcal{H}$. Based on the introduction to many-body lattice quantum systems provided in Sec. 5.1.1, $\mathcal{H}$ is the global Hilbert space resulting from the composition of the single-particle Hilber spaces $\mathcal{H}_i$ and a tensor is the generic many-body wave function $|\psi\rangle$.

Moving from the previous definition to a more general and descriptive level, a tensor can be seen as a container of numbers, each one uniquely addressed by specific values taken by the tensor indexes. The number of indexes, N, equals the number of vector spaces $\mathcal{H}_1, \ldots \mathcal{H}_N$ that originated $\mathcal{H}$, and it is called the *rank* of the tensor. Therefore, a generic tensor will be written as follows:

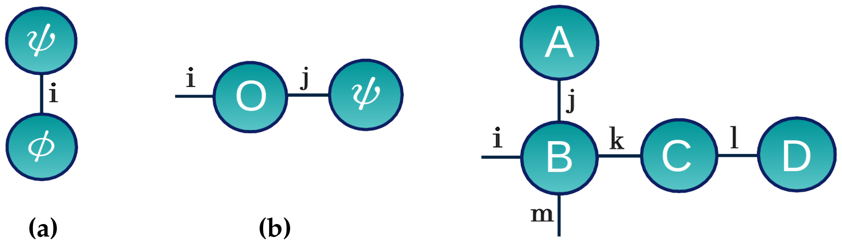
$$T_{\alpha_1, \ldots, \alpha_N}. \tag{5.17}$$

Each index α_i can take integer values in the range $[1; dim(\mathcal{H}_i)]$. For a detailed mathematical description of tensor properties please refer to [156].

In the context of TNMs, tensors and related operations are represented using a graphical notation. In particular, a generic tensor can be thought of as a round-box with a number of *legs* equal to the rank N, as it is shown in Fig. 5.1a It follows that a scalar, which is a 0-rank tensor, will be represented as a circle with no indexes, as shown in Figure 5.1b. In the same way a vector (1-rank tensor) and a matrix (2-rank tensor) will be represented as boxes with one and two legs respectively, as it is shown in Figure 5.1c and 5.1d. Manipulating tensors with many indexes is not a trivial task in general, but it can become quite simple if we imagine to represent each operation among them as a *tensor network*, where *connections indicate contractions*.

In Figure 5.2a the following scalar product is represented:

$$\langle\phi|\psi\rangle = \sum_i \phi_i^* \psi_i \tag{5.18}$$



(a) (b) (c)

Figure 5.2: Representation of some basic operation among tensor. **(a)** Scalar product between two vectors; **(b)** Matrix-vector multiplication; **(c)** Contraction of four tensors leading to a rank 2 tensor as final result.

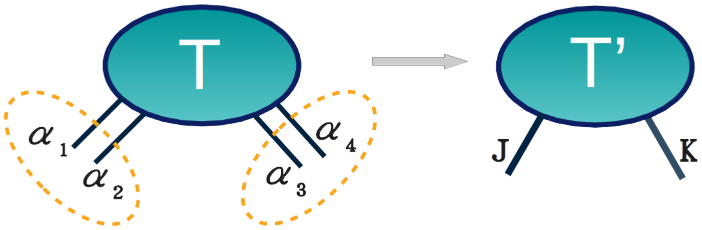


Figure 5.3: Representation of the *index fusion* procedure.

where the resulting tensor has no free links, meaning that it is a scalar, as it was expected. In the same way, in Figure 5.2b a representation of the following contraction for a matrix-vector multiplication is given:

$$(\mathcal{O}\psi)_i = \sum_j \mathcal{O}_{ij}\psi_j \tag{5.19}$$

where the final tensor has rank 1, meaning that it is a vector. One can generalize the previous arguments to any number of tensors with arbitrary ranks, obtaining a network of tensors, where the number of free indexes determines the rank of the resulting tensor after the indicated contractions. An example is shown in Figure 5.2c, where it's quite easy to see that the complex contractions involving four tensors lead to a rank 2 tensor (i.e. a matrix) as final result.

5.2.2 *Tensors manipulation*

Let's see how it is possible to reshape the representation of a tensor to a lower (higher) rank tensor by *fusing* (*splitting*) its indexes. Suppose we have a 4-rank tensor $T_{\alpha_1,\alpha_2,\alpha_3,\alpha_4}$. We can group the first two indexes into one single index as $\{\alpha_1,\alpha_2\} \succ \mathbf{J}$ and the other two as $\{\alpha_3,\alpha_4\} \succ \mathbf{K}$. The meaning of this *index fusion* process can be easily understood if we turn to the graphical notation.

Looking at Figure 5.3 we see that the initial 4-rank tensor gets transformed into a 2-rank one, which is a matrix. What we are practically

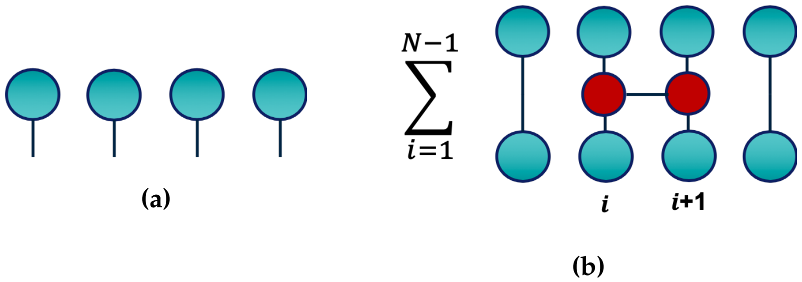


Figure 5.4: **(a)** Representation of a many-body state using the mean-field ansatz. **(b)** Energy expectation value in the mean-field approximation for the Ising in transverse field with $\lambda = 0$.

doing is to change the way information is stored. In fact, the dimension of the new index is given by the product of the dimension of the fused indexes, namely $\dim(\alpha_{1(3)}) \times \dim(\alpha_{2(4)}) = \dim(\mathbf{J}(\mathbf{K}))$.
It follows from this procedure that the tools provided by linear algebra become applicable also to tensors just by simply reshaping the latter into matrices or vector. Of course, all those manipulations have a computational cost, but it's in general very convenient to apporach the problem this way [157].

5.2.3 *Ground states via tensors networks*

When TNMs are used for the ground-state search, they act as *variational algorithms*, meaning that they aim at solving the following problem:

$$E_{GS} = \min_{\psi} \langle\psi| H |\psi\rangle \tag{5.20}$$

where H is the system Hamiltonian. One of the main features related to the use of TNMSs is that they provide different possible *ansatze* through which the wave function ψ can be represented. The method shown in this sectoin is very general and can be adapted to any tensor network ansatsz in priniciple.

The *mean-field* approximation to a 1D Ising model in transverse field is here considered. To be more general,the translational invariance can be relaxed by considering a generic product-state. The latter is the simplest tensor network (TN) ansatz, where the bond dimension is $\chi = 1$. Working in the graphical notation, the state can be represented as shown in Fig. 5.4a where all tensor are equal in case of mean-field approximation. Without loss of generality, referring to Eq (5.4), we can set $h_i = 0$ for simplicity, obtaining the following Hamiltonian, where only the two-site nearest-neighbour interactions are present:

$$H = \sum_{i=1}^{N-1} J_{ii+1}\sigma_i^x \sigma_{i+1}^x \tag{5.21}$$

The minimization problem in Eq. (5.20) correspond to minimize the energy represented in Fig. 5.4b, where the blue tensors represent the

particles, while the red ones the σ^x matrices.
If we now move to a more general description, any two-site operator can be described as a tensor $H_{\alpha_i,\alpha_j}^{\alpha'_i\alpha'_j}$ acting between particles at sites i and j. The minimization problem in case of nearest neighbour interactions can be now expressed using the method of the *Lagrange multipliers* to impose the normalization of the state as follows:

$$\mathcal{L}(\psi_1,\ldots,\psi_N,\psi_1^*,\ldots,\psi_N^*) = \langle\psi|\,H\,|\psi\rangle - \lambda\left(\langle\psi|\psi\rangle - 1\right) \equiv E - \lambda(\mathcal{N}-1) \tag{5.22}$$

with:

$$\langle\psi|\,H\,|\psi\rangle = \sum_{i=1}^{N-1} H_{\alpha_i\alpha_i+1}^{\alpha'_i\alpha'_{i+1}}\psi^*_{\alpha_i}\psi^*_{\alpha_{i+1}}\psi_{\alpha'_i}\psi_{\alpha'_{i+1}} \tag{5.23}$$

where the sum over the tensors' components is implied. Eq. (5.22) has to be minimized with respect to the components of each tensor in the network, as follow:

$$\frac{\partial\mathcal{L}}{\partial\psi_i^*} = H_{\alpha_{i-1},\alpha_i}^{\alpha'_{i-1}\alpha'_i}\psi^*_{\alpha_{i-1}}\psi_{\alpha'_{i-1}}\psi_{\alpha'_i} + H_{\alpha_i\alpha_{i+1}}^{\alpha'_i\alpha'_{i+1}}\psi^*_{\alpha_{i+1}}\psi_{\alpha'_i}\psi_{\alpha'_{i+1}} - \lambda\psi_{\alpha'_i} = 0 \tag{5.24}$$

The effect of the differentiation of the network with respect to one of its tensors is to remove that tensor from the network. If we now define an *effective Hamiltonian* $\tilde{H}$ as:

$$\tilde{H}_{\alpha_i}^{\alpha'_i} \equiv H_{\alpha_{i-1},\alpha_i}^{\alpha'_{i-1}\alpha'_i}\psi^*_{\alpha_{i-1}}\psi_{\alpha'_{i-1}} + H_{\alpha_i\alpha_{i+1}}^{\alpha'_i\alpha'_{i+1}}\psi^*_{\alpha_{i+1}}\psi_{\alpha'_{i+1}} \tag{5.25}$$

we see that Equation (5.24) can be expressed as:

$$\tilde{H}_{\alpha_i}^{\alpha'_i}\psi_{\alpha'_i} = \lambda\psi_{\alpha'_i} \tag{5.26}$$

which is a *local* eigenvalue problem for the effective Hamiltonian $\tilde{H}_{\alpha_i}^{\alpha'_i}$ and the tensor $\psi_{\alpha'_i}$. This procedure can be applied iteratively to all tensors in the network, updating their entries with the result from the minimization at each step, until convergence is reached.

As it is already been said at the beginning of the discussion, this procedure is very general and can be extended to any operator. TNMs provides two very powerful ansazte for an efficient representation of many-body operators. They are the *Matrix Product Operator* (MPO) [158] and the *Tensor Product Operator* (TPO) [157]. The former is very suited in case of the so-called *Matrix Product State* (MPS) [159] representation of the wave function, which is particularly suitable in case of one dimensional systems. Conversely, the TPO can be regarded as a generalization of the MPO which is more suitable for other TNs structures (e.g. tree tensor networks). However, their details are not discussed in the present section, which aimed at providing a general introduction to the topic.

In the next sub-section, a more detailed description of the TN ansatz used in this work is presended, namely the binary Tree Tensor Network.

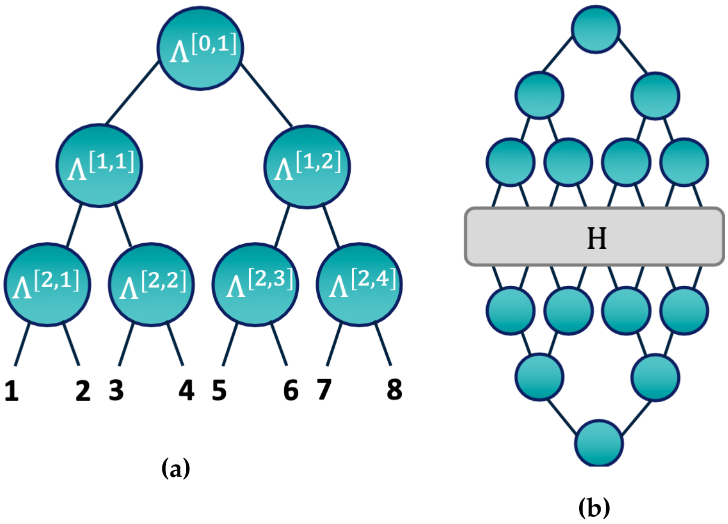


Figure 5.5: **(a)** Example of binary tree tensor network for a system of eight sites. **(b)** Graphical representation of the energy expectation value computation with bTTN.

5.2.4 *Binary Tree Tensor Networks*

Let's consider a one dimensional lattice with $N = 2^L$ sites, where each site has an associated local Hilbert space of dimension d and L is the number of layers. A *binary tree tensor network* (bTTN) [157, 160, 161] is a loop-less structure built on the top of a one dimensional lattice and it's fully composed by tensors with three links. We see from Fig. 5.5a, the n^{th} tensor of the l^{th} layers is referred to as $\Lambda^{[l,n]}$; layers are enumerated from the top to the bottom. The links in the last layer (the lowest one) are the so-called *physical links*, they represent the sites of the lattice and their dimension equals the local dimension of the associated single-particle Hilbert space. All other links are called *virtual links*.

The tree structure has the property of mapping two sites of the $(l+1)^{th}$ layer into one coarse-grained site of the l^{th} layer. The full Hilbert spaces of the sites at layer l have dimension $M(l) = d^{2^{L-l}}$, which increases exponentially with the number of physical sites which are blocked together in layer l. An efficient numerical representation of the many-body state requires to introduce a cutoff in the dimension of the virtual links. It can be done by setting a *maximum bond dimension* χ_{max} so that all virtual links must have a dimension $\chi_l \leq \min(\chi_{max}, d^{2^{L-l}})$. The number of variational parameters in a bTTN scales as $\mathcal{O}(N\chi^3)$.

5.2.5 *Ground state search with bTTN*

It's now interesting to see how bTTN can be used to find the ground state of the system. The procedure explained below can be thought

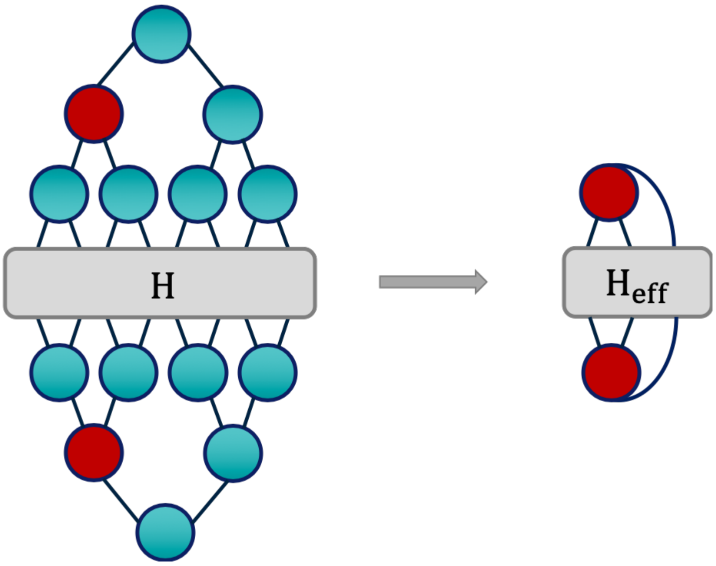


Figure 5.6: Introduction of the effective Hamiltonian for the local optimization of the tensors in the network.

of as a more efficient version of the one explained in Sec. 5.2.3 and which is also the one implemented in the algorithm. The complete description of this procedure can be found in [157]; here we'll provide only general details to give to the reader the essential knowledge.
For the bTTN, the variational problem in Eq (5.20) aims at minimizing the energy in Fig. 5.5b with respect to each tensor. Even in this case, the underlying idea is to solve local optimization problems for each tensor in the network, defining an effective Hamiltonian. Let's look at Figure 5.6 and suppose we want to solve the variational problem for the red tensor. It means that all other tensor entires are kept fixed. In this way we can contract the Hamiltonian with all other tensors in the network and apply the procedure in Eq. (5.22) based on the Lagrange Multipliers to obtain an eigenvalue problem involving the red tensor and the effective Hamiltonian H_{eff}. Once it has been solved, the red tensor gets replaced by the eigenvector corresponding to the lowest energy eigenvalue. This is done for all tensors in the network and a single optimization of all tensors is called a *sweep*. This procedure can be applied iteratively until convergence to the global ground state is reached.
As already mentioned in the remarks at the end of previous section, all this procedure can be performed very efficiently if the Hamiltonian H is represented using the TPO anstatz.

5.2.6 *Computation of expectation values with bTTN*

At the end of the energy minimization process, the ground state of the system can be used to compute the expectation values of observables. As we will see in the next chapter a case of particular interest for the

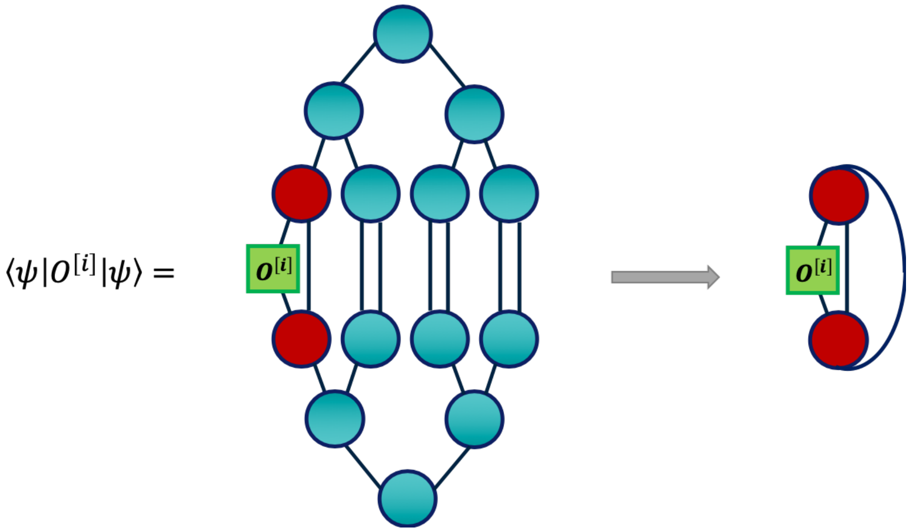


Figure 5.7: Computation of the expectation value of a local observable for the generic site i of the lattice. The introduction of the isometries in the network allows to reduce the number of relevant contractions.

purposes of this work is the computation of the expectation value of a local observable with support only on the i^{th} single-particle space, as shown in Fig. 5.7. An efficient way to perform this operations is to isometrize (gauge) the network with respect to the red tensor so that most of the contractions lead to identities and only those represented on the right-hand side of Fig. 5.7 remain relevant [157]. A bTTN can be isometrized with respect to anyone of its tensor, so that the number of contractions needed to compute the expectation value of local observables is $\mathcal{O}(1)$.

6

OPTIMIZING RADIOTHERAPY PLANS WITH TENSOR NETWORK METHODS

In Ch. 1, the role of radiotherapy in the treatment of solid tumors was presented. It was stressed out that one of the hardest challenges encountered while treating patients with ionizing radiation is to deliver an optimal dose to the target tumor while keeping the radiation dose to the surrounding healthy tissues as low as possible. In Sec. 1.1 an overview of the most advanced techniques used in clinics nowadays was provided, including a description of the IMRT technique[162–165]. We have seen that, in IMRT, the beam intensity is modulated through a mechanical collimator, the MLC, to reach the desired dose distribution inside the patient. As discussed, this optimal modulation is obtained by solving a non-trivial numerical optimization problem, known as inverse-planning problem, with a high number of optimization parameters and numerous constraints on the final radiation dose distribution. Since the introduction of IMRT techniques in clinics, several algorithms have been developed to address this challenge [163, 166–169], with some of them currently available in the TPSs used every day in clinics.

In 2015, Nazareth and Spaans proposed to solve a fixed-gantry IMRT beam fluence optimization problem using the D-Wave annealer [170], and more recently Pakela *et al.* proposed an approach exploiting the simulated quantum tunneling effect [171], opening to the possibility of using quantum computers for this task in the future. Despite these remarkable novel approaches paved the way for future applications of quantum computation in radiotherapy, their application is still limited. On the one hand, this is due to the current lack of scalable quantum hardware. On the other hand, it is not straightforward to extend what is done on classical computers to quantum ones due to the lack of a robust and clear strategy to map the classical IMRT problem to quantum hardware. With the aim of overcoming these limitations as well as to further investigate the applicability of quantum-inspired techniques to the solution of classical optimization problems and foster future applications of quantum technologies to medicine, here we apply TNMs to a fixed-gantry IMRT fluence optimization problem.

It has already been discussed in Ch 5 that TNMs are among the most successful algorithms for simulating quantum many-body systems on classical computers. Indeed, whenever possible, they efficiently represent quantum many-body wavefunctions in a compact and efficient form on classical computers [172–176]. In the last few decades, TNMs have proven their effectiveness in the research and analysis of quantum many-body systems, especially for low-dimensional ground-state [177–184]. In addition to that, thanks to the properties they share with quantum hardware, TNMs may play the role of test benches for the development of quantum algorithms [185–187].

Hereafter, we show how to solve an IMRT optimization problem with TNMs. We first introduce how the classical cost function is mapped into an Ising-like Hamiltonian, where the optimization variables are represented as a set of long-range interacting spins. Finally, we solve the classical optimization problem by finding the ground-state for this Hamiltonian using TNM. We present the application of TNMs to two different toy models and to a more realistic anatomical scenario simulating a prostate cancer treatment by using the TG-119 phantom [111]. We show that TNMs results are compatible with other classical techniques, Quadratic Programming (QP) and Simulated Annealing (SA). This chapter repeats paper [15].

6.1 THE FIXED-GANTRY IMRT FLUENCE OPTIMIZATION PROBLEM

The goal of IMRT is to create a personalized dose distribution for each patient's anatomy that ensures the appropriate dose to the tumor while saving the Organs At Risk (OARs) as much as possible. In particular, as already described in Sec. 1.1, to achieve this treatment goal, the IMRT optimization process begins with the splitting of each beam into a grid of pencil-beams called beamlets. Only beamlets that traverse the target must be optimized, and the intensity of the j-th beamlet can be modulated independently with a weight x_j. The volume of the patient considered for the optimization get divided in 3D finite-size elements called *voxels*.

The sum of the dose contributions from each beamlet produces the total dose in each voxel i and lead to the global dose distribution $D(\mathbf{x})$. Therefore, the total dose delivered to the voxel i can be expressed as:

$$D_i(\vec{x}) \equiv D_i(x_1, \ldots, x_{N_B}) = \sum_{j=1}^{N_B} a_{ij} x_j \;, \tag{6.1}$$

where N_B is the total number of beamlets and a_{ij} describes the elements of the influence matrix A which gives the unmodulated contribution of the j-th beamlet to the i-th voxel, namely the voxels dose for unitary beamlets values.

The IMRT planning procedure is then solved through an iterative inverse planning process: the intensity of the delivered photon beams are optimized towards the desired dose distribution $D^{(P)}$ inside the

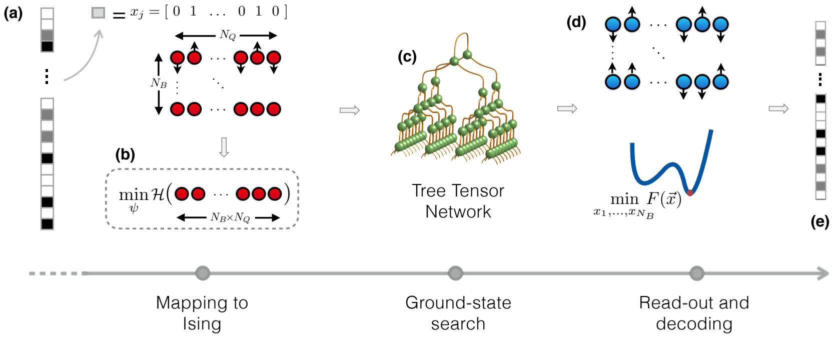


Figure 6.1: Solving the classical IMRT optimization problem as a quantum Hamiltonian with TNs. The beamlet weights, x_j, are represented as a set of long-range pairwise interacting spins (a). The initial problem of minimizing a cost function is thus mapped into a ground-state search problem for the Ising-like Hamiltonian (b) which can be efficiently solved using the Tree Tensor Network algorithm (c). After the minimization the final spin configuration we read out the results by reconstructing the optimal values for the beamlet weights (d, e).

patient. In this work, the inverse-planning problem was described through the following quadratic cost function:

$$F(\mathbf{x}) = \sum_{r=0}^{R} \sum_{i=1}^{\mathcal{V}_r} \gamma_i \left[D_i(\mathbf{x}) - D_i^{(P)} \right]^2 \tag{6.2}$$

where r is an index running over all the volumes (the targeted tumor and the OARs), and γ_i a weight assigned to the i^{th} voxel in order to prioritize certain volumes during the treatment. $\mathcal{V}_r$ gives the total number of voxels belonging to the object r.

6.2 SOLVING CLASSICAL PROBLEMS WITH TENSOR NETWORKS

6.2.1 *Mapping the problem to a classical Hamiltonian*

In the following, we describe how to rewrite the optimization problem in Eq. (6.2) into a ground-state search of a quantum many-body Hamiltonian. In particular, we here propose a procedure based on the binary-decimal conversion to map the cost function $F(\mathbf{x})$ into an Ising-type Hamiltonian, $\mathcal{H}_{IMRT}$. The procedure is summarized in Fig. 6.1.

We first discretize the weights x_j for each beamlet by a set of N_Q bits. N_Q is called the *bit-depth*. Thus, we represent

$$x_j \approx \frac{1}{(B/2)} \sum_{n=1}^{N_Q} 2^{n-1} b_n^{(j)} \tag{6.3}$$

with the bits $b_n^{(j)} = \{0, 1\}$, introducing a normalization constant B to set the range of the beamlets weights such that $x_j \in \left[0; \frac{2^{N_Q-1}}{(B/2)}\right]$.

With increasing number of bits N_Q we can increase the resolution of the discretization. Then, we map the binary values $b_n^{(j)}$ into spin variables $s_n^{(j)} = \{-1; +1\} = 2b_n^{(j)} - 1$ for each site j. Consequently, we construct a $N_B \times N_Q$-dimensional many-body Hamiltonian with the first dimension running over the different beamlets and the second representing the discretized space for each beamlet. Inserting the spin variables together with Eq. (6.3) into the cost function of Eq. (6.2), we obtain this Ising-type Hamiltonian with the following general expression:

$$\mathcal{H}_{IMRT} = \mathcal{H}_{SP} + \mathcal{H}_{INT}^{(a)} + \mathcal{H}_{INT}^{(b)} \tag{6.4}$$

where the first term describes the single particle terms with

$$\mathcal{H}_{SP} = \sum_{j,n} \left[\sum_i \frac{\gamma_i}{B} \left(a_{ij} \sum_k a_{ik} - 2D_i^{(P)} a_{ij} \right) 2^{n-1} \right] s_n^{(j)} \ , \tag{6.5}$$

the second term captures spins interacting in the same beamlet

$$\mathcal{H}_{INT}^{(a)} = \sum_j \sum_{m \neq n} \left[\sum_i \gamma_i \frac{a_{ij}^2}{B^2} 2^{n-1} 2^{m-1} \right] s_n^{(j)} s_m^{(j)} \ , \tag{6.6}$$

and the last term represents the interactions between different beamlets

$$\mathcal{H}_{INT}^{(b)} = \sum_{j \neq k} \sum_{n,m} \left[\sum_i \gamma_i \frac{a_{ij} a_{ik}}{B^2} 2^{n-1} 2^{m-1} \right] s_n^{(j)} s_m^{(k)} \ . \tag{6.7}$$

We point out, that this Hamiltonian describes a two-dimensional fully-connected lattice of long-range interacting spins with N_B sites on one direction and N_Q on the other. At this point, we can solve the initial problem of minimizing Eq. (6.2) by finding the ground-state for the classical Ising-type Hamiltonian in Eq. (6.4).

The ground state of the system is given by the spin-configuration which provides the lowest energy E_0, that is the configuration that minimizes Eq. (6.2). The case $E_0 = 0$ would correspond to the case where the beamlets setup returns exactly the desired dose distribution within the patient, a situation which is never achievable in practice. In principle, we can find the solution of the initial optimization problem by going through all possible spin-configurations of the Ising-type Hamiltonian. However, this classical search rapidly becomes unfeasible, as the number of spin-configurations grows exponentially with increasing system size and the number of beamlets in a clinical scenario can easily get to the order of a few thousand. For this reason, in the following section we introduce an approach based on TNMs to address this optimization task.

6.2.2 *Description of TTN for solving quantum many-body problems*

Considering the classical Hamiltonian $\mathcal{H}_{IMRT}$, we further allow each spin to be a quantum variable by representing $s_n^{(j)}$ with the Pauli matrix σ_z. In this way, the quantum Ising-type Hamitonian $\mathcal{H}_{IMRT}$ is a diagonal matrix with its entries corresponding to the energies of all possible spin combinations of the classical Hamiltonian. Here, TNMs are a vital tool to find ground states and their physical properties despite the exponentially growing Hilbert space [172, 174, 188]. In the following, the idea of TNMs to address the challenge of investigating complex quantum many-body systems is presented. For a more in depth introduction on TNMs, the interested reader should refer to dedicated literature [172–174, 176]

As discussed in Ch. 5, TNMs are used to efficiently represent quantum many-body wavefunctions $|\psi\rangle$, by decomposing the complete rank-N tensor (containing all d^N coefficients) into a set of local tensors with smaller rank, connected with auxiliary indices. We control the dimension of the auxiliary indices with the bond-dimension χ and thereby the amount of captured information. Thus, tuning this parameter χ, TNMs interpolate between a product state, where quantum correlations are neglected, and the exact, but inefficient representation.

In this analysis, the TTN anstatsz descibed in Sec. 5.5a is used. The TTN offers better connectivity between long-range interactions (with a logarithmic distance through the network), while for the simpler MPS the distance by connecting tensors within the network is linear.

Due to the bond-dimension $\chi > 1$, the quantum ground state search is performed within the subspace limited by χ. Thus, in contrast to the classical optimization routines, within one optimization step several classical solutions are explored as they are superposed in the quantum representation of the TTN. This also allows further to tunnel through higher, but reasonably thin, potentials within the optimization landscape.

After the TTN analysis, the obtained ground state of the quantum Hamiltonian might, in general, not be classical but rather a superposition of classical states. However, we know, that all solutions to the initial IMRT problem are classical, thus we expect that each of the superposed classical solutions separately has the same ground state energy E_0 once the TTN algorithm is fully converged to the quantum ground state. Therefore, we enforce the “classicality" of the solution as a constraint after the TTN minimisation problem by selecting one of the classical solutions from the converged TTN by truncating the bond-dimension to $\chi = 1$, resulting in a separable, product-state solution

$$|\Psi_{\chi=1}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_L\rangle \ ,$$

in which the only superposition can be local (such as a local site being $|\psi_n\rangle = 1/\sqrt{2}(|\uparrow\rangle + |\downarrow\rangle)$). From here on, we measure the quantum observable $\langle\sigma_n^z\rangle$ for each site n, resulting in $\langle\sigma_n^z\rangle = -1$ for a spin down, $\langle\sigma_n^z\rangle = 1$ for a spin up and in between those in case of a local superposition. In the latter case, we project the spin to the classical one with the highest probability by using the sign. The resulting spin configuration is further mapped back to the binary encoded solution for each voxel as described in the previous section. In fact, the analysis presented herin confirms that in most of the times the resulting TTN ground state is indeed a classical state ($\chi = 1$), or in some scenarios that the obtained state is, as expected, a superposition of degenerate classical solutions, each of them with the same energy E_0.

In the case of the IMRT optimization problem, we are dealing with a non-trivial quantum Spin-Glass Hamiltonian type with over 32000 long-range interactions, hence the TTN-algorithm can be quite sensitive to the initialization procedure. Therefore, each run is repeated several times by randomly initializing the TTN-algorithm each time. At the end, the best run can be selected by comparing the resulting energies.

6.3 CLASSICAL SOLVERS

In order to validate the solution returned by the TTN algorithm, two different classical optimization algorithms have been employed. They are briefly described in this section.

6.3.1 *Simulated annealing*

Simulated annealing (SA) is a widespread combinatorial optimization method based on randomization techniques [189], in particular based on the Metropolis-Hastings algorithm. By varying a control parameter T, called *temperature*, it is possible to explore the landscape of a target free-energy function in order to find its global minima. SA is typically applied to large-scale non-convex optimization problems where the number of local minima in the energy landscape is very high. The algorithm always requires a starting point which in practice is either a random one or the best one known for a specific problem. From the initial starting point, a rule to generate new configurations is given: Configurations with an energy lower than the previous configuration are always accepted (with probability one). On the other hand, a move towards configurations with a higher energy is accepted with a certain probability which significantly reduces the risk of getting stuck into a local minimum. In this case, the acceptance probability decreases dynamically with the temperature T, which itself decreases during the optimization from a given value T_{max} to T_{min} according to a pre-defined annealing schedule.

In this work, SA is applied to find the ground state of the classical Ising spin-glass problem representing the initial IMRT optimization problem. We always start the SA from a random configuration for the spins in the lattice. New configurations are generated by flipping a randomly chosen spin in the prior configuration. Due to the intrinsic stochastic nature of SA, the same problem is solved several times. The number of repetition is called N_{runs}. For both, the sphere and the prostate cancer case, N_{runs} was set to 100.

The code used in this study to perform SA is based on the Python library in [190].

6.3.2 *Quadratic programming*

Quadratic programming (QP) refers to a set of widespread methods for solving (non-linear) quadratic optimization problems subject to linear constraints. A general QP problem can be formulated as follows:

$$\min_{x} f(\mathbf{x}) = \min_{x} \mathbf{q}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T Q \mathbf{x} \quad s.\ t. \quad \begin{cases} A\mathbf{x} = \mathbf{a}, \\ B\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases} \tag{6.8}$$

where $f(\mathbf{x})$ is the target function, $\mathbf{q}$ its gradient and Q its Hessian matrix. The equation $A\mathbf{x} = \mathbf{a}$ contains all the equality constraints, while $B\mathbf{x} \leq \mathbf{b}$ all the inequality constraints.

In the QP formulation, the initial IMRT optimization problem in Eq. (6.2) can be written as follows:

$$\min_{\mathbf{x}} \ 2\sum_{r} \left(\mathbf{x}^T \tilde{A}_r^T \tilde{A}_r \mathbf{x} - \mathbf{D}_r^{(P)} \tilde{A}_r \mathbf{x} \right), \ 0 \leq \mathbf{x} \leq 1 \tag{6.9}$$

where r is an index running over the different volumes or organs (the targeted tumor and the OARs), $\tilde{A}_r$ is the influence matrix for the r-th volume whose entries are weighted by the pre-assigned priorities $\gamma_{i(r)}$ to each voxel, i, and $\mathbf{D}_r^{(P)}$ a vector containing the dose objectives for the voxels in the r-th volume.

`MATLAB`®'s *Optimization Toolbox*™ provides the function *quadprog* which exploits different solvers to attack a wide class QP problems. In this work, this methods is used to solve the initial quadratic problem and produce results to be compared to those obtained with TTN.

6.4 APPLICATION TO THE QUADRATIC IMRT PROBLEM

In the following, we show the results of TTN for two different toy models to validate our approach, followed by a more realistic anatomical scenario simulating a prostate IMRT treatment. We first compare the TTN algorithm with the analytical solution for a 3D box toy-model scenario. For the second, more complex two-sphere model and the more realistic IMRT phantom, we lack a general analytical solution. Therefore, we test the results of our TTN approach by comparing it with two different optimization methods, Quadratic Programming (QP) and Simulated Annealing (SA). On the other hand, the problem in the Ising-type formulation in Eq. (6.4) is addressed using both TTN and SA. QP and SA are used to validate the results obtained with the TTN approach.

It is important to point out, that both SA and TNs algorithms contain elements of randomness: samples of independent and randomly initialized simulations are always collected and the best solution is considered. This also allows us to calculate the standard deviation of the samples and have a quantitative idea of the general behavior of the algorithms.

6.4.1 *The bipartite box*

In this section we describe a simplified analytical model used to validate the correctness of the mapping discussed in Sec. 6.2.1 and further our TN approach.

The model consists of a 3D box subdivided into two different regions for which we assign specific dose prescriptions as shown in Fig. 6.2. The red number on the front of the box corresponds to the desired beamlet weights for each bipartition. We radiate the cube with two beams from two opposite directions (i.e. $\theta_1 = 0°$ and $\theta_2 = 180°$), here represented as two rectangles, with a variable number of beamlets N_B for each beam, two in this example. The radiation beam is modelled as an *ideal beam* which releases the same amount of dose to each voxel. Moreover, we neglect scattering effects limiting the interactions among different beamlets to those acting geometrically on the same voxels only. The dose prescriptions and the number of voxels are chosen in a way to ensure the exact ground state with energy $E_0 = 0$. Therefore, for instance, we choose dose values which are compatible with the number of discretization levels 2^{N_Q} used in Eq. (6.3).

We point out, that these conditions are not fulfilled in general cases. Anyhow, we introduce them to obtain an exactly solvable model for the sake of validation. Later on, we present a more realistic anatomical scenario of cancer treatment. The model is then described by the influence matrices A_r for each region r, the priorities γ_i and the dose prescriptions $D_i^{(P)}$ for each voxel. These are the same information that can be extracted from a real TPS, as we will see also in the following.

The number of bits used to represent each beamlet is set to $N_Q = 4$ and will be kept constant through the whole study. This number is choosen following the IMRT literature where the effect of beamlet intensities discretization has been widely studied in the past [191].

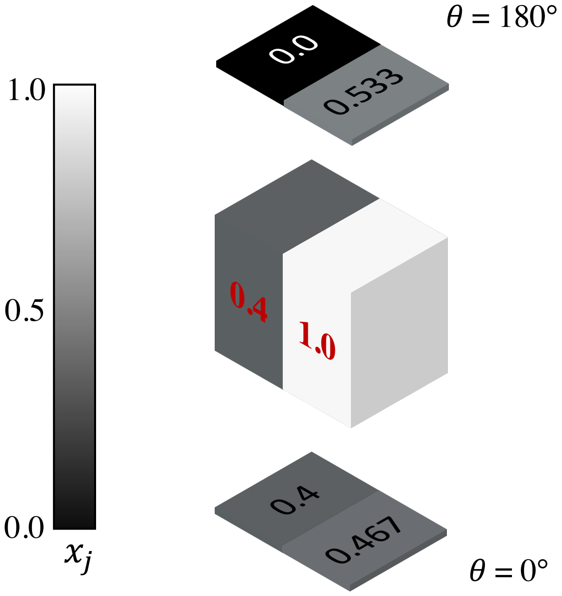


Figure 6.2: Optimization on the 3D box toy model. The box is radiated from two opposite angles $\theta = 0°$ and $\theta = 180°$. The red numbers on the box are the desired beamlet weights; the number on the upper and lower rectangles are the beamlet weights obtained with TTN. Their sum for each partition equals the desired values.

The target dose prescriptions for the two partitions are $D^{(P)}_{left} = 6.0$ and $D^{(P)}_{right} = 15.0$, which are arbitrary number chosen according to the number of discretization levels for the x_j, 16 in this case. The influence matrix, mapping the beamlets intensities to the voxels, is defined to have a uniform dose release into the box. Thereby, as we simplified physical effects of the beam, we end up with the influence matrix consisting of either zero and otherwise constant entries mapping the beamlet intensities to the voxels.

The values contained inside the upper and lower rectangles correspond to the beamlet weights obtained with the TTN algorithm. It can be observed that their sum equals the desired values for each bipartition, showing a perfect agreement between the analytical and numerical solution. This demonstrates the effective operation of the algorithm and the accuracy of the mapping procedure.

We point out, that, in general, there can be more than one configuration satisfying the constraints: indeed, depending on the system parameters, the ground state of the Ising-Hamiltonian can be degenerate. Anyhow, for the treatment, we are satisfied obtaining anyone of the degenerate ground states.

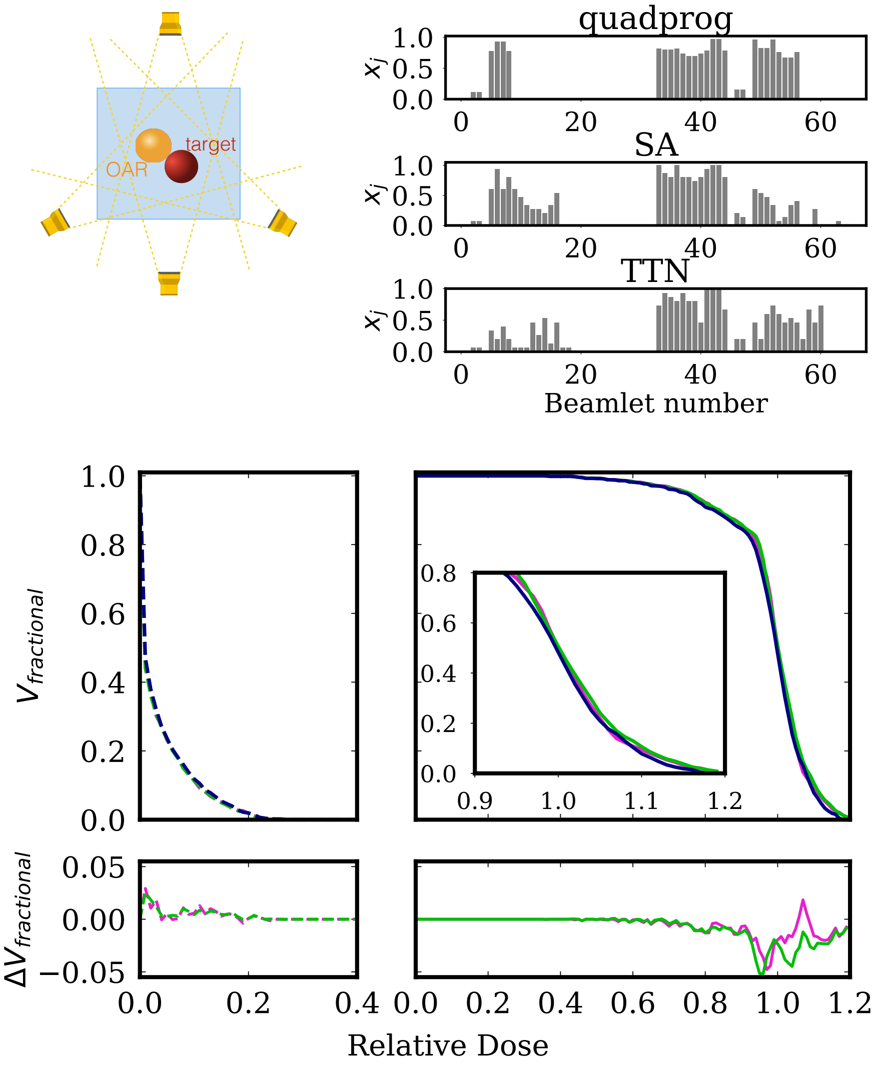


Figure 6.3: Schematic representation of the model with the red sphere as target and the orange sphere as OAR;(b) best final beamlet configuration for the three algorithms; (c, top) Cumulative DVHs obtained with the three algorithms: QP (pink), SA (green) and TTN (blue). On the left panel the orange sphere (dashed line); on the right panel the red sphere (solid line). (c, bottom) Difference in $V_{fractional}$ between TTN and QP (pink line), and TTN and SA (green line).

6.4.2 *The sphere*

In the following, the application of the TTN algorithm to a more realistic clinical scenario. In this case, instead of assuming an *ideal beam* we model the photon beam using the Matlab-based software CERR [192]. This software generates radiotherapy plans and can be used to obtain the influence matrixes a_{ij}. It allows working with physical effects introduced by the beams such as scattering and further allows for higher freedom in choosing the geometry, the number of beams and other typical model parameters. Thereby, the underlying medical images are in the standardized DICOM format.

In Fig. 6.3a we illustrate the model analyzed in this paragraph. The model described consists of cubic box dimensions $(50 \times 50 \times 100)$ cm^3 filled with water in which two spherical regions, shown in red and in orange respectively, of diameter $d = 3$ cm. The box is irradiated using four beams at 0°, 120°, 180°and 240° and $N_B = 64$ beamlets (16 beamlets/beam). The influence matrices (A_{red} and A_{orange}) are obtained using the CERR's dose calculation algorithm QIB [193] with the default settings; the dimensions of the beamlets are set to (1×1) cm^2.

The optimization goals are set to $D^{(P)}_{red} = 50$ Gy for the red sphere and $D^{(P)}_{orange} = 0$ Gy for the orange one, considering the first as the targeted tumor and the second as an OAR. Each sphere is weighted equally in the cost function with $\gamma = 1.0$. During the optimization procedure, the influence matrixes are always normalized to keep the final beamlets intensities in the interval $[0, 1]$.

In the mapping to the discrete problem, the bit-depth was fixed to $N_Q = 4$ bits, resulting in a fully-connected lattice of 256 sites for the underlying Hamiltonian. The total number of non-zero interaction terms was $n_{int} = 32640$. Thus, the underlying quantum many-body system is a challenging long-range spin-glass Ising model to be solved with the TTN algorithm.

In practice, it is unfeasible to obtain $E_0 = 0$ and thereby to reach exactly the prescribed dose distribution. Thus, in this example, the optimization balances the different goals for each organ according to their priorities γ. For this reason, to evaluate the quality of the results returned by the TTN algorithm, the same optimization task was tackled using the `MATLAB`® built-in function *quadprog*, which exploits QP to optimize the cost function, and SA. We recall that the optimization with *quadprog* was directly performed on the function in Eq. (7.11), while SA and the TTN were applied to the discretized problem in Eq. (6.4).

A standard method used for plan quality evaluation is the cumulative DVH histogram, which shows a 2D projection of the 3D dose distribution inside a given volume. It represents the fractional volume receiving at least a given value of dose. Given a generic volume, r, we can easily build the dose vector $\mathbf{D}_r(\mathrm{x})$ as described in Eq. (6.1) by applying its influence matrix, A_r, to the beamlets vector $\vec{x}$. The resulting vector contains the total dose delivered to each one of the voxels in the volume r. By subdividing the dose interval $[0, D_i^{(max)}]$ into n_b (dose) bins, for each of them we can count how many voxels receive a dose greater or equal than the corresponding dose value. In other words, the number of entries in the $k - th$ bin indicates the number of voxels receiving at least the corresponding dose. The obtained distribution results in a cumulative DVH with the fractional volume represented on the y-axis and the dose values on the x-axis. We point out, that in this representation of the dose distribution we lose the spatial information of the problem.

On the top panels of Figure 6.3, we show the DVHs obtained with the three methods. It is clearly visible, that the three methods show a very good qualitative agreement in the resulting DVHs for each organ. This agreement is further quantitatively confirmed by the energy E_0 - or cost - after the minimization: Within the statistical uncertainty, all three methods result in a final energy $E_0 = (1.81 \pm 0.08) \times 10^{-2}$. By looking at the bottom panels of Figure 6.3 we see that the difference between TTN and the other two methods in the relative volume coverage is globally very close to zero, with only a few peaks at about 2-5 %

In Figure 6.3, the final beamlets configurations for each of the three methods are presented. Despite their global consistency, we observe that local differences arise. First, this is due to the fact the final configurations have slightly different energies, despite are all consistent. However, in general, there may exist more than one configuration satisfying the constraints and minimizing the energy E_0 for the underlying system and this effect is further amplified when comparing the optimization on the discrete space (SA and TTN) to that on the continuous one (quadrog). This happens because the energy landscape may be altered by the discretization procedure. Thus, in this case, we have that the ground state of the underlying Hamiltonian is either degenerate or its energy gap is reasonably small. In fact, the underlying quantum spin-glass Hamiltonian has a non-trivial spectrum with many local optima and, depending on the system parameters, degenerate ground-states.

6.4.3 *Prostate cancer treatment with TG-119 IMRT phantom*

We now show the results obtained on a standard IMRT phantom provided by the American Association of Physicists in Medicine Task Group 119 for use in institutional IMRT commissioning [194, 195]. This dataset contains several segmented structures and we chose the following with the aim of simulating a prostate cancer cases: prostate as the targeted tumor, bladder and rectum as the OARs.

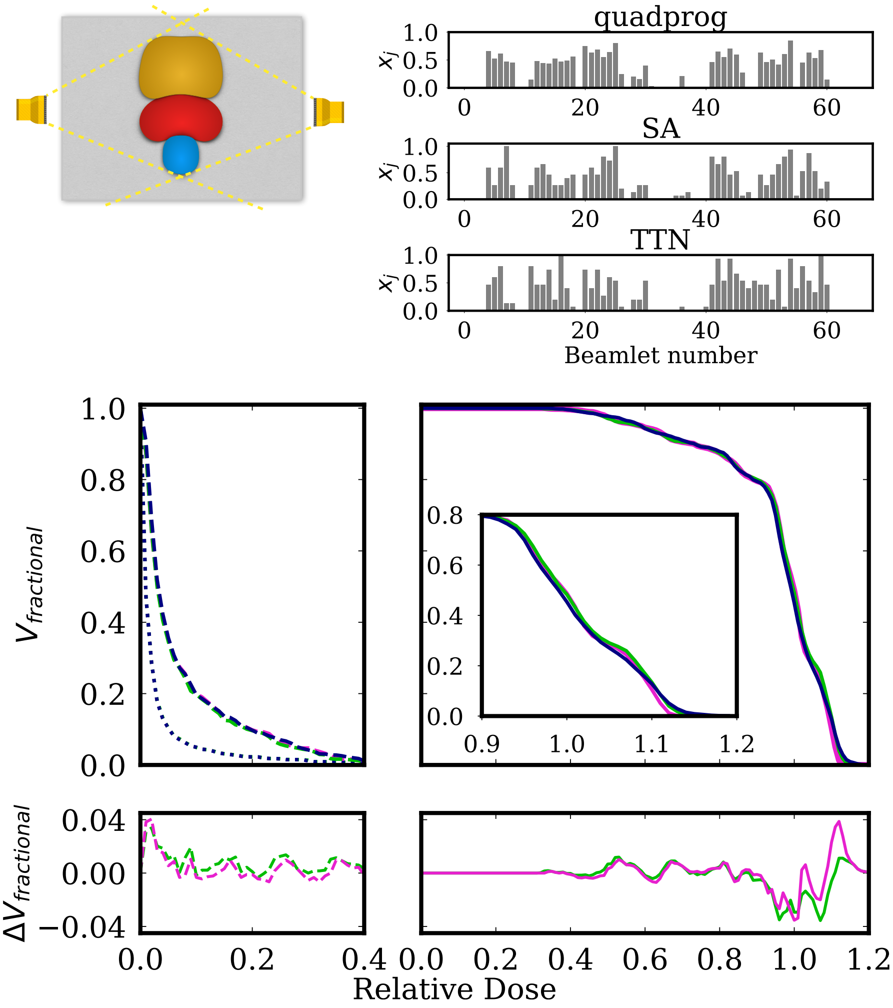


Figure 6.4: Optimization on the TG119 IMRT phantom. (a) Schematic representation of the model with the the prostate as target (red) and the bladder (brown) and the rectum (blue) as OARs;(b) best final beamlets configuration for the three algorithms; (c, top) Cumulative DVHs obtained with the three algorithms: quadprog (pink), SA (green) and TTN (blue). On the left panel the two OARs: bladder (dotted line) and rectum (dashed line). On the right panel the prostate (solid line). (c, bottom) Difference in $V_{fractional}$ for the rectum (dashed line) and the prostate (solid line) between TTN and quadprog (pink line), and TTN and SA (green line).

The geometry we used is characterized by two beams (33 and 31 beamlets) placed at 90°, 270° as shown in Fig. 6.4a, resulting in a total amount of $N_B = 64$ beamlets. The dimensions of each beamlet were (0.9×0.9) cm^2. The dose prescriptions were set to $D^{(P)}_{prostate} = 50$ Gy and $D^{(P)}_{bladder} = D^{(P)}_{rectum} = 0.0$ Gy and the priority assigned to the different structures was $\gamma = 1.0$ for all of them. The dose calculation algorithm used was CERR's QIB algorithm in the default settings. The fraction of non-zero elements in the influence matrixes was 0.55 for the $A_{prostate}$, 0.28 for $A_{bladder}$ and 0.45 for A_{rectum}. For the discrete problem, the system obtained is again characterized by 256 fully-connected long-range interacting spins and a total amount of $n_{int} = 32640$ interaction terms.

The top panels of Fig. 6.4c show the comparison between the DVHs for the three optimization methods. The bottom panels show that the difference in $V_{fractional}$ between TTN and the other two methods for the rectum and the prostate is within the 4%, proving a very good quantitative agreement between them. The results for the bladder are not shown since the differences were negligible. This agreement is additionally confirmed by the obtained energy E_0 - or cost - after the minimization: within the statistical uncertainty, the three methods result in a final energy $E_0 = (4.3 \pm 0.1) \times 10^{-2}$. This result further confirms what found from the study on the toy models.

6.5 SIGNIFICANCE OF THE RESULTS AND PROSPECTIVES

This feasibility study provided the first evidence of the applicability of TNMs to an IMRT optimization problem, fostering new applications of quantum-inspired techniques to the solution of classical optimization problems. Along the road, we defined a clear strategy to map the classical problem to simulated quantum-type hardware.

The results presented herein indicate that the TNMs approach can achieve results compatible with other optimization techniques such as QP and SA. However, at the end of this investagations there are three essential aspects that are noteworthy: (1) a reduced number of beamlets was used to reduce the complexity of the resulting models (2) the cost function describing the IMRT problem was kept convex for sake of simplicity and (3) the TNMs code "out-of-the-box" originally engineered for typical quantum systems with significantly fewer interactions but higher entanglement. Further software developments will allow to address these three points, increasing the TNMs approach efficiency. In particular, by extending this study to non-convex and non-differentiable functions further significant steps forward can be made towards real-world scenarios and their use in everyday medical care.

In the next chapter, a novel strategy developed and implemented during PhD project for the optimization of functions with a general mathematical definition using TNs and, more generally, quantum hardware will be described.

7

EFFICIENTLY OPTIMIZING CONTINUOUS VARIABLES PROBLEMS ON QUANTUM HARDWARE

In Ch. 6, an approach for the optimization of quadratic functions using qubits-based resources was presented. We have seen that, it is always possibile to estrablish a one-to-one mapping between a quadratic polynomial cost-function and a fully connected Ising-Hamiltonian. However, the presence of continuos variables might require using a large number of qubits to represent each variable with the needed precision to get accurate solutions to the optimization problem. Thereby, the scalability of the approach is highly affected by this fact, hence more efficient strategies are needed for the solution of real-world problems. Indeed, the first generation of future quantum computers, the so called Noisy Intermediate-scale Quantum (NISQ) computers will be equipped with a limited number of qubits, and this sets a hard constraints on the types of problems that can be solved [196, 197]. Therefore, the possibility of applying quantum computers in radiotherapy and more in general to real-world continuous variables problems in the near future also passes through the development of novel efficient variable discretization strategies. Additionally, several real-world problems are described by non-convex, non-polynomial and possibly non-differentiable functions. Therefore, usual approaches well-grounded approaches such as the one presented in Ch. 6 as well as the QUBO or PUBO methods commonly used in combinatorial optimization might not the best choice [198, 199].

THIS CHAPTER presented an original approach developed within this PhD project which provides a complete framework for the optimization of general continuous variable functions on quantum-hardware. The proposed approach is herein applied to different problems, including a fixed-gantry IMRT beamlet optimization task.

7.1 DINAMICALLY ADAPTED DISCRETIZATION

When solving a continous variables problems using qubits-based resources, the number of qubits N_Q used to represent a variable x_i determines automatically the number of values x_i can take. This is the idea behind the binary-decimal conversion strategy described in Eq.(6.3). However, this has two main limitations:

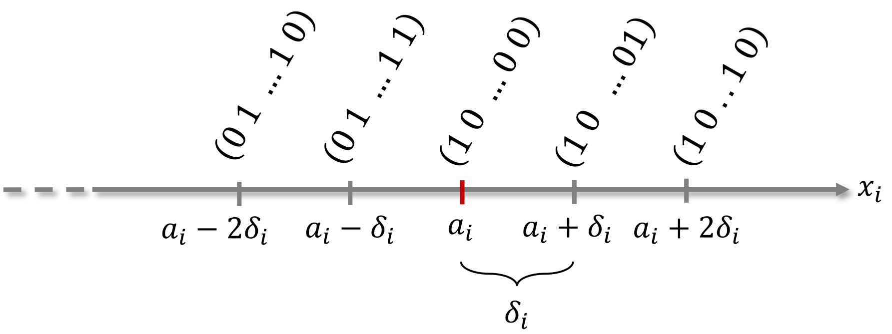


Figure 7.1: Representation of the variable discretization procedure implemented in this work.

1. The domain of each variable is bounded between 0 and a maximum value $2^{N_Q} - 1/K$, with K a normalization constant. More in general, a translation can be applied so that the domain becomes bounded between $x_{i,min}$ and $x_{i,max}$.
2. Within $[x_{i,min}, x_{i,max}]$, the values that the variable x_i can take are automatically induced by the discretization procedure. Within the standard framework, the discretization of each variable is set at the beginning of the optimization and left unvaried during the whole process.

When solving an optimization problem we might not have information about the region where the global minimum is located. Therefore, starting from this well-grounded approach, a more flexible discretization strategy that allows for a dynamic adaption during the discretization is proposed to overcome the aforementioned limitations.

As depicted in Fig. 7.1, let's consider a configuration $x_i = a_i$ and associate the sequence (1 0 0...). Since the use of this mapping allows representing 2^{N_Q} discrete values for the variable x_i, the remaining $2^{N_Q} - 1$ values are taken around a_i according to a spacing δ_i, called *discretisation step*. It should be noticed that this approach is not left-right symmetric with respect to a_i. Mathematically, this is described by the following equation:

$$x_i = a_i + \left(\sum_{n=1}^{N_Q} 2^{n-1} b_n - 2^{N_Q - 1} \right) \delta_i \tag{7.1}$$

which is an extension of Eq. (6.3) and allows for a better fitting representation of the problem variables. Let's consider an example with $N_Q = 4$ and try to apply Eq. (7.1) to the sequences (1 0 0 0), (0 1 1 1), (0 0 0 0) and (1 1 1 1) which should result in a_i, $a_i - \delta_i$, $a_i - 8\delta_i$ and $a_i + 7\delta_i$ respectively. Recalling that qubits are counted from right to left, hence the right-most and left-most qubits are addressed with $n = 1$ and $n = N_Q$, respectively, it reads:

$$
\begin{aligned}
&(1\,0\,0\,0) \rightarrow x_i = a_i + (2^0 \cdot 0 + 2^1 \cdot 1 + 2^2 \cdot 0 + 2^3 \cdot 1 - 2^3)\delta_i = a_i \\
&(0\,1\,1\,1) \rightarrow x_i = a_i + (2^0 \cdot 1 + 2^1 \cdot 1 + 2^2 \cdot 1 + 2^3 \cdot 0 - 2^3)\delta_i = a_i - \delta_i \\
&(0\,0\,0\,0) \rightarrow x_i = a_i + (2^0 \cdot 0 + 2^1 \cdot 0 + 2^2 \cdot 0 + 2^3 \cdot 0 - 2^3)\delta_i = a_i - 8\delta_i \\
&(1\,1\,1\,1) \rightarrow x_i = a_i + (2^0 \cdot 1 + 2^1 \cdot 1 + 2^2 \cdot 1 + 2^3 \cdot 1 - 2^3)\delta_i = a_i + 7\delta_i
\end{aligned}
$$

which confirm the expected results. However, this approach just solves the first of the two aforementioned limitations. Therefore, a second feature was introduced, namely the possibility of adapting the discretization step dynamically during the optimization.

This is done by considering the optimization history. Indeed, δ_i can be either increased or decreased to explore a differents sets of values around a given configuration, which is a significant improvement compared to the original approach where everything is frozen since the beginning. Additionally, in this implementation, the δ_i is made variable-dependent, namely each variable has its own associated discretization step which is different from all the others' and, therefore, can have its own history. Therefore, as it will be shown in the following sections, this approach allows for a better fit of the qubits around a given configuration depending on the optimization history of each variable independently.

In this work, the strategy adopted to vary δ_i consists of the following three steps:

- **Initialization.** An initial arbitrary value of the spacing δ_i is set at iteration $t = 0$ which might be the same for each variable. There is no pre-defined rule to determine a good starting value for this parameter and represents yet another instance of the usual exploitation vs exploration issue common to any complex optimization. In general, intuition suggests that a large δ_i might prevent the algorithm to get stacked into local minima since the early optimization phases, but this unavoidably increases the optimization time. Conversely, small value of δ_i may let the algorithm to converge faster to a solution but this increases the probability to get stacked into local minima.

- **Update.** Let's consider a configuration of I variables $\vec{x}(t_{n+1}) = (x_i(t_n) \ldots x_i(t_n) \ldots x_I(t_n))$ at iteration $t = t_n$ and a generic function $F(\vec{x}(t_n)) \equiv F_{t_n}$. If $x_i(t_{n+1}) = x_i(t_n)$, the step size of variable x_i is updated according to the following rule:

$$
\begin{cases}
\delta_i(t_{n+1}) < \delta_i(t_n) & if\ F_{t_{n+1}} = F_{t_n} \\
\delta_i(t_{n+1}) > \delta_i(t_n) & if\ F_{t_{n+1}} < F_{t_n}
\end{cases}
\tag{7.2}
$$

  Otherwise, in case $x_i(t_{n+1}) \neq x_i(t_n)$, δ_i is left unchanged.

  The rationale behind this updating rule is the following. The spacing δ_i can be interpreted as the *confidence* of the algorithms in exploring the neighborhood. Therefore, it is decreased when both the variable and the function value do not change between two iterations, namely, the algorithm loses confidence and moves closer to the current value to search for new solutions within a smaller neighborhood of x_i. Conversely, δ_i is increased if the variable value does not change but a lower value of the function is found (i.e. due to a change of another variable) since this also increases the confidence of the unchanged variable to explore a larger neighborhood of the new local function landscape. Finally, if $x_i(t_{n+1}) \neq x_i(t_n)$, δ_i is left unchanged since the current step is supposed to be on a good scale which might have resulted from a history of changes, therefore the algorithm tries to get the best from this situation.

- **Refresh.** The value of δ_i for a given variable can be restored to an arbitrary value if it becomes smaller than a predefined threshold. This is called the *refresh phase* and may help avoid the algorithm to get stacked into local minima because of a too-small discretization step. Similarly, after an arbitrary number of iterations, all the δ_i could be restored to their initial value. This is called *hard refresh* and it corresponds to starting a new optimization session from the best configuration found. This also might help improve the convergence.

In Fig. 7.2, an example of the evolution of the sub-region size $\delta_i(t_n)$ for three different variables over 100 iterations of the algorithm is shown. We can observe the different trend for each of the three variables due to the variable-dependent dynamic adaptation of the sub-region size.

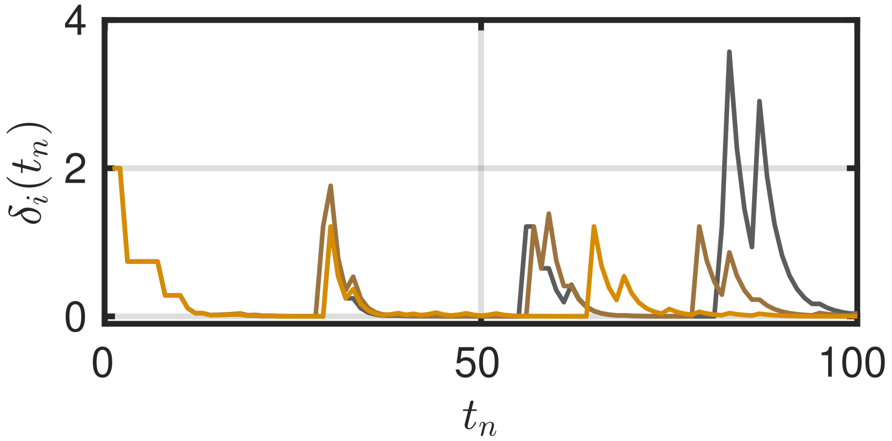


Figure 7.2: Example of evolution of the step size for three variables over 100 iterations.

7.2 BUILDING A LOCAL APPROXIMATION MODEL

A general scalar optimization problem can be described as follows:

$$\vec{x}_{opt} = argmin_{\vec{x}}(F(\vec{x})) \tag{7.3}$$

where F is a possibly non-convex, non-differentiable and non-polynomial scalar function defined on the variables $\vec{x} = (x_1, ..., x_I)$, with I the size of the problem. In general, each x_i can take values in the real numbers. The configuration $\vec{x}_{opt}$ is the optimal solution to the problem and it is what we are interested in.

As it was shown in Ch. 6, mapping a quadratic problem to a quantum-type Hamiltonian is quite straightforward, provided a variable discretization strategy. Indeed, as we have already seen, a one-to-one mapping between the quadratic function and a fully-connected (Spin glass) Hamiltonian can be established.

Conversely, in case of more generll functions this direct mapping does not exist and a different strategy is needed. Therefore, following and extending the approach in [170], an all-to-all spin-glass Hamiltonian is built around a given configuration $\vec{x}(t_n)$ at iteration t_n, namely:

$$H(t_n) = \sum_k h_k(t_n)s_k + \sum_k \sum_{l>k} J_{kl}(t_n)s_k s_l \tag{7.4}$$

where s_k and s_l are spin (or qubits) terms associated with sites k and l with values in $\{-1, 1\}$, while $h_k(t_n)$ and $J_{kl}(t_n)$ are the corresponding single-particle and two-body coupling terms, respectively. The Hamiltonian in Eq. (7.4) is meant to be a local quadratic approximation of the cost function around $\vec{x}(t_n)$.

It is important to stress the fact that, sparser interaction schemes than the fully-connected one might be needed depending on the connectivity degree of the specific qubits-based resource used. Figure 7.3 displays two instances of interaction schemes emerging from the local approximation procedure centred on the same configuration of the 8-variable Gramacy-Lee cost function (refer to Sec. 7.4). As seen, denser interaction schemes emerge by using a higher number of qubits to discretize the problem variables, making the underlying Hamiltonian problem more difficult to solve. Consequently, the selection of the number of qubits necessitates evaluation for each problem.

Following the discretization strategy proposed in Sec. 7.1, to each variable x_i in the original problem a set of N_Q qubits in the Ising Hamiltonian is associated. The relation between the q_b in Eq. (6.3) and s_b is the following:

$$q_b = \frac{s_b + 1}{2}. \tag{7.5}$$

However, differently from convex (i.e. quadratic) problems where an exact mapping can be found between the initial function and the Ising-type Hamiltonian in Eq. (7.4) [200], in this case, the model serves as

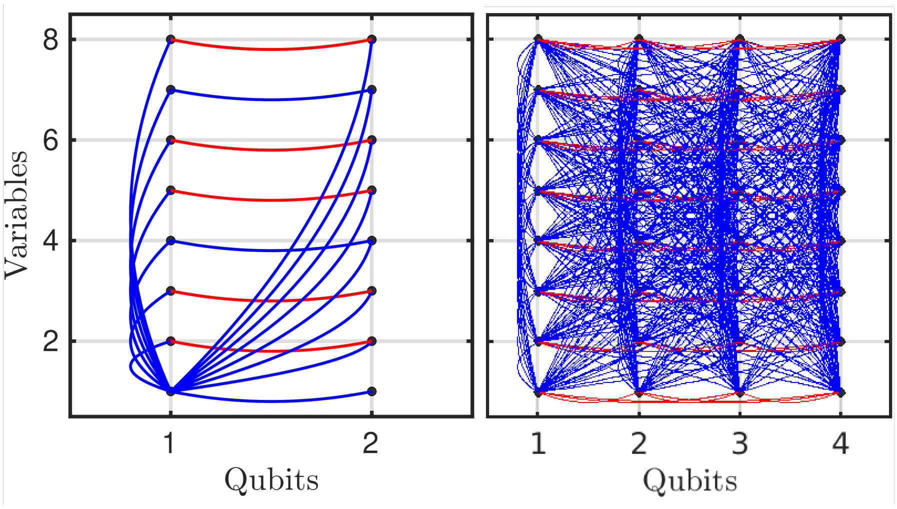


Figure 7.3: Example of interaction schemes arising from the local approximation around a given configuration $x(t_n)$ during the optimization of the same problem of 8 variables discretized using $N_Q = 2$ (left) and $N_Q = 4$ (right). Black points mark the qubits of the lattice, while blue and red lines indicate positive and negative couplings, respectively.

a local-approximation of the function around the configuration $\vec{x}(t_n)$ and needs to be updated iteratively.

In order to make Eq. (7.4) to be a good surrogate of $F(\vec{x})$ around a given configuration $\vec{x}(t_n)$, the set of parameters $\{h_k(t_n), J_{kl}(t_n)\}$ need to be tuned. The adopted strategy is the following:

1. Given a configuration $\vec{x}(t_n)$ a number of neighbor configurations equal to the number of coupling terms in the surrogate Ising model is generated. This number depends on the interaction scheme chosen and it is equal to $\frac{N_V N_Q (N_V N_Q + 1)}{2}$ for a fully-connected model, with N_V the number of variables in the problem.

2. The original function is evaluated on each configuration;

3. For each configuration and corresponding function value, the following equation is written:

$$F(\vec{x}_i(t_n)) = \sum_k h_k(t_n) s_k^{(i)} + \sum_k \sum_{l>k} J_{kl}(t_n) s_k^{(i)} s_l^{(i)} \tag{7.6}$$

   This results in a system of linear equations in the variables $\{h_k(t_n), J_{kl}(t_n)\}$. The terms $s_k^{(i)}$ and $s_k^{(i)} s_l^{(i)}$ are determined by the discretization of the variables according to the adopted procedure and play the role of the coefficients of the linear system of equations. The solution of such systems of equation returns the set of $\{h_k(t_n), J_{kl}(t_n)\}$ that should make the Ising model locally similar to the function, namely:

$$F(\vec{x}_i(t_n)) = H(\vec{x}_i(t_n)) \qquad x_i \in \left[a_i - \delta_i \frac{2^{N_Q}}{2}; a_i + \delta_i \left(\frac{2^{N_Q}}{2} - 1\right)\right] \tag{7.7}$$

The described procedure is performed at each iteration and solutions to the original problem are sampled from the low-energy configurations of the local approximation model. This is the actual step where the qubits-based solver acts. Therefore, at the end of each iteration t_n, the new candidate solutions generated by the qubits-based solver are evaluated on the original function to seek the one with the lowest associated function value.

More in detail, at the end of each iteration there are two sets of solutions available:

- $\mathcal{N}$: the set of neighbor of a given configuration that were used to build the surrogate;
- $\mathcal{Q}$: the set of configurations generated by the solver;

The number of function evaluations, n_{eval}, is recorded during the optimization and can be used to set a stopping criterion, namely:

$$if \quad n_{eval} > eval_{max} \quad \Rightarrow \quad stop! \tag{7.8}$$

with $eval_{max}$ the maximum number of function evaluations allowed. The stopping criterion based on the number of evaluations is useful when the solutions returned by different solvers need to be compared because it is solver-agnostic. In fact, although a criterion based on the maximum number of iterations might be used, the meaning of *iteration* strongly depends on the specific solver used and might lead to unfair comparisons.

7.3 QUBITS-BASED SOLVERS

In this work, the algorithm was coupled with two different qubits-based solvers:

1. The TTN algorithm already described in Ch. 6;
2. the Simulated Annealing Sampler (SA) contained in the *dwave-ocean-sdk* by D-Wave [201].

In case of TTN, the associated variational problem described in Ch. 5 is solved to find the ground-state of system. As discussed also for the quadratic problem in Ch. 6, the final ground-state might be a coherent superposition of states. However, in this case, low-energy

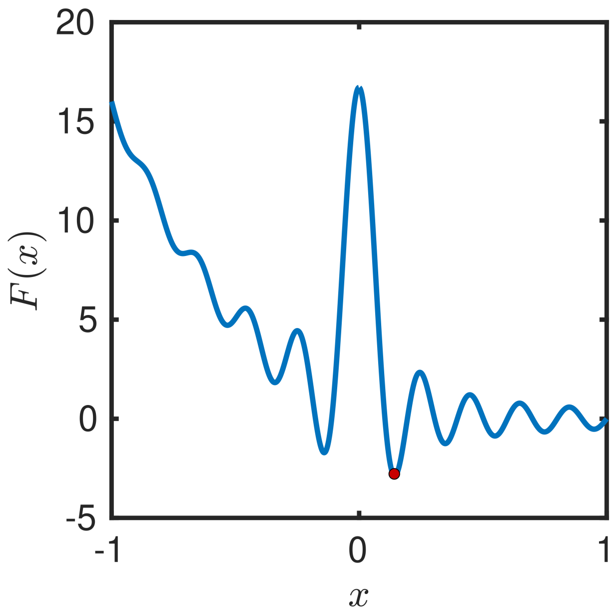


Figure 7.4: The one-dimensional Gramarcy-Lee function considered in this work. The red point marks the global minimum.

configurations were sampled from the ground-state instead of truncating the bond-dimension and taking the most-probabile one, only. In particular, the TTN is first mapped onto a MPS; then, configurations are sampled from the MPS ground-state as those populating it with the highest probability. This sampling approach allows for the evaluation of several solutions after every single optimization iteration.

For the simulated annealing sampler, repeated minimimizations are performed around the same configurations to sample low-energy configurations. This procedure is handled automatically be the package and the user can only set the maximum number of repetitions, which was set to 10000 in the present work.

Additionally, to test the effectiveness of using the local approximation models, also the best solutions provided by looking at the neighbor configurations (i.e. those used to build the local approximation) was considered. This will be referred to as best neighbor (BN) approach in the following.

7.4 APPLICATION TO THE GRAMACY-LEE FUNCTION

7.4.1 *Unconstrained optimization problem*

To validate the proposed approach, we considered the following function:

$$F(\vec{x}) = \sum_{i=1}^{n} \frac{\sin(\pi w_i\, x_i)}{2x_i} + (x_i - 1)^4 \tag{7.9}$$

which is an n-dimensional extension of the so-called *Gramacy-Lee* (GL) function [202], with $w_i > 0$ that will be referred to as GLN in the following. The GL function has a non-convex landscape characterized by several local minima and it is commonly used for testing optimization algorithms [203]. In Fig. 7.4, an example of the GLN function in Eq. (7.9) for $n = 1$ and $w_i = 10$ is shown [203]. We observe that it has a unique global minimum at $x = 0.14$. Furthermore, we observe the high peak at $x = 0$ where a non-differentiable point is located. The n-dimensional extension will still have a global minimum located in $\vec{x} = (0.14, 0.14, ..., 0.14)$. The analytical expression of the GL function makes it not suitable to be optimized using quadratic optimization strategies and therefore it is a good test bench for our algorithm.

7.4.2 *Comparison between qubits-based solvers*

The optimization problem in Eq. (7.9) was solved for increasing numbers of variables N_V. In this case, the number of qubits used for each variable was fixed to $N_Q = 2$. In each case, a set of $n_{samples} = 10$ optimization sessions was collected. Each optimization session was initialized from a different starting point and each starting point was sampled from a normal distribution randomly. For a given value of N_V the same set of starting points was used not to bias the comparison between samplers. Following Eq. (7.8), a stopping criterion based on the maximum number of function evaluations was set as $\max_{eval} = 150 \cdot (\frac{N_V \cdot N_Q (N_V \cdot N_Q + 1)}{2})$, where the term in the brackets correspond to number of configurations that are used to generate the local surrogate when a fully-connected interaction scheme is used (see as Sec. 7.1), while 150 is the maximum number of iterations allowed. In other words, exactly 150 iterations are completed at most when no candidate solutions are found by the sampler during the whole optimization session and thus only neighbor solutions are evaluated. The described procedure resulted in $\max_{eval} = \{2.04 \times 10^4, 8 \times 10^4, 3 \times 10^5, 1.2 \times 10^6\}$ for $N_V = \{8, 16, 32, 64\}$, respectively.

In Fig. 7.5, the evolution of the function value throughout the optimization is shown for different numbers of variables and solvers. The results obtained with TTN and SA are compared to the ones obtained using BN.

Among the three approaches, the best results were obtained using TTN either in terms of convergence speed and final function value. Additionally, the advantage of using TTN seems to become more significant for increasing N_V as it is suggested by the increasing gap between the two lines (TTN and SA). Finally, the larger slope of the TTN line compared to the SA seems to suggest that a further improvement could have been expected if the optimization had continued.
In the case of *BN*, we can observe that the algorithm get stucked into a local minimum since the early phases of the optimization resulting

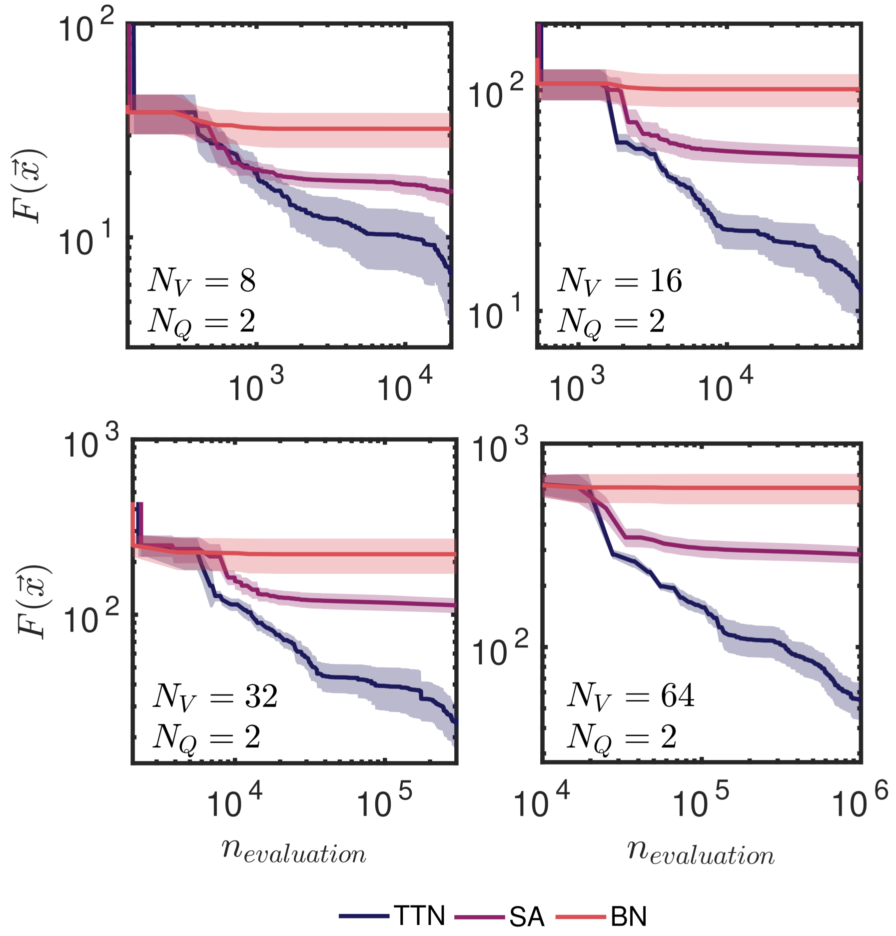


Figure 7.5: Evolution of the cost function value during the optimization (log-log scale). Four different values of N_V are considered, namely $\{8, 16, 32, 64\}$. For each solver, the solid line represent the average trend of the function value computed over $n_{sample} = 10$ optimization sessions. Shaded regions around the average line indicate the 95% confidence interval computed assuming a t-student distribution.

in poor performance. This proves the effectivness of the two solvers in providing good solutions to the problem other than the neighbour ones.

7.4.3 *The effect of the qubit-depth*

In Fig. 7.6, a comparison between the results obtained with $N_Q = 2$ and $N_Q = 4$ for the same problems of $N_V = 8$ and $N_V = 32$ variables is shown. As we can observe, in this case the use of $N_Q = 4$ does not show an improvement in the solution after the same number of function evaluations. It is worth noticing that this behavour could have been expected. In fact, using a larger N_Q requires a larger number of neighbor configurations to build the surrogate (this fact is particularly

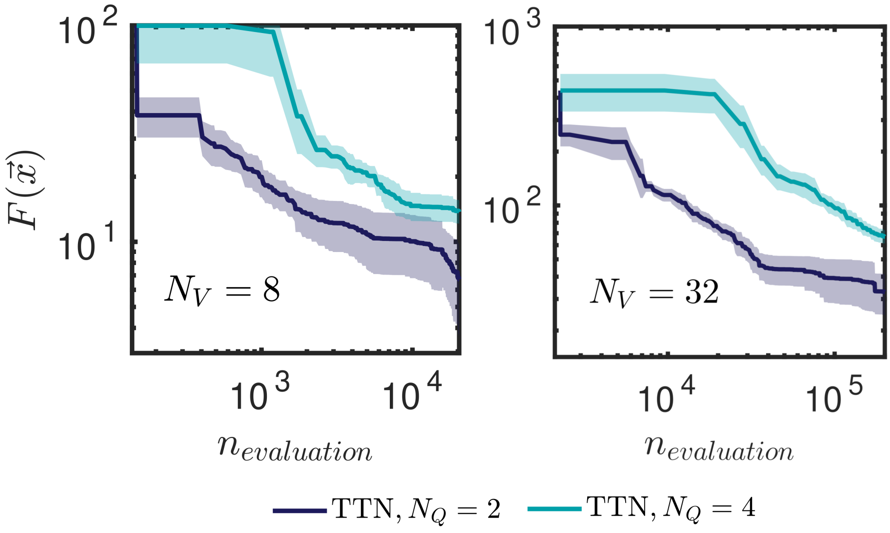


Figure 7.6: Comparison of results obtained on the GL function for $N_V = 8$ and $N_V = 32$ using our approach couple with the TTN for $N_Q = 2$ and $N_Q = 4$.

evident in the case of a fully-connected interactions scheme as the present one). Therefore, the sample $\mathcal{N}$ of neighbor solutions evaluated at each iterations is larger for $N_Q = 4$ than for $N_Q = 2$ and thus, after the same n_{eval}, the fraction of neighbor configuration is larger for the higher N_Q. On the other hand, these results suggest that the design of our approach make it intrinsically suitable to work with small N_Q and to scale up the problem.

7.4.4 *Comparison with a classical solver*

A classical optimization algorithm, the *Particle Swarm* (PS) [204] was considered to challenge the the solutions obtained with our novel optimization approach. Since the time it was first proposed in 1995, PS has been successfully applied to a variety of problems and has shown to be an effective and flexible algorithm. Its functioning is inspired by the social behaviour of birds. A group of *particles* is considered which are potential solutions to the optimization problem. Each particle starts from a position in the configuration space. Despite none of the particles knows were the minimum is located, information is shared across the particles during the optimization so that they can cooperate to find the best solutions. The number of particles used is the so-called *swarm size* (SZ). There are two main advantages in the use of particle swarm. First of all, it is not a gradient-based algorithms, a fact that make PS suitable for non-differentiable problems. Second, due to its working principle it may be easily parallelized. Over the past year, several variants of the PS algorithm have been proposed. In this work, we used its `MATLAB`® implementation [205].

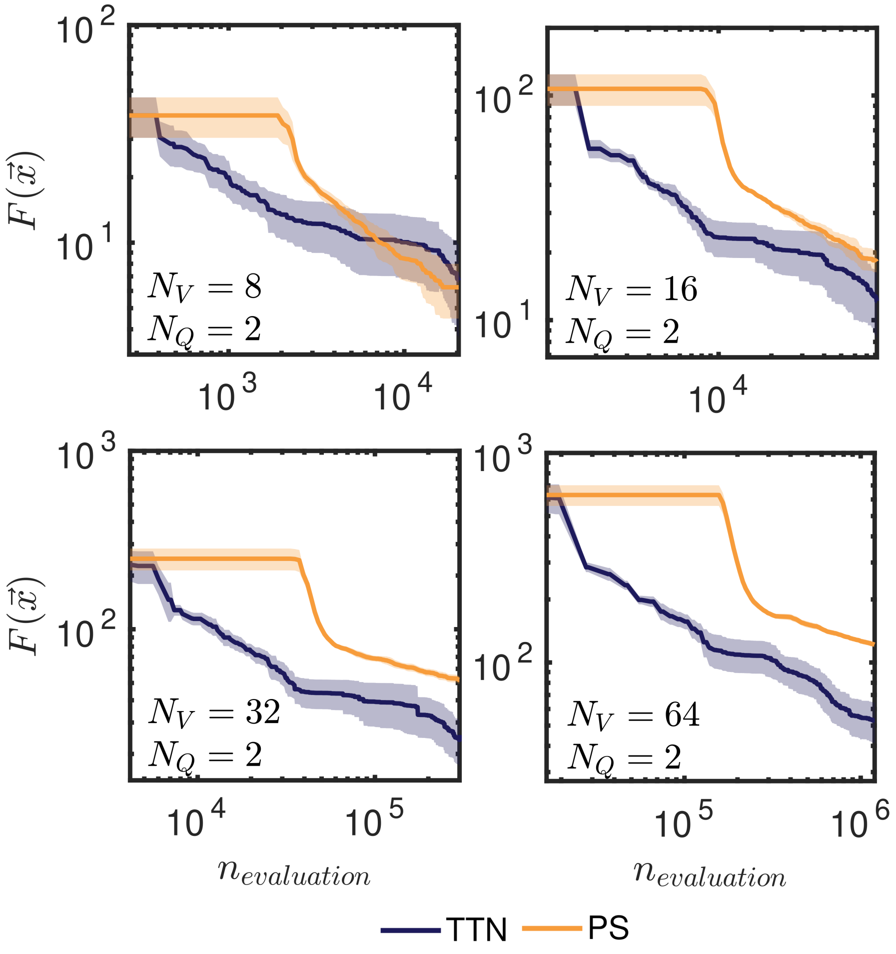


Figure 7.7: Comparison between our approach and PS for increasing number of variables. Each solid line represents the average trend of the function value computed over $n_{sample} = 10$ optimization sessions. The shaded areas around the average lines denote the 95 % confidence interval computed assuming a t-student distribution.

As it was done in Sec. 7.4.2, the comparison was performed for increasing number of variable. The same values of N_V were used and a set of $n_{samples} = 10$ optimization sessions was collected for each of them. In order to make the PS working in the same conditions as our approach, we proceeded as follows. We considered a swarm size (SZ) equal to the number of configurations used to build the local surrogate. Furthermore, the particles in swarm were initialized to the same positions as the configurations used to create the surrogate. Therefore, the two approaches were guaranteed to start exactly under the same condition. Finally, the comparison was performed for the same number of function evaluations using the same criterion adopted in Sec. 7.4.2.

The results of the comparison are shown in Fig. 7.7. The first thing we can observe is that the two algorithms have slightly different behavior, especially in the first part of the optimization process. Generally, PS stalled for a quite long time and then started converging very fast, while our approach showed a faster convergence since the early optimization phases. However, for $N_V = 8$ this difference seems not impact the final solution. Moreover, the final gap between the two lines increases in favor of the TTN approach as the size of the problem does. However, we believe that this might depend on the specific problem considered herein and should not be seen as general behavior. In any case, these results also show the effectiveness of our novel optimization approach compared to a state-of-art classical algorithm such as PS. Furthermore, following the considerations done in 7.4.3, it is worth noticing that these results were obtained for $N_Q = 2$.

7.4.5 *Constrained optimization*

The problem we have considered in the previous section falls among the so-called unconstrained problems. In fact, the variables could take any value, without restrictions. However, it happens for several real-world problems that the values that the problem variables can take are limited to a specific region of the space. This is, for instance, the case of the beamlet weights in the IMRT fluence optimization problems, which are required to be positive.

Therefore, with the aim of applying the aforementioned optimization approach to the IMRT optimization problem, a constrained version of the optimization problem described in Eq.7.9 was considered to test the functioning of the optimization approach.

In particular, Eq. (7.9) was modified as follows:

$$F(\vec{x}) = \sum_{i=1}^{n} \frac{\sin(10\pi x_i)}{2x_i} + (x_i - 1)^4 + K \sum_{i=1}^{m} \frac{x_i}{\|x_i\|} \tag{7.10}$$

The last term in the equation is meant to add a positive penalty (i.e. increased function value) for all positive variable values, and a negative penalty (i.e. decreased function values) for all negative variable values. In other words, using this definition positive variable values are disadvantaged compared to negative ones. Therefore, the algorithm should not find the global minimum as the best solution to the problem. Instead, it should find the local minimum located in $\vec{x} = (-0.14, -0.14, \ldots - 0.14)$, which becomes the global minimum of the constrained optimization problem. The weight K is used to adjust the impact of the penalty term. In Fig. 7.8, the comparison of the quantum-based approach coupled with TTN, and the PS algorithm for $N_V = 64$ is shown. The effectivness of the algorithm is confirmed also in this case, showing that it can return solutions that are compatible with those found by PS.

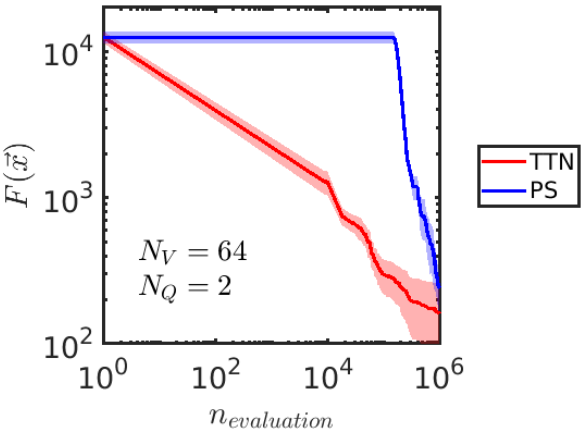


Figure 7.8: Comparison between TTN and PS in the optimization of the constrained GLN function.

7.5 APPLICATION TO FIXED-GANTRY IMRT

7.5.1 *Adding non-polynomial terms to the problem.*

In Ch. 6 results on the optimization of a quadratic cost function describing an IMRT fluence optimization problem were presented. As it was already mentioned, in real case applications, the cost functions used implemented in the TPSs are more sophisticated than the one presented in Eq. (7.11). In particular, the so-called *dose-volume* constraints are added to cost function's definition to account for more clinical objectives during the optimization process than single-voxel dose values. Such constraints generally correspond to specific points on the DVH which are related to clinical requests made by the physician when prescribing a radiotherapy treatment and they are generally based on evidences arising from large collections of data concerning the possibile side effects.

In the case presented herein, the cost function in Eq. (7.11) was modified by adding the following term:

$$F_{NP}(\vec{x}) = \underbrace{\sum_{r=0}^{R} \alpha_r \max\left(0, D^{(r)}(\vec{x}) - D^{(r)}_{max}\right)}_{A} + \underbrace{\sum_{r=0}^{R} \beta_r \max\left(0, D^{(r)}_{min} - D^{(r}(\vec{x})\right)}_{B} \tag{7.11}$$

where *NP* stands for *non-polynomial*. In particular, the new added terms have the following meaning:

- *A* adds a penalty on all voxels of structure r for whom a maximum allowed dose value $D^{(r)}_{max}$ is exceeded. Based also on the formulation provided in Ch. 1, the factor α_r is quantifies the importance of each objective;

- B adds a penalty on all voxels of structure r for whom a minimum allowed dose value $D_{min}^{(r)}$ is not reached. Again, β_r quantifies the importance of the objective.

Additionally, as the beamlets weights are required to take only positive values, an additional penalty term was added to the final cost function following the strategy that was presented and validated in Sec. 7.4.5.

7.5.2 *Application to the TG-119 IMRT phantom*

Consistently with what was presented in Ch.6, the prostate case of the TG-119 phantom was considered. The beam geometry is the same shown in Fig. 6.4a, with two opposite fields at 90° and 270°. Similarly to that case, $N_B = 64$ was set. Each beamlet weight was discretized using 2 qubits which resulted in a lattice of $N_Q = 128$ possibly fully-connected qubits. The actual degree of connectivity the local approximation model around a given configuration depends on the local structure of the evaluated function. The maximum allowed bond-dimension was set to $\chi_{max} = 10$. In Tab. 7.1, the list of the objectives and corresponding priorities used for each structure is summarized. It is important to underline that the choice of the indicated objectives and priorities was not driven by any clinical consideration or protocol being the aim of the present feasibility study showing the ability of the developed algorithm to optimize such kind of non-polynomial functions. This is also based on the reduced number of beamlets considered compared to the hundreds or thousands generally used in real clinical cases.

Table 7.1: List of objectives and constrained set for the optimization of the non-convex inverse planning problem on the TG-119 prostate case.

Structure	**Objective**	**Priority**
Prostate (PTV)	$D_i = 1.0$	$\gamma = 10.0$
	$D_{max} = 1.07$	$\alpha = 90.0$
	$D_{min} = 0.98$	$\beta = 50.0$
Rectum	$D_i = 0.0$	$\gamma = 5.0$
	$D_{max} = 0.3$	$\alpha = 10.0$
Bladder	$D_i = 0.0$	$\gamma = 4.0$
	$D_{max} = 0.5$	$\alpha = 1.0$

In Fig. 7.9 optimization result obtained with the TTN approach is presented. The algorithm was run for a maximum of 150 iterations

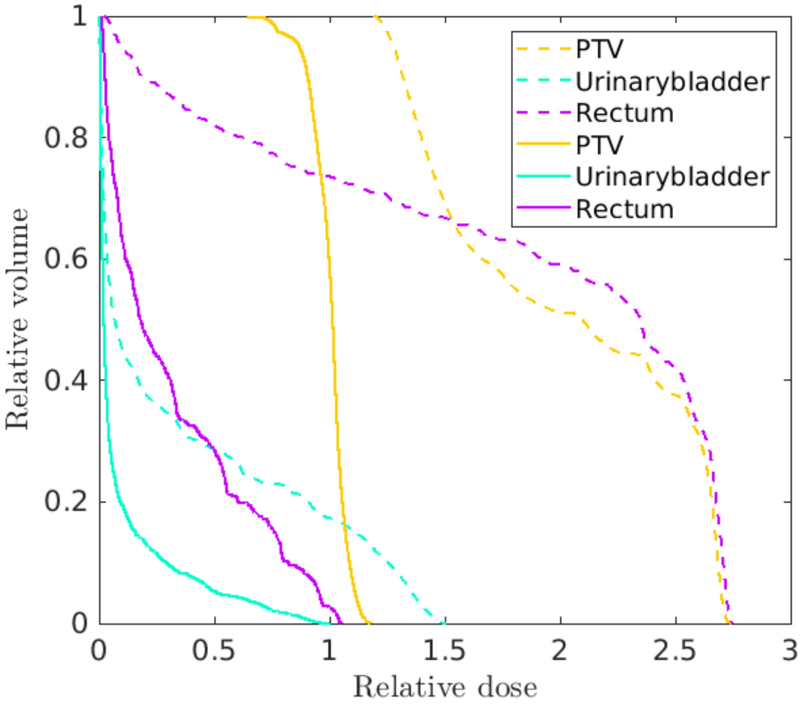


Figure 7.9: Comparison between the DVHs of the three structures before (dashed lines) and after (solid lines) the optimization with TTN.

and was allowed to perform 1.2×10^6 function evaluation. The optimization began with all the beamlets weights set to 1, being it also the condition under which the influence matrices are computed. From the comparison shown in the plot, we see that the solution provided by the algorithm tries to meet the objectives set in the function and provides a significant improvement compared to the starting point. In generale, the objectives for the prostate are closer to be meet due to the higher priorities set for the latter.

In Fig. 7.10, the comparison between the solution returned by the TTN-based approach and the PS is shown. For both algorithms, the same starting point and the same number of function evaluations were used, which are those reported in the previous lines. The first thing we notice is that the two algorithms returned comparable solutions, confirming the effectiveness of the proposed quantum-based approach also on this problem. Furthermore, despite evaluating the reported DVHs from a clinical point of view is far beyond the aim of this feasibility investigation, we observe that the obtained with TTN is slightly better fitted to the objectives than the PS's one. This can be noted by looking at the lower dose to the rectum and the lower D_{max} for the prostate.

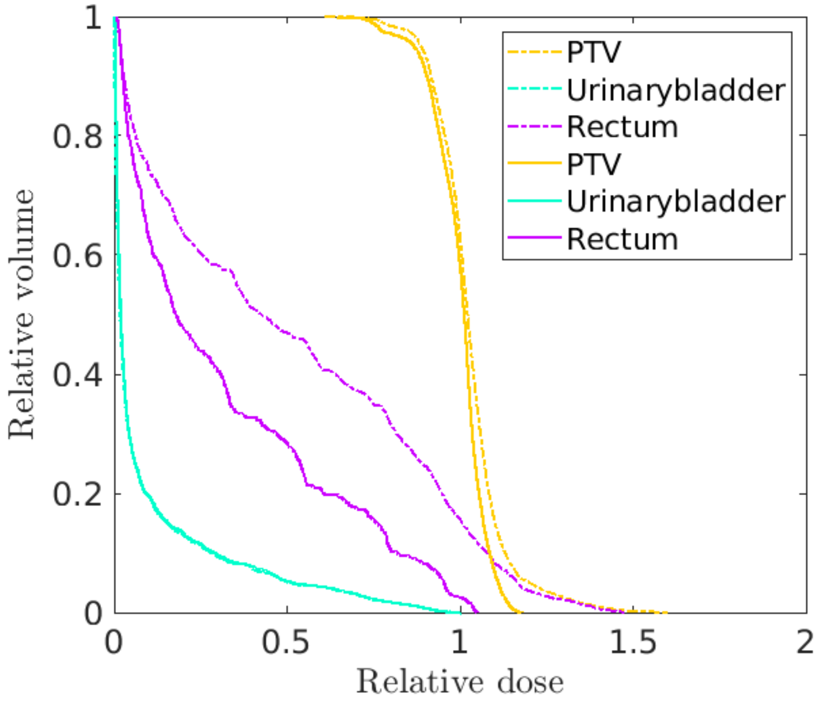


Figure 7.10: Comparison of the DVHs obtained with TTN (solid lines) and PS (dashed lines).

7.6 IMPACT OF THE RESULTS AND PROSPECTIVES

This chapter presented a new approach to the optimisation of functions with a general mathematical expression and defined on continuous variables that on quantum hardware. This approach consists of two main elements. On the one hand, a new approach for the efficient discretisation of continuous variables in terms of qubits, which allows a better fitting representation. On the other hand, the construction of a local approximation model that allows the treatment of any type of function, regardless of its mathematical expression. The proposed approach has been validated on three different optimisation problems, showing a performance comparable to that of Particle Swarm.

On the one hand, the results obtained here extend what was already shown in the previous chapter, providing for the first time, to our knowledge, a complete and efficient framework for the optimisation of any type of function on quantum hardware. The results obtained thus open up interesting prospects for the possible use of quantum computers in classical optimisation in the near future. Although, as discussed above, the first generation of quantum processors will be limited in number, the approach presented here could fasten their application to real-world problems. On the other hand, the generality of the approach overcomes the current limitations of the QUBO and

PUBO formulations, which are already widely used in the literature and better suited for polynomial and combinatorial problems.

Regarding the application to fixed-gantry IMRT, the main limitation of the present work is related to the reduced dimensionality of the investigated problem. As anticipated in the previous chapter, TNMs are used in this context outside their natural scope. Therefore, future research directions may focus on the optimisation and parallelization of the TNMs code for this specific task. This, combined with the use of appropriate supercomputing platforms, could enable optimisation problems of real cinical relevance to be solved in the medium to short term even before the acutal availability of quantum computers. Additionally, this would also open up the possibility of applying and validating the present approach to the optimisation of VMAT and Helical Tomotherapy plans.

Therefore, although the current clinical relevance of the approach presented here is limited, this feasibility study provides for the first time a solid basis and starting point for the development of quantum hardware-based optimisation methods and contributes to bridge the gap to their future introduction in radiotherapy.

CONCLUSIONS

This doctoral project investigated the development of novel computational tools aimed at enhancing the efficiency of two processes that are part of the radiotherapy workflow. Overall, at the end of this project three main achievements have been reached.

First, with the development and release of the `UCoMX` software package, a common and platform-agnostic tool to extract complexity metrics from radiotherapy plans is now available. This should make complexity assessment more accessible to those without the resources to develop tools in-house and should be seen as a first step towards the definition of universal standards for the calculation and extraction of complexity metrics and the identification of a robust and reliable set of indicators. Furthermore, this should bridge the gap towards the availability of tools for the evaluation of plan complexity in clinic. In this context, integrating the metrics extraction engines within the TPSs is a natural and necessary evolution of the present work that is worth being explored in the future.

Second, the implementation of complexity indicators to reduce the PSQA workload has resulted in noteworthy progress. In summary, the decision-support tools developed for VMAT and Helical Tomotherapy could lead to an estimated 50% workload reduction at the center where this research was carried out. Future investigations will focus on clinical prospective validations of the developed tools, also on data collected at multiple institutions.

Third, the methods crafted to optimize radiotherapy plans on quantum hardware offer favorable prospects for the future. Although it is still not clear when large-scale quantum computers will become available and if they will be implemented in radiotherapy, it is crucial to begin laying the groundwork today. Therefore, future investigations will focus on the application of the developed methodologies to more realistic clinical scenarios, with the aim of progressively filling the gap towards their usability in clinic.

As a matter of fact, all the investigations performed in this work fall in a healthcare realm where expressions like 'precision medicine' or 'personalized medicine' are now commonplace among professionals. In fact, thanks to the advances in imaging and diagnostic methodologies of the recent years, the availability of *-omics* data and tools to analyze them, and the growing availability of effective and efficient Artificial Intelligence tools, it will be conceivable to craft tailored therapeutic regimens that are highly sophisticated and personalized for each individual patient in the future. In radiotherapy, the efficient implementation of those sophisticated and personalized care paths should also account for the delivery and dosimetric accuracy of the resulting treatments. However, handling multivariate input data efficiently will require increasing computational power, and this is why the revolution that computation is undergoing due to development of quantum computers should not be ignored. Therefore, it is of primary importance fostering these research lines.

In conclusion, it is not possible to predict what and where radiotherapy will be in 10, 20 or 50 years or if the aforementioned tools will become part of the future every-day clinical practice. However, every single, even small, advancement made today is a brick that will build the radiotherapy of the future, and that will provide medical professionals with the possibility of achieving unprecedented levels of accuracy and precision in administering the optimal treatment, to the right patient at the right time, leading therapeutic possibilities that are completely unfeasible nowdays. All this, with the ultimate goal of reducing the cancer burden worldwide and improving people lives.

CONCLUSIONI

Questo progetto di dottorato si è focalizzato sullo sviluppo di strumenti computazionali orientati a migliorare l'efficienza di due processi che sono parte integrante del flusso di lavoro radioterapico. Complessivamente, al termine di questo progetto sono stati raggiunti tre obiettivi principali.

In primo luogo, mediante lo sviluppo e il rilascio del pacchetto software `UCoMX` è finalmente disponibile uno strumento universale e macchina-TPS indipendente per l'estrazione di metriche di complessità dai piani di radioterapia. Tale strumento dovrebbe rendere la valuazione della complessità dei piani di trattamento radioterapico accessibile anche a coloro che non hanno le risorse per sviluppare software in-house. Inoltre, costituisce un primo fondamentale traguardo verso la definizione di standard universali per il calcolo e l'estrazione delle metriche di complessità e l'identificazione di indicatori robusti e affidabili. Infine, esso contribuisce a ridurre il gap verso l'introduzione di strumenti per la valutazione della complessità in clinica. In questo contesto, l'integrazione dei motori di estrazioni sviluppati nell'ambito di questo progetto all'interno di uno o più TPS disponibili in clinica è da considerarsi come la naturale e necessaria procecuzione del lavoro qui presentato e sarà una direzione di ricerca che verrà esplorata nel prossimo futuro.

In secondo luogo, l'implementazione di strategie basate sugli indicatori di complessità ai fini della riduzione del carico di lavoro associato al PSQA ha portato a dei risultati degni di nota e con una potenziale rilevanza clinica. Infatti, complessivamente, si prevedere la possibilità di ridurre tale carico di lavoro di circa il 50% presso il centro in cui è stato svolto questo progetto. Pertanto, future linee di ricerca si concentreranno su una validazione clinica prospettica degli strumenti ad oggi disponibili, che coinvolgerà anche test su dati raccolti a livello multicentrico.

Infine, il metodi sviluppati per l'ottimizzazione dei piani radioterapici su hardware quantistico prospettano delle evoluzioni interessanti per il futuro. Infatti, nonostante non sia ancora chiaro quando i primi computer quantistici su larga scala saranno disponibili e tantomeno se e quando questi verranno implementati in radioterapia, è di fondamentale importanza iniziare fin da oggi a porre delle solide basi. Pertanto, future direzioni di ricerca si focalizzeranno sull'applicazione delle metodologie svilupate durante questo lavoro a scenari clinici più realistici, con lo scopo ultimo di renderli adatti ad una potenziale introduzione in clinica.

Alla luce di quanto detto finora, non si può prescindere dal fatto che il presente lavoro si inserisce in una panorama sanitario in cui espressioni quali 'medicina di precisione' o 'medicina personalizzata' stanno diventando di uso comune tra i professionisti del settore. Infatti, grazie all'evoluzione delle metodologie di imaging degli ultimi anni, alla disponibilità dei dati -omici e strumenti per analizzarli, oltre che alla crescente disponibilità di strumenti di intelligenza artificiale, ci si sta muovendo in una direzione tale per cui, in un futuro, si prevede sarà possibile somministrare ai pazienti trattamenti altamente personalizzati. In radioterapia, l'implentezione efficiente di percorsi di cura personalizzati, tuttavia, è intimamente legata sia alla capacità di garantire una elevata efficienza globale del processo radioterapico, che alla realizzazione di piani di trattamento dotati di elevati standard di accuratezza dosimetrica e di erogabilità. Inoltre, una efficiente manipolazione e gestione di dati multivariati richiederà una potenza computazionale via via crescente, ed è per questo motivo che non possiamo ingnorare la rivoluzione che la computazione sta vivendo grazie allo sviluppo dei computer quantistici. Pertanto, risulta di primaria importanza continuare a perseguire con costanza le due linee di ricerca oggetto del presente progetto di dottorato.

In conclusione, è difficile prevedere come e dove sarà la radioterapia tra 10, 20 o 50 anni, così come non è possibile sapere fin da ora se gli strumenti sopra descritti diventeranno mai parte della pratica clinica futura. Tuttavia, ogni singolo traguardo o avanzamento raggiunto oggi, piccolo o grande che sia, è un mattone su cui la radioterapia del futuro si ergerà e che potenzialmente fornirà ai professionisti sanitari la possibilità di raggiungere livelli di accuratezza, precisione e personalizzazione dei trattamenti ad oggi inimmaginabili. Tutto ciò, con lo scopo ultimo di ridurre l'impatto del cancro a livello mondiale e garantire alle persone una migliore qualità di vita.

APPENDIX

A

ANALYSIS OF CLINICAL PATIENT-SPECIFIC PRE-TREATMENT QUALITY ASSURANCE WITH THE NEW HELICAL TOMOTHERAPY PLATFORM, FOLLOWING THE AAPM TG-218 REPORT

THIS CHAPTER presents the paper by Fusella *et al.* to which I have contributed as co-author.

A.1 INTRODUCTION

In early 2018 a new version of the `TomoTherapy`® platform was commercially released. This new machine is characterized by a higher dose rate of 1000 cGy/min, improved jaw dynamics, faster ring rotation (up to 10 rotations per minute), couch movement with separated axis and a couch catcher to reduce couch sag effect (to less than 2 mm) toward the gantry. The new platform comes with a redesigned version of the TPS: Precision TPS v1.0.02 which incorporates the graphics processing unit (GPU) based optimizer, `VoLO`™ and a Collapsed Cone Convolution Superposition dose calculation algorithm [110, 120]. This new system grants the capability of faster treatments and more freedom in the combination of the modulation factor, jaw width, pitch and gantry period. Thanks to these advanced features, the use of this system opens to the possibility of achieving more modulated and complex plans.

At our department, a `Radixact`® unit machine has become clinical in June 2018 and since the first clinical plan a detailed QA program has been implemented following The long term dosimetric stability of this new delivery system has already been studied and reported by Smilowitz [206] and the new TPS version has been validated by Chen [120]. In order to thoroughly assess the new capability of the system, a detailed and complete analysis on a large database of clinical plans is also needed the recommendations of AAPM TG-218 [23]. Approximately 395 clinical helical plans were delivered and measured with `ArcCheck`® (Sun Nuclear, Melbourne, FL). Measured and calculated dose distributions were compared by means of 2D gamma analysis. Within this work, we propose the clinic and machine specific tolerance limit (TLcs) and action limit (ALcs), obtained through statistical process control techniques, to be adopted for the new `Radixact`®platform coupled to the `Precision` TPS v1.0.02. We also analyse and report which planning parameters are more related to QA failures.

A.2 MATERIALS AND METHODS

A.2.1 *Dataset*

Between June 2018 and October 2020 a total of 395 treatments have been planned with the `Precision`TPS v1.0.02 using the GPU-based optimizer `VoLO`™, which incorporates the Collapsed Cone Convolution Superposition dose calculation algorithm [110, 120]. The treatments were related to a large number of different sites and diseases; therefore, within the scope of this work they have been grouped in the following categories: abdomen (28 pts), brain (32 pts), head and neck (77 pts), lungs (71 pts), pelvis (84 pts), prostate (52 pts) and others (41 pts). The dose per fraction (D/fr) ranged from 1.6 to 5.0 Gy.

A.2.2 *Plan characteristics*

To characterize the plans, the following parameters were collected: actual modulation factor (MF), pitch, field width (FW), gantry period, total delivery time, couch speed, couch travel distance and TTDF (ratio between the treatment time and the prescribed dose per fraction). Moreover, some descriptors of the leaf open time (LOT) distribution were collected: minimum (min-LOT), mean (mean-LOT), maximum (max-LOT) and LOT standard deviation (SD-LOT). All treatments utilized dynamic jaws [207].

A.2.3 *QA delivery and analysis*

All the patient-specific QA measurements were collected with `ArcCheck`® (Sun Nuclear, Melbourne, FL) and performed without the 15 *cm* diameter homogeneous PMMA plug. The plans were recomputed on the synthetic `ArcCheck`® CT provided by the vendor with an imposed density of 1.1836 g/cm3 following manufacturer's recommendations. Acquisition and analysis were performed with Sun Nuclear SNC Patient™ version 6.7. The absolute dose calibration of `ArcCheck`® was periodically controlled against a calibrated farmer-type ionisation chamber. Before each measurement session, the output of the `Radixact`® unit was checked using the cheese phantom and ionization chambers, and a correction applied to the `ArcCheck`® analysis. QA plans were computed in high resolution mode which grants a 1.87 *mm* spaced dose grid.

The QA plan was computed placing the centre of the planning PTV at the machine's isocenter. This procedure reduces the possibility for high dose areas to fall at the border of the `ArcCheck`® limiting the dependence of the gamma passing rate ($PR_\gamma\%$) on the maximum calculated dose [208]. According to the AAPM TG-218 report, the dose comparison was performed through 2D gamma analysis using a 10% dose threshold as well as 3% dose difference(DD) and 2 mm distance-to-agreement (DTA) criteria. In order to compare our results with previous works [123, 207–209], 3%/3 mm gamma analysis was also collected. Both local and global normalization were used.

A.2.4 *QA program*

Following the AAPM TG-218 report, the clinical deliverability of a plan was evaluated on the basis of the universal tolerance limit (TL) and action limit (AL), which are 95% and 90% respectively, on the $PR_\gamma\%$ computed with $(3\%, 2\ mm)$, global normalization criteria. Statistical process control techniques recommended in TG-218 have been employed to define a clinic specific TLcs and ALcs [23] as follows:

$$AL_{cs} = 100 - 3\sqrt{\sigma^2 + (\bar{x} - 100)^2} \tag{A.1}$$

$$TL_{cs} = \bar{x} - 2.660 \times m\bar{R} \tag{A.2}$$

where x is the $PR_\gamma\%$ average over the investigation period, σ is its standard deviation and mR the moving average. These values were computed on the first 40 collected measurements. The IMRT treatment process has been monitored and investigated by periodically computing these values to verify whether the process was under control. For the subsequent evaluations, periods of approximately 6 months were considered. Only clinically deliverable plans, those fulfilling $GP\%(3\%G, 2\ mm) > 90\%$, were considered in this analysis and 95% confidence intervals on TL_{cs} and AL_{cs} have been computed through a bootstrap approach with 10,000 samples [210, 211].

A.2.5 *Statistical analysis*

To spot possible differences of $PR_\gamma\%$ among the different treatment sites a Kruskal-Wallis test was performed followed by a post-hoc analysis where Bonferroni correction was applied [212]. N-way analysis of variance (n-way ANOVA) was performed to spot the influence of the planning parameters on $PR_\gamma\%$ [213]. Only parameters that strictly bore no correlations among themselves were included in the test ($p > 0.05$ and/or Pearson's $r < 0.8$) [214, 215]. All tests were performed using `MATLAB`® R2020b (MathWorks, Inc., Natick, MA, USA). All the p-values reported are two-sided and $p < 0.05$ is considered as statistically significant.

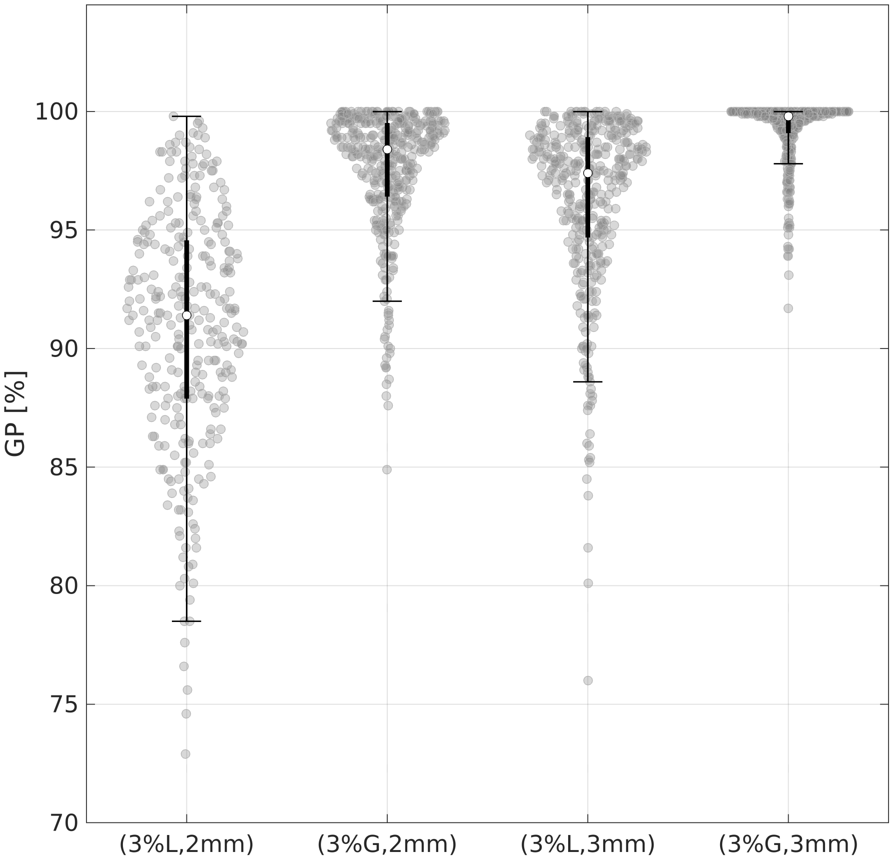


Figure A.1: Violin plot of the complete database of collected measurements distinguished among the different gamma analysis criteria. The central white dot marks the median, the edges of the box correspond to the 25th and 75th percentiles, the whiskers extend to the adjacent values which are the most extreme data values that are not outliers, and data outside the whiskers are outliers. L stands for local gamma normalization, G for global one

A.3 RESULTS

Over the entire collection period, only 10 out of 385 plans failed to meet the universal AL defined by the TG-218 report. These failing plans were re-planned to meet the universal AL and added to the original set, thus resulting in a total of 395 plans that were actually measured. Figure A.1 depicts the whole population of collected QA results. The measured $PR_\gamma\%(3\%, 2\ mm)$ (both global and local normalization) of the whole population $(n = 395)$ was distinguished among treatment sites and the information is presented in Fig. A.2. Data include both clinically and non-clinically $(PR_\gamma\%(3\%G, 2\ mm) < 90\%)$ delivered plans. The Kruskal-Wallis test confirmed that some differences among the groups does exist $(p - value < 0.001)$. The post-hoc test showed that these differences are statistically significant only when the pelvis plans are compared to abdomen, lungs and others groups. Table A.1

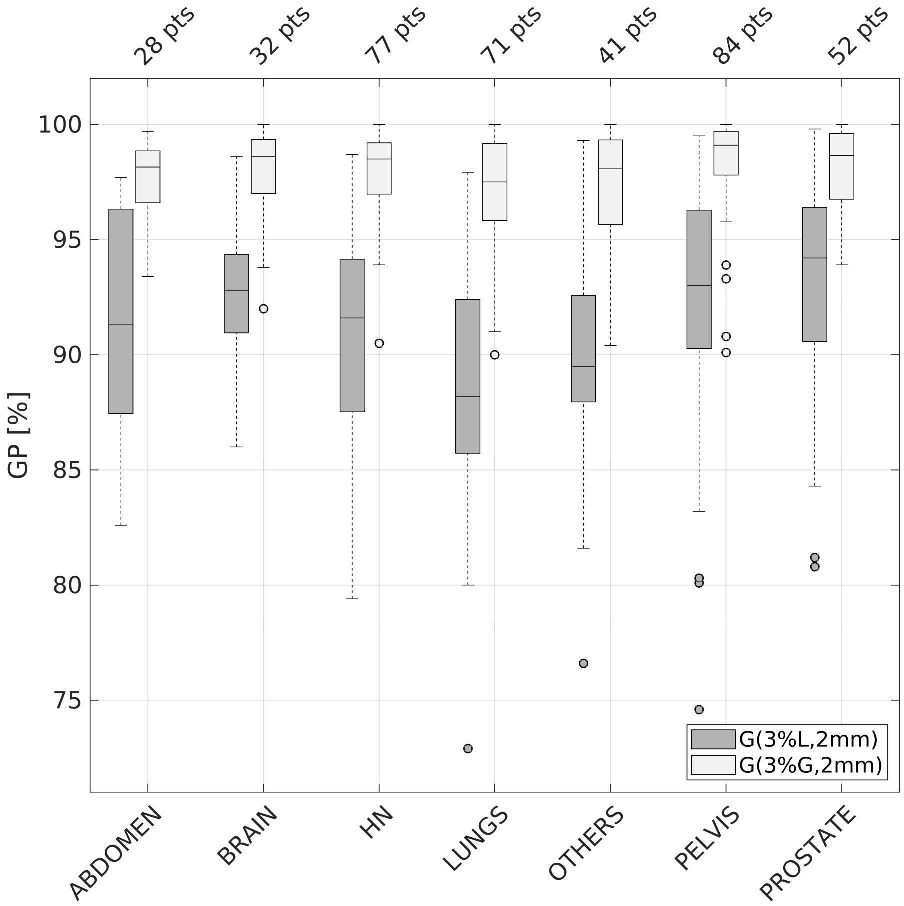


Figure A.2: Whiskers box plot of $PR_\gamma(3\%G, 2\ mm)$ (global and local normalization) of all measured plans distinguished among the different treatment sites. The central line marks the median, the edges of the box are the 25^{th} and 25^{th} percentiles, the whiskers extend to the adjacent values, which are the most extreme data values that are not outliers, and the circles represent the outliers.

reports the distribution of the collected planning parameters for our database of plans.

Figure A.3 reports the periodical evaluation of the AL_{cs} and TL_{cs}; non-clinical plans are not included. The first 40 plans used for the first calculations (as proposed in TG-218), are also included in period "1". Among the first 40 planned treatments, the calculated ALcs and TLcs are, respectively, 93.2% and 93.7%, with all the plans satisfying $PR_\gamma\%(3\%G, 2\ mm) > 93.3\%$. The final computation of ALcs and TLcs on the last set of 78 clinical plans yielded $AL_{cs} = 89.4\%$ and $TL_{cs} = 91.1\%$. A Kruskal-Wallis test has been performed to compare AL_{cs} and TL_{cs} samples of periods 2, 3 and 4 and no statistically significant differences exist. In these three periods not even a single plan lies below the estimated AL_{cs}, following the recommendations of TG-218, these evidences allow to state that the process is settled and under control. Using data from these 3 periods a n-way ANOVA was performed to detect which parameters might predict QA failures are summarized for both local and global normalization of $PR_{\gamma}\%(3\%G, 2\ mm)$ criteria. Results are given in Table 2. In Tab. A.2, the results of the n-way ANOVA performed The analysis has been conducted for the data of the last three periods, where the process seems to become settled and under control (Fig. A.3).

Table A.1: Collected planning parameters for the entire population of 395 collected QA deliveries.

Metric	**Mean ± SD**	**Range [min; max]**
Dose/Fraction [cGy]	219.8 ± 62.6	[160.0; 500.0]
MF	1.68 ± 0.28	[1.10; 2.79]
Pitch	0.39 ± 0.06	[0.12; 0.48]
Gantry Period [s]	20.1 ± 7.3	[11.8; 52.8]
Treatment Time [s]	299.7 ± 141.0	[66.3; 1241.3]
TTDF [s/cGy]	1.41 ± 0.69	[0.36; 5.43]
minLOT [ms]	18.2 ± 1.1	[18.0; 36.4]
mLOT [ms]	236.3 ± 82.7	[57.8; 617.1]
maxLOT [ms]	391.4 ± 144.0	[123.2; 1034.2]
sdLOT [ms]	107.9 ± 44.2	[21.1; 316.1]
Couch speed [mm/s]	0.56 ± 0.20	[0.11; 1.84]
Couch travel [mm]	160.2 ± 88.9	[31.1; 895.2]
FW (10/25/50 mm)	2.1/92.0/5.9%	

Table A.2: Results of the n-way ANOVA test on $PR_{\gamma}(3\%, 2\ mm)$ with respect to all the collected planning parameters in terms of p-values. Underlined values mark significant tests.

Variable	$PR_{\gamma}(3\%L, 2\ mm)$	$PR_{\gamma}(3\%G, 2\ mm)$
Site	$\underline{0.002}$	0.063
MF	0.294	0.199
TTDF	$\underline{0.007}$	0.373
mLOT	0.642	0.437
maxLOT	0.804	0.859
Pitch	0.884	0.659
GantryPeriod	0.820	0.841
Couch speed	0.189	0.055

A.4 DISCUSSION

In this work, we assessed the performances of the `TomoTherapy`® platform coupled with a new dedicated TPS in terms of the gamma index

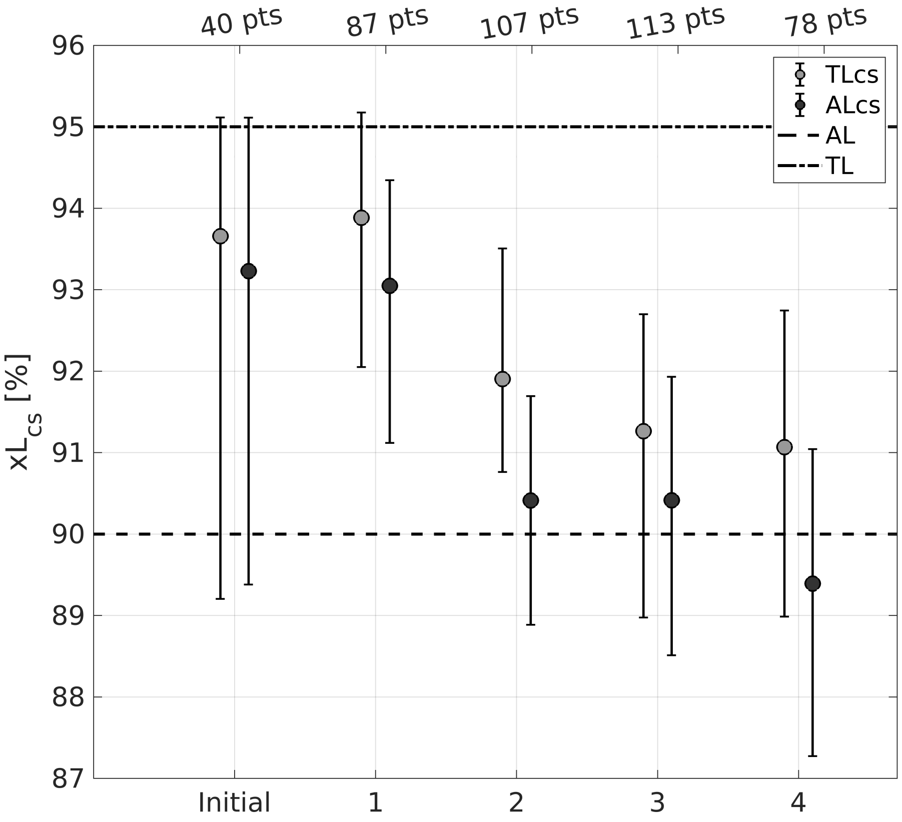


Figure A.3: Evaluation of the AL_{cs} and TL_{cs}. The circles mark the obtained values while the whiskers extend to the 95% confidence interval.

passing rate measured with a 3D dosimeter, extending the analysis of previous works on earlier versions of `TomoTherapy`® and associated TPS [123, 208]. The entire scope of work was performed following AAPM TG-218 recommendations. We considered the $PR_\gamma\%$ metric with $3\%/2\ mm$ criteria, analysing also the impact of local and global normalization. Our results showed that the mean $PR_\gamma\%/3\%G, 2\ mm)$ was $97.6 \pm 2.6\%$ with some differences arising among the different sites used to stratify the sample.

Any previous study on the `TomoTherapy`® unit has not reported results using the $PR_\gamma\%$ metric defined by AAPM TG-218. For a comparison with existing literature we used the $PR_\gamma\%$ with 3% $3mm$ criteria. The results here reportedly outperformed those obtained with different versions of tomotherapy machines and the same 3D dosimeter (`ArcCheck`®) in terms of $PR_\gamma\%$. We measured systematically higher mean values with respect to Bresciani [123], Binny [216] and Yue [217]: 99.2% (SD 1.3%) vs. 96.1% (SD 4.4%), 95.9% (SD 2.9%) and 97.5% (SD 2.7%), respectively.

Applying the statistical process control techniques proposed by the Task Group, we periodically computed the centre-specific action level and tolerance limit (see Fig. 3). The first evaluation yielded values comparable to the universal ones indicated in the report (93.7% and

93.2%, versus 95% and 90% respectively). The periodical evaluation of the two control limits showed that a change occurred in the process between the first and the second re-evaluation. In fact, both TL_{cs} and A_{cs} decreased and remained approximately constant afterwards (Fig. A.3). This probably happened since the start of 2019 and after six months of initial training, new treatments have been introduced into the clinical practice (e.g. mesotheliomas, craniospinalirradiations, Hodgkin's and non-Hodgkin's lymphomas) that can be considered as inherently more complex; moreover, moving up the learning curve, the planners have begun to achieve increasingly complex and modulated dose distribution; thus probably inducing a decrease of treatment deliverability. All this sources of variation forced to follow TG-218 recommendations, going through a periodically monitoring of the process, investigating the $PR_{gamma}\%$ over a long period.

To spot possible planning parameters closely related to QA results, and their trend, a n-way ANOVA test was performed (see Table 2). Only the treatment site and the TTDF (treatment time divided by the prescribed dose per fraction) were highlighted as possible predictors of QA failures, only when $PR_{gamma}\%(3\%L, 2\ mm)$ is considered. In fact, the TTDF can be considered as a simple indicator of complexity, since it is somehow related to the longitudinal extension of the target volume and to the pitch. A similar behaviour was recently reported by Santos et al. [69].

Previous studies reported the effect of mean-LOT, max-LOT, MF, pitch and gantry period on plan deliverability [68, 123, 208, 216]. In our database, none of the collected parameter can be considered relevant when the $PR_{gamma}\%(3\%G, 2\ mm)$ is used, and this might be due to different reasons. First of all, the presence of the TTDF might mask all the other parameters, since mathematical relationships hold between the TTDF and most of the aforementioned parameters. Secondly, our planning practice strictly follows the recommendation of the Accuray planning manual [1] which suggests to prepare plans with max-LOT higher than 241 ms, and mean-LOT higher than 100 ms. Moreover, approximately 95% of our plans show the MF within 1.4 and 2.5, which Binny et al. [68] have previously shown tobe associated to acceptable deliverability.

Since no dependence from anatomical site has been found for $PR_{gamma}\%(3\%G, 2\ mm)$, it is possible to use for all treatments the same AL_{cs} and TL_{cs} here reported. When the more stringent local criteria is considered a significant difference among the anatomical site arouse. In this case group specific AL_{cs} and TL_{cs} would have been required and, following AAPM TG-218 recommendation, should be computed. The assessment of the TPS's performances in regards to complexity of the calculated sinograms is beyond the scope of this work, as well as the detailed analysis of the relationship between $PR_{gamma}\%$ and clinically relevant deviations [121, 208, 218, 219]. A more detailed analysis on the predictive value of the analysed parameters on QA trend and failures is still ongoing.

The results of this study highlight the capability of the `Radixact`® system to accurately deliver complex dose distributions over a large variety of treatment sites. We set the TL and AL for this new machine to be used as reference for comparison for other centres.

B

UPDATING A CLINICAL KNOWLEDGE-BASED PLANNING PREDICTION MODEL FOR PROSTATE RADIOTHERAPY

THIS CHAPTER presents the paper by Scaggion *et al.* to which I have contributed as co-author. For the sake of brevity, only the main text is presented. For any supplementary material, please refer to the journal article [220].

B.1 INTRODUCTION

The performance of all prediction systems based on prior knowledge strongly depends on the consistency, quality, and vastness of such knowledge. When these systems are adopted in a clinical environment, their possible benefits should be weighed against the amount of human resources required for their set-up, maintenance and updating. In fact, such systems require an intense workload of data farming and mining, which may last even beyond their initial clinical implementation.

A substantial body of literature demonstrated the effectiveness of Knowledge-Based Planning (KBP) systems in assisting and improving the clinical management of external beam radiation therapy treatment plans. Improvements in the quality of treatments, increased consistency, reduced variability, and reduction in planners' workloads are just a few of the major benefits associated with the implementation of such systems in a clinical setting [221–228]. On the other hand, many research groups have reported labour-intensive, iterative and time-consuming processes associated with KBP systems such as: population selection, model revision, model refinement and model validation [225, 229–231]. In general, some maintenance is necessary to ensure KBP models have proper clinical usability, relevance and quality. This leads to the necessity of a periodical repetition of the aforementioned process, especially in cases when the clinical practice changes significantly [232].

Different groups have recently reported their endeavors to reduce this maintenance effort: the performance of KBP models could be improved with an iterative learning process for head and neck (HN) and prostate cases [233–236]; data-driven methods can be employed to periodically improve KBP automated planning routines and KBP model performance could be incrementally improved with supervised machine learning [224, 237].

However, regardless of the strategy used to update KBP models, a clinical validation should be carried out to prove the benefit associated with their introduction in a real-world clinical environment where clinical requirements, patient management and planning practices can change over time. As a matter of fact, none of the aforementioned studies conducted a long-term longitudinal evaluation with a large number of cases, and only some of them have reported multiple updates of the same model. Furthermore, to our knowledge, there is not any study comparing different updating strategies explicitly.

In this study, we performed a paired comparison of two antithetical KBP models update strategies and tested how these impact the quality of prostate treatments plans through a three-year long continuous observation. In particular, we performed multiple subsequent updates of the first clinically adopted prostate model to assess whether an updating procedure should prioritize a larger sample size or a higher mean sample quality.

B.2 MATERIALS AND METHODS

B.2.1 *Enrolled patient sample*

Within the period from 01/01/2018 to 31/12/2020, a total of one hundred and one patient records were recruited for this work. Inclusion criteria were: men with localized histologically confirmed prostate adenocarcinoma (all risk groups with clinical stage T1b-T3a, N0, M0 with no clinical evidence of pelvic lymphadenopathy, and an estimated risk of nodal involvement <15 % based on the Roach formula [238]); radical external beam radiotherapy conducted with a moderate hypofractionation scheme to the prostate only. Clinical Target Volume (CTV) was contoured following the NCCN v.3.2016 prostate guidelines. The penile bulb was added whenever a recent magnetic resonance imaging study was available. The contouring procedure was completed by two dedicated radiation oncologists. The planning goals were to cover the PTV with 95 % of the prescribed dose (V95% > 98 %) while limiting overdosage to 107 % of the prescribed dose (D1cc < 107 %).

Until the end of 2017 all of prostate treatments were delivered on a Unique linac (Varian Medical Systems, Palo Alto, CA) with VMAT technique with an IGRT scheme based on MV imaging and X-ray triggering of gold fiducials based on stereoscopic 2D image [239]. The standard treatment schedule was 70 Gy in 28 fractions at the beginning of the study [240]. During this period the Planning Target Volume (PTV) was obtained by expanding the CTV with a posterior margin of 5 mm and 7 mm margin in all other directions.

Since the beginning of 2018 a new TrueBeam STX unit (Varian Medical Systems, Palo Alto, CA) with CBCT capability started its clinical operations. This introduction led to change the IGRT scheme for prostate treatments to daily CBCT. Such introduction led to gradually change the treatment schedule to 60 Gy in 20 fractions following recent evidence [150–152]. From September 2019 X-ray triggering of gold fiducials based on OBI X-ray imaging was introduced in the clinical routine for a frac- tion of patients. The PTV margins recipe remained unaltered for patient undergoing daily CBCT while it was reduced to 4 mm posteriorly and 5 mm in all other directions for the patients implanted with gold fiducials.

For all the patients included in the study rectum, bladder and femoral heads were delineated as Organ-at-Risks (OARs). The planning goals were maintained constant over the entire period of inquiry. All plans were optimized according to our department's prostate radical treatment protocol which is based on the ASTRO/ ASCO/AUA guidelines [153].

B.2.2 *Treatment plans*

All patients were treated with Volumetric Modulated Arc Treatment (VMAT) using two complete arcs either with 6X or 10X, on a couple of matched TrueBeam STX machines equipped with HD Millenium MLC (Varian Medical Systems, Palo Alto, CA). 10× was only used for overweight patients when a planning comparison against 6× was considered significantly favourable. All treatment plans were obtained with Eclipse TPS, optimized with the PO algorithm v15.5.11 and computed with Acuros XB v15.5.11 (Varian Medical Systems, Palo Alto, CA) with dose-to-medium reporting. All treatment plans resulted from a specific strategy adopted to limit plan complexity, which was developed within our institution. This strategy has proven to guarantee high levels of plan deliverability without compromising plan quality or clinical acceptability [153]. In detail, the plans were optimized by setting the Aperture Shape Controller (ASC) priority to Very High and using a monitor units (MU) limit objective which aims to constrain the total number of MU close to a MU/cGy value of 3 [32, 241]. ASC is a leaf sequencer that simplifies MLC configuration by minimizing the curva- ture of the beam-eye-view shapes. ASC is implemented as a multiplica- tive penalty term in the optimizer cost function and can be modified by the user by choosing among six discrete penalty levels. In rare cases when clinical goals were not met, the MU limit was adjusted until clinical requirements were achieved.

B.2.3 *Plan quality assessment*

Table B.1: Numerosity and quality characteristics of models' library population. Statistically significant differences in comparison to M0 are marked with asterisks.

Model	**Sample**	**APQM%**	
		mean ± SD	**Range [min;max]**
M0	73	87.3 ± 5.3	[75.8; 96.6]
M1	94	88.2 ± 5.5	[76.1; 97.8]
M2	117	88.9 ± 5.7	[76.1; 97.8]
M3	141	90.1 ± 5.6*	[76.1; 97.8]
M4	164	90.6 ± 5.5*	[76.1; 97.8]
R1	73	90.4 ± 4.4*	[83.3; 97.8]
R2	73	92.5 ± 3.4*	[84.2; 97.8]
R3	73	94.7 ± 2.3*	[88.5; 97.8]
R4	73	95.7 ± 1.4*	[92.5; 97.8]
M0	73	88.2 ± 6.3	[64.2; 95.0]
M1	94	87.1 ± 7.7	[64.2; 96.4]
M2	117	87.5 ± 7.0	[64.8; 96.4]
M3	141	87.7 ± 7.1	[64.8; 96.4]
M4	164	87.4 ± 7.4	[61.7; 96.4]
R1	73	87.3 ± 7.0	[69.2; 96.4]
R2	73	89.3 ± 4.4	[75.8; 96.4]
R3	73	90.9 ± 3.7*	[78.1; 96.4]
R4	73	92.2 ± 2.5*	[86.2; 96.4]
M0	73	86.9 ± 11.9	[56.0; 100.0]
M1	94	89.8 ± 11.3	[60.0; 100.0]
M2	117	90.9 ± 11.2	[60.0; 100.0]
M3	141	93.2 ± 10.2*	[60.0; 100.0]
M4	164	94.4 ± 9.6	[60.0; 100.0]
R1	73	94.3 ± 7.3*	[75.2; 100.0]
R2	73	96.3 ± 5.5*	[77.3; 100.0]
R3	73	99.0 ± 2.2*	[88.9; 100.0]
R4	73	99.6 ± 0.9*	[93.9; 100.0]

The Plan Quality Metric (PQM) was adopted as a global measure of quality in order to assess overall plan quality and simplify the plan comparison process [112]. PQM is a user-defined metric designed to compare the quality of competing treatment plans and it is useful to limit the subjectivity of judgment. It gathers into a single number the judgment of quality expressed by a clinical team on the basis of its knowledge and experience. It is built through a list of sub-metrics (e.g. Dose-Volume Histogram (DVH) metrics), which should summarize the

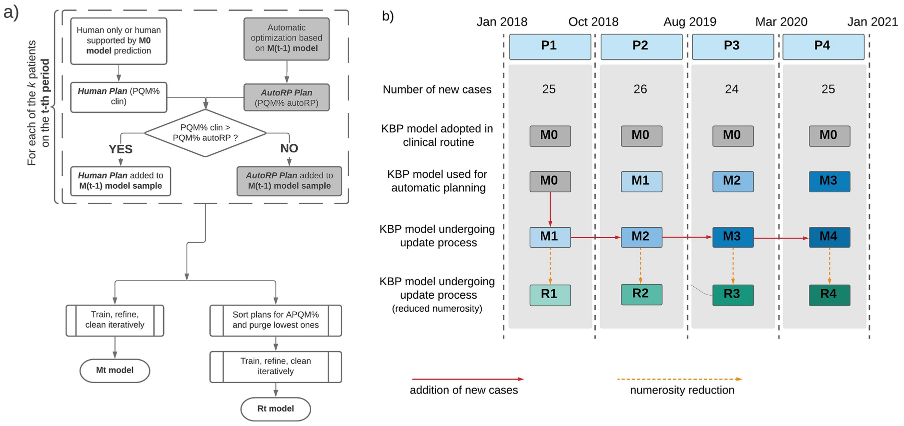


Figure B.1: a) The workflow for plan generation and model upgrade for period pt. b) The timetable for the entire study is presented along with its subdivision into periods, the models in uses and the models obtained as a result of the updating procedures.

treatment's specific goals. Each metric is associated with a numerical scoring func- tion to model clinician's judgment criteria as accurately as possible. The PQM is the sum of the scores obtained by each sub-metric and measures the extent to which the plan adheres to the list of identified goals. First introduced by Nelms, it is now adopted in many studies [242–246]. The details of the PQM algorithm used in this work are described elsewhere and are reported in detail in the Supplementary Material [134, 247]. The PQM% algorithm was initially developed for treatments delivering 70 Gy in 28 fractions. To be used for 60 Gy in 20 fraction treatments, all the sub-metrics and the scoring functions has been linearly scaled down with the total dose.

In this study, the PQM% is used in conjunction with its adjusted version (APQM%) to allow for plan quality ranking across a cohort of patients [144]. APQM% tailors the PQM algorithm to the anatomical characteristics of each patient. In order to assess whether plans were obtained with different trade-offs between target coverage and OARs sparing, the PQM% was also split into two complementary measures: the $\text{PQM}_{target}\%$ which gathers together all of the sub-metrics related to target coverage, homogeneity and conformance; and the $\text{PQM}_{OARs}\%$ which represents the sum of the sub-metrics related to OARs sparing. The same subdivision was applied to the APQM%.

B.2.4 *Generation of models*

The first KBP model for prostate treatments was adopted in our clinical practice at the beginning of 2018. It was based on a prior sample of 73 historical prostate plans (gathered from January 2015 to December 2016) treated with a Unique machine equipped with Millennium MLC with dose prescription of either 78 Gy/39 fx or 70 Gy/28 fx [134, 247]. It was generated with RapidPlan v15.1 (Varian Medical Systems, Palo

Alto, CA) and will be referred to as M0. A continuous program of updates was implemented after its clinical introduction, with a model update performed approximately every 25 new prostate cases. To fulfil Wang and colleagues' recommendation, the initial model was updated in an attempt to enlarge the model sample size [236].

The investigation lasted approximately 36 months and ended with four subsequent updates performed at the end of four time intervals referred to as Pt with t = 1,...,4. The whole process is visually summrized in Fig. B.1. A description of the patients sample accrued in the four periods is given in the Supplementary Material. During the entire investigation, the clinical planning was conducted by human planners assisted by the prediction of the first clinical KBP model M0 (plans).

During this planning effort, the planner started from the KBP predictions and was left free to manually refine the optimization objective and iterate the optimization process until a satisfactory result was reached. In the case of unreliable DVH prediction the planner was free to perform a completely manual optimization without relying on the KBP predictions. This is the usual clinical approach to planning and would remain unaltered through time if the KBP model did not get updated. To compare the effectiveness of the two updating strategies in challenging conditions, each Human plan was compared to a plan generated through a fully automated optimization based on the prediction of the most recently updated model M(t-1) (AutoRP plans). Thereby, the updated model outcome did not benefit of the human interaction devoted to refine the plan, making the comparison with the human-generated plans more challenging [221, 248]. The set of optimization objectives automatically generated by RapidPlan remained unaltered throughout the study and is reported in the Supplementary Material.

The two competing plans (Human and AutoRP), sharing the same geometry and energy, were reviewed by a clinician to judge their clinical soundness and were also ranked through the PQM%. The plan judged clinically appropriate with the higher quality (i.e. the higher PQM%) was added to the sample forming the M(t-1) model population so to obtain an expanded library which would have become the population of the next update, i.e. model Mt. At the end of each period, after approximately 25 cases, the new Mt model was trained (see Fig. B.1a). This updating procedure aimed to improve the model's generality by increasing the number of cases the model was trained on. After the training, each Mt model was cleaned and refined to remove plans or single OARs identified as outliers or largely influential points, and un- derwent a single round of internal and external validation [248–250].

The second updating procedure was based on the efficient selection method proposed by Fusella et al [247]. This procedure was designed to increase the mean quality of model samples while keeping the same li- brary size as the ancestor. The plans comprising the Mt library were sorted according to the APQM% and only the 73 top-ranked plans (i.e. those with the highest quality) were retained. A new model was trained on this reduced sample and is referred to as Rt. Rt models were subjected to internal and external validation as Mt models, with outlier removal limited to cases that were highly influential on the model's predictions (those classified as outlier by the Cook distance) [247].

B.2.5 *Models comparison through open-loop tests*

A thorough comparison of model performance was accomplished by means of an open-loop test conducted on an external validation set (i.e. not used for training the models) consisting of 30 historical patients not included in any of the model libraries [134, 247]. This group was first used to validate the Mo model. All patients in this group were treated with 70 Gy in 28 fractions in late 2016. A set of plans generated by an experi- enced human planner, without any support from the KBP predictions, was taken as a reference. A completely automated planning procedure was undertaken for each patient with each of the KBP models and compared to the relative reference plans.

The RapidPlan DVH estimation algorithm generates an alert if the geometrical characteristics of the patient whose DHV is to be estimated fall outside or at the library sample's extreme borders. Each alert is an indication that the DVH estimation might not be reliable [249]. To assess the models' generality, the number of alerts returned by RapidPlan were collected for each patient and each involved OAR. The number of alerts collected was used as a rough measure of model reliability, with fewer alerts indicating a more reliable model.

To assess the KBP models' general quality the goodness-of-fit statistics (R2 and X2) and the goodness-of-estimation parameter (mean square error, MSE) were collected and compared. R2 represents the coefficient of determination of regression model parameters and X2 is the average chi square of regression model parameters. Better results are expected as the former approaches 1 while the latter approaches 0. The MSE describes how well the model is able to estimate the original DVH in a training plan, and the closer it is to 0, the better the model's esti- mation capability for plans that are not part of the training set [233, 247, 249].

B.2.6 *Evaluation of plan complexity*

In line with the existing literature, plan complexity was evaluated using 4 complexity metrics, computed with a Matlab routine developed in-house. These metrics were chosen on the basis of their complementarity and their reported correlation with plan deliverability [32, 135]. These complexity metrics consist of: the ratio of the total number of Monitor Units to the dose per fraction (MU/cGy); the Edge Metric (EM), which measures the complexity of the MLC aperture shapes as the ratio of MLC side edge length to aperture area [61]; the VMAT adapted Modulation Complexity Score (vMCS), which represents modulation complexity taking into account the relative variation on leaf positions, beam aperture area, and MUs between control points [47]; and the total modulation index (MIt) which reflects the speed and ac- celeration of modulating parameters such as MLC movements, dose-rate, and gantry speed [51].

B.2.7 *Statistical analysis*

Statistical differences arising from the anatomical and dosimetric features, quality and complexity of plans generated using the subsequent updates of the initial model M0 were assessed through the two-tailed Student's t-test when the tested variables were normally distributed. Data normality was checked through the Shapiro-Wilk test. The significance level was set to 0.05 throughout the whole study, with Bonferroni's correction when multiple comparisons were performed.

The whole study has been conducted following the recommendations of the Radiotherapy Treatment plannINg study Guidelines (RATING) [251], the completed RATING scoresheet is given as Supplementary Material.

B.3 RESULTS

B.3.1 *Models characteristics*

The average plan quality of Rt models, as measured by the PQM%, increased with each model update and the PQM% distribution of Rt model sample nudged towards 100% as a direct consequence of the update approach. The average quality of Mt models increased while their spread remained relatively constant. The detailed characteristics of all obtained KBP models in terms of plan numerosity and plan quality are given in Tab. B.1 1. A Bonferroni corrected two-tailed Student's t-test proved that the APQM% of Rt models increased significantly by approximately 2% after every update, while for Mt models a statistically significant increase of approximately 3% was reached at the third up- date. For Mt models, such statistically significant differences were almost entirely the result of increased OARs sparing, while target coverage and homogeneity remained relatively unchanged.

The overall quality of Mt models increased monotonically (an increase in R2 and a reduction in X2 and MSE) for all OARs, while it fluctuated for Rt models with overall indicators that were not always better than the M0 model. The anatomical characteristics of patient samples varied across the different models: PTV volume decreased

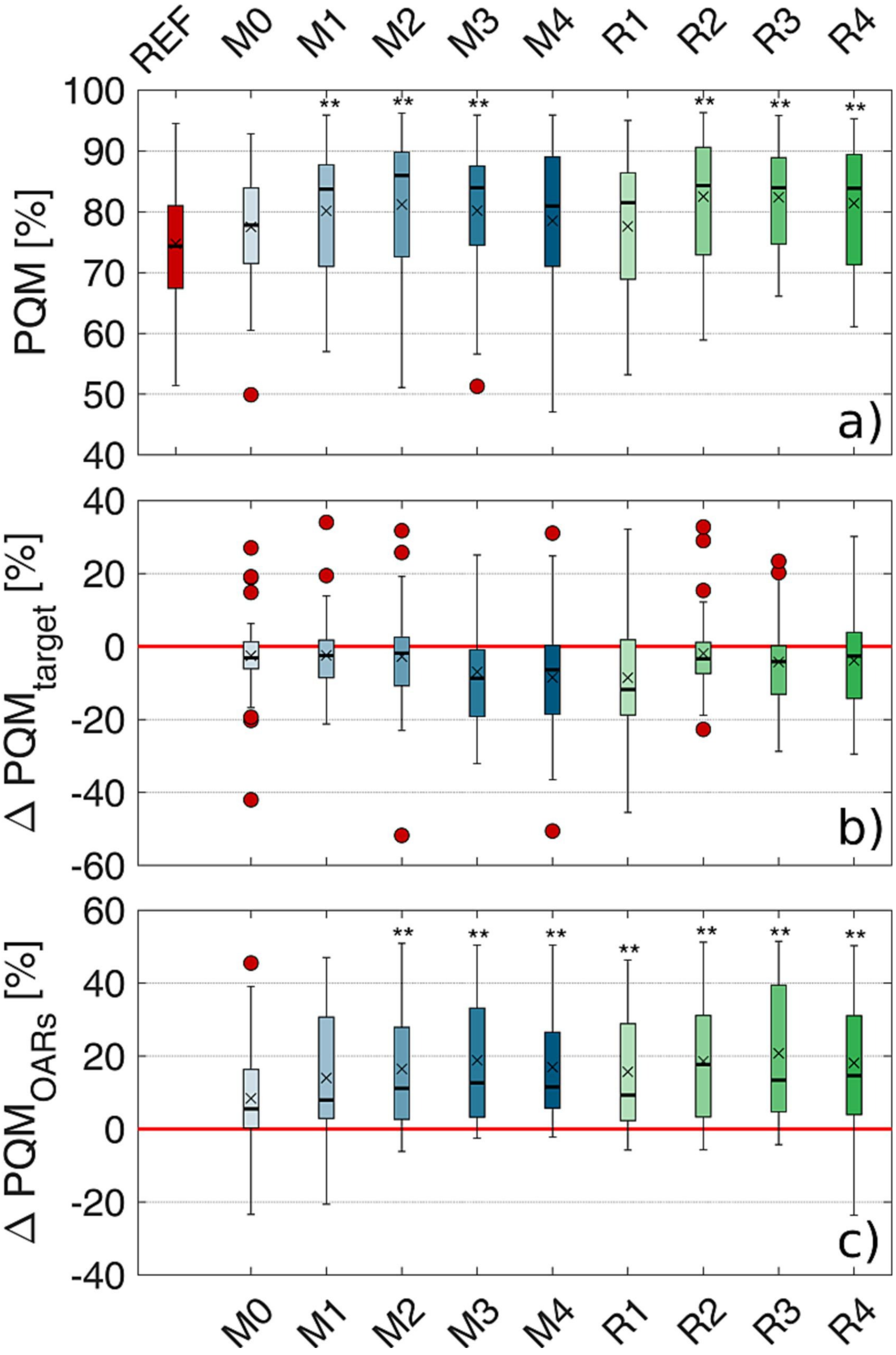


Figure B.2: Plan quality of automated plan model outcomes. a) Overall quality for all the models compared to the reference plan set. b) Target coverage quality difference (model-reference). c) OARs sparing quality difference (model-refer- ence). The central line marks the median, the cross marks the mean, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the adjacent values, which are the most extreme data values that are not outliers, and the circles represent the outliers. The asterisks mark models that are significantly different to the reference set.

monotonically across updates for both Mt and Rt models. More in detail, it was 131 ± 42 cc for M0 and became 111 ± 39 cc and 106 ± 37 cc for R3 and R4, respectively, a difference that was statistically significant to a Bonferroni corrected two-tailed paired Student's t-test. PTV shrinkage was closely coupled with a reduction in PTV-bladder overlap, which was significant for R3 and R4. Rectum and bladder volumes were constant across all the models. Details are reported in the Supplementary Material.

B.3.2 *Model outcomes*

AutoRP plans produced with Mt models were compared to Human plans on the set of newly enrolled patients in each specific period. The quality of Human plans remained constant over the entire duration of the investigation. On the other hand, the quality of AutoRP plans increased as the model was updated and the associated variability (here repre- sented in term of sample standard deviation) tended to decrease (Tab. B.2). The average quality difference is null at period P1 and it becomes positive and statistically significant to a two-tailed paired Student's t-test in P3 and P4.

Table B.2: Comparison of Human vs AutoRP plans along the four periods of investigation. Differences in plan quality ($APQM\%$) between Human and AutoRP plans are reported (mean ± standard deviation) and marked with an asterisk if statistically significant. Complexity metrics are also reported (mean ± standard deviation) significant paired differences are marked with an asterisk. EM: Edge Metric; vMCS: VMAT adapted Modulation Complexity Score; MIt: Total Modulation Index.

Period	**Features**	**Human**		**AutoRP**
P1	Model	M0		M0
	APQM%	85.7 ± 7.0		85.1 ± 9.1
	Difference		-0.6 ± 10.7	
	In next model	13		12
	MU/cGy	3.73 ± 0.76		3.34 ± 0.65
	EM	0.23 ± 0.12		0.19 ± 0.11
	VMCS	0.40 ± 0.07		0.38 ± 0.04
	MI_t	25.5 ± 6.0		29.2 ± 4.8
P2	Model	M0		M1
	APQM%	85.7 ± 7.6		89.6 ± 6.4
	Difference		3.9 ± 9.3	

Continued on next page

Table B.2: Comparison of Human vs AutoRP plans along the four periods of investigation. Differences in plan quality ($APQM\%$) between Human and AutoRP plans are reported (mean ± standard deviation) and marked with an asterisk if statistically significant. Complexity metrics are also reported (mean ± standard deviation) significant paired differences are marked with an asterisk. EM: Edge Metric; vMCS: VMAT adapted Modulation Complexity Score; MIt: Total Modulation Index. (Continued)

Period	Features	Human		AutoRP
	In next model	9		14
	MU/cGy	3.42 ± 0.8		3.54 ± 1.04
	EM	0.17 ± 0.05		0.17 ± 0.06
	VMCS	0.44 ± 0.08		0.42 ± 0.07
	MI_t	25.9 ± 7.3		26.3 ± 6.6
P3	Model	M0		M2
	APQM%	84.3 ± 8.2		90.8 ± 6.5
	Difference		$6.5 \pm 8.8^*$	
	In next model	4		20
	MU/cGy	3.18 ± 0.68		3.18 ± 0.59
	EM	0.15 ± 0.04		0.16 ± 0.04
	VMCS	0.44 ± 0.07		$0.41 \pm 0.06^*$
	MI_t	23.2 ± 5.9		25.5 ± 6.4
P4	Model	M0		M3
	APQM%	86.1 ± 6.1		91.4 ± 5.5
	Difference		$5.3 \pm 6.2^*$	
	In next model	2		23
	MU/cGy	2.89 ± 0.32		2.97 ± 0.34
	EM	0.14 ± 0.03		$0.15 \pm 0.03^*$
	VMCS	0.44 ± 0.04		0.44 ± 0.04
	MI_t	22.9 ± 3.8		23.5 ± 4.2

Similarly, also the number of human plans winning the comparison against the AutoRP ones decreased monotonically from P1 to P4, resulting in 2 human plans and 23 AutoRP (M3) plans accepted in P4. There was no consistent discernible difference in plan complexity. EM, VMCS and MIt remained nearly constant throughout the whole investigation period at approximately 0.16, 0.42 and 25 respectively with only sporadic differences between Human and AutoRP plans. MU/cGy showed a slightly but constant reduction for both classes of

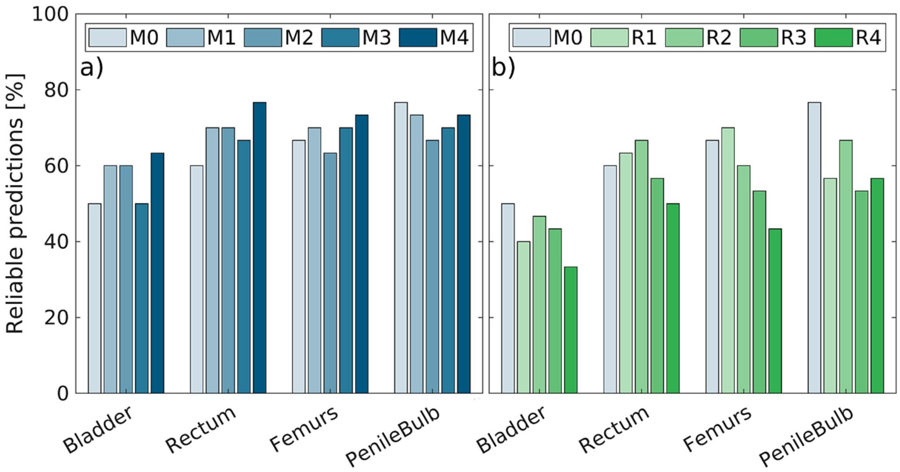


Figure B.3: Number of robust predictions for the open-loop comparison set: a) M1–M4 models; b) R1–R4 models. The M0 model is always given as a reference.

plans throughout the whole investigation period passing from ≈3.5 cGy - 1 in P1 to ≈ 3.1 cGy - 1 in P4 for both classes of plans. A detailed graph can be found in the Supplementary Material.

B.3.3 *Models comparison through open-loop test*

To compare model performances, we measured the differences in plan quality (ΔAPQM%) of an automatic optimization guided by the models' prediction against a set of reference plans on the external validation set. The results are shown in Fig. B.2. The mean ΔPQMOARs% ranged within 3.9 and 19.2 (Fig. B.2c): all models showed a significantly higher OAR sparing than M0 when compared to the reference plans, with the exception of M1. Target coverage and homogeneity slight reduced: mean ΔPQMtarget% was within 0.2 % and 14.7 % (Fig. B.2b). None of these reduction were statistitically significant to a Student's paired t-test.

The number of reliable DVH predictions, reported in Fig. B.3, increased with updates for Mt models but decreased for Rt models. In general, with subsequent model upgrades, plans tended to be more complex: larger MU/cGy, more complex MLC shaping and movement and larger modulation of gantry speed and dose rate, but significant differences were consistently observed only for MIt when compared to the reference set of plans (see Fig. B.4). The increased complexity is thus mainly related to an increased degree of modulation of gantry speed and dose rate.

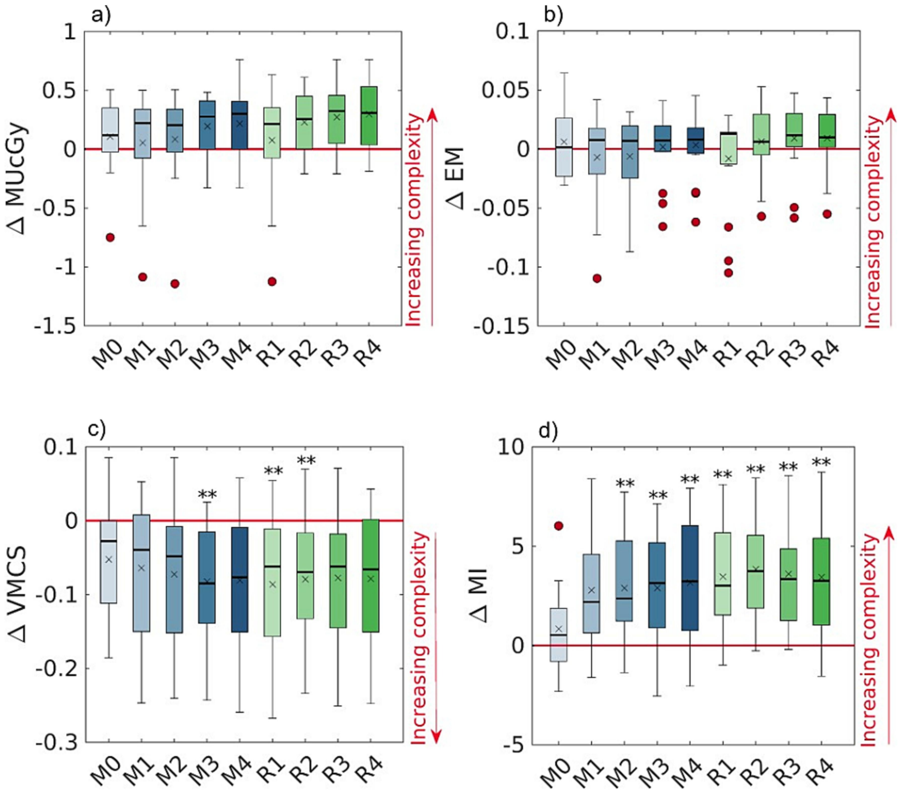


Figure B.4: Plan complexity of AutoRP plans for the validation sample. Differences are taken with respect to set of reference plans (model – reference). Panels report: a) MU/cGy, b) EM, c) vMCS, d) MIt. The central line marks the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the adjacent values, which are the most extreme data values that are not outliers, and the circles represent the outliers. No statistically significant differences appear. EM: Edge Metric; vMCS: VMAT adapted Modulation Complexity Score; MIt: Total Modulation Index.

B.4 DISCUSSION

In this study, we compared two antithetical approaches to periodi- cally update a KBP model dedicated to prostate radiotherapy and tested how they perform on a longitudinal clinical validation characterized by changes in the clinical settings. In particular, compared to the sample of patients used to train the ancestor KBP model (M0), three main clinical differences were introduced in the case sample considered during the investigation period. A high-definition MLC was used for all treatments, the fraction scheduling was gradually changed from 70 Gy/28 fx to 60 Gy/20 fx and the CTV-PTV margin recipe was reduced for a fraction of the cases towards the end of the investigation period. KPB updates were performed approximately every 25 new treatments and the benefits of such periodic updates, based on two different strategies, are reported.

Expanding the population of plans used to train a KBP model (Mt models) monotonically increases the overall model quality (an increase in R2 and a reduction in X2 and MSE), improves its predictions reliability and robustness, and caused an increase in the quality of its outcomes over the four time periods. The latter should not be considered as a consequence of the increasing sample size only. In fact, this was also due to the plan selection procedure followed in this work which was particularly challenging for the automatically generated plans due to the lack of human intervention for plan refinement. On the other hand, training a KBP model on a fixed number of more recent high-quality plans (Rt models) can at most improve the quality of its outcomes while undermining the model's quality as well as the robustness of its predictions.

The plan selection procedure adopted to populate the Rt models did not preserve the sample's anatomical homogeneity: the PTV average volume and the percentage of in-field bladder volume decreased across the updates (Supplementary material). Thus, selecting the population of plans only on the basis of their quality, as measured by the PQM%, leads to the discarding of unfavourable cases, e.g. patients with larger prostate glands and higher OARs volumes involved in the treatment field. If such a selection procedure is to be adopted for KBP model update, attention should be posed to reduce or minimize such artificial reduction of the geometrical clinical domain of the patient's characteristics. On the other hand, the narrower distributions of the predictors used by the model to generate a prediction may increase the probability of classifying a plan as a geometric outlier which may be also an explanation of the lower reliability of the predictions of Rt models compared to Mt ones.

Both the model generation phase (Tab. B.2) and the open-loop comparison (Fig. B.2) revealed an increase in the quality of model outcomes, when compared to human plans supported by the ancestor model M0. For Mt models, consistent increases are seen after the second update. For Rt models, one round of updates was already enough to induce consistent improvements. The quality increase was two-folded. On the one hand, the model upgrade seemed to be beneficial in our evolving clinical routine. In fact, in P4, when all the patients are treated with the new treatment schedule and a fraction of them are treated with reduced CTV- PTV margin, the KBP ancestor model was proven not to be as performant as the automatically generated one. In fact, from Tab. B.2 we observed that AutoRP (M3) plans had a significantly higher quality than Human ones, with 23 out of 25 automatically generated plans winning the selection procedure, even under the challenging conditions of lack of human refinement. On the other side, the open-loop comparison shows that the updated models outperform the ancestor one even on a set of cases which dates back to the same period of M0 population.

The average increase in the quality of the outcomes of Rt models is not necessarily in contrast with the reduction of the reliability of their predictions. In fact, the raw measure of model reliability introduced in this study is a good surrogate for the extent of knowledge each model contains, but the quality of the model should be measured on the basis of goodness-of-fit statistics (R2 and X2) and the goodness-of-estimation parameter (mean square error, MSE). In other words, a good algebraic relationship between the geometric and dosimetric characteristics of a reduced amount of good plans representing only a limited portion of the clinical geometric domain can still give reasonable predictions outside that limited geometric domain [230, 252].

The automated planning procedure, with a fixed set of optimization objectives, drove an increased organ sparing with a slight loss in target coverage and homogeneity (Fig. B.2). Our set of objectives probably tends to favour OARs sparing rather than PTV coverage and conformality. To correct for such trend, the set of optimization objectives should have been updated along with the model. This action was outside the aim of this work and was already proved useful in the work of Kaderka and colleagues [224].

The real improvement induced by the updating procedure might be underestimated, as a certain amount of skilled manual interventions are needed to achieve the highest quality results, even when KBP-generated objectives drive the optimization [221, 232, 248, 250, 253]. This work confirmed previous results by other groups: expanding KBP libraries induces an increase in model quality, while it does not guarantee the improved clinical quality of the model outcomes. On the other hand, reducing the model sample only to increase its mean quality might result in a less robust and less general model [223, 234, 236, 247].

As regards to plan complexity, several studies have reported that the improved plan quality due to KBP planning also resulted in increased plan complexity [237, 254, 255]. Our results demonstrated that when the plan complexity is appropriately constrained, the improvement in plan quality induced by KBP model predictions does not necessarily imply an unnecessary increase in plan complexity. Only one out of four metrics (MIt) showed significant increases which are mostly related to larger variations in the instantaneous dose rate and gantry speed.

The results of this work strictly apply only to VMAT prostate treatments, but the methods proposed herein are easily applicable to any other treatment sites and treatment techniques. The prostate case limits the need of trade-offs between OAR sparing and target coverage, but a well prepared quality metric should help the clinical staff to rank the plans on the basis of their quality. The PQM% can be replaced by any other measure of overall plan quality or by a detailed clinical evaluation by a team of clinicians [256, 257].

In conclusion, the longitudinal evaluation proposed in the present study showed the benefit associated with the update of a KBP model in a real clinical scenario characterized by changes happening over time. This work suggests how the KBP periodical update should be carried out in order to be both clinically and cost-effective. Pros and cons associated with the two updating strategies proposed were discussed. In general, the update routine should not only focus on increasing the sample library, but also on defining the criteria for the inclusion of new plans. Such criteria should be focused on selecting plans with high-quality removing unacceptable low-quality ones. The growing number of cases in the model sample and the elimination of clearly underperforming plans should result in: an increase in plan quality, an improved model reliability and also an improvement in the model's general quality [236]. In comparison to iterative approaches, this approach should save human resources and does not require the use of an external validation tool [231, 233, 234, 236, 237, 248].

C

QUANTUM ALGORITHM FOR THE CLASSIFICATION OF SUPERSYMMETRIC TOP QUARK EVENTS

THIS CHAPTER presents the paper by Bargassa *et al.* to which I have contributed as co-author. For the sake of brevity, only the main text is presented. For any supplementary material, please refer to the journal article [258].

C.1 INTRODUCTION

After attaining its nominal energy, the Large Hadron Collider (LHC) will reach an unprecedented collision rate during its high luminosity phase, opening the stage to discoveries beyond the standard model (SM) of particle physics. One of the most challenging tasks in searches taking place at the LHC is the capacity to categorize events of new phenomena (signal) and those of SM processes (background) which mimic the signal. Machine learning (ML) tools are among the most powerful means for separating signal from background events, having been key to the discovery of *eg.* the Higgs boson [259, 260]. More recently, quantum annealing for machine learning (QAML) [261] and its zooming variant (QAML-Z) [262] represent the first examples of a quantum approach to a classification problem in high energy physics (HEP).

In this paper, we study the application of the QAML-Z algorithm to the selection of supersymmetric top quark (stop) versus SM events. It is important to test this algorithm on a new classification problem where both the abundance of signal versus background events, and their overlap in the experimental observables are different from [262], therefore allowing to have a better assessment of its classification capability. A result on the stop search based on the data accumulated by the LHC in 2016 (35.9 fb^{-1}) has been published [263]. It is based on a classical ML tool which will serve as a reference for gauging the performance of the new classifiers. The variables discriminating between the stop signal and the SM background, which are used to train the QAML-Z algorithm, are based on the same ones as in the classical ML tool [263]. We present results of the QAML-Z algorithm for different schemes of zoomed quantum annealing, and various sets of variables used in the annealer. Also, we introduce a pre-processing of the data through a principal component analysis [264] (PCA) before feeding it to the annealer.

C.2 SEARCH FOR SUPERSYMMETRIC TOP QUARK

One of the main objectives of the physics program at the LHC are searches for supersymmetry (SUSY) [265–270], one of the most promising extensions of the SM. SUSY predicts superpartners of SM particles (sparticles) having the same gauge interactions, and whose spin differs by one-half unit with respect to their SM partners. The search for SUSY has special interest in view of the recent discovery of a Higgs boson [259, 260] as it naturally solves the problem of quadratically divergent loop corrections to the Higgs boson mass. In this study we describe the classification aspect of a search for the pair production of the lightest supersymmetric partner of the top quark $\tilde{t}_1$ at the LHC machine at $\sqrt{s}$ = 13 TeV, where each stop decays in four objects, see Fig. C.1. The lightest neutralino $\tilde{\chi}_1^0$ is considered to be stable as the lightest supersymmetric particle. The final states considered contain jets, missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$), and a lepton which can be either a muon or an electron.

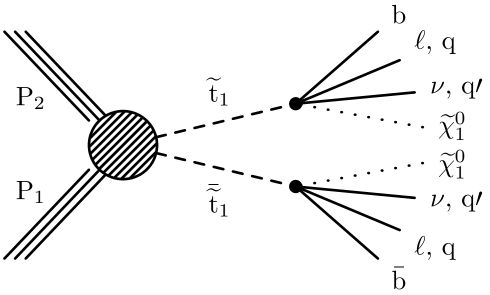


Figure C.1: Stop pair production at the LHC with four-body decays.

The sensitivity of this type of search is dominated by the capacity to distinguish the stop signal from background events, whose production dominates the signal by several orders of magnitude, and whose observables overlap with the ones of the stop signal. In this search, the main background processes are the $t\bar{t}$ and W+jets productions. In the search based on a classical ML tool [263], a pre-selection is first applied to decrease the overwhelming background stemming from the SM. In a second stage, boosted decision trees (BDTs) [271, 272] are used to classify events as signal and background. To find which variables are the most discriminating and should be fed as input to the BDT, different sets of variables are tested as input to a BDT whose output is used to maximize a figure of merit (FOM) [273]:

$$\mathrm{FOM} = \sqrt{2\Big((S+B)\ln\Big[\frac{(S+B)\cdot(B+\sigma_B^2)}{B^2+(S+B)\cdot\sigma_B^2}\Big] - \frac{B^2}{\sigma_B^2}\ln\Big[1+\frac{\sigma_B^2\cdot S}{B\cdot(B+\sigma_B^2)}\Big]\Big)} \tag{C.1}$$

where S and B respectively stand for the expected signal and background yields for an integrated luminosity of 35.9 fb^{-1} at the LHC. The term $\sigma_B = (f \cdot B)$ represents the expected systematic uncertainty on the background with f being the relative uncertainty of the background yield, taken to be $f = 20\%$ as in [263]. The set of variables which maximize the FOM is chosen as the final set of input variables to the BDT. This metric captures the full information of the statistical and systematic uncertainties of a given selection, as it is important to account for the actual conditions of a search. The approach based on the maximization of this FOM has been very effective to find the smallest and most efficient set of discriminating variables in several searches [263, 274]. A description of the BDT parameters as well as its development with a FOM maximization procedure as used in [263] are provided in Sec. **??**. The list of variables is presented in Tab. C.1 and their distribution for signal and background is provided in Fig. C.2. To render the results of the classification based on the QAML-Z algorithm as comparable as possible to the one of [263], we use the same pre-selection of events before training (see Supplementary Material). Furthermore, since the FOM as defined in Eq. (C.1) represents a complete and single-number measure of the power of a selection, we evaluate the performance of the QAML-Z algorithm by a maximization of the FOM versus a cut on its output. Finally, for the comparison of performances to reflect only the difference of a quantum based versus a classical tool, we train the QAML-Z algorithm with different sets made of the same discriminating variables as in the BDT based search [263] (see Tab. C.1).

C.3 QUANTUM ANNEALING & ZOOMING

From the distribution of each variable i in signal and background events, we construct a weak classifier χ_i as in [261] which retains the discriminant character of each variable while adapting it to an annealing

process. We then construct an Ising problem as follows. For each training event $\tau \in [1, S]$, we consider the vector x_τ of the values of each variable of index i we use, and a binary tag y_τ labeling the event τ as either signal ($+1$) or background (-1). The value of the i-th weak classifier for the event τ is given by the sign of the corresponding weak classifier χ_i: $c_i(x_\tau) = \text{sgn}(\chi_i(x_\tau))/N = \pm 1/N$, where N is the number of weak classifiers. In the QAML algorithm, the optimization

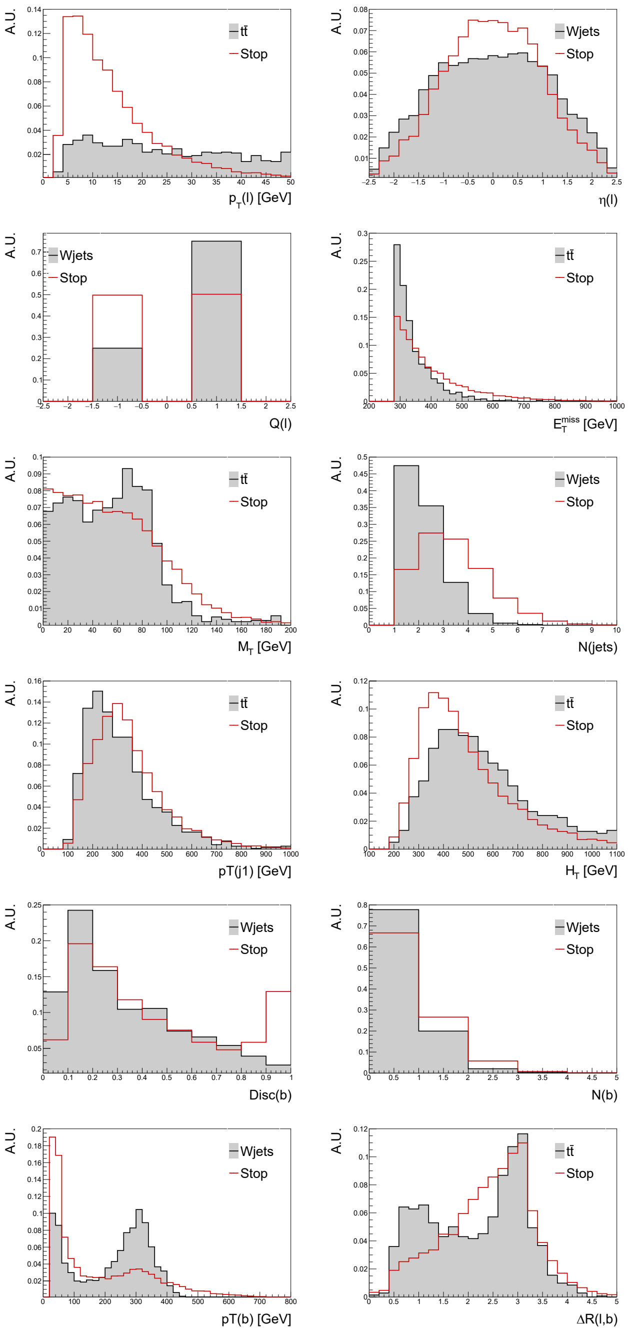


Figure C.2: Distribution of the discriminating variables for the stop signal with $\Delta m = 30$, W+jets and $t\bar{t}$, used as input to a BDT in [263]. Starting from top-left to bottom-right: $p_T(l)$, $\eta(l)$, $Q(l)$, E_T^{miss}, M_T, N(jets) , $p_T(j_1)$, H_T, Disc(b) , $N(b)$, $p_T(b)$, $\Delta R(l,b)$. Distributions are normalized to the same area and shown at preselection.

Variable	Description
$p_{\mathrm{T}}(l)$	p_{T} of the lepton l
$\eta(l)$	Pseudorapidity of the lepton l
$Q(l)$	Charge of the lepton l
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy
M_{T}	Transverse invariant mass of the ($E_{\mathrm{T}}^{\mathrm{miss}}$,$p_{\mathrm{T}}(l)$) system
$N(\text{jets})$	Multiplicity of selected jets
$p_{\mathrm{T}}(j_1)$	p_{T} of the leading jet
H_{T}	$\sum_i p_{\mathrm{T}}(jet(i))$
Disc(b)	Maximum b-quark tagging discriminant of the jets
$N(b)$	Number of b-tagged jets
$p_{\mathrm{T}}(b)$	p_{T} of the jet with the highest b-discriminant
$\Delta R(l,b)$	Distance between the lepton and the jet with the highest b-discriminant

Table C.1: List of discriminating variables used as input to a BDT in [263].

of the signal-background classification problem is expressed in terms of the search for the set of spins s_i minimizing the Ising Hamiltonian:

$$\begin{aligned} H_{\text{Ising}} &= \sum_{i=1}^{N} h_i s_i + \sum_{i=1}^{N}\sum_{j>i}^{N} J_{ij} s_i s_j \\ &= \sum_{i=1}^{N} \left(\lambda - C_i + \frac{1}{2}\sum_{j>i}^{N} C_{ij}\right) s_i + \frac{1}{4}\sum_{i=1}^{N}\sum_{j>i}^{N} C_{ij} s_i s_j \end{aligned} \tag{C.2}$$

where h_i is the local field on spin s_i, and J_{ij} is the coupling between spins s_i and s_j. The factor λ is a regularization constant, and the terms C_i and C_{ij} are defined as function of weak classifier values and event tags as:

$$C_i = \sum_{\tau=1}^{S} c_i(x_\tau) y_\tau, \quad C_{ij} = \sum_{\tau=1}^{S} c_i(x_\tau) c_j(x_\tau). \tag{C.3}$$

A strong classifier R is then built as a linear combination of all weak classifiers and the spins, merging for each event the discriminating power provided by all c_i's and the spins s_i obtained from the quantum annealing process. The minimization of the classification error is performed by the minimization of the Euclidean distance between the binary tag of each event and its classification R as obtained by the annealing:

$$||y-R||^2 \quad = \quad \sum_{\tau=1}^{S} |y_\tau - \sum_{i=1}^{N} s_i c_i(x_\tau)|^2. \tag{C.4}$$

In the QAML-Z approach, the quantum annealing is operated iteratively while a substitution is made to the spin s_i:

$$s_i \quad \longrightarrow \quad \mu_i(t) + s_i \cdot \sigma(t) = \mu_i(t+1), \tag{C.5}$$

where:

- $\mu_i(t)$ is the mean value of qubit i at time t. We have: $\forall i \ \mu_i(0) = 0$.
- $\sigma(t)$ is the search width at each annealing iteration t. We have: $\sigma(t) = b^t$, where $b = \frac{1}{2}$ and $t \in [0, T-1]$.

This iterative procedure effectively shifts and narrows the region of search in the space of spins. It updates the vector μ_i which is collected at the final iteration to form the strong classifier:

$$R(x_\tau) \quad = \quad \sum_{i=1}^{N} \mu_i(T-1) c_i(x_\tau), \tag{C.6}$$

where the use of the weak classifiers isn't limited to the binary choice $\{0, 1\}$, but is extended to the continuous interval $[-1, 1]$ via the use of the vector μ_i. The classification capacity of the QAML-Z algorithm is further enhanced by an augmentation scheme applied on the weak classifiers. For each h_i, several new classifiers c_{il} are created:

$$c_{il}(x_\tau) \quad = \quad \frac{\text{sgn}(h_i(x_\tau) + \delta l)}{N}, \tag{C.7}$$

where $l \in \mathbb{Z}$ is the offset: $-A \leq l \leq A$, and δ is the step size. While the value c_i of the old classifier has only a binary outcome for each h_i, the new classifiers c_{il} have similar but $(2A+1)$ different outcomes depending on the very distribution of h_i. We therefore have a better discrimination because a more continuous, thus more precise representation of the spectrum of h_i with c_{il} than with c_i. Applying the substitution of Eq.(C.5) in Eq. (C.4), omitting spin independent and quadratic self-spin interaction terms, and defining new indices I as $\{il\}$ and J as $\{jl'\}$, we obtain the Hamiltonian (see Sec. **??**):

$$\begin{aligned} H(t) \quad = \quad & \sum_{I=1}^{N(2A+1)} \left(-C_I + \sum_{J=1}^{N(2A+1)} \mu_J(t) C_{IJ} \right) \sigma(t) s_I \\ & + \frac{1}{2} \sum_{I=1}^{N(2A+1)} \sum_{J \neq I}^{N(2A+1)} C_{IJ} \sigma^2(t) s_I s_J, \end{aligned} \tag{C.8}$$

with:

$$C_I = \sum_{\tau=1}^{S} c_{il}(x_\tau) y_\tau, \quad C_{IJ} = \sum_{\tau=1}^{S} c_{il}(x_\tau) c_{jl'}(x_\tau). \tag{C.9}$$

The terms C_I and C_{IJ} are the input to the classification problem. The Hamiltonian $H(t)$ is iteratively optimized for t, with the vector μ_I updated similarly to Eq.((C.5)). The information about the iterative quantum annealing, corresponding parameters, and control results are provided in Sec. **??**, where we ensure that the Ising model energy decreases and stabilizes for the chosen parameters. The output of the optimization procedure is a strong classifier built as in Eq.((C.6)), and whose distribution is used to discriminate signal from background.

C.4 CLASSIFICATION OF STOP WITH THE QAML-Z ALGORITHM

As in [263], only the main background processes W+jets and $t\bar{t}$ are used for training the QAML-Z algorithm. To realistically represent the SM in the training, a background sample is formed where events of these two processes are present proportionally to their production rate at the LHC. We divide this sample in two equal parts, one being used by the QAML-Z algorithm and one to assess the performance of the strong classifier through the maximization of the FOM: N(Sample)=N(QA)+N(Assess). The QA sample is further divided in two equal parts, one to train the annealer and another one to test for over-training in the annealer: N(QA)=N(Train)+N(Test). It should be noted that only the Train sample is involved in the annealing process. Having shown [263] that the kinematic properties of all signal points $(m(\tilde{t}_1), m(\tilde{\chi}_1^0))$ are quasi identical along the line $\Delta m = m(\tilde{t}_1) - m(\tilde{\chi}_1^0)$, we use all signal events with $\Delta m = 30$ except the signal point $(550, 520)$ as QA sample, while entirely using this latter signal as Assess sample. This organization of samples allows the usage of a maximal number of both signal and background events for assessing the performance of the classification as well as testing the annealing process.

The data is run on the 2000Q quantum annealer of D-Wave Systems Inc. [275], where the time to solution is $O(\mu s)$, *ie.* the time of the annealing (see Sec. **??**). This computer is based on the Chimera graph which has 2048 qubits and 5600 couplers. To embed the Ising Hamiltonian in the annealer, qubits of the graph are ferromagnetically coupled into a chain to represent a single spin of the Hamiltonian $H(t)$. While the Hamiltonian in Eq. (C.8) is fully connected, the Chimera graph is not, thus limiting the hardware implementation of the classification problem. The number of J_{ij} couplers is given by: $N(J_{ij}) = N_\mathrm{v} \cdot (N_\mathrm{v} - 1)/2$ with $N_\mathrm{v} = N_\mathrm{var} \cdot (2A + 1)$, where N_var is the number of variables used to train the QAML-Z algorithm, and A is a parameter of the augmentation scheme (see Eq. (C.7)). Given the number of variables and the augmentation schemes used, the limit of 5600 couplers can be exceeded by $N(J_{ij})$; typically, for N_{v}=12 and for an augmentation with A=5, the needed number of couplers is 8646. We therefore prune the elements of the J_{ij} matrix, retaining the largest $(1-C)$ elements, where C is a cutoff percentage. Different cutoff values are expected to optimize the performance for different sets of variables and different augmentations schemes (A, δ). As a further option to reduce the size of the Ising model to be encoded on the annealer, we use the polynomial-time variable fixing scheme of the D-Wave API. This scheme is a classical procedure to fix the value of a portion of the input variables to values that have a high probability of being optimal. An illustration of the effect of the cutoff C, the use of variable fixing, and the augmentation scheme is given in Sec. (**??**).

In order to compare the performance of a quantum annealing with a classical ML counterpart, we explore various *settings* of the QAML-Z algorithm, namely different augmentation schemes, cutoffs, and variable fixing options, reporting only the performance of the best setting for each tested set of variables (see Sec. C.5). Despite averaging out the random errors on the annealing and mitigating the possible effects of over-fitting due to zooming (see Sec. C.3), the outcome of the annealing (the vector μ_I) can vary due to the probabilistic nature of these schemes and to the variations of the machine itself (e.g. low-frequency flux noise of the qubits), leading to an uncertainty on the performance. In order to estimate this uncertainty, we run the annealing ten times with the same input variables, in the very same setting, and on the same sample of events, and we consider the standard deviation of the corresponding maximal FOMs as uncertainty of the performance for a given set of variables and setting. In Fig. C.3 we report the performance of the QAML-Z algorithm with the variables of Tab. C.1 as input and with a given augmentation scheme and cutoff as a function of the number of events used in the training. The performance of the annealer increases with N(Train), witnessing a clear rise for rather small number of events and a more moderate increase for larger numbers of events, confirming the results of [261] with another signal. Henceforth, we will present all results for N(Train)=N(Test)=$50{\cdot}10^3$ where signal and background events respectively represent 40% and 60% of these two samples. We therefore benefit from a large sample size to train the QAML-Z algorithm, while observing a quasi identical evolution of the Hamiltonian energy for the Train and Test samples (see Fig. **??** in Sec. **??**). The Assess sample contains approximately $200{\cdot}10^3$ background, and $7{\cdot}10^3$ signal events. In Fig. C.4 we present the distribution of the strong classifier for signal, and the two main background processes. As can be observed, there is no over-training of the QAML-Z algorithm because the response of the strong classifier is statistically very similar for events which are used to train the annealer and those not exposed to the training. Also shown in Fig. C.4 is the evolution of the FOM in the Assess sample as function

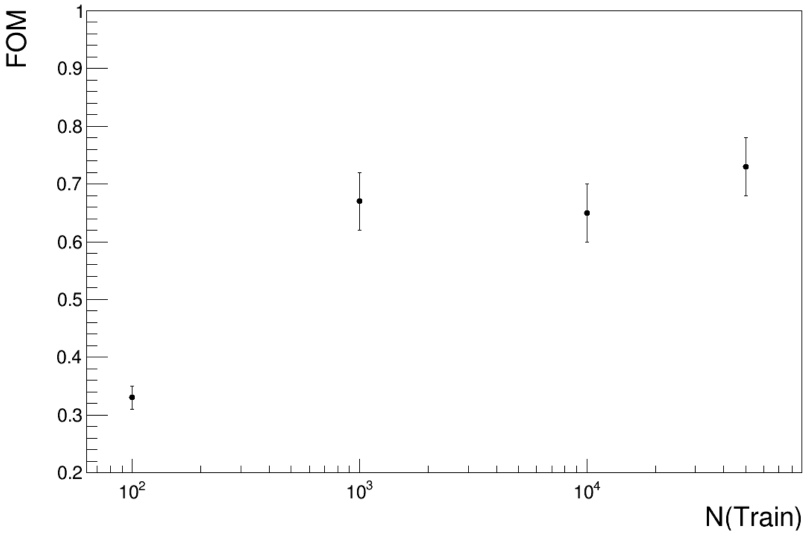


Figure C.3: Evolution of the FOM as a function of the number events used for training. The QAML-Z algorithm uses the variables of Tab. C.1 transformed in weak classifiers, with an augmentation scheme of (δ,A)=(0.025,3), and with a cutoff C=85%, without using a variable-fixing procedure.

of the cut applied on the output of the strong classifier. Henceforth, all the reported values of maximal FOM are checked to correspond to a cut where there are enough events in both signal and background samples.

C.5 APPROACHES AND RESULTS

We define the main sets of tested variables in Tab. C.2. For each set and the different approaches to test it, we perform an extensive study of the performance of the QAML-Z algorithm for different augmentation schemes, cutoffs, and the use (or not) of variable fixing, as illustrated for the sets A and B respectively in the Supplementary Material. For each set and approach, we report the optimal setting and the corresponding performance in Tab. C.3.

The set α contains the variables defined in Tab. C.1 where the discriminating variables are not transformed into weak classifiers, being only normalized to the $[-1,+1]$ interval. The set β consists of the same variables where these are transformed into weak classifiers. As can be seen in Tab. C.3, the performance of the set β is expectedly higher than for the set α, where the weak classifiers are scaled as a function of the initial distribution of the discriminating variables to better reflect the separation between signal and background. The transformation into weak classifiers is performed for all subsequent tests.

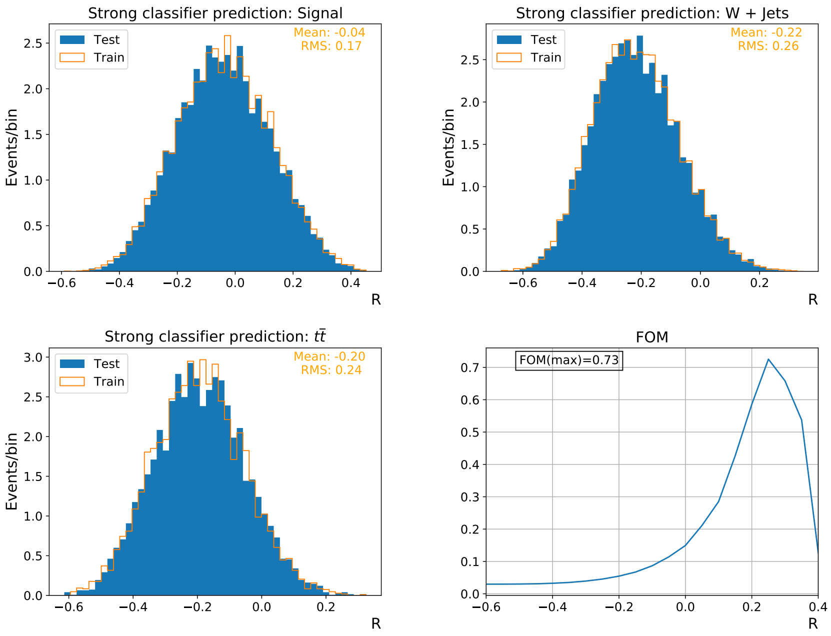


Figure C.4: The output of the strong classifier for the signal (top-left), W+jets (top-right) and $t\bar{t}$ background (bottom-left) in the train (orange) and test (blue) events within the QA sample. The evolution of the FOM as a function of the cut applied on the strong classifier's output is illustrated in the plot in the bottom-right. The QAML-Z algorithm uses the variables of Tab. C.1 transformed in weak classifiers, with an augmentation scheme of (δ,A)=(0.025,3), and with a cutoff C=85%, without using a variable-fixing procedure. The number of events used for training is N(Train)=$50{\cdot}10^3$.

We explore in a second step the effect of additional discriminating variables built from the same initial set of Tab. C.1. The methodology followed to built these new variables in explained in Sec. **??**, where the discriminating power of each variable is appraised via its maximal FOM. Two new sets of variables are constructed based on these new variables, as reported in Tab. C.2: the set A including the variables of the Tab. C.1 and new variables with the highest FOMs, and the set B including those of set A and additional variables with the lower FOMs (see Tab. **??**). As can be observed in Tab. C.3, the addition of variables with higher maximal FOMs in the set A increases the performance of the QAML-Z algorithm, while the further addition of variables with lower FOM in the set B doesn't significantly improve the quality of the classification.

The results of the search [263] are based on the use of BDT where the discriminating variables are diagonalized before being fed to the training [271, 272]. This step better prepares the data for classification because the original discriminating variables do not necessarily constitute the optimal basis in which signal and background are optimally

Variable set name	List of variables	Use of weak classifiers
α / β	Tab. C.1	No / Yes
A	Tab. C.1 and: $p_{\mathrm{T}}(l)/E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}(l)/p_{\mathrm{T}}(j_1)$, (Disc(b) $-1)p_{\mathrm{T}}(b)$, $\lvert(E_{\mathrm{T}}^{\mathrm{miss}}-280)(M_{\mathrm{T}}-80)\rvert$, $\lvert(E_{\mathrm{T}}^{\mathrm{miss}}-280)(H_{\mathrm{T}}-400)\rvert$	Yes
B	Variables of set A and: $\Delta R(l,b)-(M_{\mathrm{T}}/40)$, $H_{\mathrm{T}}{}^2/N(\mathrm{jets})$, $p_{\mathrm{T}}\ +3.5\eta(l)^2$, $p_{\mathrm{T}}\ /H_{\mathrm{T}}$	Yes

Table C.2: Definition of different variable sets as a function of the used variables.

separated. In order to render our approach as comparable as possible to the one followed with a BDT [263], we pass our data through the procedure of PCA [264] before feeding it to the QAML-Z algorithm. It must be noted that the use of PCA is only one method for diagonalizing the data, other methods also being applicable to this end. The application of PCA on the data before the quantum annealing further improves the results for the set of variables A, and to a lesser extent for B, as can be seen in Tab. C.3. We note a larger uncertainty of the QAML-Z algorithm where the data is prepared with the PCA, this for the same sets of variables. In the PCA basis, the weak classifiers are more decorrelated from each other, rendering the corresponding weights μ_I more independent from one another. When a μ_I fluctuates (e.g. because of the state of the machine), the strong classifier R (see Eq. (C.6)) is sensitive to the variations of a larger number of μ_I's, hence a larger variation of its outcome. It is noticeable that the QAML-Z algorithm, once put on a footing as similar as possible to the BDT based approach [263], can reach an equivalent, possibly better performance. It is interesting to observe that the best result is achieved without using the variable fixing scheme, where the annealing is put at full use.

C.6 SUMMARY

We studied the capability of the quantum annealing, where the zoomed and augmented QAML-Z approach is applied to a new classification problem, namely the discrimination of stop versus SM background

Variable set	Fixing variable	C [%]	(δ,A)	FOM
α	False	85	$(2.50\cdot10^{-2},3)$	0.48 ± 0.03
β	False	85	$(2.50\cdot10^{-2},3)$	0.73 ± 0.03
A	True	95	$(0.90\cdot10^{-2},5)$	0.88 ± 0.04
B	True	85	$(0.70\cdot10^{-2},3)$	0.91 ± 0.05
PCA(A)	False	95	$(1.45\cdot10^{-2},3)$	1.57 ± 0.24
PCA(B)	True	95	$(0.70\cdot10^{-2},5)$	1.09 ± 0.17
BDT	NA	NA	NA	1.44 ± 0.06

Table C.3: Best performance obtained for different sets of variables as defined in Tab. C.2, and for different approaches applied on some sets. The corresponding use (or not) of variable fixing, cutoff and augmentation scheme are reported. All results are provided for N(Train)=50000. For comparison, the performance of the BDT of [263] is also reported, where "NA" stands for non-applicable.

events. The classification is based on well motivated variables whose discriminating power has been tested with a FOM maximization procedure. The use of this latter metric constitutes a novel and reliable assessment of the performance of a selection as it includes its full statistical and systematic uncertainties. We systematically tested each set of variables used by the QAML-Z algorithm for different augmentation schemes and percentages of pruning on the couplers of the annealer as to find the optimal setting. The performance of different settings is assessed for large training samples which are observed to yield the best performances, and are also more adapted to the needs of experimental particle physics where very large data samples are used. We observe an improvement of the classification performance when adding variables with a high FOM. To put the annealing approach and the classical BDT approach on the same footing, we pass the data through a PCA procedure before feeding it to the quantum annealer. For the first time in HEP, we show that for large training samples the QAML-Z approach running on the Chimera graph reaches a performance which is at least comparable to the best-known classical ML tool. With more recent graphs there is the prospect that the larger number of connected qubits will yield a better correspondence between the Ising Hamiltonian and the system of qubits of the annealer. The larger number of available couplers in the machine will allow a more complete use of the information contained in the couplers of the Hamiltonian; it will render each chain more stable, thus less prone to be broken, where the discriminating information of the classification will be more effectively used.

ACKNOWLEDGEMENTS

We acknowledge the authors of [261, 262] for their help which has been key for this study. We specially thank the Center for Quantum Information Science & Technology of the University of Southern California in Los Angeles for granting us access to the 2000Q quantum annealer of D-Wave Systems Inc. [275]. Finally, we acknowledge the Instituto de Telecomunicações of Lisbon. PB thanks the support from Fundação para a Ciência e a Tecnologia (Portugal), namely through project UIDB/50008/2020, as well as from project *QuantHEP – Quantum Computing Solutions for High-Energy Physics*, supported by the EU H2020 QuantERA ERANET Cofund in Quantum Technologies, and by FCT (QuantERA/0001/2019).

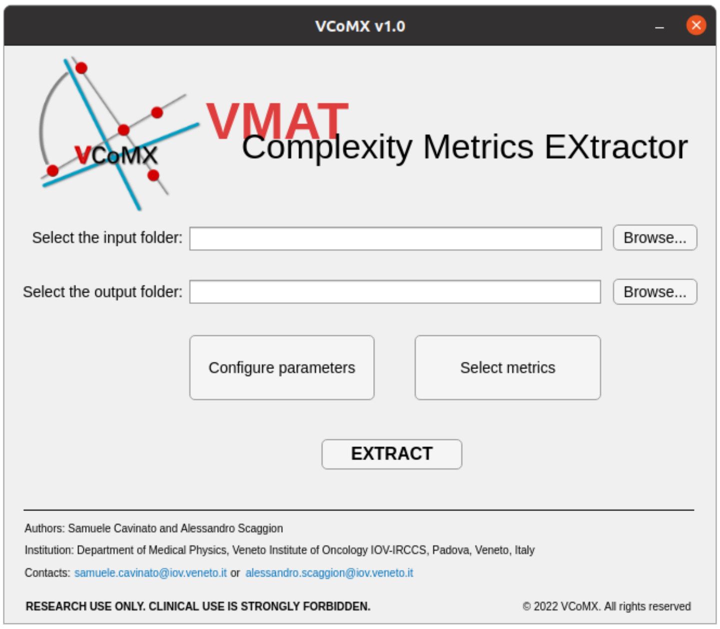
D

UCOMX USER MANUAL

THIS CHAPTER presents the functioning of the `UCoMX` package.

D.1 OVERVIEW AND FUNCTIONING OF VCOMX

D.1.1 *Input section*



Authors: Samuele Cavinato and Alessandro Scaggion

Institution: Department of Medical Physics, Veneto Institute of Oncology IOV-IRCCS, Padova, Veneto, Italy

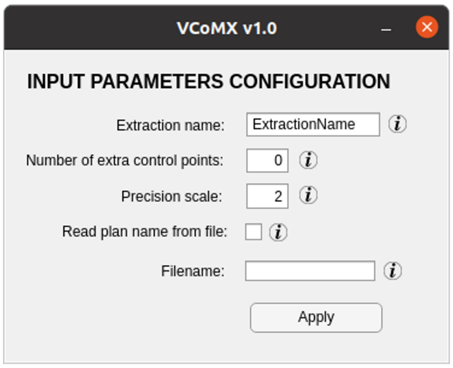
Contacts: samuele.cavinato@iov.veneto.it or alessandro.scaggion@iov.veneto.it

RESEARCH USE ONLY. CLINICAL USE IS STRONGLY FORBIDDEN.

© 2022 VCoMX. All rights reserved

Figure D.1: Main panel of the VCoMX extraction engine. From this panel, the input and output paths can be set.

The input data for the VCoMX engine is taken from two specific text files: CONFIG.in and METRICS.in. In the GUI-based version, the package automatically generates these files when the user fills in the GUI fields. Conversely, in the Matlab command-line version, the user must manually enter the data into these files. A pre-filled version is provided with the package. The definitions of the different fields are explained in the following subsections.



VCoMX v1.0

INPUT PARAMETERS CONFIGURATION

Extraction name: ExtractionName

Number of extra control points: 0

Precision scale: 2

Read plan name from file: ☐

Filename:

Figure D.2: Panels for configuration of the additional field of the CONFIG.in file.

CONFIG.in

This file is used to set all the necessary input parameters. The overall structure of the file is the following:

CONFIG.in

 # Analysis ID name
(OPTIONAL) Type here a user's define name to assign to the current session

 # Dataset input folder
Type here the full path where your DICOM RTPLAN files or a file containing the list of plans (see below) are stored.

 # Results output folder
Type here the full path where the results of the extraction process will be saved

 # Precision scale (1: 1mm, 2: 0.5mm, 10:0.1mm
Spatial Resolution for the computation of some MLC related metrics (DEFAUL: 10)

 # Read from file
True or *False*. If it is set to *true*, the library will read the filenames from a dedicated .xlsx file stored in the 'Dataset input folder'. The name of the .xlxs file is specified in the 'List file name' field (see below). An example of this file is provided together with the reference dataset.

 # List file name
.xlsx file containing the list of plans. It is used only if 'Read from file' is set to *true*.

In Fig. D.1 and Fig. D.2 the GUI panels used for setting the values of fields of the CONFIG.in file in the GUI-based versions are shown.

WARNING/1

If the CONFIG.in is filled manually (i.e. not using the GUI), be sure of:

- **Leaving** the fields with the # as they are in the pre-filled version
- **Writing** the value of a given field must be written immediately below the name of the field as shown in the CONFIG.in box above.

machinename.txt

There are some needed parameters that are not stored in the DICOM RT-Plan files and which are needed in order to get an accurate computation of the complexity indicators: the maximum gantry speed and maximum leaf speed. These two parameters depend on the particular Linac used. The user can create one file for each Linac by naming it as *machinename*.txt, where machinename should be replace with the name of the machine as store in the TPS. This file must be placed .../UCoMX/VCoMX/database/LINACInfo/. The file structure to use is the following:

machinename.txt

```
##Optional comments such as the machine name

##

# Maximum gantry speed [deg/s]
value

# Maximum MLC leaf speed [cm/s]
value
```

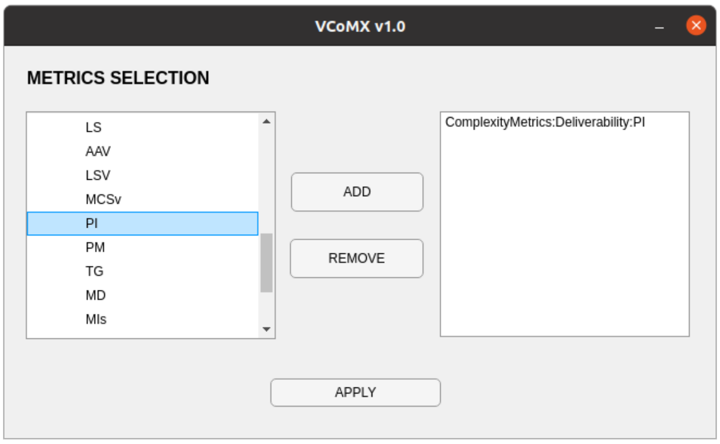
In case machinename.txt is not provided, default parameters are used. In particular, the maximum gantry speed will be set to $6deg/s$ and the maximum leaf speed to $2.5cm/s$.

WARNING/2

Before setting the default paramters, VCoMX looks for the machinename.txt file using the machinename stored in the DICOM RT-Plan file. Therefore, check that machinename is spelled correctly, otherwise VCoMX will not find it.

D.1.2 *METRICS.in*

This file is used to select the metrics to extract and to set their extraction parameters. The general structure of the file is the following:



ComplexityMetrics:Deliverability:PI

Figure D.3: Panel of the VCoMX interface for the selection of the complexity metrics to extract.

METRICS.in

```
FIRST CATEGORY NAME:
    FIRST SUBCATEGORY NAME:
        METRIC NAME 1
        METRIC NAME 2
        ...
    SECOND SUBCATEGORY NAME:
        METRIC NAME 1
        ...
SECOND CATEGORY NAME:
    ...
```

TIPS

- The indentation of the input file can help to visualize the content, but it is not fundamental for correct execution.
- **Punctuation marks are part of the syntax** and must be used properly. In particular:
  - Each category and subcategory is followed by semi-colons ":"
  - No punctuation has to be but after the names of the metrics.

D.1.3 *How to use VCoMX*

Configuration and execution of the stand-alone GUI-based version

The stand-alone version of UCoMX can be downloaded from:

https://doi.org/10.5281/zenodo.8276838

This version was created for Windows 10 a requires the Matlab Run-Time r2022b for its execution (available at this link[1]). If stand-alone version for a different operative system is needed, please email us at samuele.cavinato@iov.veneto.it or alessandro.scaggion@iov.veneto.it.

Once you have UCoMX available on your PC and you have the Matlab Runtime installed, please follow these steps for the execution:

1. Move to the UCoMX download folder;
2. Double click on UCoMX;
3. Select VCoMX.
4. Select the input folder where the DICOM RT-Plan files are located;
5. Select the output folder where the results will be stored;
6. (Optional) Click on 'Configure parameters'.
7. From the new panel you can:
    a) Give a name to the execution.
    b) Select the precision scale of the MLC BEVs.
    c) Select if you want to read the list of plans from a .xlsx file by enabling the checkbox. If so, write the filename. In this case, the file must be place in the input folder indicated in the previous panel.

    REMEMBER

    The VCoMX library can handle three different types of input:

    i. Folder containing all the DICOM RT-Plan files;
    ii. Folder containing a subfolder for each DICOM RT-Plan file;
    iii. File containing the list of plans to extract the metrics from.
8. Close the 'Configure parameters' panel;
9. Click on 'Select metrics', a new panel appears.The metrics implemented in the VCoMX engine are listed on the left box. The metrics that will be extracted are listed in the right You can perform the following operations:

---

[1] https://ssd.mathworks.com/supportfiles/downloads/R2022a/Release/7/deployment_files/installer/complete/win64/MATLAB_Runtime_R2022a_Update_7_win64.zip

   a) Add metrics to the right box. To do it, select a metric on the left box and click 'ADD'.
   b) Remove a metric from the left box. To do it, select a metrics a click 'REMOVE'.

10. Once the metrics have been selected, close the panel.

11. From the main panel, click 'EXTRACT' to start the execution.

Configuration and execution of the `MATLAB`® GUI-based version

The `MATLAB`® GUI version of UCoMX can be downloaded from this link[2]. The execution of this version requires a valid Matlab licence installed. The use of this version is very similar to that of the stand-alone versione, with the first three steps replaced by the following ones:

1. Open Matlab
2. Move to .../UCoMX
3. Add .../UCoMX and its subfolders to the Matlab path;
4. Type 'UCoMX' on the command line;
5. Continue from point 4 of the instructions for the stand-alone version.

Configuration and execution of the `MATLAB`® command-line version

This version requires a valid `MATLAB`®licence installed. It is the same as the Matlab GUI-based version (download), but it is run from the Matlab command line without using the GUI. This extends its compatibility with older Matlab versions. The steps for the execution are the following:

1. Open Matlab
2. Move to .../UCoMX/
3. Add .../UCoMX/VCoMX to the Matlab path
4. Move to .../UCoMX/VCoMX/database/input
5. Open *'CONFIG.in'*. Opening it with Matlab is warmly suggested since it keeps the formatting.
   a) (Optional) Give a name to the execution. Type it below "Analysis ID name";

---

[2] 10.5281/zenodo.8276838

   b) Add the full path to the folder containing the DICOM RT-Plan files or the file containing the list of plans

REMEMBER

The TCoMX library can handle three different types of input:

- Folder containing all the DICOM RT-PLAN files;
- Folder containing a subfolder for each DICOM RT-PLAN file.
- File containing the list of plans to extract the metrics from.

   c) Add the full path of the results folder where you want save the result. If it does not exist, VCoMX will create it for you
   d) Select if you want to read the list of plans from a .xlsx file (*True*) or not (*False*).
   e) If you chose *True* at point 1d, write the filename.

6. Close *'CONFIG.in'*.

1. Open METRICS.in.
2. You can perform the following operations:
   a) Leave everything as it is and compute all the metrics;
   b) Remove some metric;
   c) Remove a whole subcategory;
   d) Remove a whole category.
3. Save and close.
4. Move to .../UCoMX, type VCoMX on your Matlab console, and press return.

WARNING/3

The following operation are not allowed and will comprimise the correct execution of the routines in the library:

- Remove a category without removing the corresponding subcategories
- Remove a subcategory without removing the corresponding metrics

REMEMBER

The provided input files has been tested many times and their syntax was found to be robust over different Matlab versions and operative systems. Every time you make a change in the metrics to extract, be sure to keep the correct syntax.

D.1.3.1 *Reading the results*

The results of the extraction will be stored within the results folder provided during the Configuration. The following files will be stored:

- dataset.xlsx: this file contains general metadata information about the analyzed plans, the metrics and some basic descriptive statistics about the metrics over the whole sample such as mean, standard deviation, mode, median.
- dataset.mat: this file contains the same information as dataset.xlsx but it is in .mat format.
- CONFIG.in
- METRICS.in
- logfile.log: a summary containing information about the extraction such as the unique id associated by VCoMX and the total number of metrics extracted.
- vmatplan.mat: a matlab structure containing the values of every metric at each CP. This can be used for a more in-depth analysis.

D.2 OVERVIEW AND FUNCTIONING OF TCOMX

D.2.1 *Input files*

The input data for the TCoMX engine is taken from two specific text files: CONFIG.in and METRICS.in. In the GUI-based version, the package automatically generates these files when the user fills in the GUI fields. Conversely, in the Matlab command-line version, the user must manually enter the data into these files. A pre-filled version is provided with the package. The definitions of the different fields are explained in the following subsections.

CONFIG.in

This file is used to set all the necessary input parameters. The overall structure of the file is the following:

CONFIG.in

 # Analysis ID name
(OPTIONAL) Type here a user's define name to assign to the current session

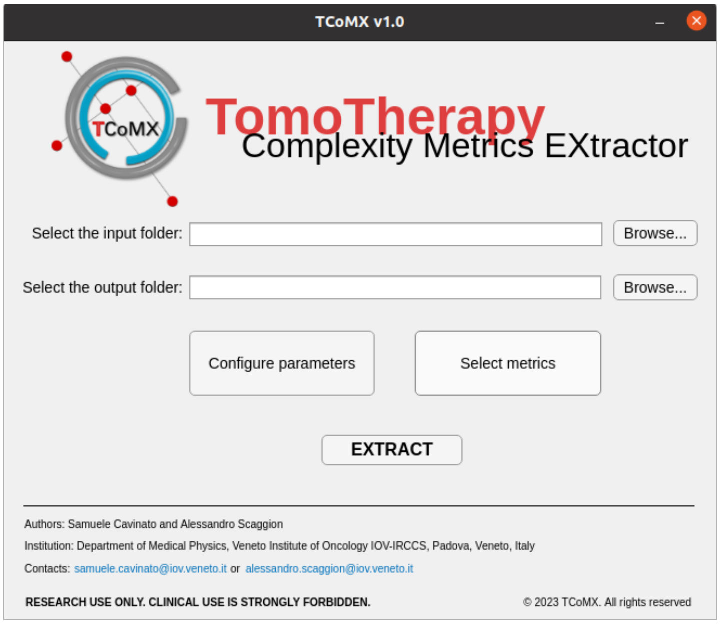
 # Dataset input folder
Type here the full path where your DICOM RTPLAN files or a file containing the list of plans (see below) are stored.

 # Results output folder
Type here the full path where the results of the extraction process will be saved

 # Read from file
True or *False*. If it is set to *true*, the library will read the filenames from a dedicated .xlsx file stored in the the 'Dataset input folder'. The name of the .xlxs file is specified in the 'List file name' field (see below). An example of this file is provided together with the reference dataset.

 # List file name
.xlsx file containing the list of plans. It is used only if 'Read from file' is set to *true*.

In Fig. D.4 and Fig. D.5 the GUI panels used for setting the values of fields of the CONFIG.in file in the GUI-based versions are shown.



Authors: Samuele Cavinato and Alessandro Scaggion

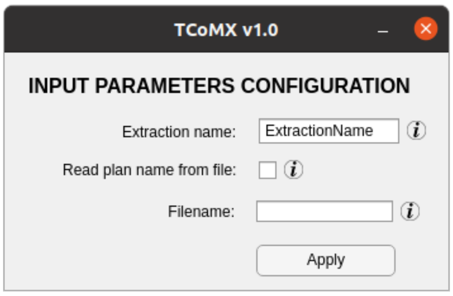
Institution: Department of Medical Physics, Veneto Institute of Oncology IOV-IRCCS, Padova, Veneto, Italy

Contacts: samuele.cavinato@iov.veneto.it or alessandro.scaggion@iov.veneto.it

RESEARCH USE ONLY. CLINICAL USE IS STRONGLY FORBIDDEN.

© 2023 TCoMX. All rights reserved

Figure D.4: Main panel of the TCoMX extraction engine. From this panel, the input and output path can be set.



Extraction name: ExtractionName

Read plan name from file: ☐

Filename:

Figure D.5: Panels for configuration of the additional field of the CONFIG.in fil

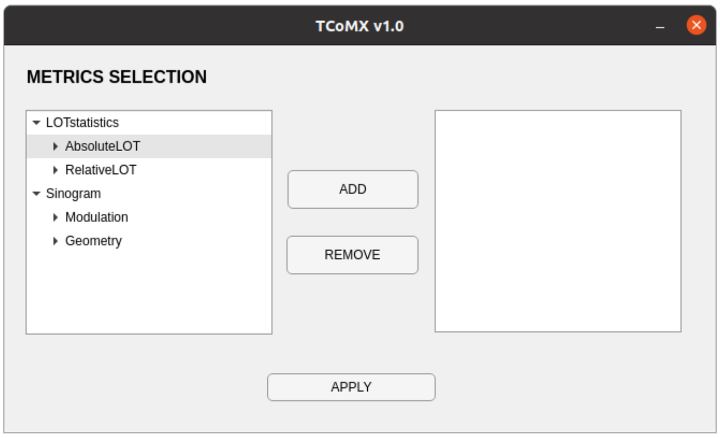


Figure D.6: Panel of the VCoMX interface for the selection of the complexity metrics to extract.

WARNING/1

If CONFIG.in is filled manually (i.e. not using the GUI), be sure of:

- **Leaving** the fields with the # must as they are.
- **Writing** the value of a given field immediately after the corresponding field name as shown in the CONFIG.in box above.

METRICS.in

This file is used to select the metrics to extract and to set their extraction parameters. The general structure of the file is the following:

METRICS.in

```
FIRST CATEGORY NAME:
    FIRST SUBCATEGORY NAME:
        METRIC NAME 1
        METRIC NAME 2
        ...
    SECOND SUBCATEGORY NAME:
        METRIC NAME 1
        ...
SECOND CATEGORY NAME:
    ...
```

The METRICS NAME field can have **three** different input formats, depending on the corresponding definition:

1. **Metrics with no input parameters.** The metric has to be written using the corresponding name as:

   METRIC NAME

2. **Metrics with one input parameter.** The metric has to be reported as:

   METRIC NAME → par1, par2, ..., parN

3. **Metrics with two input parameters.** The metric has to be reported as:

   METRIC NAME → [par11;par12], [par21;par22], ..., [parN1;parN2]

REMEMBER

For cases 2 and 3, a given metric is computed for all the different (pairs of) input parameters. The total number of metrics computed starting from a given metric will be equal to the number of (pairs of) parameters N.

TIPS

- The indentation of the input file can help to visualize the content, but it is not fundamental for a correct execution.
- Punctuation marks **are part of the syntax** and must be used properly. In particular:
  - Each category and subcategory is followed by semicolons ":"
  - No punctuation has to be but after the names of the metrics with no tunable parameters.
  - Do not add a → after the metrics that are not endowed with tunable parameters. The library will not recognize the metric and will not extract it.
  - Tunable parameters needs to be reported after the → and separated by a comma ","
  - Vectors of tunable parameters needs to be written after the → as $[a; b]$, and separated by commas ","

D.2.2 *How to use TCoMX*

Configuration and execution of the stand-alone GUI-based version

The stand-alone version of UCoMX can be downloaded from:

https://doi.org/10.5281/zenodo.8276838.

This version was created for Windows 10 a requires the Matlab Run-Time R2022b for its execution (available at this link[3]). If stand-alone version for a different operative system is needed, please email us at samuele.cavinato@iov.veneto.it or alessandro.scaggion@iov.veneto.it.

Once you have UCoMX available on your PC and you have the Matlab Runtime installed, please follow these steps for the execution:

1. Move to the UCoMX download folder;
2. Double click on UCoMX;
3. Select TCoMX.
4. Select the input folder where the DICOM RT-Plan files are located;
5. Select the output folder where the results will be stored;
6. (Optional) Click on 'Configure parameters'.

---

[3] https://ssd.mathworks.com/supportfiles/downloads/R2022a/Release/7/deployment_files/installer/complete/win64/MATLAB_Runtime_R2022a_Update_7_win64.zip

7. From the new panel you can:
    a) Give a name to the execution.
    b) Select if you want to read the list of plans from a .xlsx file by enabling the checkbox. If so, write the filename. In this case, the file must be place in the input folder indicated in the previous panel.

REMEMBER

The TCoMX library can handle three different types of input:

i. Folder containing all the DICOM RT-Plan files;

ii. Folder containing a subfolder for each DICOM RT-Plan file;

iii. File containing the list of plans to extract the metrics from.

8. Close the 'Configure parameters' panel;

9. Click on 'Select metrics', a new panel appears.The metrics implemented in the TCoMX engine are listed on the left box. The metrics that will be extracted are listed in the right You can perform the following operations:
    a) Add metrics to the right box. To do it, select a metric on the left box and click 'ADD'.
    b) Remove a metric from the left box. To do it, select a metrics a click 'REMOVE'.
    c) Some metrics have configurable parameters (See. Sec.**??**) which can be configured by right-clicking on the metrics before adding it.

10. Once the metrics have been selected, close the panel.

11. From the main panel, click 'EXTRACT' to start the execution.

Configuration and execution of the `MATLAB`® *GUI-based version*

The `MATLAB`® GUI version of UCoMX can be downloaded from:

https://doi.org/10.5281/zenodo.8276838.

The execution of this version requires a valid Matlab licence installed. The use of this version is very similar to that of the stand-alone versione, with the first three steps replaced by the following ones:

1. Open Matlab

2. Move to .../UCoMX
3. Add .../UCoMX and its subfolders to the Maltab path;
4. Type 'UCoMX' on the command line;
5. Continue from point 4 of the instructions for the stand-alone version.

D.2.2.1 *Configuration and execution of the* `MATLAB`® *command-line version*

1. Move to .../UCoMX/TCoMX/input/
2. Open *'CONFIG.in'*. Opening it with Matlab is warmly suggested since it keeps the formatting.
    a) (Optional) Give a name to the execution. Type it below "Analysis ID name";
    b) Add the full path to the folder containing the DICOM RT-Plan files or the file containing the list of plans.

REMEMBER

The TCoMX library can handle three different types of input:

- Folder containing all the DICOM RT-Plan files;
- Folder containing a subfolder for each DICOM RT-Plan file.
- File containing the list of plans to extract the metrics from.

    c) Add the full path of the results folder where you want save the result. If it does not exist, TCoMX will create it for you
    d) Select if you want to read the list of plans from a .xlsx file (*True*) or not (*False*).
    e) If you chose *True* at point 1d, write the filename.
3. Close *'CONFIG.in'*.
4. Open METRICS.in.
5. You can perform the following operations:
    a) Leave everything as it is and compute all the metrics;
    b) Remove some metric;
    c) Remove a whole subcategory;
    d) Remove a whole category.
6. Save and close.

7. Move to …/UCoMX/, type TCoMX on your Matlab console, and press return.

WARNING/2

The following operation are not allowed and will compromise the correct execution of the routines in the library:

- Remove a category without removing the corresponding subcategories
- Remove a subcategory without removing the corresponding metrics

WARNING/3

- In the current version, the metrics belonging to the subcategory *delivery* are always computed and do not need to be added to the *'METRICS.in'* file.
- The computation of the TTDF requires the dose per fraction. This information is not contained in the DICOM RT-PLAN files create by Precision. In this case, you need to specify the fraction dose for each plan in a file called 'DoseFraction.xlsx' which has to be stored in the 'Dataset input folder'. An example of this file is provided together with the reference dataset.

REMEMBER

The provided input files has been tested many times and their syntax was found to be robust over different Matlab versions and operative systems. Every time you make a change in the metrics to extract, be sure to keep the correct syntax.

D.2.3 *Reading the results*

The results of the extraction will be stored within the results folder provided during the Configuration. The following files will be stored:

- dataset.xlsx: this file contains general metadata information about the analyzed plans, the metrics and some basic descriptive statistics about the metrics over the whole sample such as mean, standard deviation, mode, median.
- dataset.mat: this file contains the same information as dataset.xlsx but it is in .mat format.
- CONFIG.in

- METRICS.in
- logfile.log: a summary containing information about the extraction such as the unique id associated by VCoMX and the total number of metrics extracted.

BIBLIOGRAPHY

[1] Reid F. Thompson et al. "Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation?" In: *Radiotherapy and Oncology* 129.3 (2018), pp. 421–426. ISSN: 0167-8140. DOI: https://doi.org/10.1016/j.radonc.2018.05.030.

[2] Kristy K. Brock. "Adaptive Radiotherapy: Moving Into the Future." In: *Seminars in Radiation Oncology* 29.3 (2019). Adaptive Radiotherapy and Automation, pp. 181–184. ISSN: 1053-4296. DOI: https://doi.org/10.1016/j.semradonc.2019.02.011.

[3] Lukas Schaub et al. "Particle therapy in the future of precision therapy." In: *The British Journal of Radiology* 93.1114 (2020). PMID: 32795176, p. 20200183. DOI: 10.1259/bjr.20200183.

[4] John Ng et al. "MRI-LINAC: A transformative technology in radiation oncology." In: *Frontiers in Oncology* 13 (2023). ISSN: 2234-943X. DOI: 10.3389/fonc.2023.1117874.

[5] Marco Schwarz et al. "Treatment planning for Flash radiotherapy: General aspects and applications to proton beams." In: *Medical Physics* 49.4 (2022), pp. 2861–2874. DOI: https://doi.org/10.1002/mp.15579.

[6] Zengfu Zhang et al. "Radiotherapy combined with immunotherapy: the dawn of cancer treatment." In: *Signal Transduction and Targeted Therapy* 7.1 (2022), p. 258. ISSN: 2059-3635. DOI: 10.1038/s41392-022-01102-y.

[7] Yifan Wang et al. "Combining Immunotherapy and Radiotherapy for Cancer Treatment: Current Challenges and Future Directions." In: *Frontiers in Pharmacology* 9 (2018). ISSN: 1663-9812. DOI: 10.3389/fphar.2018.00185.

[8] Yuting Ma et al. "Chemotherapy and radiotherapy: Cryptic anticancer vaccines." In: *Seminars in Immunology* 22.3 (2010). Cancer Vaccines: The state of the art, pp. 113 –124. ISSN: 1044-5323. DOI: https://doi.org/10.1016/j.smim.2010.03.001.

[9] IAEA. *The IAEA Directory of Radiotherapy Centres (DIRAC)*. (accessed on 08/2023). URL: https:/dirac.iaea.org.

[10] Cyrus Chargari et al. "Brachytherapy: An overview for clinicians." In: *CA: A Cancer Journal for Clinicians* 69.5 (2019), pp. 386–401. DOI: 10.3322/caac.21578.

[11] J. Staffurth. "A Review of the Clinical Evidence for Intensity-modulated Radiotherapy." In: *Clinical Oncology* 22.8 (2010). The Technical Aspects of Radiotherapy, pp. 643–657. ISSN: 0936-6555. DOI: https://doi.org/10.1016/j.clon.2010.06.013.

[12] V.S. Khoo. "Radiotherapeutic Techniques for Prostate Cancer, Dose Escalation and Brachytherapy." In: *Clinical Oncology* 17.7 (2005). Urological Oncology, pp. 560–571. ISSN: 0936-6555. DOI: https://doi.org/10.1016/j.clon.2005.07.006.

[13] K. Odrazka et al. "Clinical results of intensity-modulated radiation therapy (IMRT) for tumors of the head and neck region." In: *Neoplasma* 52.2 (2005). Cited by: 11, 85 – 94.

[14] S Webb. "The physical basis of IMRT and inverse planning." In: *The British Journal of Radiology* 76.910 (2003). PMID: 14512327, pp. 678–689. DOI: 10.1259/bjr/65676879.

[15] Samuele Cavinato et al. "Optimizing radiotherapy plans for cancer treatment with Tensor Networks." In: *Physics in Medicine & Biology* 66.12 (2021), p. 125015. DOI: 10.1088/1361-6560/ac01f2.

[16] Jalil ur Rehman et al. "Intensity modulated radiation therapy: A review of current practice and future outlooks." In: *Journal of Radiation Research and Applied Sciences* 11.4 (2018), pp. 361–367. ISSN: 1687-8507. DOI: https://doi.org/10.1016/j.jrras.2018.07.006.

[17] D J Convery et al. "The generation of intensity-modulated fields for conformal radiotherapy by dynamic collimation." In: *Physics in Medicine & Biology* 37.6 (1992), p. 1359. DOI: 10.1088/0031-9155/37/6/012.

[18] Karl Otto. "Volumetric modulated arc therapy: IMRT in a single gantry arc." In: *Medical Physics* 35.1 (2008), pp. 310–317. DOI: https://doi.org/10.1118/1.2818738.

[19] James S. Welsh et al. "Helical Tomotherapy: An Innovative Technology and Approach to Radiation Therapy." In: *Technology in Cancer Research & Treatment* 1.4 (2002). PMID: 12625791, pp. 311–316. DOI: 10.1177/153303460200100413.

[20] Claudio Fiorino et al. "Significant improvement in normal tissue sparing and target coverage for head and neck cancer by means of helical tomotherapy." In: *Radiotherapy and oncology* 78.3 (2006), pp. 276–282.

[21] Rufus A. Scrimger et al. "Reduction in radiation dose to lung and other normal tissues using helical tomotherapy to treat lung cancer, in comparison to conventional field arrangements." In: *American Journal of Clinical Oncology: Cancer Clinical Trials* 26.1 (2003). Cited by: 70; All Open Access, Bronze Open Access, 70 – 78. DOI: 10.1097/00000421-200302000-00014.

[22] Hidefumi Aoyama et al. “Integral radiation dose to normal structures with conformal external beam radiation.” In: *International Journal of Radiation Oncology*Biology*Physics* 64.3 (2006), pp. 962–967. ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2005.11.005.

[23] Moyed Miften et al. “Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218.” In: *Medical Physics* 45.4 (2018), e53–e83. DOI: https://doi.org/10.1002/mp.12810.

[24] M. Hussein et al. “Challenges in calculation of the gamma index in radiotherapy – Towards good practice.” In: *Physica Medica* 36 (2017), pp. 1–11. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2017.03.001.

[25] Josef Bogner Ludwigand Scherer et al. “Verification of IMRT: Techniques and Problems.” In: *Strahlentherapie und Onkologie* 180.6 (2004), pp. 340–350. ISSN: 1439-099X. DOI: 10.1007/s00066-004-1219-0.

[26] J.Van Dyk et al. “Commissioning and quality assurance of treatment planning computers.” In: *International Journal of Radiation Oncology*Biology*Physics* 26.2 (1993), pp. 261–273. ISSN: 0360-3016. DOI: https://doi.org/10.1016/0360-3016(93)90206-B.

[27] William B. Harms Sr. et al. “A software tool for the quantitative evaluation of 3D dose calculation algorithms.” In: *Medical Physics* 25.10 (1998), pp. 1830–1836. DOI: https://doi.org/10.1118/1.598363.

[28] Daniel A. Low et al. “A technique for the quantitative evaluation of dose distributions.” In: *Medical Physics* 25.5 (1998), pp. 656–661. DOI: https://doi.org/10.1118/1.598248.

[29] Daniel A. Low et al. “Evaluation of the gamma dose distribution comparison method.” In: *Medical Physics* 30.9 (2003), pp. 2455–2464. DOI: https://doi.org/10.1118/1.1598711.

[30] Marco Fusella et al. “Analysis of clinical patient-specific pretreatment quality assurance with the new helical tomotherapy platform, following the AAPM TG-218 report.” In: *Radiation Oncology* 16 (2021), p. 226. DOI: 10.1186/s13014-021-01952-w.

[31] Mohammad Hussein et al. “A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems.” In: *Radiotherapy and Oncology* 109.3 (2013), pp. 370–376. ISSN: 0167-8140. DOI: https://doi.org/10.1016/j.radonc.2013.08.048.

[32] Alessandro Scaggion et al. “Limiting treatment plan complexity by applying a novel commercial tool.” In: *Journal of Applied Clinical Medical Physics* 21.8 (2020), pp. 27–34. DOI: https://doi.org/10.1002/acm2.12908.

[33] Sophie Chiavassa et al. “Complexity metrics for IMRT and VMAT plans: a review of current literature and applications.” In: *The British Journal of Radiology* 92.1102 (2019). PMID: 31295002, p. 20190270. DOI: 10.1259/bjr.20190270.

[34] Mikaël Antoine et al. “Use of metrics to quantify IMRT and VMAT treatment plan complexity: A systematic review and perspectives.” In: *Physica Medica* 64 (2019), pp. 98–108. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2019.05.024.

[35] Radhe Mohan et al. “The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy.” In: *Medical Physics* 27.6 (2000), pp. 1226–1237. DOI: https://doi.org/10.1118/1.599000.

[36] S Webb. “Use of a quantitative index of beam modulation to characterize dose conformality: illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy.” In: *Physics in Medicine & Biology* 48.14 (2003), p. 2051. DOI: 10.1088/0031-9155/48/14/301.

[37] P. Zygmanski et al. “Method of identifying dynamic multileaf collimator irradiation that is highly sensitive to a systematic MLC calibration error.” In: *Medical Physics* 28.11 (2001), pp. 2220–2226. DOI: https://doi.org/10.1118/1.1408284.

[38] Jorge Llacer et al. “Comparative behaviour of the Dynamically Penalized Likelihood algorithm in inverse radiation therapy planning.” In: *Physics in Medicine & Biology* 46.10 (2001), p. 2637. DOI: 10.1088/0031-9155/46/10/309.

[39] Martha M. Coselmon et al. “Improving IMRT delivery efficiency using intensity limits during inverse planning.” In: *Medical Physics* 32.5 (2005), pp. 1234–1245. DOI: https://doi.org/10.1118/1.1895545.

[40] S B Crowe et al. “Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results.” In: *Physics in Medicine and Biology* 60.6 (2015), p. 2587. DOI: 10.1088/0031-9155/60/6/2587.

[41] Weiliang Du et al. “Quantification of beam complexity in intensity-modulated radiation therapy treatment plans.” In: *Medical Physics* 41.2 (2014), p. 021716. DOI: https://doi.org/10.1118/1.4861821.

[42] Nicolini Giorgia et al. “What is an acceptably smoothed fluence? Dosimetric and delivery considerations for dynamic sliding window IMRT.” In: *Radiation Oncology* 2.1 (2007), p. 42. ISSN: 1748-717X. DOI: 10.1186/1748-717X-2-42.

[43] Marcel Nauta et al. “Fractal analysis for assessing the level of modulation of IMRT fields.” In: *Medical Physics* 38.10 (2011), pp. 5385–5393. DOI: `https://doi.org/10.1118/1.3633912`.

[44] So-Yeon Park et al. “Texture analysis on the edge-enhanced fluence of VMAT.” In: *Radiation Oncology* 10.1 (2015), p. 74. ISSN: 1748-717X. DOI: `10.1186/s13014-015-0382-z`.

[45] Thomas LoSasso et al. “Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy.” In: *Medical Physics* 25.10 (1998), pp. 1919–1927. DOI: `https://doi.org/10.1118/1.598381`.

[46] James R. Kerns et al. “A multi-institution evaluation of MLC log files and performance in IMRT delivery.” In: *Radiation Oncology* 9.1 (2014), p. 176. ISSN: 1748-717X. DOI: `10.1186/1748-717X-9-176`.

[47] Laura Masi et al. “Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy.” In: *Medical Physics* 40.7 (2013), p. 071718. DOI: `https://doi.org/10.1118/1.4810969`.

[48] Andrea L. McNiven et al. “A new metric for assessing IMRT modulation complexity and plan deliverability.” In: *Medical Physics* 37.2 (2010), pp. 505–515. DOI: `https://doi.org/10.1118/1.3276775`.

[49] H. Miura et al. “Predicting Delivery Error Using a DICOM-RT Plan for Volumetric Modulated Arc Therapy.” In: *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology* 3 (2014), pp. 82–87. DOI: `10.4236/ijmpcero.2014.32013`.

[50] Lanxiao Shen et al. “Multidimensional correlation among plan complexity, quality and deliverability parameters for volumetric-modulated arc therapy using canonical correlation analysis.” In: *Journal of Radiation Research* 59.2 (Feb. 2018), pp. 207–215. ISSN: 0449-3060. DOI: `10.1093/jrr/rrx100`.

[51] Jong Min Park et al. “Modulation indices for volumetric modulated arc therapy.” In: *Physics in Medicine and Biology* 59.23 (2014), pp. 7315–7340. DOI: `10.1088/0031-9155/59/23/7315`.

[52] J M Park et al. “The effect of MLC speed and acceleration on the plan delivery accuracy of VMAT.” In: *The British Journal of Radiology* 88.1049 (2015). PMID: 25734490, p. 20140698. DOI: `10.1259/bjr.20140698`.

[53] Lei Xing et al. “Computer verification of fluence map for intensity modulated radiation therapy.” In: *Medical Physics* 27.9 (2000), pp. 2084–2092. DOI: `https://doi.org/10.1118/1.1289374`.

[54] Christopher J. Anker et al. “Evaluation of fluence-smoothing feature for three IMRT planning systems.” In: *Journal of Applied Clinical Medical Physics* 11.2 (2010), pp. 33–61. DOI: `https://doi.org/10.1120/jacmp.v11i2.3035`.

[55] Iori Sumida et al. “Organ-specific modulation complexity score for the evaluation of dose delivery.” In: *Journal of Radiation Research* 58.5 (2017), pp. 675–684. ISSN: 0449-3060. DOI: `10.1093/jrr/rrw129`.

[56] Christina E. Agnew et al. “Correlation of phantom-based and log file patient-specific QA with complexity scores for VMAT.” In: *Journal of Applied Clinical Medical Physics* 15.6 (2014), pp. 204–216. DOI: `https://doi.org/10.1120/jacmp.v15i6.4994`.

[57] T Kairn et al. “Predicting the likelihood of QA failure using treatment plan accuracy metrics.” In: *Journal of Physics: Conference Series* 489.1 (2014), p. 012051. DOI: `10.1088/1742-6596/489/1/012051`.

[58] Ben Heijmen et al. “Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy – An international multi-center validation for prostate cancer.” In: *Radiotherapy and Oncology* 128.2 (2018), pp. 343–348. ISSN: 0167-8140. DOI: `https://doi.org/10.1016/j.radonc.2018.06.023`.

[59] Joel N K Carlson et al. “A machine learning approach to the accurate prediction of multi-leaf collimator positional errors.” In: *Physics in Medicine & Biology* 61.6 (2016), p. 2514. DOI: `10.1088/0031-9155/61/6/2514`.

[60] Julia Götstedt et al. “Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings.” In: *Medical Physics* 42.7 (2015), pp. 3911–3921. DOI: `https://doi.org/10.1118/1.4921733`.

[61] Kelly C. Younge et al. “Penalization of aperture complexity in inversely planned volumetric modulated arc therapy.” In: *Medical Physics* 39.11 (2012), pp. 7160–7170. DOI: `https://doi.org/10.1118/1.4762566`.

[62] M. Mathot et al. “6 VMAT complexity metrics can reduce patient QA workload.” In: *Physica Medica* 56 (2018). Abstracts of the 57èmes Journées Scientifiques de la Société Française de Physique Médicale, pp. 3–4. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2018.09.019`.

[63] Weiguang Yao et al. “Determining the optimal dosimetric leaf gap setting for rounded leaf-end multileaf collimator systems by simple test fields.” In: *Journal of Applied Clinical Medical Physics* 16.4 (2015), pp. 65–77. DOI: `https://doi.org/10.1120/jacmp.v16i4.5321`.

[64] Cecile J. A. Wolfs et al. "Should dose from small fields be limited for dose verification procedures?: uncertainty versus small field dose in VMAT treatments." English. In: *Physics in Medicine and Biology* 63.20 (Oct. 2018). ISSN: 0031-9155. DOI: `10.1088/1361-6560/aae338`.

[65] Robert Boyd et al. "Determining efficient helical IMRT modulation factor from the MLC leaf-open time distribution on precision treatment planning system." In: *Journal of Applied Clinical Medical Physics* 20.5 (2019), pp. 64–74. DOI: `https://doi.org/10.1002/acm2.12581`.

[66] David C. Westerly et al. "Treatment Planning to Improve Delivery Accuracy and Patient Throughput in Helical Tomotherapy." In: *International Journal of Radiation Oncology*Biology*Physics* 74.4 (2009), pp. 1290–1297. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/j.ijrobp.2009.02.004`.

[67] Steffen Lissner et al. "A method for testing the performance and the accuracy of the binary MLC used in helical tomotherapy." In: *Zeitschrift für Medizinische Physik* 23.2 (2013), pp. 153–161. ISSN: 0939-3889. DOI: `https://doi.org/10.1016/j.zemedi.2012.08.001`.

[68] Diana Binny et al. "Effects of changing modulation and pitch parameters on tomotherapy delivery quality assurance plans." In: *Journal of Applied Clinical Medical Physics* 16.5 (2015), pp. 87–105. DOI: `https://doi.org/10.1120/jacmp.v16i5.5282`.

[69] Tania Santos et al. "On the complexity of helical tomotherapy treatment plans." In: *Journal of Applied Clinical Medical Physics* 21.7 (2020), pp. 107–118. DOI: `https://doi.org/10.1002/acm2.12895`.

[70] Landon S. Wootton et al. "Error Detection in Intensity-Modulated Radiation Therapy Quality Assurance Using Radiomic Analysis of Gamma Distributions." In: *International Journal of Radiation Oncology*Biology*Physics* 102.1 (2018), pp. 219–228. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/j.ijrobp.2018.05.033`.

[71] Matthew J. Nyflot et al. "Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks." In: *Medical Physics* 46.2 (2019), pp. 456–464. DOI: `https://doi.org/10.1002/mp.13338`.

[72] Dal A Granville et al. "Predicting VMAT patient-specific QA results using a support vector classifier trained on treatment plan characteristics and linac QC metrics." In: *Physics in Medicine & Biology* 64.9 (2019), p. 095017. DOI: `10.1088/1361-6560/ab142e`.

[73] Chaoqiong Ma et al. “The structural similarity index for IMRT quality assurance: radiomics-based error classification.” In: *Medical Physics* 48.1 (2021), pp. 80–93. DOI: `https://doi.org/10.1002/mp.14559`.

[74] Madoka Sakai et al. “Detecting MLC modeling errors using radiomics-based machine learning in patient-specific QA with an EPID for intensity-modulated radiation therapy.” In: *Medical Physics* 48.3 (2021), pp. 991–1002. DOI: `https://doi.org/10.1002/mp.14699`.

[75] Nicholas J. Potter et al. “Error detection and classification in patient-specific IMRT QA with dual neural networks.” In: *Medical Physics* 47.10 (2020), pp. 4711–4720. DOI: `https://doi.org/10.1002/mp.14416`.

[76] Jeremy Braun et al. “Machine learning-generated decision boundaries for prediction and exploration of patient-specific quality assurance failures in stereotactic radiosurgery plans.” In: *Medical Physics* 49.3 (2022), pp. 1955–1963. DOI: `https://doi.org/10.1002/mp.15454`.

[77] Caroline Noblet et al. “Implementation of volumetric-modulated arc therapy for locally advanced breast cancer patients: Dosimetric comparison with deliverability consideration of planning techniques and predictions of patient-specific QA results via supervised machine learning.” In: *Physica Medica* 96 (2022), pp. 18–31. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2022.02.015`.

[78] Sangutid Thongsawad et al. “Predicting gamma evaluation results of patient-specific head and neck volumetric-modulated arc therapy quality assurance based on multileaf collimator patterns and fluence map features: A feasibility study.” In: *Journal of Applied Clinical Medical Physics* 23.7 (2022), e13622. DOI: `https://doi.org/10.1002/acm2.13622`.

[79] Cecile J.A. Wolfs et al. “What is the optimal input information for deep learning-based pre-treatment error identification in radiotherapy?” In: *Physics and Imaging in Radiation Oncology* 24 (2022), pp. 14–20. ISSN: 2405-6316. DOI: `https://doi.org/10.1016/j.phro.2022.08.007`.

[80] Savino Cilla et al. “Prediction and classification of VMAT dosimetric accuracy using plan complexity and log-files analysis.” In: *Physica Medica* 103 (2022), pp. 76–88. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2022.10.004`.

[81] Noémie Moreau et al. “Deep Hybrid Learning Prediction of Patient-Specific Quality Assurance in Radiotherapy: Implementation in Clinical Routine.” In: *Diagnostics* 13.5 (2023). ISSN: 2075-4418. DOI: `10.3390/diagnostics13050943`.

[82] Paulo Quintero et al. "Evaluation of the dataset quality in gamma passing rate predictions using machine learning methods." In: *The British Journal of Radiology* 96.1147 (2023). PMID: 37129359, p. 20220302. DOI: 10.1259/bjr.20220302.

[83] Yuto Kimura et al. "Development of a deep learning-based error detection system without error dose maps in the patient-specific quality assurance of volumetric modulated arc therapy." In: *Journal of Radiation Research* 64.4 (May 2023), pp. 728–737. ISSN: 1349-9157. DOI: 10.1093/jrr/rrad028.

[84] Ce Han et al. "Integrating plan complexity and dosiomics features with deep learning in patient-specific quality assurance for volumetric modulated arc therapy." In: *Radiation Oncology* 18.1 (2023), p. 116. ISSN: 1748-717X. DOI: 10.1186/s13014-023-02311-7.

[85] G. Valdes et al. "A mathematical framework for virtual IMRT QA using machine learning." In: *Medical Physics* 43.7 (2016), pp. 4323–4334. DOI: https://doi.org/10.1118/1.4953835.

[86] Gilmer Valdes et al. "IMRT QA using machine learning: A multi-institutional validation." In: *Journal of Applied Clinical Medical Physics* 18.5 (2017), pp. 279–284. DOI: https://doi.org/10.1002/acm2.12161.

[87] Yannet Interian et al. "Deep nets vs expert designed features in medical physics: An IMRT QA case study." In: *Medical Physics* 45.6 (2018), pp. 2672–2680. DOI: https://doi.org/10.1002/mp.12890.

[88] Tomohiro Ono et al. "Prediction of dosimetric accuracy for VMAT plans using plan complexity parameters via machine learning." In: *Medical Physics* 46.9 (2019), pp. 3823–3832. DOI: https://doi.org/10.1002/mp.13669.

[89] Dao Lam et al. "Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning." In: *Medical Physics* 46.10 (2019), pp. 4666–4675. DOI: https://doi.org/10.1002/mp.13752.

[90] Phillip D.H. Wall et al. "Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning." In: *Informatics in Medicine Unlocked* 18 (2020), p. 100292. ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2020.100292.

[91] Jéssica Caroline Lizar et al. "Patient-specific IMRT QA verification using machine learning and gamma radiomics." In: *Physica Medica* 82 (2021), pp. 100–108. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2021.01.071.

[92] Ting Hu et al. “Deep Multimodal Neural Network Based on Data-Feature Fusion for Patient-Specific Quality Assurance.” In: *International Journal of Neural Systems* 32.01 (2022). PMID: 34895106, p. 2150055. DOI: `10.1142/S0129065721500556`.

[93] Yao Hao et al. “Improvement of IMRT QA prediction using imaging-based neural architecture search.” In: *Medical Physics* 49.8 (2022), pp. 5236–5243. DOI: `https://doi.org/10.1002/mp.15694`.

[94] Heling Zhu et al. “Patient-specific quality assurance prediction models based on machine learning for novel dual-layered MLC linac.” In: *Medical Physics* 50.2 (2023), pp. 1205–1214. DOI: `https://doi.org/10.1002/mp.16091`.

[95] Elahheh Salari et al. “Using machine learning to predict gamma passing rate in volumetric-modulated arc therapy treatment plans.” In: *Journal of Applied Clinical Medical Physics* 24.2 (2023), e13824. DOI: `https://doi.org/10.1002/acm2.13824`.

[96] Kah Seng Lew et al. “Prediction of portal dosimetry quality assurance results using log files-derived errors and machine learning techniques.” In: *Frontiers in Oncology* 12 (2023). ISSN: 2234-943X. DOI: `10.3389/fonc.2022.1096838`.

[97] R Levin et al. “Patient-specific quality assurance failure prediction with deep tabular models.” In: *Biomedical Physics & Engineering Express* 9.4 (2023), p. 045007. DOI: `10.1088/2057-1976/acd255`.

[98] Jiaqi Li et al. “Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy.” In: *International Journal of Radiation Oncology*Biology*Physics* 105.4 (2019), pp. 893–902. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/j.ijrobp.2019.07.049`.

[99] Hideaki Hirashima et al. “Improvement of prediction and classification performance for gamma passing rate by using plan complexity and dosiomics features.” In: *Radiotherapy and Oncology* 153 (2020). Physics Special Issue: ESTRO Physics Research Workshops on Science in Development, pp. 250–257. ISSN: 0167-8140. DOI: `https://doi.org/10.1016/j.radonc.2020.07.031`.

[100] Le Wang et al. “Multi-task autoencoder based classification-regression model for patient-specific VMAT QA.” In: *Physics in Medicine & Biology* 65.23 (2020), p. 235023. DOI: `10.1088/1361-6560/abb31c`.

[101] Phillip D.H. Wall et al. “Prospective Clinical Validation of Virtual Patient-Specific Quality Assurance of Volumetric Modulated Arc Therapy Radiation Therapy Plans.” In: *International Journal of Radiation Oncology*Biology*Physics* 113.5 (2022),

pp. 1091–1102. ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2022.04.040.

[102] Samuele Cavinato et al. “Prediction models as decision-support tools for virtual patient-specific quality assurance of helical tomotherapy plans.” In: *Physics and Imaging in Radiation Oncology* 26 (2023), p. 100435. ISSN: 2405-6316. DOI: https://doi.org/10.1016/j.phro.2023.100435.

[103] Shizhen Bin et al. “Study of the prediction of gamma passing rate in dosimetric verification of intensity-modulated radiotherapy using machine learning models based on plan complexity.” In: *Frontiers in Oncology* 13 (2023). ISSN: 2234-943X. DOI: 10.3389/fonc.2023.1094927.

[104] Seied Rabie Mahdavi et al. “Use of artificial neural network for pretreatment verification of intensity modulation radiation therapy fields.” In: *The British Journal of Radiology* 92.1102 (2019). PMID: 31317765, p. 20190355. DOI: 10.1259/bjr.20190355.

[105] Phillip D.H. Wall et al. “Quality assurance-based optimization (QAO): Towards improving patient-specific quality assurance in volumetric modulated arc therapy plans using machine learning.” In: *Physica Medica* 87 (2021), pp. 136–143. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2021.03.017.

[106] Mengyu Jia et al. “Deep learning-enabled EPID-based 3D dosimetry for dose verification of step-and-shoot radiotherapy.” In: *Medical Physics* 48.11 (2021), pp. 6810–6819. DOI: https://doi.org/10.1002/mp.15218.

[107] Changfei Gong et al. “Efficient dose–volume histogram–based pretreatment patient-specific quality assurance methodology with combined deep learning and machine learning models for volumetric modulated arc radiotherapy.” In: *Medical Physics* 49.12 (2022), pp. 7779–7790. DOI: https://doi.org/10.1002/mp.16010.

[108] Takaaki Matsuura et al. “A synthesized gamma distribution-based patient-specific VMAT QA using a generative adversarial network.” In: *Medical Physics* 50.4 (2023), pp. 2488–2498. DOI: https://doi.org/10.1002/mp.16210.

[109] SA Yoganathan et al. “Virtual pretreatment patient-specific quality assurance of volumetric modulated arc therapy using deep learning.” In: *Medical Physics* n/a.n/a (2023), pp. 1–13. DOI: https://doi.org/10.1002/mp.16567.

[110] Accuray. “Physics essential guide.” In: (2018).

[111] Dinesh Kumar Mynampati et al. “Application of AAPM TG 119 to volumetric arc therapy (VMAT).” In: *Journal of Applied Clinical Medical Physics* 13.5 (2012), pp. 108–116. DOI: https://doi.org/10.1120/jacmp.v13i5.3382.

[112] Victor Hernandez et al. “What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans.” In: *Radiotherapy and Oncology* 153 (2020). Physics Special Issue: ESTRO Physics Research Workshops on Science in Development, pp. 26–33. ISSN: 0167-8140. DOI: `https://doi.org/10.1016/j.radonc.2020.09.038`.

[113] Laura Patricia Kaplan et al. “Plan quality assessment in clinical practice: Results of the 2020 ESTRO survey on plan complexity and robustness.” In: *Radiotherapy and Oncology* 173 (2022), pp. 254–261. ISSN: 0167-8140. DOI: `10.1016/j.radonc.2022.06.005`.

[114] Madoka Sakai et al. “Detecting MLC modeling errors using radiomics-based machine learning in patient-specific QA with an EPID for intensity-modulated radiation therapy.” In: *Medical Physics* 48.3 (2021), pp. 991–1002. DOI: `https://doi.org/10.1002/mp.14699`.

[115] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping.” In: *Radiology* 295.2 (2020). PMID: 32154773, pp. 328–338. DOI: `10.1148/radiol.2020191145`.

[116] Andrea Bettinelli et al. “Technical Note: An IBEX adaption toward image biomarker standardization.” In: *Medical Physics* 47.3 (2020), pp. 1167–1173. DOI: `https://doi.org/10.1002/mp.13956`.

[117] Andrea Bettinelli et al. “A Novel Benchmarking Approach to Assess the Agreement among Radiomic Tools.” In: *Radiology* 303.3 (2022). PMID: 35230182, pp. 533–541. DOI: `10.1148/radiol.211604`.

[118] Samuele Cavinato et al. “Quantitative assessment of helical tomotherapy plans complexity.” In: *Journal of Applied Clinical Medical Physics* 24.1 (2023), e13781. DOI: `https://doi.org/10.1002/acm2.13781`.

[119] David Sevillano et al. “Measurement and correction of leaf open times in helical tomotherapy.” In: *Medical Physics* 39.11 (2012), pp. 6972–6980. DOI: `https://doi.org/10.1118/1.4762565`.

[120] Quan Chen et al. “Validation of GPU based TomoTherapy dose calculation engine.” In: *Medical Physics* 39.4 (2012), pp. 1877–1886. DOI: `https://doi.org/10.1118/1.3693057`.

[121] Benjamin E. Nelms et al. “Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors.” In: *Medical Physics* 38.2 (2011), pp. 1037–1044. DOI: `https://doi.org/10.1118/1.3544657`.

[122] Benjamin E. Nelms et al. "Evaluating IMRT and VMAT dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels." In: *Medical Physics* 40.11 (2013), p. 111722. DOI: https://doi.org/10.1118/1.4826166.

[123] Sara Bresciani et al. "A pre-treatment quality assurance survey on 384 patients treated with helical intensity-modulated radiotherapy." In: *Radiotherapy and Oncology* 118.3 (2016), pp. 574–576. ISSN: 0167-8140. DOI: https://doi.org/10.1016/j.radonc.2015.12.021.

[124] Samuele Cavinato et al. *TCoMX: Tomotherapy Complexity Metrics EXtractor*. 2021.

[125] Rebecca M. Seibert et al. "Verification of helical tomotherapy delivery using autoassociative kernel regressiona)." In: *Medical Physics* 34.8 (2007), pp. 3249–3262. DOI: https://doi.org/10.1118/1.2754059.

[126] Inhwan Jason Yeo et al. "Dose reconstruction for intensity-modulated radiation therapy using a non-iterative method and portal dose image." In: *Physics in Medicine & Biology* 54.17 (2009), p. 5223. DOI: 10.1088/0031-9155/54/17/010.

[127] John Balog et al. "Helical Tomotherapy Quality Assurance." In: *International Journal of Radiation Oncology*Biology*Physics* 71.1, Supplement (2008). Quality Assurance for Radiation Therapy, S113–S117. ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2007.10.001.

[128] J. M. Kapatoes et al. "A feasible method for clinical delivery verification and dose reconstruction in tomotherapy." In: *Medical Physics* 28.4 (2001), pp. 528–542. DOI: https://doi.org/10.1118/1.1352579.

[129] *XGBoost: A Scalable Tree Boosting System*. New York, NY, USA: Assiciation for Computing Machinery, 2016, pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785.

[130] David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambrige University Press, 2003.

[131] *Python API Reference - xgboost 2.0.0-dev documentation*. Accessed: 10-11-2022. URL: https://xgboost.readthedocs.io/en/latest/python/python_api.html.

[132] S. Kraus K. M.and Kampfer et al. "Helical tomotherapy: Comparison of Hi-ART and Radixact clinical patient treatments at the Technical University of Munich." In: *Scientific Reports* 10.1 (2020), p. 4928. ISSN: 2045-2322. DOI: 10.1038/s41598-020-61499-w.

[133] Dayna Bodensteiner. "RayStation: External beam treatment planning system." In: *Medical Dosimetry* 43.2 (2018). Special Issue: 3D Treatment Planning Systems, pp. 168–176. ISSN: 0958-3947. DOI: https://doi.org/10.1016/j.meddos.2018.02.013.

[134] Alessandro Scaggion et al. "Reducing inter- and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution." In: *Physica Medica* 53 (2018), pp. 86–93. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2018.08.016.

[135] Victor Hernandez et al. "Comparison of complexity metrics for multi-institutional evaluations of treatment plans in radiotherapy." In: *Physics and Imaging in Radiation Oncology* 5 (2018), pp. 37–43. ISSN: 2405-6316. DOI: https://doi.org/10.1016/j.phro.2018.02.002.

[136] Alexander F. I. Osman et al. "Applications of machine and deep learning to patient-specific IMRT/VMAT quality assurance." In: *Journal of Applied Clinical Medical Physics* 22.9 (2021), pp. 20–36. DOI: https://doi.org/10.1002/acm2.13375.

[137] Yuto Kimura et al. "Error detection using a convolutional neural network with dose difference maps in patient-specific quality assurance for volumetric modulated arc therapy." In: *Physica Medica* 73 (2020), pp. 57–64. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2020.03.022.

[138] Seiji Tomori et al. "A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance." In: *Medical Physics* 45.9 (), pp. 4055–4065. DOI: https://doi.org/10.1002/mp.13112.

[139] Diana Binny et al. "Investigating the use of aperture shape controller in VMAT treatment deliveries." In: *Medical Dosimetry* 45.3 (2020), pp. 284–292. ISSN: 0958-3947. DOI: https://doi.org/10.1016/j.meddos.2020.02.003.

[140] Tatsuya Kamima et al. "[Impact of Aperture Shape Controller on Volumetric Modulated Arc Therapy Treatment Planning for Nasopharyngeal Cancer]." In: *Nihon Hoshasen Gijutsu Gakkai zasshi* 76.4 (2020), 356—365. ISSN: 0369-4305. DOI: 10.6009/jjrt.2020_jsrt_76.4.356.

[141] Takaaki Ito et al. "[Impact of Aperture Shape Controller on Knowledge-based VMAT Planning of Prostate Cancer]." In: *Nihon Hoshasen Gijutsu Gakkai zasshi* 77.1 (2021), 23—31. ISSN: 0369-4305. DOI: 10.6009/jjrt.2021_jsrt_77.1.23.

[142] Maija Rossi et al. "The use of aperture shape controller and convergence mode in radiotherapy treatment planning." In: *Journal of Radiotherapy in Practice* 21.2 (2022), 171–178. DOI: 10.1017/S1460396920001028.

[143] Eric C. Lobb et al. “Comparison of VMAT complexity-reduction strategies for single-target cranial radiosurgery with the Eclipse treatment planning system.” In: *Journal of Applied Clinical Medical Physics* 21.10 (2020), pp. 97–108. DOI: https://doi.org/10.1002/acm2.13014.

[144] Saeed Ahmed et al. “A method for a priori estimation of best feasible DVH for organs-at-risk: Validation for head and neck VMAT planning.” In: *Medical Physics* 44.10 (2017), pp. 5486–5497. DOI: https://doi.org/10.1002/mp.12500.

[145] Julian Biau et al. “Selection of lymph node target volumes for definitive head and neck radiation therapy: a 2019 Update.” In: *Radiotherapy and Oncology* 134 (2019), pp. 1–9. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2019.01.018.

[146] Vincent Grégoire et al. “Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines.” In: *Radiotherapy and Oncology* 126.1 (2018), pp. 3–24. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2017.10.016.

[147] Anne W. Lee et al. “International guideline for the delineation of the clinical target volumes (CTV) for nasopharyngeal carcinoma.” In: *Radiotherapy and Oncology* 126.1 (2018), pp. 25–36. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2017.10.032.

[148] G. Peyraga et al. “Radiothérapie des cancers des cavités nasosinusiennes.” In: *Cancer/Radiothérapie* 20 (2016). Recorad : Recommandations pour la pratique de la radiothérapie externe et de la curiethérapie, S99–S103. ISSN: 1278-3218. DOI: https://doi.org/10.1016/j.canrad.2016.07.035.

[149] Ursula Nestle et al. “ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-small cell lung cancer.” In: *Radiotherapy and Oncology* 127.1 (2018), pp. 1–5. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2018.02.023.

[150] Charles N. Catton et al. “Randomized Trial of a Hypofractionated Radiation Regimen for the Treatment of Localized Prostate Cancer.” In: *Journal of Clinical Oncology* 35.17 (2017). PMID: 28296582, pp. 1884–1890. DOI: 10.1200/JCO.2016.71.7397.

[151] David Dearnaley et al. “Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial.” In: *The Lancet Oncology* 17.8 (2016), pp. 1047–1060. ISSN: 1470-2045. DOI: 10.1016/S1470-2045(16)30102-4.

[152] Luca Incrocci et al. “Hypofractionated versus conventionally fractionated radiotherapy for patients with localised prostate cancer (HYPRO): final efficacy results from a randomised, multicentre, open-label, phase 3 trial.” In: *The Lancet Oncology* 17.8 (2016), pp. 1061–1069. ISSN: 1470-2045. DOI: `10.1016/S1470-2045(16)30070-5`.

[153] Scott C. Morgan et al. “Hypofractionated Radiation Therapy for Localized Prostate Cancer: An ASTRO, ASCO, and AUA Evidence-Based Guideline.” In: *Journal of Clinical Oncology* 36.34 (2018), pp. 3411–3430. DOI: `10.1200/JCO.18.01097`.

[154] Karl Blum. *Density Matrix Theory and Applications*. Springer, 2012. DOI: `https://doi.org/10.1007/978-3-642-20561-3`.

[155] Richard M. Johnson. “On a theorem stated by eckart and young.” In: *Psychometrika* 28.3 (1963), pp. 259–263. ISSN: 1860-0980. DOI: `10.1007/BF02289573`.

[156] K. Dullemond et al. *Introduction to Tensor Calculus*. 2010. URL: `"http://www.ita.uni-heidelberg.de/~dullemond/lectures/tensor/tensor.pdf"`.

[157] P. Silvi et al. “The Tensor Networks Anthology: Simulation techniques for many-body quantum lattice systems.” In: *SciPost Physics Lecture Notes* (Mar. 2019). DOI: `10.21468/scipostphyslectnotes.8`.

[158] B. Pirvu et al. “Matrix product operator representations.” In: *New Journal of Physics* 12.2 (Feb. 2010). DOI: `10.1088/1367-2630/12/2/025012`.

[159] Stellan Östlund et al. “Thermodynamic Limit of Density Matrix Renormalization.” In: *Phys. Rev. Lett.* 75 (Nov. 1995), pp. 3537–3540. DOI: `10.1103/PhysRevLett.75.3537`.

[160] M. Gerster et al. “Unconstrained tree tensor network: An adaptive gauge picture for enhanced performance.” In: *Physical Review B* (Sept. 2014). DOI: `10.1103/physrevb.90.125154`.

[161] S. Montangero. *Introduction to tensor network methods*. 1st ed. Springer, 2018.

[162] A Taylor et al. “Intensity-modulated radiotherapy–what is it?” In: *Cancer imaging : the official publication of the International Cancer Imaging Society* 4.2 (Mar. 2004), pp. 68–73. DOI: `10.1102/1470-7330.2004.0003`.

[163] Thomas Bortfeld. “IMRT: a review and preview.” In: *Physics in Medicine and Biology* 51.13 (2006), R363–R379. DOI: `10.1088/0031-9155/51/13/r21`.

[164] Jalil ur Rehman et al. “Intensity modulated radiation therapy: A review of current practice and future outlooks.” In: *Journal of Radiation Research and Applied Sciences* 11.4 (2018), pp. 361–367. DOI: 10.1016/j.jrras.2018.07.006.

[165] Craig Elith et al. “An Introduction to the Intensity-modulated Radiation Therapy (IMRT) Techniques, Tomotherapy, and VMAT.” In: *Journal of Medical Imaging and Radiation Sciences* 42.1 (2011), pp. 37–43. DOI: 10.1016/j.jmir.2010.11.005.

[166] C Cotrutz et al. “A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy.” In: *Physics in Medicine and Biology* 46.8 (2001), pp. 2161–2175. DOI: 10.1088/0031-9155/46/8/309.

[167] Steve Webb. “Optimizing radiation therapy inverse treatment planning using the simulated annealing technique.” In: *International Journal of Imaging Systems and Technology* 6.1 (1995), pp. 71–79. DOI: 10.1002/ima.1850060110.

[168] Yair Censor et al. “From analytic inversion to contemporary IMRT optimization: Radiation therapy planning revisited from a mathematical perspective.” In: *Physica Medica* 28.2 (2012), pp. 109 –118. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2011.04.002.

[169] Gary A. Ezzell. “Genetic and geometric optimization of three-dimensional radiation therapy treatment planning.” In: *Medical Physics* 23.3 (1996), pp. 293–305. DOI: 10.1118/1.597660.

[170] Daryl P Nazareth et al. “First application of quantum annealing to IMRT beamlet intensity optimization.” In: 60.10 (2015), pp. 4137–4148. DOI: 10.1088/0031-9155/60/10/4137.

[171] Julia M. Pakela et al. “Quantum-inspired algorithm for radiotherapy planning optimization.” In: *Medical Physics* 47.1 (2020), pp. 5–18. DOI: 10.1002/mp.13840.

[172] Ulrich Schollwöck. “The density-matrix renormalization group in the age of matrix product states.” In: *Annals of Physics* 326.1 (2011), 96–192. ISSN: 0003-4916. DOI: 10.1016/j.aop.2010.09.012.

[173] Pietro Silvi et al. “The Tensor Networks Anthology: Simulation techniques for many-body quantum lattice systems.” In: *SciPost Phys. Lect. Notes* (2019), p. 8. DOI: 10.21468/SciPostPhysLectNotes.8.

[174] Simone Montangero. *Introduction to Tensor Network Methods*. Springer International Publishing, 2018. ISBN: 978-3-030-01408-7.

[175] G. Evenbly et al. “Algorithms for entanglement renormalization.” In: *Phys. Rev. B* 79 (14 2009), p. 144108. DOI: 10.1103/PhysRevB.79.144108.

[176] Román Orús. “A practical introduction to tensor networks: Matrix product states and projected entangled pair states.” In: *Annals of Physics* 349 (2014), 117–158. ISSN: 0003-4916. DOI: 10.1016/j.aop.2014.06.013.

[177] T. Felser et al. “Augmented Tree Tensor Network.” In: *to be published* ().

[178] Ian P McCulloch. “From density-matrix renormalization group to matrix product states.” In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.10 (2007), P10014–P10014. DOI: 10.1088/1742-5468/2007/10/p10014.

[179] Sukhwinder Singh et al. “Global symmetries in tensor network states: Symmetric tensors versus minimal bond dimension.” In: *Phys. Rev. B* 88 (11 2013), p. 115147. DOI: 10.1103/PhysRevB.88.115147.

[180] M. Dalmonte et al. “Lattice gauge theory simulations in the quantum information era.” In: *Contemporary Physics* 57.3 (2016), pp. 388–412. DOI: 10.1080/00107514.2016.1151199.

[181] M. Gerster et al. “Fractional quantum Hall effect in the interacting Hofstadter model via tensor networks.” In: *Phys. Rev. B* 96 (19 2017), p. 195123. DOI: 10.1103/PhysRevB.96.195123.

[182] M. C. Bañuls et al. *Simulating Lattice Gauge Theories within Quantum Technologies*. 2019.

[183] Timo Felser et al. *Two-dimensional quantum-link lattice Quantum Electrodynamics at finite density*. 2019.

[184] Mari Carmen Bañuls et al. “Review on novel methods for lattice gauge theories.” In: *Reports on Progress in Physics* 83.2 (2020), p. 024401. DOI: 10.1088/1361-6633/ab6311.

[185] William Huggins et al. “Towards quantum machine learning with tensor networks.” In: *Quantum Science and Technology* 4.2 (2019), p. 024001. ISSN: 2058-9565. DOI: 10.1088/2058-9565/aaea94.

[186] Isaac H. Kim et al. *Robust entanglement renormalization on a noisy quantum computer*. 2017.

[187] Yiqing Zhou et al. *What limits the simulation of quantum computers?* 2020.

[188] M. Nielsen et al., eds. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000. ISBN: 978-0-521-63503-5.

[189] Emile H. L. Aarts et al. “Simulated Annealing: A Pedestrian Review of the Theory and Some Applications.” In: *Pattern Recognition Theory and Applications*. Ed. by Pierre A. Devijver et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 1987, pp. 179–192. ISBN: 978-3-642-83069-3.

[190] Richard J. Wagner. *Python module for simulated annealing*. 2019. URL: https://github.com/perrygeo/simanneal.

[191] Xuepeng Sun et al. "Effects of the intensity levels and beam map resolutions on static IMRT plans." In: *Medical Physics* 31.9 (2004), pp. 2402–2411. DOI: https://doi.org/10.1118/1.1783551.

[192] A. Apte et al. *CERR*. 2010. URL: https://github.com/cerr/CERR.

[193] Y Hardiyanti et al. "The Comparison Study of Quadratic Infinite Beam Program on Optimization Instensity Modulated Radiation Therapy Treatment Planning (IMRTP) between Threshold and Exponential Scatter Method with CERR® In The Case of Lung Cancer." In: *Journal of Physics: Conference Series* 739 (2016), p. 012144. DOI: 10.1088/1742-6596/739/1/012144.

[194] David Craft et al. "Shared data for intensity modulated radiation therapy (IMRT) optimization research: the CORT dataset." In: *GigaScience* 3.1 (Dec. 2014), pp. 37–37. DOI: 10.1186/2047-217X-3-37.

[195] Elena Villaggi et al. "Plan quality improvement by DVH sharing and planner's experience: Results of a SBRT multicentric planning study on prostate." In: *Physica Medica* 62 (2019), pp. 73–82. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2019.05.003.

[196] John Preskill. "Quantum Computing in the NISQ era and beyond." In: *Quantum* 2 (Aug. 2018), p. 79. ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79.

[197] Jonathan Wei Zhong Lau et al. "NISQ computing: where are we and where do we go?" In: *AAPPS Bulletin* 32.1 (2022), p. 27. ISSN: 2309-4710. DOI: 10.1007/s43673-022-00058-z.

[198] Jonas Stein et al. "Evidence that PUBO outperforms QUBO when solving continuous optimization problems with the QAOA." In: *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*. ACM, 2023. DOI: 10.1145/3583133.3596358.

[199] Fred Glover et al. *A Tutorial on Formulating and Using QUBO Models*. 2019.

[200] Samuele Cavinato et al. "Optimizing radiotherapy plans for cancer treatment with Tensor Networks." In: *Physics in Medicine and Biology* 66.12 (2021), p. 125015. DOI: 10.1088/1361-6560/ac01f2.

[201] D. W. Systems. *Simulated Annealing Sampler*. 2017. URL: https://docs.ocean.dwavesys.com/projects/neal/en/latest/reference/sampler.html.

[202] Robert B. Gramacy et al. “Cases for the nugget in modeling computer experiments.” In: *Statistics and computing* 22 (2012), pp. 713–722. DOI: `https://doi.org/10.1007/s11222-010-9224-x`.

[203] Sonja Surjanovic et al. *Virtual library of simulation experiments*. 2013. URL: `https://www.sfu.ca/~ssurjano/optimization.html`.

[204] R. Eberhart et al. “A new optimizer using particle swarm theory.” In: *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*. 1995, pp. 39–43. DOI: `10.1109/MHS.1995.494215`.

[205] MathWorks. *Particle swarm optimization*. 2020. URL: `https://it.mathworks.com/help/gads/particleswarm.html`.

[206] Jennifer B. Smilowitz et al. “Long-term dosimetric stability of multiple TomoTherapy delivery systems.” In: *Journal of Applied Clinical Medical Physics* 18.3 (2017), pp. 137–143. DOI: `https://doi.org/10.1002/acm2.12085`.

[207] Florian Sterzing et al. “Dynamic Jaws and Dynamic Couch in Helical Tomotherapy.” In: *International Journal of Radiation Oncology*Biology*Physics* 76.4 (2010), pp. 1266–1273. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/j.ijrobp.2009.07.1686`.

[208] Sara Bresciani et al. “Tomotherapy treatment plan quality assurance: The impact of applied criteria on passing rate in gamma index method.” In: *Medical Physics* 40.12 (2013), p. 121711. DOI: `https://doi.org/10.1118/1.4829515`.

[209] M. Geurts et al. “Longitudinal study using a diode phantom for helical tomotherapy IMRT QA.” In: *Medical Physics* 36.11 (2009), pp. 4977–4983. DOI: `https://doi.org/10.1118/1.3238153`.

[210] Thomas J. DiCiccio et al. “Bootstrap confidence intervals.” In: *Statistical Science* 11.3 (1996), pp. 189 –228. DOI: `10.1214/ss/1032280214`.

[211] Kwanghee Jung et al. “Comparison of Bootstrap Confidence Interval Methods for GSCA Using a Monte Carlo Simulation.” In: *Frontiers in Psychology* 10 (2019). ISSN: 1664-1078. DOI: `10.3389/fpsyg.2019.02215`.

[212] William H. Kruskal et al. “Use of Ranks in One-Criterion Variance Analysis.” In: *Journal of the American Statistical Association* 47.260 (1952), pp. 583–621. DOI: `10.1080/01621459.1952.10483441`.

[213] O.J. Clark Dunn. *Applied Statistics:Analysis of Vaariance and Regression*. New York:Wiley, 1974.

[214] Jirayus Jiarpakdee et al. “The Impact of Correlated Metrics on the Interpretation of Defect Models.” In: *IEEE Transactions on Software Engineering* 47.2 (2021), pp. 320–331. DOI: `10.1109/TSE.2019.2891758`.

[215] M Kuhn et al. *Applied predictive modeling*. New York:Springer, 2013.

[216] Diana Binny et al. “Tomotherapy treatment site specific planning using statistical process control.” In: *Physica Medica* 53 (2018), pp. 32–39. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2018.08.003`.

[217] Q Yue et al. “Systematic Analysis of the ArcCheck Diode Arrays for Tomotherapy Delivery Verification.” In: *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology* 3 (2014), pp. 218–225. DOI: `https://doi.org/10.4236/ijmpcero.2014.34028.`.

[218] Mohammad Hussein et al. “A critical evaluation of the PTW 2D-ARRAY seven29 and OCTAVIUS II phantom for IMRT and VMAT verification.” In: *Journal of Applied Clinical Medical Physics* 14.6 (2013), pp. 274–292. DOI: `https://doi.org/10.1120/jacmp.v14i6.4460`.

[219] M. Stasi et al. “Pretreatment patient-specific IMRT quality assurance: A correlation study between gamma index and patient clinical dose volume histogram.” In: *Medical Physics* 39.12 (2012), pp. 7626–7634. DOI: `https://doi.org/10.1118/1.4767763`.

[220] Alessandro Scaggion et al. “Updating a clinical Knowledge-Based Planning prediction model for prostate radiotherapy.” In: *Physica Medica* 107 (2023), p. 102542. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2023.102542`.

[221] Amy T.Y. Chang et al. “Comparison of Planning Quality and Efficiency Between Conventional and Knowledge-based Algorithms in Nasopharyngeal Cancer Patients Using Intensity Modulated Radiation Therapy.” In: *International Journal of Radiation Oncology*Biology*Physics* 95.3 (2016), pp. 981–990. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/j.ijrobp.2016.02.017`.

[222] Mariel Cornell et al. “Noninferiority Study of Automated Knowledge-Based Planning Versus Human-Driven Optimization Across Multiple Disease Sites.” In: *International Journal of Radiation Oncology*Biology*Physics* 106.2 (2020), pp. 430–439. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/j.ijrobp.2019.10.036`.

[223] Mohammad Hussein et al. “Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations.” In: *The British Journal of Radiology* 91.1092 (2018). PMID: 30074813, p. 20180270. DOI: 10.1259/bjr.20180270.

[224] Robert Kaderka et al. “Wide-Scale Clinical Implementation of Knowledge-Based Planning: An Investigation of Workforce Efficiency, Need for Post-automation Refinement, and Data-Driven Model Maintenance.” In: *International Journal of Radiation Oncology*Biology*Physics* 111.3 (2021), pp. 705–715. ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2021.06.028.

[225] Nan Li et al. “Highly Efficient Training, Refinement, and Validation of a Knowledge-based Planning Quality-Control System for Radiation Therapy Clinical Trials.” In: *International Journal of Radiation Oncology*Biology*Physics* 97.1 (2017), pp. 164–172. ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2016.10.005.

[226] Shadab Momin et al. “Knowledge-based radiation treatment planning: A data-driven method survey.” In: *Journal of Applied Clinical Medical Physics* 22.8 (2021), pp. 16–44. DOI: https://doi.org/10.1002/acm2.13337.

[227] Vanessa Panettieri et al. “Development of a multicentre automated model to reduce planning variability in radiotherapy of prostate cancer.” In: *Physics and Imaging in Radiation Oncology* 11 (2019), pp. 34–40. ISSN: 2405-6316. DOI: https://doi.org/10.1016/j.phro.2019.07.005.

[228] Geert Wortel et al. “Characterization of automatic treatment planning approaches in radiotherapy.” In: *Physics and Imaging in Radiation Oncology* 19 (2021), pp. 60–65. ISSN: 2405-6316. DOI: https://doi.org/10.1016/j.phro.2021.07.003.

[229] Justin J. Boutilier et al. “Sample size requirements for knowledge-based treatment planning.” In: *Medical Physics* 43.3 (2016), pp. 1212–1221. DOI: https://doi.org/10.1118/1.4941363.

[230] Yaorong Ge et al. “Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches.” In: *Medical Physics* 46.6 (2019), pp. 2760–2775. DOI: https://doi.org/10.1002/mp.13526.

[231] Tamura Mikoto et al. “Influence of Cleaned-up Commercial Knowledge-Based Treatment Planning on Volumetric-Modulated Arc Therapy of Prostate Cancer.” In: *Journal of Medical Physics* 45.2 (2020), pp. 71–77. DOI: https://doi.org/10.4103/jmp.JMP_109_19.

[232] Jim P. Tol et al. “A longitudinal evaluation of improvements in radiotherapy treatment plan quality for head and neck cancer patients.” In: *Radiotherapy and Oncology* 119.2 (2016), pp. 337–343. ISSN: 0167-8140. DOI: `https://doi.org/10.1016/j.radonc.2016.04.011`.

[233] A. Fogliata et al. “RapidPlan knowledge based planning: iterative learning process and model ability to steer planning strategies.” In: *Radiation Oncology* 14.1 (2019), p. 187. ISSN: 1748-717X. DOI: `10.1186/s13014-019-1403-0`.

[234] Johanna Austrheim Hundvin et al. “Clinical iterative model development improves knowledge-based plan quality for high-risk prostate cancer with four integrated dose levels.” In: *Acta Oncologica* 60.2 (2021), pp. 237–244. DOI: `10.1080/0284186X.2020.1828619`.

[235] Kenji Nakamura et al. “An updating approach for knowledge-based planning models to improve plan quality and variability in volumetric-modulated arc therapy for prostate cancer.” In: *Journal of Applied Clinical Medical Physics* 22.9 (2021), pp. 113–122. DOI: `https://doi.org/10.1002/acm2.13353`.

[236] Meijiao Wang et al. “An interactive plan and model evolution method for knowledge-based pelvic VMAT planning.” In: *Journal of Applied Clinical Medical Physics* 19.5 (2018), pp. 491–498. DOI: `https://doi.org/10.1002/acm2.12403`.

[237] Hajime Monzen et al. “Dosimetric evaluation with knowledge-based planning created at different periods in volumetric-modulated arc therapy for prostate cancer: a multi-institution study.” In: *Radiological Physics and Technology* 13.4 (2020), pp. 327–335. ISSN: 1865-0341. DOI: `10.1007/s12194-020-00585-0`.

[238] Mack Roach et al. “Predicting the risk of lymph node involvement using the pre-treatment prostate specific antigen and gleason score in men with clinically localized prostate cancer.” In: *International Journal of Radiation Oncology*Biology*Physics* 28.1 (1994), pp. 33–37. ISSN: 0360-3016. DOI: `https://doi.org/10.1016/0360-3016(94)90138-4`.

[239] A. Scaggion et al. “Delivering RapidArc®: A comprehensive study on accuracy and long term stability.” In: *Physica Medica* 32.7 (2016), pp. 866–873. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2016.05.056`.

[240] W. Robert Lee et al. “Randomized phase III noninferiority study comparing two radiotherapy fractionation schedules in patients with low-risk prostate cancer.” In: *Journal of Clinical Oncology* 34.20 (2016). Cited by: 424; All Open Access, Green Open Access, 2325 – 2332. DOI: `10.1200/JCO.2016.67.0448`.

[241] Varian Medical System. "Eclipse Photon and Electron Reference Guide v15.5." In: (2017).

[242] Benjamin E. Nelms et al. "Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems." In: *Practical Radiation Oncology* 2.4 (2012), pp. 296–305. issn: 1879-8500. doi: `https://doi.org/10.1016/j.prro.2011.11.012`.

[243] Irfan Ahmad et al. "Plan quality assessment of modern radiotherapy delivery techniques in left-sided breast cancer: an analysis stratified by target delineation guidelines." In: *BJR|Open* 2.1 (2020), p. 20200007. doi: `10.1259/bjro.20200007`.

[244] Angelia Landers et al. "Automated 4π radiotherapy treatment planning with evolving knowledge-base." In: *Medical Physics* 46.9 (2019), pp. 3833–3843. doi: `https://doi.org/10.1002/mp.13682`.

[245] Motoharu Sasaki et al. "Analysis of prostate intensity- and volumetric-modulated arc radiation therapy planning quality with PlanIQTM." In: *Journal of Applied Clinical Medical Physics* 22.4 (2021), pp. 132–142. doi: `https://doi.org/10.1002/acm2.13233`.

[246] Savino Cilla et al. "Personalized automation of treatment planning in head-neck cancer: A step forward for quality in radiation therapy?" In: *Physica Medica* 82 (2021), pp. 7–16. issn: 1120-1797. doi: `https://doi.org/10.1016/j.ejmp.2020.12.015`.

[247] Marco Fusella et al. "Efficiently train and validate a RapidPlan model through APQM scoring." In: *Medical Physics* 45.6 (2018), pp. 2611–2619. doi: `https://doi.org/10.1002/mp.12896`.

[248] Mohammad Hussein et al. "Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy." In: *Radiotherapy and Oncology* 120.3 (2016), pp. 473–479. issn: 0167-8140. doi: `https://doi.org/10.1016/j.radonc.2016.06.022`.

[249] Varian Medical System. "Eclipse Photon and Electron Reference Guide v13.7." In: (2015).

[250] Antonella Fogliata et al. "On the pre-clinical validation of a commercial model-based optimisation engine: Application to volumetric modulated arc therapy for patients with lung or prostate cancer." In: *Radiotherapy and Oncology* 113.3 (2014), pp. 385–391. issn: 0167-8140. doi: `https://doi.org/10.1016/j.radonc.2014.11.009`.

[251] Christian Rønn Hansen et al. “Radiotherapy Treatment planniNg study Guidelines (RATING): A framework for setting up and reporting on scientific treatment planning studies.” In: *Radiotherapy and Oncology* 153 (2020). Physics Special Issue: ESTRO Physics Research Workshops on Science in Development, pp. 67–78. ISSN: 0167-8140. DOI: `https://doi.org/10.1016/j.radonc.2020.09.033`.

[252] Avishek Chatterjee et al. “Novel knowledge-based treatment planning model for hypofractionated radiotherapy of prostate cancer patients.” In: *Physica Medica* 69 (2020), pp. 36–43. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2019.11.023`.

[253] Binbin Wu et al. “Cross-institutional knowledge-based planning (KBP) implementation and its performance comparison to Auto-Planning Engine (APE).” In: *Radiotherapy and Oncology* 123.1 (2017), pp. 57–62. ISSN: 0167-8140. DOI: `https://doi.org/10.1016/j.radonc.2017.01.012`.

[254] Phillip D. H. Wall et al. “Evaluation of complexity and deliverability of prostate cancer treatment plans designed with a knowledge-based VMAT planning technique.” In: *Journal of Applied Clinical Medical Physics* 21.1 (2020), pp. 69–77. DOI: `https://doi.org/10.1002/acm2.12790`.

[255] Mikoto Tamura et al. “Mechanical performance of a commercial knowledge-based VMAT planning for prostate cancer.” In: *Radiation Oncology* 13.1 (2018), p. 163. ISSN: 1748-717X. DOI: `10.1186/s13014-018-1114-y`.

[256] Marie-Claude Biston et al. “Evaluation of fully automated a priori MCO treatment planning in VMAT for head-and-neck cancer.” In: *Physica Medica* 87 (2021), pp. 31–38. ISSN: 1120-1797. DOI: `https://doi.org/10.1016/j.ejmp.2021.05.037`.

[257] Hilary Akpati et al. “Unified dosimetry index (UDI): a figure of merit for ranking treatment plans.” In: *Journal of Applied Clinical Medical Physics* 9.3 (2008), pp. 99–108. DOI: `https://doi.org/10.1120/jacmp.v9i3.2803`.

[258] Pedrame Bargassa et al. “Quantum algorithm for the classification of supersymmetric top quark events.” In: *Phys. Rev. D* 104 (9 2021), p. 096004. DOI: `10.1103/PhysRevD.104.096004`.

[259] S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.” In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2012.08.021`.

[260] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.” In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020.

[261] Alex Mott et al. “Solving a Higgs optimization problem with quantum annealing for machine learning.” In: *Nature* 550.7676 (2017), pp. 375–379. ISSN: 1476-4687. DOI: 10.1038/nature24047.

[262] Alexander Zlokapa et al. “Quantum adiabatic machine learning by zooming into a region of the energy surface.” In: *Phys. Rev. A* 102 (6 2020), p. 062405. DOI: 10.1103/PhysRevA.102.062405.

[263] A. M. Sirunyan et al. “Search for top squarks decaying via four-body or chargino-mediated modes in single-lepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: *Journal of High Energy Physics* 2018.9 (2018), p. 65. ISSN: 1029-8479. DOI: 10.1007/JHEP09(2018)065.

[264] Jonathon Shlens. *A Tutorial on Principal Component Analysis*. 2014.

[265] J. Wess et al. “Supergauge transformations in four dimensions.” In: *Nuclear Physics B* 70.1 (1974), pp. 39–50. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(74)90355-1.

[266] R. Barbieri et al. “Gauge models with spontaneously broken local supersymmetry.” In: *Physics Letters B* 119.4 (1982), pp. 343–347. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(82)90685-2.

[267] H.P. Nilles. “Supersymmetry, supergravity and particle physics.” In: *Physics Reports* 110.1 (1984), pp. 1–162. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(84)90008-5.

[268] H.E. Haber et al. “The search for supersymmetry: Probing physics beyond the standard model.” In: *Physics Reports* 117.2 (1985), pp. 75–263. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(85)90051-1.

[269] Sally Dawson et al. “Search for supersymmetric particles in hadron-hadron collisions.” In: *Phys. Rev. D* 31 (7 1985), pp. 1581–1637. DOI: 10.1103/PhysRevD.31.1581.

[270] STEPHEN P. MARTIN. “A SUPERSYMMETRY PRIMER.” In: *Perspectives on Supersymmetry*. WORLD SCIENTIFIC, 1998, pp. 1–98. DOI: 10.1142/9789812839657_0001.

[271] L. Rokach et al. *Data mining with decision trees: theory and applications*. ISBN: 978-981-277-171-1. World Scientific, 2008.

[272] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2009.

[273] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics.” In: *The European Physical Journal C* 71.2 (2011), p. 1554. ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-011-1554-0`.

[274] V. Khachatryan et al. “Search for direct pair production of scalar top quarks in the single- and dilepton channels in proton-proton collisions at $\sqrt{s}$=8 TeV.” In: *Journal of High Energy Physics* 2016.7 (2016), p. 27. ISSN: 1029-8479. DOI: `10.1007/JHEP07(2016)027`.

[275] Paul I. Bunyk et al. “Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor.” In: *IEEE Transactions on Applied Superconductivity* 24.4 (2014), pp. 1–10. DOI: `10.1109/TASC.2014.2318294`.