UNIVERSIDADE D
COIMBRA

Rafael Alexandre Cerqueira Lobão

CHARACTERIZING THE QUALITY AND COMPLEXITY OF PROSTATE RADIOTHERAPY TREATMENT PLANS TOWARDS AUTOMATED PLANNING

Dissertação no âmbito do Mestrado em Física Médica orientada pela Professora Doutora Maria do Carmo Carrilho Calado Antunes Lopes e pelo Doutor Tiago Emanuel Pereira Ventura, co-orientada pela Doutora Tânia Filipa Sobrinho dos Santos e apresentada ao Departamento da Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Julho de 2024

UNIVERSIDADE D
COIMBRA

Characterizing the Quality and Complexity of Prostate Radiotherapy Treatment Plans Towards Automated Planning

Orientadores:
Professora Doutora Maria do Carmo Carrinho Calado Antunes Lopes
Doutor Tiago Emanuel Pereira Ventura

Co-orientadora:
Doutora Tânia Filipa Sobrinho dos Santos

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Física Médica.

Coimbra, July 2024

This work was developed in collaboration with:

Instituto Português de Oncologia de Coimbra Francisco Gentil, E.P.E.

Agradecimentos

Gostaria de aproveitar esta página para expressar os meus mais sinceros agradecimentos a todos que contribuíram, direta ou indiretamente, para a realização desta tese.

O meu primeiro agradecimento vai para todos os físicos médicos e dosimetristas do IPO de Coimbra que, ao longo dos últimos 16 meses, sempre me fizeram sentir bem-vindo e privilegiado por poder aprender com eles. Em especial, gostaria de agradecer aos meus orientadores, Professora Maria do Carmo Lopes, Tânia Santos e Tiago Ventura, a quem a minha dívida de "bolinhos da Vénus" tem, por motivos misteriosos, aumentado consideravelmente nos últimos dias. Como prometido, agradeço à Mariana Marques pela ajuda na escolha da paleta de cores para alguns gráficos desta tese.

Gostaria de agradecer também a todos os professores da Universidade de Coimbra com os quais me cruzei duranto o mestado. Em especial, gostaria de agradecer à Professora Isabel Lopes, que sempre fomentou o meu espírito crítico e se mostrou disponível para responder às minhas inúmeras questões, o que começou mesmo antes de eu descobrir que Coimbra não era uma cidade plana

Agradeço também ao André, Carolina, Felipe, Mariana, Miguel, Renata, Rita e Rúben, pelo apoio durante o terceiro ano da licenciatura, sem dúvida o mais desafiante.

Ao João, Ana, Diogo, Inês 2, Joana, Inês 1 e Mafalda (ordenados alfabeticamente, Mafalda), agradeço por terem tornado a minha estadia em Coimbra mais agradável ao longo dos últimos dois anos.

Ao Luís e ao Renato, agradeço pelas múltiplas discussões que tivemos sobre tópicos completamente distintos ao longo da minha licenciatura no Porto. Sendo ambos de medicina, não é de se estranhar que tenham tido um impacto significativo na minha decisão de seguir a área da saúde.

Por último, o meu maior agradecimento vai para os meus pais e família, que sempre me apoiaram ao longo dos meus 17 anos de estudos, o que me permitiu alcançar o sucesso que tão desejei.

A todos os que mencionei, em palavras ou pensamentos, Obrigado!

Resumo

Com o aumento de técnicas avançadas de radioterapia, nomeadamente a radioterapia de intensidade modulada e a arcoterapia volumétrica modulada (VMAT), têm vindo a ser obtidos planos de tratamento mais conformais e com maior poupança dos órgãos de risco. Todavia, o planeamento dosimétrico é um processo complexo, laborioso e demorado e depende da destreza e experiência do planeador. Estas desvantagens podem ser superadas com a implementação de plataformas de planeamento automático. Um passo importante na configuração dessas ferramentas é a caracterização da qualidade e complexidade dos planos de tratamento produzidos localmente.

Este projeto teve como principal objetivo estabelecer uma base de dados para a configuração de uma plataforma de planeamento automático, RapidPlan, disponível no sistema de planeamento Eclipse. Desta forma, foram restrospectivamente selecionados planos VMAT de doentes com cancro da próstata, tratados num acelerador linear Halcyon, desde a sua instalação, em junho de 2021, até abril de 2024. Para avaliar a qualidade dos planos, a ferramenta SPIDERplan, desenvolvida no IPO Coimbra, foi configurada de acordo com as preferências clínicas dos radioncologistas. A sua apresentação gráfica foi igualmente melhorada e um novo módulo de análise estatística implementado. Quanto à complexidade dos planos foram calculadas várias métricas propostas na literatura. Os módulos de análise da qualidade e de complexidade dos planos foram integrados numa plataforma eletrónica denominada qualComp. Esta plataforma permite o cálculo automático da qualidade e da complexidade de cada novo plano, bem como a sua comparação gráfica com os planos existentes. De maneira abrangente e dinâmica, a qualComp permite, por exemplo, identificar planos com excesso de modulação, sem que isso corresponda necessariamente a uma melhoria de qualidade. No período considerado, o estudo retrospetivo mostrou uma redução constante na complexidade, que estabilizou após o plateau da curva de aprendizagem de planeamento, no Eclipse, ter sido atingido e terem sido introduzidas normas de planeamento, tendentes à uniformização das práticas. Verificou-se, ainda, que a qualidade dos planos não foi afetada pela redução da complexidade dos mesmos.

De um modo geral, a caracterização automática da qualidade e da complexidade dos planos permite a padronização de soluções de otimização do planeamento dosimétrico e abre caminho para a configuração e validação de soluções de planeamento automático.

Abstract

With the increased use of advanced radiotherapy treatment delivery techniques, including intensity modulated radiation therapy and volumetric modulated arc therapy (VMAT), more conformal dose distributions have been achieved with a high degree of sparing of organs at risk. However, dose planning is labor and time intensive, and highly dependent on the planner skills and experience. The implementation of automated planning solutions can certainly contribute to reduce these drawbacks. An important step in setting up such tools is the characterization of the quality and the complexity of local treatment plans.

The aim of this project was establishing a database for the configuration of RapidPlan, the Eclipse TPS knowledge-based auto-planning solution. Plans from prostate patients treated with VMAT in a Halcyon linac were retrospectively selected from the installation in June 2021 till April 2024. To assess plan quality, SPIDERplan, a tool developed at IPO Coimbra, was configured according to the clinical preferences of the radiation oncologists. Its graphical presentation was improved and statistical analysis was included. Plan complexity was evaluated through the calculation of multiple metrics proposed in the literature. Both plan quality and complexity analyses were integrated into an electronic platform called qualComp. This platform allows the automatic calculation of the quality and complexity of each new plan, as well as the graphical comparison with the existing database. In a comprehensive and dynamic way, qualComp namely allows identifying plans with overmodulation, not necessarily with a quality improvement. The retrospective study showed a steady decrease in complexity which then stabilized, after the Eclipse reaching the learning curve plateau and the introduction of planning guidelines aiming to clinical practice standardization. It was also observed that the quality of the plans was not affected by the decrease in complexity.

Overall, the characterization of plan quality and complexity in an automated manner enables planning standardization, and paves the way towards a reliable configuration, validation and testing of automated planning solutions.

Contents

List of Figures

List of Tables

List of Abbreviations

3DCRT	Three-Dimensional Conformal Radiation Therapy
AAA	Anisotropic Analytical Algorithm
AAPM	American Association of Physicists in Medicine
AAV	Aperture Area Variability
ASC	Aperture Shape Control
CAS	Cross Axis Score
CLS	Close Leaf Score
CN	Conformity Number
COIN	Conformity Index
CT	Computed Tomography
CTV	Clinical Target Volume
DICOM	Digital Imaging and Communications in Medicine
DVH	Dove-Volume Histogram
EAM	Edge Area Metric
EM	Edge Metric
EPID	Electronic Portal Imaging Device
ESTRO	European Society for Radiotherapy and Oncology
gEUD	Generalized Equivalent Uniform Dose

GI	Gradient Index
GPD	Groups Plan Diagram
GPR	Gamma Passing Rate
GPS	Global Plan Score
GTV	Gross Target Volue
GUI	Graphical User Interface
ICRU	International Commission on Radiation Units and Measurements
IMRT	Intensity-Modulated Radiation Therapy
KBP	Knowledge-Based Planning
Linac	Linear Accelerator
LSV	Leaf Sequence Variability
LT	Leaf Travel
LTMCS	Leaf Travel Modulation Complexity Score
MAD	Mean Asymmetry Displacement
MCS	Modulation Complexity Score
MIt	Total Modulation Index
MLC	Multi-Leaf Collimator
MRI	Magnetic resonance imaging
MU	Monitor Unit
MUd	Monitor Units per fraction dose in centigrays
NTCP	Normal Tissue Complication Probability
OAR	Organ at Risk
PA	Plan Area
PET	Positron Emission Tomography

PI	Plan Irregularity
PM	Plan Modulation
PSQA	Patient-Specific Quality Assurance
PQI	Plan Quality Index
PTV	Planning Target Volume
RO	Radiation Oncologist
RT	Radiotherapy
RTOG	Radiation Therapy Oncology Group
SAS	Small Aperture Score
SGD	Structures Group Diagram
SPD	Structures Plan Diagram
TCP	Tumor Control Probability
TPS	Treatment Planning Score
UL	Uncovered-Layer Score
VMAT	Volumetric Modulated Arc Therapy

1 Introduction

1.1 Motivation

Cancer is one of the leading causes of morbidity and mortality worldwide. According to the World Health Organization [1], there were 20 million new cancer diagnosis and almost 9.7 million cancer deaths reported in 2022, with the number of new diagnosis expected to increase 77% by 2050. Portugal specifically, where 1 in 3 men and 1 in 4 women develop cancer before reaching the age of 75, saw 69 thousand new cancer cases and 33 thousand deaths in 2022 [1].

Radiotherapy plays a crucial role in cancer treatment, with approximately 50% of cancer patients expected to undergo radiotherapy at some point during their treatment [2]. Modern techniques, such as Intensity Modulated Radiotherapy (IMRT), have significantly improved treatment outcomes and reduced toxicities [3]. Despite these advancements, treatment planning remains a labor and time consuming process, involving iterative, trial-and-error adjustments of optimization parameters [4]. Additionally, the quality of the resultant plan is highly dependent on the skills and experience of the planner [5].

Automated planning techniques have been proposed to streamline the plan creation and mitigate the variability between planners, with most treatment planning systems (TPS) incorporating some form of automation solution [6]. Among these, knowledge-based planning (KBP) solutions utilize models trained on clinical plans to generate planning objectives tailored to the patient's anatomy. KBP solutions have been shown to successfully reduce planning time, minimize inter-planner variation, and improve plan quality, without significantly deviating from other clinical plans in terms of complexity [7–9].

The performance of this solution is highly dependent on the quality of the treatment plans used to train the KBP model [10]. Therefore, a database of high-quality plans is essential for KBP configuration. Additionally, a local database allows for validating whether the automated plans align with the clinical practices of the institution.

This dissertation aims to establish and characterize a local database of prostate radiotherapy plans previously delivered at IPO Coimbra using the Varian Halcyon linac, with the future goal of utilizing this database for the training and validation of RapidPlan, the Varian KBP solution. During the characterization of the database, plan quality was assessed using SPIDERplan, a

graphical tool for comparing and scoring each plan according to their dosimetric performance, while plan complexity was evaluated using 15 complexity metrics adapted to the Halcyon linac. Both analyses were automatically performed using an electronic platform called qualComp, which was developed during this thesis.

1.2 Organization of the Thesis

This work is organized into five chapters:

1. Introduction: the motivation and outline of the dissertation are presented.

2. Background Knowledge: key concepts of radiotherapy, plan quality, plan complexity, and knowledge-based planning are briefly discussed as to allow the reader to follow the dissertation.

3. Materials and Methods: the qualComp platform is described, as well as the SPIDERplan configuration and the complexity metrics used. The methodologies followed for plan quality and complexity analysis are summarized.

4. Results and Discussion: the main results are presented and discussed in relation to the objectives of this dissertation.

5. Conclusion: the conclusions of this dissertation are presented.

2 Background Knowledge

In this chapter, key concepts of radiotherapy (section 2.1), plan quality (section 2.2), plan complexity (section 2.3), and knowledge-based planning (section 2.4) are briefly discussed as to allow the reader to follow the rest of the dissertation.

2.1 Radiotherapy

Radiotherapy (RT) is a cancer treatment modality that employs ionizing radiation to well-defined volumes while minimizing damage to surrounding healthy tissues. It can be used alone (monotherapy) or in conjunction with other treatment modalities, including surgery and chemotherapy, serving both curative and palliative purposes [11]. There are three forms of RT [12]:

- External beam radiotherapy: the radiation is delivered from outside of the patient's body, generally using a linear accelerator (linac).
- Brachytherapy: the radiation source is placed temporarily or permanently near the tumor site.
- Systemic radiation therapy: a soluble radioactive substance is delivered by ingestion, catheter infusion, or intravenous administration.

Hereafter, Radiotherapy will be referencing external beam radiotherapy as delivered by a linac.

2.1.1 Conventional Linear Accelerators

Linear Accelerators (linacs) operate by accelerating electrons in a linear accelerating waveguide, being able to produce electron beams used to treat superficial lesions in a very reduced number of clinical cases or photon beams, when a heavy-metal target is placed in the electron beam path [13].

A conventional linac is an isocentric C-arm machine that through the movement of the gantry, collimator and couch allows delivering radiation beams from multiple directions. The rotation axis of the gantry, collimator and couch intercept at the machine isocenter. In conventional linacs, the photon beam is shaped and modulated in the linac head by a set of collimators.

These include the primary collimator, which defines the maximum circular field size, secondary collimators (jaws), composed of independently adjustable shielding blocks, and the multi-leaf collimator (MLC) [13]. The MLC consists of two banks of parallel movable leaves with a leaf width from a few millimetres up to 1 cm at the isocenter level. The MLC leaves can move during treatment, allowing modulation of the beam intensity [14]. The design of the collimators varies greatly among linac models, with some lacking some collimators or having them arranged differently [15].

Additionally, the linac head may also contain a series of flattening filters intended to provide a uniform beam profile (Figure 2.1) [16]. The appropriate filter is selected based on the beam energy, when a uniform beam profile is intended [13]. Still, flattening filter free beams can also be obtained, allowing for higher dose rates (Figure 2.1).

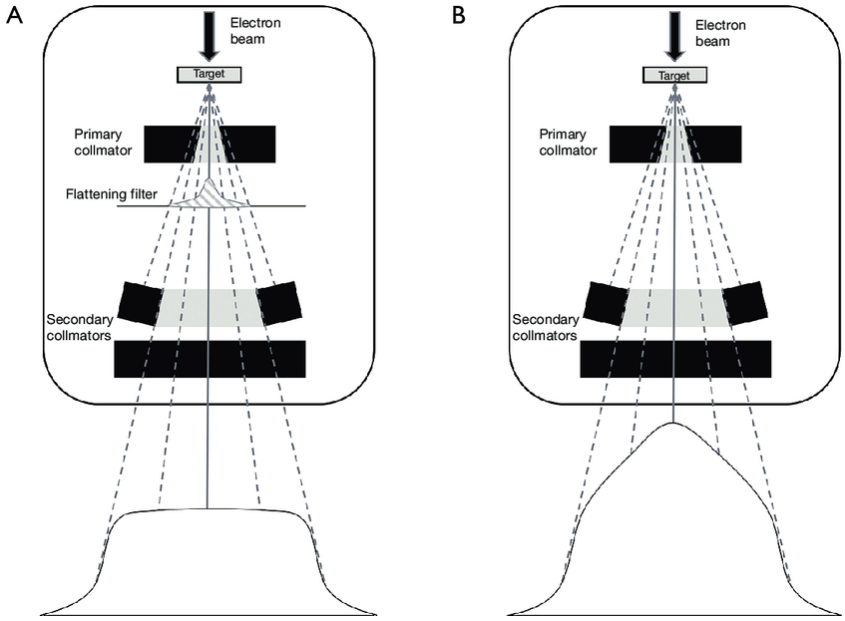


Figure 2.1: Beam profile with (A) flattening filter or (B) flattening filter free. The target, primary collimator, and second collimator are also depicted.

Also present in the linac head are two sealed ionization chambers that monitor the radiation output, known as monitor chambers. The machine output is given in monitor units (MU), where 1 MU corresponds to a specific delivered dose at a given depth of a water phantom at a specified field size and distance from the radiation source [13, 17].

The above-mentioned components are installed in the gantry, a structural component that holds the linac head. Also mounted in the gantry, modern linacs feature a 2D flat-panel detector in the beam exit, known as an EPID (Electronic Portal Imaging Device) and, in the most recent models, a retractable kilovoltage system at 90º to the linac head, enabling 2D/3D patient positioning verification [13, 18]. Besides, EPID serves also other purposes such as machine and patient-specific quality assurance, and in-vivo dosimetry [18].

Linacs are also equipped with a treatment couch where patients are positioned during treatment.

The couch provides up to six degrees of freedom: three translational and three rotational [16].

There are also different designs of radiotherapy treatment machines, including Tomotherapy, Cyberknife, and Halcyon.

Halcyon

The Halcyon linac from Varian Medical Systems (Figure 2.2) is an O-ring gantry linac that utilizes flattening filter-free 6 MV photon beams (6FFF) at a maximum output rate of 600 MU/s.

The Halcyon collimation configuration stands out from conventional linacs (Figure 2.3). The primary collimator defines a maximum square field size of 28×28 cm^2, while the secondary collimator, instead of the conventional pairs of jaws, is a fixed collimator that mirrors the beam limiting function of the primary collimator and rotates alongside the MLC. The MLC comprises two layers of 1 cm wide leaves, offset by 5 mm. The proximal layer, positioned closer to the radiation source, consists of 29 leaf pairs, while the distal layer consists of 28.

Halcyon offers some advantages over conventional linacs, such as higher gantry rotation and MLC leaf speeds, lower MLC penumbra, and no restrictions on leaves over-travel or interdigitation [19]. On the other hand, the Halcyon treatment couch only provides three translation degrees of freedom and only 6 MV photon beams are available [19].

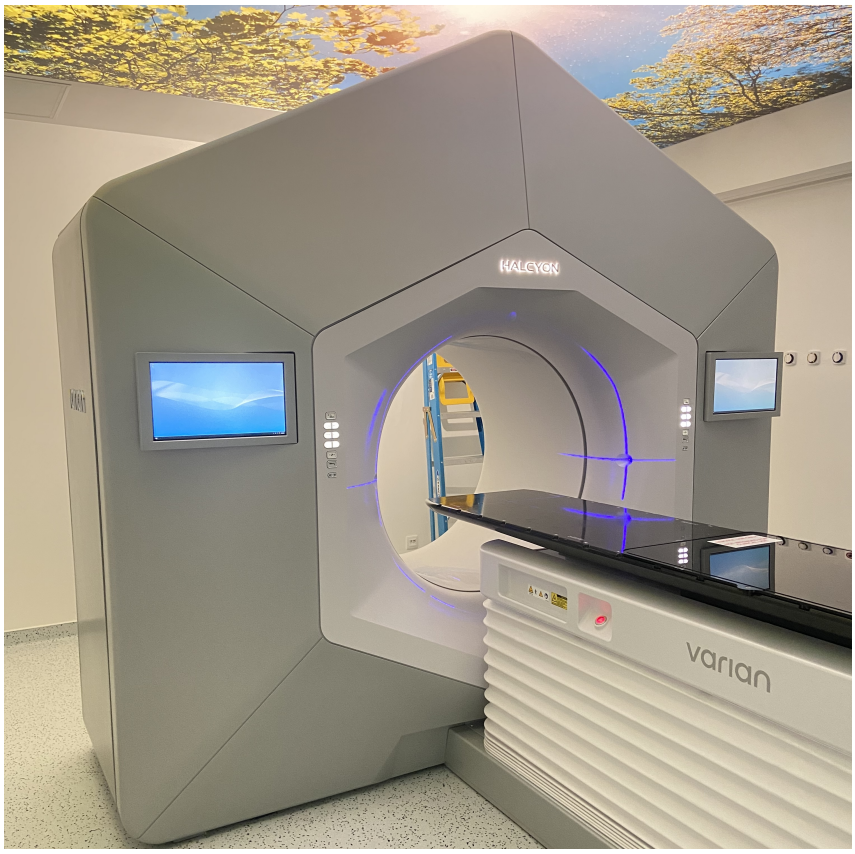


Figure 2.2: Halcyon linac at IPO Coimbra.

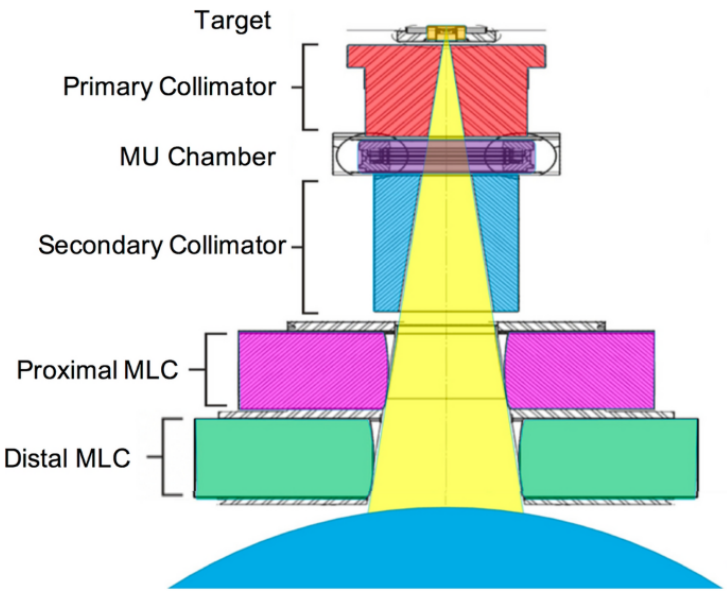


Figure 2.3: Schematic of the Halcyon collimators. Adapted from [20].

2.1.2 Radiotherapy Delivery Techniques

Among the various RT delivery techniques, Three-Dimensional Conformal Radiation Therapy (3DCRT) and Intensity-Modulated Radiation Therapy (IMRT) are two of the most widely used techniques.

Three-Dimensional Conformal Radiation Therapy (3DCRT) employs radiation beams shaped by multileaf collimators (MLCs) from different directions that conform to the target as viewed by the specified gantry angle [13, 21]. With the MLC shape fixed at each angle, 3DCRT employs uniform intensity beams or beams with constantly varying intensity using wedge filters [4]. In contrast, IMRT varies the MLC shape to achieve modulation of the beam intensity. Modulation can be performed at fixed angles using multiple segmented fields (step-and-shoot IMRT) or by varying the MLC leaves position during irradiation (sliding-window IMRT) [4]. Beam intensity modulation can also be performed while the gantry rotates around the patient, continuously varying the leaves position, dose rate and gantry speed, in a IMRT technique known as Volumetric Modulated Arc Therapy (VMAT).

Compared to 3DCRT, IMRT offers significant advantages in terms of dose conformity to the targets and the sparing of critical organs, although at the cost of higher low-dose radiation exposure and a greater number of monitor units [22]. In turn, VMAT is more efficient than fixed-gantry IMRT, requiring fewer monitor units and shorter treatment times while maintaining similar conformity and organ sparing [22].

2.1.3 Radiotherapy Treatment Workflow

The treatment workflow involves a multidisciplinary team of professionals such as radiation oncologists, nurses, radiotherapy technologists, medical physicists, and dosimetrists [23]. The workflow generally encompasses the first radiotherapy medical appointment, immobilization and

planning CT, structures delineation, prescription, treatment planning, patient specific quality assurance, treatment delivery, and follow-up.

First Radiotherapy Medical Appointment

The radiotherapy treatment workflow begins with the first radiotherapy medical appointment, following a therapeutic decision made by a multidisciplinary team. During this initial consultation, the radiation oncologist conducts a thorough clinical evaluation of the patient's condition, discusses the chosen treatment modality, informs the patient about their clinical prognosis, and addresses any concerns or questions the patient may have.

Immobilization and Planning CT

To ensure consistent patient positioning and immobilization throughout the radiotherapy treatment course and during treatment sessions, appropriate immobilization devices for the treatment site are selected [13, 16]. With the immobilization devices in place, the patient then undergoes a computed tomography (CT) scan in the exact position they will assume during treatment [16]. The resulting CT images provide electron density information required for dose calculation [13].

Structures Delineation

The radiation oncologist delineates the target volumes to be irradiated and organs at risk of interest on the planning CT, using the CT images alone or in conjunction with other diagnostic modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET) [13].

The International Commission on Radiation Units and Measurements (ICRU) defines three main volume concepts for target delineation: gross tumor volume (GTV), clinical target volume (CTV), and planning target volume (PTV). The GTV corresponds to the macroscopic disease, which sometimes has been surgically removed. The CTV, which includes the GTV plus a margin, accounts for tissues with a certain probability of subclinical malignancies. Finally, the PTV encompasses the CTV along with an additional margin to compensate for patient setup errors and organ motion.

Prescription

For each case, the radiation oncologist determines the treatment prescription, specifying details such as the dose to be delivered to the targets, dose constraints for organs at risk, and the number of treatment fractions [4]. In conventional radiotherapy treatments, five daily sessions per week are performed over a span of 3 to 8 weeks [24], delivering 1.8 to 2 cGy per fraction. In hypofractionated treatments, higher doses per fraction are delivered in a shorter time period (few days or weeks).

Treatment Planning

Once the prescription is defined, a medical physicist or dosimetrist uses a treatment planning system (TPS) to create the treatment plan, aiming to achieve an optimal dose distribution that meets all objectives and constraints set by the radiation oncologist [4, 23].

There are two planning techniques: forward-planning, used in 3DCRT, and inverse-planning, used in IMRT [16].

- Forward-planning: the planner manually adjusts the characteristics of each radiation beam until a desirable dose distribution is achieved [4]. This includes setting iteratively the gantry and multi-leaf collimator (MLC) angles, shaping the MLC, and determining the number of monitor units (MUs).

- Inverse-planning: the dose distribution goals are defined for each delineated structure and the TPS works to achieve them by minimizing cost functions. These cost functions quantify the discrepancies between the current and the desired dose distribution, incorporating planning objectives and other dose distribution descriptors [4]. During planning, the descriptors and their weights (priorities) can be adjusted to ensure the resulting plan is clinically acceptable. This process is done in a time-consuming, trial-and-error manner that it is highly dependent on the planner's experience [4, 5].

The final treatment plan must be evaluated and approved by a radiation oncologist before clinical use [13]. For that, the radiation oncologist generally visually assesses the 3D dose distribution overlapped on CT planning images and analyses dose statistics (e.g., minimum dose, maximum dose, mean dose, median dose) and specific dose-volume metrics for each delineated structure. Dose-volume metrics can be easily evaluated using dose-volume histograms (DVHs). There are two types of DVHs: differential and cumulative (Figure 2.4), with their relation given by equation 2.1.

$$\mathrm{DVH}_{\mathrm{cum.}}(D) = 1 - \int_0^D \mathrm{DVH}_{\mathrm{diff.}}(D')\,\mathrm{d}D' \tag{2.1}$$

In its differential form, the DVH represents the dose frequency distribution within the structure volume [13]. However, it is the cumulative DVH that is more useful for assessing dose-volume metrics, as D_V is the dose coordinate of the point along the DVH line with volume V, and V_D is the volume coordinate of the point with dose D. While cumulative DVH are used in plan evaluation during planning, spatial information of the dose distribution is lost [26]. For instance, DVHs are not sensitive to low and high-dose regions. Therefore, DVH analysis should not a substitute to visual inspection of the dose distribution [4]. Hereafter, cumulative DVH will be referred as DVH.

Given the large number of dosimetric parameters to be assessed for each target volume and organ

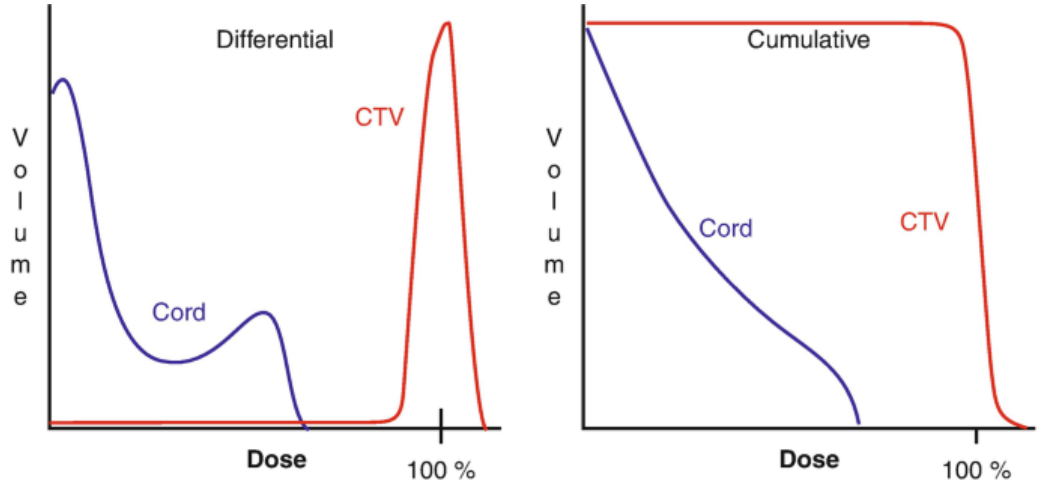


Figure 2.4: Differential and cumulative dose-volume histograms for the spinal cord (blue) and clinical target volume (red). Adapted from [25].

at risk, plan evaluation is a subjective and time-consuming task, particularly when selecting among multiple plans. To assist radiation oncologists, various decision support tools of differing sophistication have been developed, which will be further discussed in section 2.2.

Patient Specific Quality Assurance

Dose discrepancies between the predicted dose distribution and the delivered dose by the linac might occur due to, for example, dose calculation inaccuracies or machine delivery errors [27]. IMRT plans are particularly susceptible to these discrepancies, prompting various professional organizations to strongly recommend detecting clinically relevant discrepancies before IMRT treatment delivery, in a process known as patient specific quality assurance (PSQA) [28, 29].

The measurement-based PSQA employs measurement systems such as ionization chambers, diode arrays, and the linac EPID to measure the dose distribution delivered by the linac [28]. The measured distribution is then compared to a dose distribution calculated for the measurement device based on the technical data of the treatment plan [28]. If the distributions agree within the institution-defined criteria based on international recommendations ([28], for instance), the plan is considered deliverable; if not, a medical physicist investigates the deviation and decides if replanning is warranted [28, 30, 31]. This comparison is often performed using the dose difference or the gamma index, which considers both dose difference and spatial agreement [32]. The percentage of measurement points within the gamma index criteria is termed gamma passing rate (GPR).

Treatment Delivery

In each treatment session, the patient is placed on the linac couch with the immobilization devices used during the planning CT. The couch position is then adjusted so that room-mounted lasers intersecting the linac isocenter align with previously tattooed dots on the patient's skin [16]. Following laser alignment, the treatment area is imaged using the linac kilovoltage system. Based on the misalignment between the planning CT and kilovoltage images, finer adjustments

are made to the patient's position by translating and rotating the linac couch [13, 33]. These steps ensure accurate and reproducible positioning between and during treatment sessions.

Follow-up

After treatment completion, the patient has several periodic medical examinations to evaluate overall health and the tumor's response. These appointments also allow for monitoring side effects and detecting any recurrences early [23].

2.2 Plan Quality

According to Hernandez et al. [34], plan quality is the clinical suitability of the delivered dose distribution that can be realistically expected from a treatment plan. Therefore, evaluating plan quality encompasses quantifying the performance of the dose distribution while addressing the susceptibility to discrepancies between the delivered and planned distribution. In this work, 'plan quality' directly refers to the dosimetric performance of the plan, while the plan susceptibility to discrepancies is evaluated separately by considering its plan complexity (section 2.3).

The dosimetric performance of a plan is mainly evaluated visually [35], which is inevitably subjective. To address this, various dose metrics, including dose-volume quality indices, have been proposed, focusing on three main aspects:

- Homogeneity: these metrics assess the uniformity of the dose distribution within the target volume. Usually, they are given by ratios of dose-volume metrics, such as the difference between the near-maximum and near-minimum doses over the median dose ($(D_{2\%} - D_{98\%})/D_{50\%}$) [4] or the maximum dose over the prescription dose $D_{\max}/D_{\text{presc}}$ [36]. Alternatively, homogeneity has been assessed using the standard deviation [37] and, more recently, the ideal DVH for target volumes [38], which is also applicable for plans with multiple dose levels.

- Conformity: these metrics quantify how well a reference isodose conforms to the shape of the target volume. Early conformity metrics focused on target coverage and thus were vulnerable to false positives when overdosing of surrounding healthy tissue [39]. To prevent misleading results, metrics such as the conformity number (CN) introduced by van't Riet et al. [40] consider the extent to which the reference isodose extends beyond the target volume. Later, Baltas et al. [41] proposed the conformity index (COIN), a CN modified to penalize the overlap between the reference isodose and organs at risk (OARs).

- Gradient: these metrics describe the steepness of the dose fall-off outside the target volume, indicating the extent of the dose spillage. The Radiation Therapy Oncology Group (RTOG) 0915 report on lung SBRT schedules [42] introduced the $R_{50\%}$ metric, defined as the ratio of the 50% isodose volume to the PTV volume. Similarly, Paddick and Lippitz [43] proposed the gradient index (GI) for radiosurgery plans, given by the ratio of the 50% isodose volume to the prescription isodose volume. No gradient metrics for conventional

radiotherapy were identified in the literature.

Many conformity and gradient metrics are highly dependent on the size and shape of the target. Some proposed metrics aim to minimize this dependence by considering the distances between volumes or isodoses [44–47]. However, these metrics are difficult to compute, as they cannot be solely derived from dose-volume metrics.

Some quality metrics, such as tumor control probability (TCP) and normal tissue complication probability (NTCP), aim to predict the biological impact of a treatment plan, providing the most direct measure of clinical suitability. However, these metrics are not widely used due to concerns about the accuracy of model parameters [48]. According to a 2020 survey by the European Society for Radiotherapy and Oncology (ESTRO) on clinical practices, only 5 out of 77 institutions (6.5%) reported using biological models when evaluating plan quality.

Plan quality can also be assessed using a single plan quality index (PQI) that incorporates multiple individual score functions for quality metrics and planning objectives [5, 49–54], with the majority using scoring functions based on either quotients or linear trends. These individual score values can sometimes be visually summarized in scorecards [49] or some implementation of radar plots [50, 52, 53].

2.2.1 SPIDERplan

In 2016, Ventura et al. [50] proposed the SPIDERplan, a graphical tool for quantifying and comparing plan quality between treatment plans, which is complemented with a PQI. Plans are processed in SPIDERplan using a predefined template containing two basic components: groups, which are a set of structures (OARs or PTVs) with an associated normalized weight, and structures, which are members of a group with an assigned objective and a normalized weight within the group. The template weights reflect the relative importance given by the radiation oncologist to the different objectives when they evaluate the dose distribution [50].

For a given template, each plan is plotted as a polygon in one of three custom radar plots:

- Groups Plan Diagram (GPD): Each group is represented as a section of the diagram with a central angle proportional to the group weight. For each section, the bisector represents the group score.

- Structures Plan Diagram (SPD): Each group's section is further subdivided to represent the individual structures within that group. The central angle within the group section is proportional to the structure weight and the bisector represents the structure score.

- Structures Group Diagram (SGD): A selected group section of the SPD is projected on the entire plot.

In its typical form, each structure is associated with a dose objective and has a score given by Equation 2.2 for PTVs and by Equation 2.3 for OARs, where D_{TC} is the dose tolerance criterion

and D_{P} is the planned dose. However, SPIDERplan allows the use of any metric and associated score function [50], such as conformity quality metrics [55]. A global plan score (GPS) is given by the weighted average of the structure's scores (Equation 3.34). For all scores, a value of 1 is deemed acceptable, with 0 representing the ideal case.

$$\mathrm{Score}_{\mathrm{PTV}} = \left(\frac{D_{\mathrm{TC}}}{D_{\mathrm{P}}}\right)_{\mathrm{PTV}} \tag{2.2}$$

$$\mathrm{Score}_{\mathrm{OAR}} = \left(\frac{D_{\mathrm{P}}}{D_{\mathrm{TC}}}\right)_{\mathrm{OAR}} \tag{2.3}$$

$$\text{Global Plan Score} = \sum_i w_{\mathrm{Group(i)}} \sum_j w_{\mathrm{Structure(i,j)}}\ \mathrm{Score}_{\mathrm{Structure(i,j)}} \tag{2.4}$$

SPIDERplan templates are robust to different group weights and definitions, although the robustness to different objectives was not evaluated [50, 55]

2.3 Plan Complexity

2.3.1 Definition

Despite the high plan quality often achieved with IMRT, discrepancies can arise between the planned dose distribution calculated by the Treatment Planning System (TPS) and the actual dose distribution delivered by the treatment unit [27]. These discrepancies can result from various factors, including patient setup errors, inaccurate beam modeling, inaccurate dose calculation, or machine delivery errors. The susceptibility to dose discrepancies due to dose calculation and machine delivery errors is referred to as complexity [34, 56, 57].

2.3.2 Complexity Metrics

Several metrics have been proposed to quantify IMRT plans complexity, based on identified sources of complexity such as aperture area [58], aperture irregularity [59], leaf movement [27], leaf speed [60], and dose at off-axis positions [58]. Many metrics were first proposed for step-and-shoot IMRT and later adapted to VMAT.

The first metrics focused on the modulation of photon fluence maps, under the premise that less uniform fluence maps are more difficult for the linear accelerator (linac) to accurately deliver, thereby increasing the complexity of the plan [61]. Nevertheless, these metrics have two major limitations: the linac can deliver the same fluence map with different parameter variations, which these metrics cannot discern, and fluence maps are not always available [34, 62]. Consequently, fluence map-based metrics have become less prevalent [34] and will not be further discussed in this work.

The most ubiquitous metric, due to its ease of acquisition, is the number of monitor units (MUs) of a plan, which is associated with small and irregular apertures [62]. Given that MUs are inherently dose-dependent, normalized variations are often employed [27, 62].

The majority of metrics, however, are based on the MLC aperture. Some of these include the modulation complexity score (MCS) proposed by McNiven et al. [63], which quantifies both aperture irregularity and aperture area modulation over the plan; the small aperture score (SAS) proposed by Crowe et al. [58], which is the fraction of open leaf pairs whose interval is less than a defined value; and the plan area (PA) proposed by Du et al. [62], which is the a weighted average aperture area over the treatment.

Less commonly, some metric quantify machine parameters other than the MLC configurations, such as dose rate variation and gantry acceleration. The total modulation index MIt [60] is one such example.

2.3.3 VMAT Adaptation

During VMAT optimization, the continuously delivered dose distribution over an arc is approximated by the TPS using 'control points', which include static MLC configurations at given gantry angles delivering a defined number of MUs [30, 64]. Based on this, Masi et al.[27] argued that it was reasonable to adapt the MCS metric to VMAT by using the MLC configurations encoded on control points as a replacement for IMRT segments. Following the introduction of MCS, many aperture-based complexity metrics adopted the same reasoning when applied to VMAT [58, 59, 62].

Generally, similar to the VMAT optimization, a number of MUs is assigned to each control point, with arc complexity being the average of the complexities of the individual control points, weighted by the number of MUs. However, some authors follow different approaches. Masi et al.[27] argue that, since monitor units are delivered between control points, arc complexity should weight the average complexity between adjacent control points. Gotstedt et al.[65] question whether MUs are the most appropriate weighting factor, as higher MU control points are often associated with smaller aperture areas, which contribute to a smaller irradiated volume. Instead, they apply an arithmetic average for their proposed EAM metric.

Certain complexity metrics are exclusive to VMAT plans, as they consider the treatment's dynamic nature. Examples include the total Modulation Index (MIt) [60], which considers the modulation of leaf speed, leaf accelaration, gantry speed and dose rate, and the leaf travel (LT) proposed by Masi et al. [27], which corresponds to the total distance traveled by the MLC leaves.

The metrics discussed until this point are designed to be easily adapted to every conventional linac. Non-conventional linac, however, such as the Varian Halcyon or the Accuray Tomotherapy, may require more substantial adaptations.

2.3.4 Halcyon Adaptation

Due to the dual-layer MLC present in the Halcyon linac, aperture-based metrics must be modified accordingly. Three approaches have been found in the literature [20, 66–70]:

1. The previously proposed metrics are applied to each MLC layer independently. [20, 69]

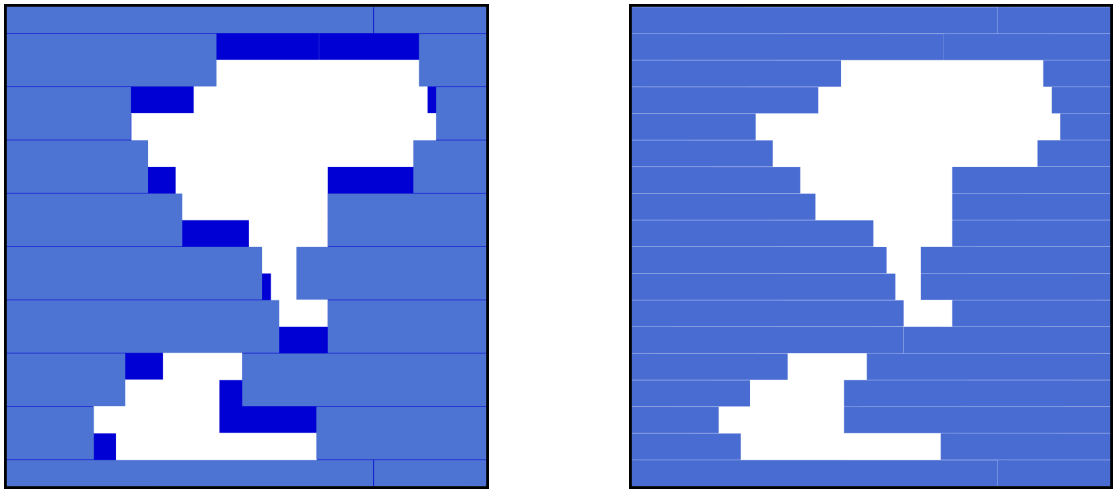
2. The previously proposed metrics are applied to the effective 5mm MLC aperture [20, 66–68]. In fact, the dual-layer MLC can be considered as an MLC with 5mm leaf width, where the leaf ends are constrained to be positioned between the two adjacent leaf ends.

3. The metrics applied to each MLC are averaged weighting the contribution of each MLC to the effective MLC aperture [20, 70].

Other metrics, however, have been proposed considering complexity sources exclusive to Halcyon. Quintero et al.[70] observed that certain segments of leaf pairs, denominated uncovered segments, are not covered by the complementary MLC layer at certain times during treatment. To quantify the presence of these uncovered segments, they developed the Uncovered-Layer Score (UL). The same behavior was also observed at IPOC, where they were further divided in two categories:

- Preparation segment: the uncovered segment will contribute to the beam aperture before closing. The uncovered segment might be stationary or moving towards a final contributing position.

- Unused segment: the uncovered segment closes without ever contributing to the beam aperture. They are generally larger than preparation segments and could be the vestige of a contributing segment in a previous iteration of the Direct Aperture Optimization algorithm used in the Varian TPS for VMAT planning.



(a) (b)

Figure 2.5: (a) Dual layer aperture defined by the proximal (light blue) and distal (dark blue) layers. (b) Associated 5 mm effective aperture.

2.3.5 Complexity Metrics Applications

Complexity metrics have found varied uses, of which three are presented in this subsection: PSQA results prediction, penalty in the cost function, and characterization.

PSQA Results Prediction

With higher complexity being more susceptible to dose discrepancies, accurate complexity assessment would allow us to predict PSQA results. This prediction can reduce workload by two means:

1. Plans with a higher probability of failing would prompt replanning before performing PSQA [71];

2. Plans with high probability of passing could skip PSQA altogether [27, 58].

One common approach involves establishing a complexity metric score threshold for a binary classifier of deliverable and undeliverable plans. Assuming that passing PSQA is the positive outcome, threshold values are usually chosen so that a specificity (point 1.) or a sensitivity (point 2.) of 100% is achieved.

Masi et al. [27] set a threshold value of 0.8 for the proposed leaf travel modulation complexity score (LTMCS) metric, a combination of the LT and MCS metrics previously presented, using a cohort of 142 VMAT plans, with the majority (N=94) being for prostate treatments. Prior to its clinical implementation, four months were devoted to validation, during which all nine of forty-one plans with LTMCS greater than 0.8 passed PSQA. Assuming that the class frequency remains constant, this corresponds to a 22% reduction in PSQA workload. Marsac et al. [72] defined a threshold value of 0.34 for MCS using a cohort of 299 pelvic plans and validating

it with an additional set of 145 plans. Over the course of a year, the team estimated that a 41% decrease in PSQA workload was achieved. These thresholds are dependent on variations in PSQA protocols, gamma index criteria, treatment site, treatment unit and TPS used [27, 31, 74]. As such, validation of reported values is required before clinical implementation.

Recently, machine learning models have been utilized and validated to predict gamma passing rate (GPR) values using an extensive number of complexity metrics. Mean absolute errors as low as 1.81 % for a gamma index with 3%/3mm criteria are reported [75]. However, this approach has two major limitations. Firstly, predicting lower GPR values is challenging due to highly imbalanced datasets that predominantly consist of high GPR plans. This imbalance, which persists even with stricter gamma index criteria [27, 60, 63], leads to low sensitivities. Lambri et al.[76] demonstrated that undersampling can enhance predictions for low GPR values, albeit at the cost of increasing the overall mean absolute error. Secondly, machine learning models must be trained or validated locally prior to clinical implementation, necessitating a large local database [76].

Penalty in Cost Function

As previously mentioned, the optimization process for inverse planning involves minimizing cost functions. Incorporating a complexity penalty into these functions can affect the final plan complexity and thus prevent replannings [71].

Younge et al. [59] integrated a complexity metric sensitive to irregular and small apertures as a penalty in their in-house TPS. By using appropriate penalty weights, they achieved VMAT plans with fewer monitor units, less complex apertures, and higher dose calculation accuracy. More recently, Varian incorporated a tool in the Eclipse TPS denominated Aperture Shape Control (ASC) that penalizes irregular apertures during optimization. The penalty weight can be set from the user from very low to very high. While this tool leads to less complex apertures, the impact on the number of MUs varies depending on the institution, treatment site, and treatment unit [70, 77, 78].

The impact of the penalty in the cost function on plan quality must be carefully evaluated, as high penalty weights may generate less favorable plans [59, 78].

Characterization

Complexity evaluation can be used to characterize local clinical practices, which is valuable for evaluating changes or updates to the TPS [56], introduction of new planning or treatment techniques, assessing a center's propensity for high-complexity plans during muli-institutional audits [31, 79, 80], and establishing limits to promote plan standardization within the clinic [9].

Such characterizations are particularly valuable given the substantial evidence suggesting an association between plan complexity and the planner's experience. Desai et al. [79] observed substantial complexity differences between clinics, even for the same patient anatomy, linac, TPS, and planning aims. Similar findings were reported by Santos et al. [80], although for a

smaller cohort. On the other hand, Lambri et al. [76], in turn, evaluated the complexity of total marrow and lymphoid irradiation plans over five years. These plans, primarily created by one professional, demonstrated a steady decrease in complexity over time, which suggests that experience influence plan complexity. They also showed statistically significant changes in plan complexity following a change of optimization algorithm within the TPS.

Higher complexities can also be a result of factors such as planning strategy, and conflicting or unrealistic planning objectives [71, 80, 81].

2.3.6 Correlation with Plan Quality

The correlation between plan quality and complexity has been evaluated in multiple institutional audits, with a general consensus that plan quality is not correlated with complexity [5, 80, 82]. As an example, Villaggi et al. [83] were able to increase the plan quality of less experienced planners without increasing complexity after sharing median DVHs. Conversely, McGarry et al. [81] found a moderate correlation between plan quality and MUs, although only when using TPSs from a company other than the linac manufacturer.

Given the general lack of a strong correlation between complexity and plan quality, efforts should be made to reduce complexity as much as possible to minimize the probability of dose delivery discrepancies. Additionally, a decrease in MUs is associated with shorter treatment times, which enhance patient comfort and reduce errors due to organ motion during treatment [59, 63]. Lower MUs also reduce collimator radiation leakage, potentially lowering the probability of radiation-induced malignancies [84].

2.4 Knowledge-Based Planning

Automated planning solutions using artificial intelligence have gained traction recently due to their proven ability to reduce variability between planners and decrease planning time while maintaining or improving plan quality [85–88]. One class of these solutions is knowledge-based planning (KBP), where a model trained on prior treatment plans is used to predict the dose distribution or the dose-volume histograms (DVHs) for delineated structures and automatically generate planning objectives tailored to the patient's anatomy. The Varian Eclipse KBP solution, intended for Halcyon plans at IPO Coimbra, is the Varian RapidPlan.

2.4.1 RapidPlan

RapidPlan employs principal component analysis to establish correlations between geometrical and dosimetric features for each structure within the model, thereby determining a prediction range for the dose-volume histograms (DVHs) of these structures (Figure 2.6). Based on the predicted DVHs, RapidPlan can be set to generate five types of planning objectives: upper and lower dose bounds, median dose values, generalized equivalent uniform dose (gEUD), and line objectives, which are a continuous set of upper dose bounds positioned slightly below the estimated DVH prediction range (Figure 2.6) [10]. Additionally, RapidPlan assigns objective

weights (priorities) in the cost function inversely proportional to the prediction uncertainty of the structure DVH [89].

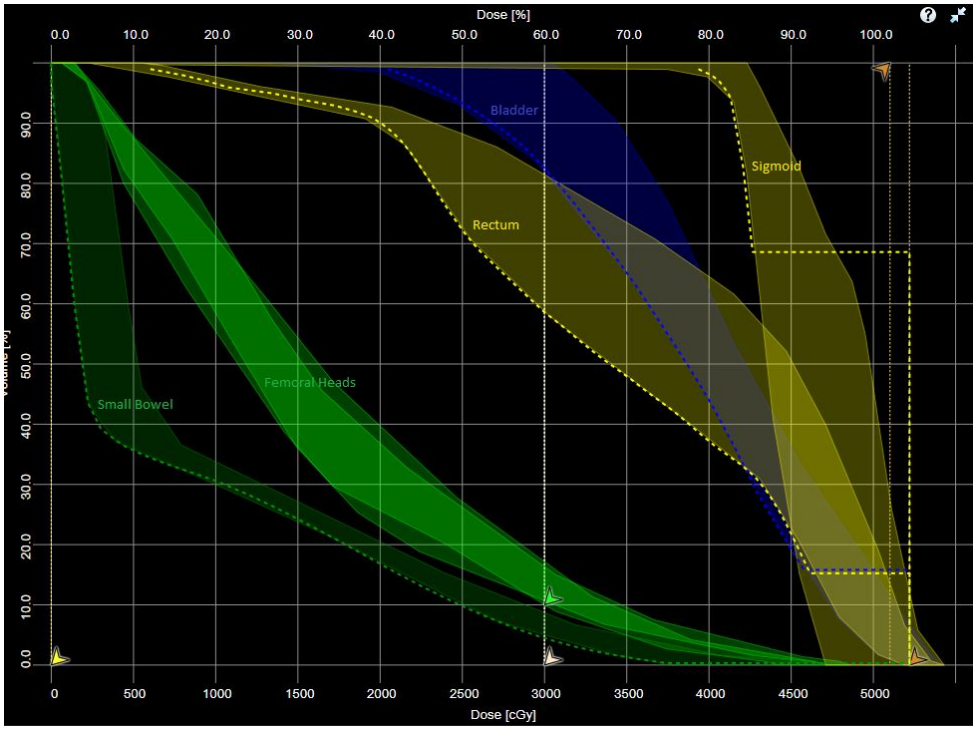


Figure 2.6: Example of RapidPlan DVH prediction ranges for the small bowel, femoral heads, rectum, bladder and sigmoid. The dashed lines represent the line objectives. Adapted from [90]

While RapidPlan allows for manual adjustments of planning objectives and objective priorities during optimization, it is recommended to reduce these interventions to maximize the benefits of KBP [91]. For instance, Kubo et al. [54] noted differences between novice and experienced planners when manual adjustments of objective priorities were permitted. Nonetheless, complicated cases may require manual intervention to achieve clinically acceptable solutions [54, 92].

Training

The efficacy of RapidPlan highly depends on both the dosimetric and geometric features of the training set, particularly the quality of the plans and the diversity of anatomical representations included [6, 10, 89]. Therefore, a thorough approach to plan selection is required, which can be very time-consuming.

To assist in model training, a Varian cloud service named Model Analytics identifies gaps in these features that hinder the prediction power of the model and flags plans with geometric or dosimetric outliers [89, 91]. While the presence of outliers is generally associated with higher prediction uncertainties, their removal must be done cautiously, as it may result in a lack of representation of certain geometric features in the model library and potentially worsen the resulting DVH predictions and subsequent plan quality [89]. Fusella et al. [93] found that using their plan quality (PQI) to filter low-quality plans for model training yielded similar or better results compared to iterative outlier removal. As noted by the authors, their PQI definition may

not generalize to different PQIs, it remains a promising approach for future studies.

Alternatively, the RapidPlan model may be trained on treatment plans provided by the manufacturer [6]. However, plans created with RapidPlan in such case may not be representative of the local clinical practice.

Complexity of KBP Plans

Three papers were found in the literature regarding the impact of the implementation of the RapidPlan KBP solution on plan complexity:

- Fanou et al., [7] found no significant changes after implementing RapidPlan for head and neck cases treated with Varian VitalBeam linacs. However, manual adjustments were performed during the second iteration of optimization.

- Lambri et al. [8] found no significant changes after implementing RapidPlan for total marrow and lymphoid irradiation cases treated with a Varian TrueBeam linac. However, extensive manual adjustments to the objectives and objective weights were made, and the cohort of plans created with RapidPlan at the time of publication was limited to only eight plans.

- Scaggion et al. [9] found a slight increase in linac mechanical strain, as measured by the MIt metric, after implementing RapidPlan for prostate cases treated with a Varian Unique linac. Extensive manual adjustments to the objectives and objective weights were also performed.

Overall, plans created with the RapidPlan KBP are comparable to previously clinical plans in terms of plan quality and complexity. It must be noted that all three papers reported systematic planner intervention during optimization. No studies were found using complexity as a filtering criterion for the selection of plans for KBP models, warranting further investigation.

3 Materials and Methods

3.1 qualComp

With the primary goal of streamlining the assessment and comparison of radiotherapy treatment plans regarding plan complexity and quality, an information system called qualComp was developed in MATLAB R2023b for this thesis and for potential use in clinical practice.

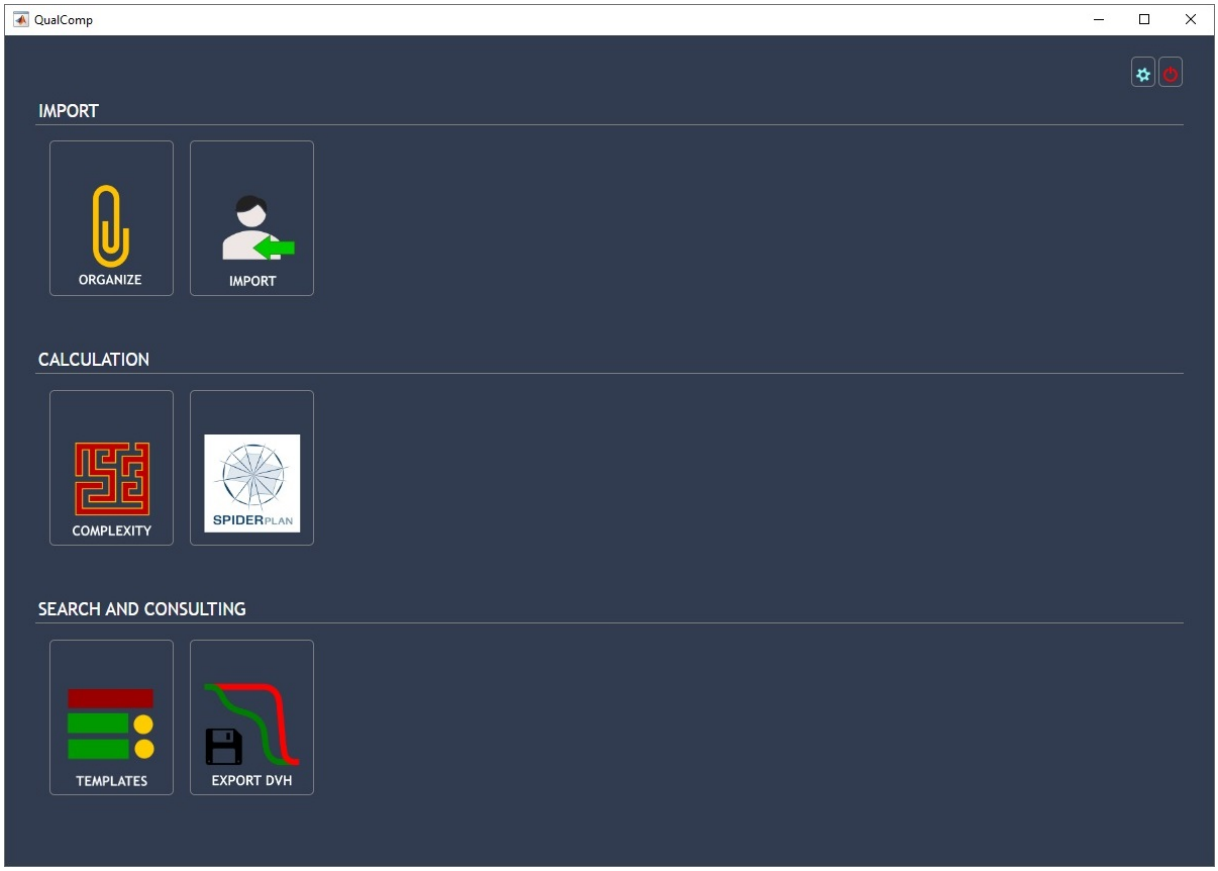


Figure 3.1: qualComp Main Window

Currently, qualComp includes three modules (Figure 3.1), allowing for: data importation and processing (Import module), plan complexity metrics computation and plan quality evaluation using SPIDERplan (Calculation module), and definition of SPIDERplan templates and the exportation of processed DVH data (Search and Consulting module). Complexity evaluation is currently implemented only for Varian Halcyon plans. However, all other features also support

plans from Varian TrueBeam and Accuray Tomotherapy.

One main design principle of qualComp was user-friendliness, ensuring all graphical user interfaces (GUI) provide feedback indicating the reason certain actions are invalid or cannot be performed. Furthermore, efforts were made so that data uniqueness, validity, consistency, and completeness is not dependent on the user's familiarity with the software.

All qualComp data is stored in a local file system, here referred as the qualComp file system.

3.1.1 Import Module

The data imported into qualComp include the dose-volume histogram (DVH) file for each delineated structure of the plan and the radiotherapy plan file from the Digital Imaging and Communications in Medicine (DICOM) standard (DICOM RT Plan). The DVH file is utilized for evaluating plan quality in SPIDERplan, while the DICOM RT Plan is used for analyzing plan complexity, providing technical details for treatment delivery such as MLC configurations and gantry angles. Both the DVH file and the DICOM RT Plan are exported from the treatment planning system (TPS) for each treatment plan.

The data importation involves two main steps corresponding to the Organize and Import sub-modules.

Organize Sub-module

In this sub-module, the files placed in the designated importation folder are stored in the qualComp file system, organized by patient OBS and plan name.

The format and content of the files is also validated. In particular, DVHs from the Varian TPS must be differential in order to compute the structure mean dose. DVHs exported from Accuray Precision are cumulative and the mean dose for each structure is separately reported within the file.

Import Sub-module

After organizing, the plan is available in the search GUI of the Import sub-module (Figure 3.2). In the GUI, all patients available for import are displayed in the left panel. When a patient is selected, the right panel lists the plans associated with that patient, as well as the corresponding DVH (.txt or .csv) and DICOM files (.dcm). The search bar at the top allows filtering the patient list by name or ID (OBS). At the bottom, six action buttons are arranged from left to right with the following functionalities:

- Refresh: the patient list is updated.
- Delete Patient: the selected patient is removed from the qualComp file system.
- Delete Plan: the selected plan is removed from the qualComp file system.

- Move DVH: allows the user to move DVH files between plans. This action is particularly useful for Accuray Precision plans, where long plan names registered in the DVH file can get truncated at the end, leading the DVH file to not get associated with the corresponding DICOM file.

- Cancel: closes the window.

- Import: proceeds to the next step if both DVH and DICOM files are present.

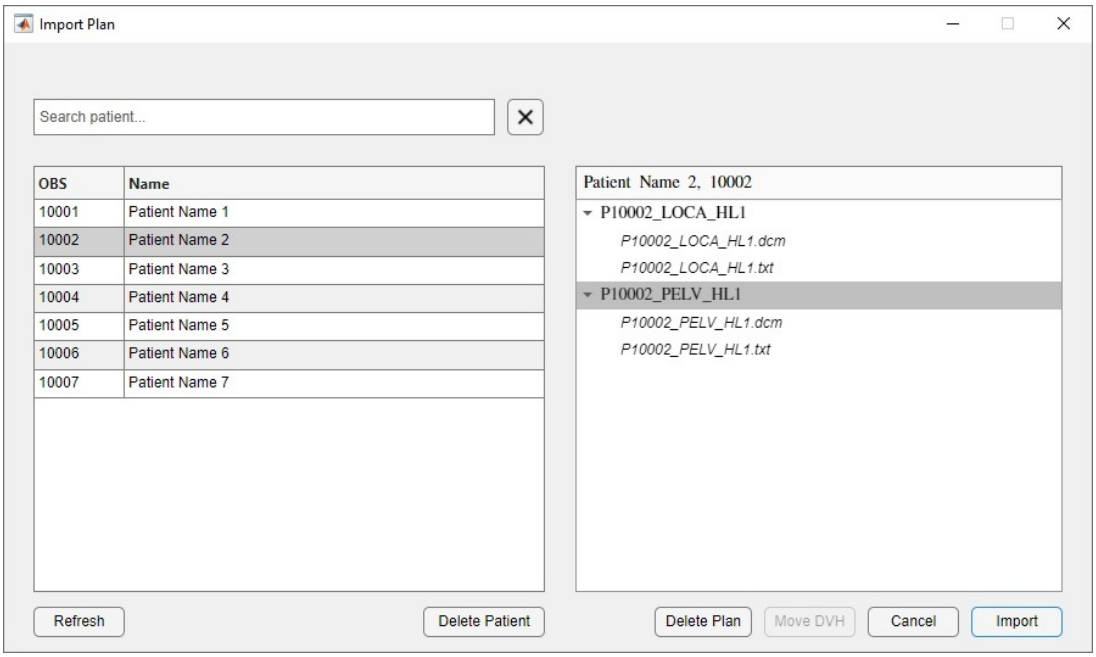


Figure 3.2: qualComp Search GUI

After hitting 'Import' button, a Structures Selection GUI (Figure 3.3) opens. In this GUI, the user fills in specific plan information and selects the delineated structures that are relevant for quality evaluation.

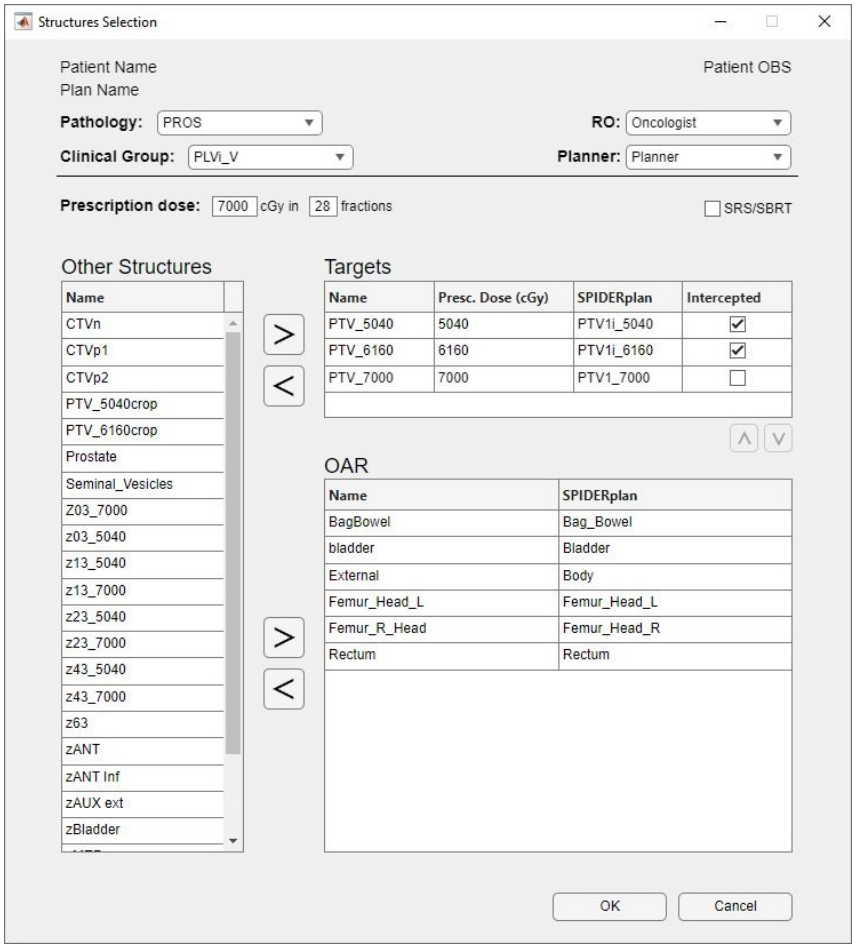


Figure 3.3: qualComp Structures Selection GUI

The patient name, OBS, and plan name are displayed at the top. Below these details, the user selects the pathology (e.g., prostate), clinical group (e.g., prostate bed plans), radiation oncologist (RO), and describes the fractionation. When possible, some of these fields are automatically filled from data in the DVH or DICOM files. There are three tables with structures in the GUI:

- Other Structures: all structures are placed in this table by default, identified by their names in the DVH file. Structures in this table will not be used for plan quality evaluation.

- Targets: contains all the plan targets and includes four columns, with the second and the fourth being editable. The first column contains the structure's name as recorded in the DVH file, the second column the prescribed dose, the third column the name by which the structure will be recorded, and the fourth column indicates whether the target is intercepted by other targets. Target names are standardized to the format 'PTVx_D', where x differentiates between targets with the same dose, ordered as listed in the table and D represents the prescribed dose in cGy. If a target intersects with others, an 'i' is added to its name, resulting in the format 'PTVxi_D'.

- OARs: contains all OARs relevant for plan quality evaluation. The first column contains the names of the structures in the DVH file. In the second column the user selects the

name by which the OAR will be stored from a dropdown list of protocoled structure names. The naming protocol at IPO Coimbra adheres to the nomenclature guidelines set by Task Group 263 of the American Association of Physicists in Medicine (AAPM) [94].

Each structure can be moved between tables using the arrow buttons ('>' and '<'). However, most structures are automatically placed in the respective tables. Additionally, each OAR is associated with a protocoled structure name based on prior associations. As an example, the structure 'Head_Femur_Right' in Figure 3.3 was recognized as 'Femur_Head_R' because this association had already been done for a previously imported plan. With these automation measures, users typically only need to fill in the pathology and clinical group fields, as well as the second and fourth columns of the target table.

At the end of the import task, the plan is saved in a Patient File within the file system, containing dosimetric and treatment delivery information (automatically extracted from the DICOM RT Plan) for all treatment plans of that patient. Dosimetric information includes DVHs, the volumes of the structures, and some dose-volume metrics (e.g., $D_{2\%}$, $D_{98\%}$ and mean dose), adhering to the recording recommendations of the ICRU 83 [4] for IMRT plans. Delivery information includes all necessary parameters for complexity evaluation, such as gantry angle, MLC leaves position, and monitor units.

3.1.2 Calculation Module - Complexity

In this sub-module, which is only implemented for Halcyon, fifteen complexity metrics proposed in the literature were implemented are computed automatically for plans in the qualComp file system. The computed metrics evaluate different complexity aspects and include:

- MLC aperture irregularity and variation: modulation complexity score (MCS)[27]
- Aperture area variation: aperture area variability (AAV)[27] and plan modulation (PM)[62]
- MLC aperture area: plan average (PA)[62], and small aperture score (SAS)[58].
- MLC aperture irregularity: leaf sequence variability (LSV)[27] and plan irregularity (PI)[62]
- Closed leafs inside the field: close leaf score (CLS)[58]
- Off-axis irradiation: cross axis score (CAS)[58] and mean asymmetry displacement (MAD)[58]
- MLC aperture area and irregularity: edge area metric (EAM)[57] and edge metric (EM)[59]
- Modulation of the machine parameters: leaf travel (LT) [27] and total modulation index (MIt) [60]
- Plan efficiency: monitor units per fraction dose (MUd)

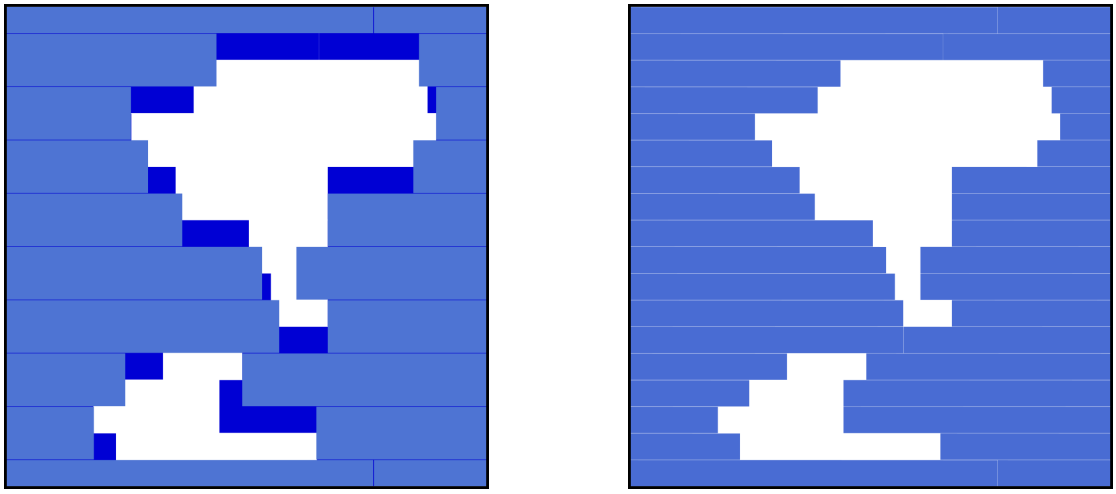
Delivery parameters required for complexity computation, such as MLC leaves position, monitor

units, and gantry angles are retrieved for each control point from the qualComp Patient files. The position of the leaves is recorded at the isocenter plane using the MLC vertical midline as a reference. Leaf ends to the left of the midline have negative positions and leaf ends to the right of the midline have positive positions.

A summary of the nomenclature used in the equations within this section is provided at the end of the section (Table 3.1).

Halcyon adaptation

Aperture-based metrics proposed in the literature for IMRT/VMAT plans cannot be directly implemented for Halcyon, given its dual-layer MLC. The most common reported implementation approach considers a 5 mm effective aperture formed by the two layers (Figure 3.4) [20, 66–68]. However, when two opposite leaves from different layers intersect, the two separate apertures (proximal and distal) can be mapped into multiple effective apertures. In this work, the ambiguity was resolved by positioning the effective leaf end halfway between the two leaf ends (Figure 3.5).



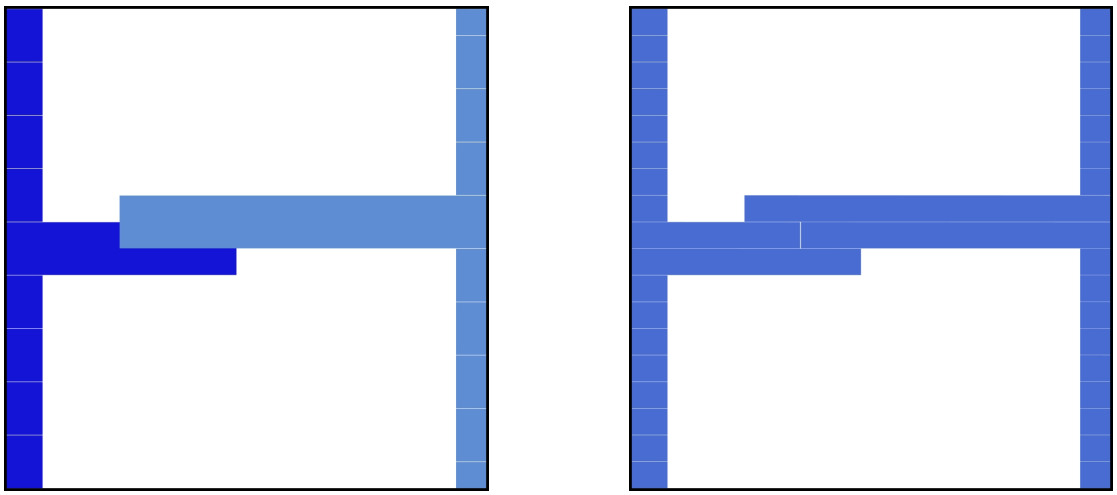
(a) (b)

Figure 3.4: (a) Dual layer aperture defined by the proximal (light blue) and distal (dark blue) layers. (b) Associated 5 mm effective aperture.

Another approach was proposed by Tamura et al. [20], where metrics are calculated for each layer individually and then combined weighting the relative contribution of each layer of the MLC to the effective aperture. These weighted metrics are identified by the suffix 'w' added to the original metric names, while metrics applied to the proximal and distal layers are identified by the prefixes 'p' and 'd', respectively. Tamura et al. employed both weighted and single layer implementations for the MCS, PA, PI, and PM metrics, which was replicated in this work.

Some aperture-based metrics proposed in the literature only consider leaf pairs not covered by the jaws. Given that the Halcyon linac does not have jaws, it was considered in this work that, at each control point, any leaf pairs outside the contour of the effective MLC are as if they were

covered by 'virtual jaws' (Figure 3.6).



(a) (b)

Figure 3.5: (a) Example of a MLC configuration that can be associated with multiple effective apertures. Proximal layers is depicted in light blue and the distal layer in dark blue. (b) When two opposite leaves from different layers intersect, the effective leaf position in the intersection is the halfway point between the ends of the two intersecting leaves.

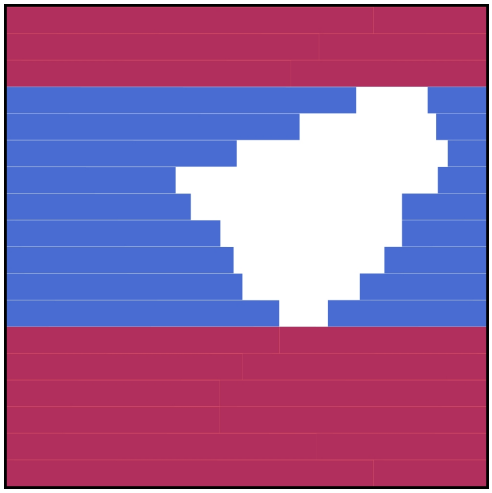


Figure 3.6: Virtual jaws depicted in red for an effective MLC aperture.

MCS, AAV, LSV

The modulation complexity score (MCS) was developed by McNiven et al. [63] for step-and-shoot IMRT and later adapted for VMAT by Masi et al. [27]. MCS is calculated, for each arc, combining the leaf sequence variability (LSV) and Aperture area variability (AAV) indices.

The Leaf Sequence Variability (LSV) (equation 3.1) considers, for each control point, the average difference in leaf ends positions for each bank relative to the maximum possible change in the control point given by equation 3.2.

$$\mathrm{LSV}_{ac} = \left(1 - \frac{\sum_{l=1}^{L_{ac}-1} \left|\mathrm{pos}_{acl} - \mathrm{pos}_{ac,l+1}\right|}{(L_{ac}-1)\cdot \mathrm{pos}_{\mathrm{max},ac}}\right)_{\mathrm{left}} \cdot \left(1 - \frac{\sum_{l=1}^{L_{ac}-1} \left|\mathrm{pos}_{acl} - \mathrm{pos}_{ac,l+1}\right|}{(L_{ac}-1)\cdot \mathrm{pos}_{\mathrm{max},ac}}\right)_{\mathrm{right}} \tag{3.1}$$

$$\left(\mathrm{pos}_{\mathrm{max},ac}\right)_{\mathrm{bank}} = \left(\max\{\mathrm{pos}_{acl}\}_{l=1}^{L_{ac}} - \min\{\mathrm{pos}_{acl}\}_{l=1}^{L_{ac}}\right)_{\mathrm{bank}} \tag{3.2}$$

In equations 3.1 and 3.2, a and c are the arc and control point indexes, L_{ac} is the number of bank leaves not covered by jaws at arc a and control point c, and pos_{acl} is the leaf end position at arc a, control point c, and bank leaf l.

Aperture Area Variability (AAV) is, for each control point, the ratio of the aperture area (AA_{ac}) to the area defined by the farthest position from the center that each leaf end took over the arc.

$$\mathrm{AAV}_{ac} = \frac{\mathrm{AA}_{ac}}{d\sum_{l=1}^{L_{ac}}(\max\{\mathrm{pos}_{acl}\}_{c=1}^{C_b})_{\mathrm{right}} - (\min\{\mathrm{pos}_{acl}\}_{c=1}^{C_b})_{\mathrm{left}}} \tag{3.3}$$

The MCS for a control point is then obtained by multiplying the average values of LSV and AAV between the current control point and the previous one (equation 3.4).

$$\mathrm{MCS}_{ac} = \frac{\mathrm{AAV}_{ac} + \mathrm{AAV}_{a,c-1}}{2} \cdot \frac{\mathrm{LSV}_{ac} + \mathrm{LSV}_{a,c-1}}{2} \tag{3.4}$$

Finally, the MCS for the treatment plan is determined by the average control point MCS weighted by the monitor units delivered between control points ($\mathrm{MU}_{ac,c-1}$) as given by equation 3.5. In equation 3.5, MU_b is the total number of monitor units delivered in arc a, MU_p is the total number of monitor units delivered in the plan, A is the number of arcs in the plan, and C_b is the number of control points for arc a.

$$\mathrm{MCS} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \cdot \mathrm{MCS}_{ac} \tag{3.5}$$

LSV and AAV are´ determined for the entire treatment plan by solely considering their associated components in MCS (equations 3.6 and 3.7).

$$\mathrm{LSV} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \cdot \frac{\mathrm{LSV}_{ac} + \mathrm{LSV}_{a,c-1}}{2} \tag{3.6}$$

$$\mathrm{AAV} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \cdot \frac{\mathrm{AAV}_{ac} + \mathrm{AAV}_{a,c-1}}{2} \tag{3.7}$$

The weighted implementation of MCS (MCSw) is obtained by extending the equation 3.5 to include the MCS calculated for each layer (pMCS and dMCS) and each layer contribution to the effective aperture (w_p and w_d), as given by equation 3.8. The contribution of a layer is quantified as the ratio of the layer leaf ends that contribute to the effective 5 mm aperture.

$$\mathrm{MCSw}_b = \sum_{c=2}^{C_{ac}} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \left(\mathrm{pMCS}_{ac} \cdot (w_p)_{ac} + \mathrm{dMCS}_{ac} \cdot (w_d)_{ac}\right) \tag{3.8}$$

LSV, AAV, MCS, and MCSw range from 0 to 1, with lower values indicating more complex plans.

PM

Plan Modulation (PM) was proposed by Du et al. [62] to quantify the aperture area modulation. In each control point the relative aperture area compared to the union area of all apertures of the arc ($U(\mathrm{AA})$) is obtained, with PM being the complement of the average relative area weighted by the monitor units delivered between control points (equation 3.9). PM ranges from 0 to 1, with lower PMs indicating plans with lower complexity.

$$\mathrm{PM} = 1 - \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU_b}} \cdot \frac{\mathrm{AA}_{ac}}{U(\mathrm{AA})_b} \tag{3.9}$$

The weighted version of PM (PMw) is obtained by extending the equation 3.9 to include the modulation of each layer and their contribution to the effective aperture (equation 3.10).

$$\mathrm{PMw} = 1 - \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \left(\frac{\mathrm{pAA}_{ac}}{U(\mathrm{AA})_a} \cdot (w_p)_{ac} + \frac{\mathrm{dAA}_{ac}}{U(\mathrm{AA})_a} \cdot (w_d)_{ac}\right) \tag{3.10}$$

PA

Plan area (PA) was proposed by Du et al. [62] as the average aperture area of the treatment plan weighted by the monitor units delivered between control points (equation 3.11). Lower PA values indicate more complex plans.

$$\mathrm{PA} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \cdot \mathrm{AA}_{ac} \tag{3.11}$$

The weighted version of PA (PAw) is obtained by extending equation 3.11 to include the aperture area of each layer and their contribution to the effective aperture (equation 3.12).

$$\mathrm{PAw} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \left(\mathrm{pAA}_{ac} \cdot (w_p)_{ac} + \mathrm{dAA}_{ac} \cdot (w_d)_{ac}\right) \tag{3.12}$$

SAS

Small Aperture Score (SAS) was proposed by Crowe et al. [58] as the average ratio of open leaf pairs whose leaf ends are within a defined distance f from each other, weighted by the monitor units delivered between control points (equation 3.13). SAS ranges from 0 to 1, with lower SAS values indicating less complex plans.

$$\mathrm{SAS}_f = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU_b}} \cdot \frac{|\{l : 0 < (\mathrm{pos}_{acl})_{\mathrm{right}} - (\mathrm{pos}_{acl})_{\mathrm{left}} < f\}|}{|\{l : (\mathrm{pos}_{acl})_{\mathrm{right}} - (\mathrm{pos}_{acl})_{\mathrm{left}} > 0\}|} \tag{3.13}$$

In this work, SAS was calculated for a distance criterion of 20 mm.

MUd

MUd is calculated as the number of monitor units delivered for each cGy of the fraction dose. Lower MUd indicate less complex plans.

PI

Plan irregularity (PI) was proposed by Du et al. [62] as the average aperture irregularity of the treatment plan weighted by the monitor units delivered between control points (equation 3.14). In equation 3.14, AP is the aperture perimeter. PI is 1 for circular apertures and higher for other shapes, as stated by the two dimensional isoperimetric inequality [95]. Not being able to generate circular apertures, the lowest PI value for a MLC is $4/\pi \approx 1.27$, obtained for square apertures. Lower PIs indicate less complex plans.

$$\mathrm{PI} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \cdot \frac{\mathrm{AP}_{ac}^2}{4\pi \mathrm{AA}_{ac}} \tag{3.14}$$

The weighted version of PI (PIw) is obtained by extending the equation 3.14 to include the aperture irregularity of each layer and their contribution to the effective aperture (equation 3.15).

$$\mathrm{PIw} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU}_b} \left(\frac{\mathrm{pAP}_{ac}^2}{4\pi \mathrm{pAA}_{ac}} \cdot (w_p)_{ac} + \frac{\mathrm{dAP}_{ac}^2}{4\pi \mathrm{dAA}_{ac}} \cdot (w_d)_{ac}\right) \tag{3.15}$$

CLS

Close Leaf Score (CLS) was proposed by Crowe et al. [58] as the average ratio of closed leaf pairs not covered by the jaws, weighted by the monitor units delivered between control points,

as given by equation 3.16. In equation 3.16, the vertical bars | | are the cardinality of the set they enclose. CLS ranges from 0 to 1, with lower CLS values indicating less complex plans.

$$\mathrm{CLS} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU_b}} \cdot \frac{|\{l : (\mathrm{pos}_{acl})_{\mathrm{right}} - (\mathrm{pos}_{acl})_{\mathrm{left}} = 0\}|}{L_{ac}} \tag{3.16}$$

CAS

Cross Axis Score (CAS) was proposed by Crowe et al. [58] as the average ratio of open leaf pairs where one of the leaves crosses the midline position (pos = 0), weighted by the monitor units delivered between control points. Considering the leaf end positions, the prior definition is equivalent to the ratio of leaf pairs whose leaf ends are both on the same side of the MLC midline (equation 3.17). In equation 3.17, sgn is the sign function. CAS ranges from 0 to 1, with lower CAS values indicating less complex plans.

$$\mathrm{CAS} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU_b}} \cdot \frac{|\{l : \mathrm{sgn}(\mathrm{pos}_{acl})_{\mathrm{right}} = \mathrm{sgn}(\mathrm{pos}_{acl})_{\mathrm{left}}\}|}{|\{l : (\mathrm{pos}_{acl})_{\mathrm{right}} - (\mathrm{pos}_{acl})_{\mathrm{left}} > 0\}|} \tag{3.17}$$

MAD

Mean Asymmetry Displacement (MAD) was proposed by Crowe et al. [58] as the plan mean distance from the open leaf pairs center at a control point to the MLC midline, weighted by the monitor units delivered between control points. Lower MAD values indicate less complex plans.

$$\mathrm{MAD} = \sum_{a=1}^{A} \frac{\mathrm{MU}_b}{\mathrm{MU}_p} \sum_{c=2}^{C_b} \frac{\mathrm{MU}_{ac,c-1}}{\mathrm{MU_b}} \sum_{l=1}^{L_{ac}} \frac{|(\mathrm{pos}_{acl})_{\mathrm{right}} - (\mathrm{pos}_{acl})_{\mathrm{left}}|}{2L_{ac}} \tag{3.18}$$

EAM

The Edge Area Metric (EAM), proposed by Gotstedt et al. [57], is the plan average ratio of complex regions in the MLC opening weighted by the monitor units delivered between control points (equation 3.19). Complex regions (R_1) are the outwards and inwards extension of the MLC perimeters by a defined distance, while non-complex regions (R_2) are beam opening areas not covered by complex regions (Figure 3.7). It can be considered as a measure of aperture irregularity that is also sensitive to the aperture area. In this work, a distance criterion of 2.5 mm was considered, so that R_1 encompasses the entirety of 5 mm-wide MLC openings, corresponding to the effective MLC leaf width. However, the criterion used holds little significance as, despite altering the amplitude of obtained values, it does not substantially influence the ranking of EAM values [65]. EAM ranges from 0 to 1, with lower EAM values indicating plans with lower complexities.

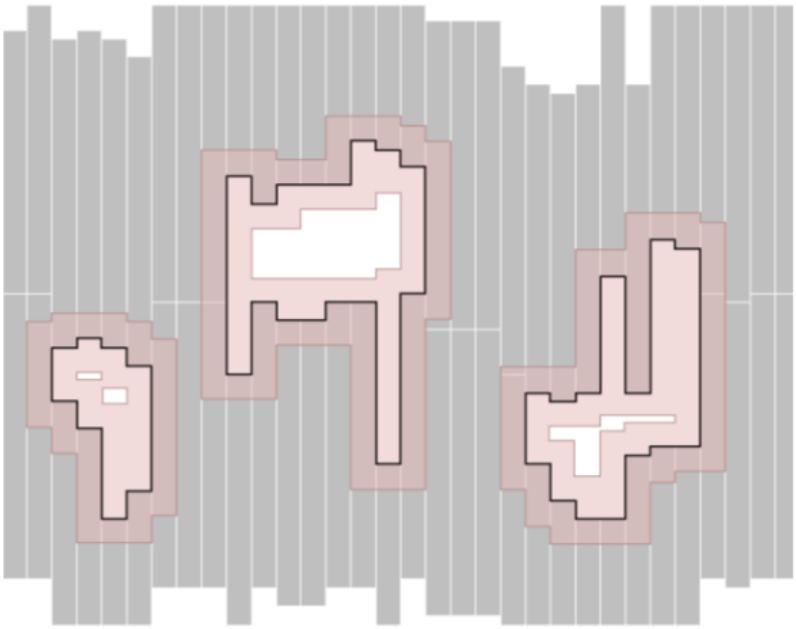


Figure 3.7: Complex and non-complex regions in EAM computation. The MLC perimeter is accentuated by a black line, the complex region R_1 is depicted by the red area extending from both sides of the MLC perimeter, and the non-complex region R_2 is the white regions not covered by R_1. Adapted from [65].

$$\text{EAM} = \frac{1}{AC} \sum_{a=1}^{A} \sum_{c=1}^{C_b} \left(\frac{R_1}{R_1 + R_2} \right)_{ac} \tag{3.19}$$

EM

Edge Metric (EM) was proposed by Younge et al. [59] as the average ratio of leaf edge perimeter to aperture area, weighted by the monitor units delivered between control points, as given by equation 3.20. As with EAM, EM can be considered as a measure of aperture irregularity that is also sensitive to the aperture area.In equation 3.20, d is the width of the MLC leaves. The leaf edge perimeter can be determined by subtracting the leaf end perimeter ($2d$) of all open leaf pairs from the total aperture perimeter. Lower EM values indicate less complex plans.

$$\text{EM} = \sum_{a=1}^{A} \frac{\text{MU}_b}{\text{MU}_p} \sum_{c=2}^{C_b} \frac{\text{MU}_{ac,c-1}}{\text{MU}_\text{b}} \left(\frac{\text{AP}_{ac} - 2d \cdot |\{l : (\text{pos}_{acl})_\text{right} - (\text{pos}_{acl})_\text{left} > 0\}|}{\text{AA}_{ac}} \right) \tag{3.20}$$

LT

Leaf Travel (LT) was proposed by Masi et al. [96] as the mean distance traveled by MLC leaves over the plan (equation 3.21). Lower LT values indicate less complex plans.

$$\mathrm{LT} = \frac{1}{A\mathcal{L}} \sum_{a=1}^{A} \sum_{\ell=1}^{\mathcal{L}} \sum_{c=2}^{C_b} |\mathrm{pos}_{ac\ell} - \mathrm{pos}_{a,c-1,\ell}| \tag{3.21}$$

MIt

MIt was proposed by Park et al. [60] to quantify the variation on leaf speed, leaf acceleration, gantry speed, and dose rate over the plan. These parameters are estimated by equations 3.22, 3.23, 3.24, and 3.25, respectably . Lower MIt values indicate less complex plans.

$$\mathrm{speed}_{ac\ell} = \frac{|\mathrm{pos}_{a,c+1,\ell} - \mathrm{pos}_{ac\ell}|}{\mathrm{Time}_{ac,c+1}} \tag{3.22}$$

$$\mathrm{accel}_{ac\ell} = \frac{|\mathrm{speed}_{a,c+1,\ell} - \mathrm{speed}_{ac\ell}|}{\mathrm{Time}_{ac,c+1}} \tag{3.23}$$

$$\mathrm{GA}_{ac} = \left| \frac{2^\circ}{\mathrm{Time}_{a,c+1,c+2}} - \frac{2^\circ}{\mathrm{Time}_{ac,c+1}} \right| \tag{3.24}$$

$$\mathrm{DRV}_{ac} = |\mathrm{DR}_{a,c+1} - \mathrm{DR}_{ac}| \tag{3.25}$$

The time elapsed between control points is not recorded in DICOM RT Plan files since it is determined during delivery by the linac control system. According to the manufacturer, the Halcyon linac sets its gantry speed at the maximum value of 2 rotations per minute while modulating the dose rate, unless the monitor units to be delivered between control points reach a threshold where slowing the gantry is required. In that case, the dose rate is set at the maximum value of 600 MU/s, and the gantry speed is modulated accordingly. This information allows to predict the time elapsed between control points with equation 3.26, where the maximum number of MUs deliverable without slowing the gantry was determined to be 1.667 MU.

$$\mathrm{Time}_{ac} = \begin{cases} \frac{2^\circ}{12^\circ\,\mathrm{s}^{-1}} & \mathrm{MU}_{ac,c-1} \leq 1.667\,\mathrm{MU} \\ \frac{\mathrm{MU}_{ac,c-1}}{10\,\mathrm{MUs}^{-1}} & \mathrm{MU}_{ac,c-1} > 1.667\,\mathrm{MU} \end{cases} \tag{3.26}$$

For each leaf, the function $z_{a\ell}(f)$ (equation 3.27) is defined, which can be seen as the ratio of control points in an arc for which the speed or acceleration are higher than a defined threshold, but also penalized at each control point by the variation of gantry speed and dose rate. For leaf speed, the threshold is f times the standard deviation (σ) of leaf speeds over the arc, while the threshold for acceleration is f times the standard deviation of leaf accelerations over the arc divided by the time elapsed between control points. The penalizations, $W_{GA,ac}$ and $W_{DRV,ac}$, have a maximum value of β and converge to β with a rate defined by γ according to equations

3.28 and 3.29. In equation 3.27, the square brackets $[x]$ represents the Iverson bracket [97], which is 1 if the condition x is true, and 0 otherwise.

$$z_{a\ell}(f) = \frac{1}{C-2} \sum_{c=1}^{C-2} W_{\mathrm{GA},ac} \cdot W_{\mathrm{DRV},ac} \cdot \left[\mathrm{speed}_{ac\ell} > f \cdot (\sigma_{\mathrm{speed}})_\ell \vee \mathrm{accel}_{ac\ell} > \frac{f \cdot (\sigma_{\mathrm{accel}})_\ell}{\mathrm{Time}_{ac,c+1}} \right] \quad (3.27)$$

$$W_{\mathrm{GA},ac} = \frac{\beta}{1 + (\beta - 1) \exp\left(\frac{-\mathrm{GA}_{ac}}{\gamma}\right)} \quad (3.28)$$

$$W_{\mathrm{DRV},ac} = \frac{\beta}{1 + (\beta - 1) \exp\left(\frac{-\mathrm{DRV}_{ac}}{\gamma}\right)} \quad (3.29)$$

The MIt for an arc a is the sum of the integrals of $z_{a\ell}(f)$ from 0 to a defined value F for each leaf ℓ. Integration was performed considering intervals of 0.01 for f. For the plan, the average arc MIt is considered.

$$\mathrm{MIt} = \frac{1}{A} \sum_{a=1}^{A} \sum_{\ell=1}^{\mathcal{L}} \int_0^F z_{a\ell}(f) \, \mathrm{d}f \quad (3.30)$$

In this work, β, γ, and F were set to 2.

Nomenclature Summary

Table 3.1: Summary of the complexity nomenclature

Variable	**Description**
A	Number of arcs of the treatment plan
C_b	Number of control points of arc a
L_{ac}	Number of bank leaves not covered by jaws at arc a and control point c
$\mathcal{L}$	Number of leaves in the dual-layer MLC
pos_{acl}	Position of leaf end at arc a, control point c, and bank leaf l
d	Width of the MLC leaves
AA_{ac}	Aperture area at control point c of arc a
AP_{ac}	Aperture perimeter at control point c of arc a
$U(\mathrm{AA})_a$	Area of the union of all apertures of arc a
$p-$	Metric calculated for the proximal layer
$d-$	Metric calculated for the distal layer
$-w$	Weighted metric as proposed by Tamura et al.
$(w_p)_{ac}$	Ratio of leaves contributing to the effective aperture of control point c from arc a belonging to the proximal layer

$(w_d)_{ac}$	Ratio of leaves contributing to the effective aperture of control point c from arc a belonging to the distal layer
$\mathrm{MU}_{ac,c-1}$	Number of monitor units delivered between control point $c-1$ and c of arc a
MU_b	Total number of monitor units for arc a
MU_p	Total number of monitor units of the treatment plan
DR_{ac}	Dose rate at control point c of arc a
DRV_{ac}	Dose rate variation at control point c of arc a
$\mathrm{Time}_{ac,c+1}$	Time elapsed between control points c and $c+1$ of arc a
GA_{ac}	Gantry acceleration at control point c of arc a

3.1.3 Calculation Module - SPIDERplan

Plans are processed in SPIDERplan using a defined 'template', which is a list composed of two basic components: groups, which are a set of structures (OARs or PTVs) with an associated normalized weight, and structures, which are members of a group with an assigned objective and a normalized weight within the group. The SPIDERplan objectives are generally based on dose-volume planning objectives, but may also be based on quality metrics such as the conformity number (COIN).

Originally, the score for each OAR was the ratio of the current value for the dose-volume metric over the threshold values (equation 3.31) and the inverse for PTVs (equation 3.32). The group score was then given by the weighted average of the structures' scores (equation 3.33) and the global plan score (GPS) by the weighted average of the groups' scores (equation 3.34). In equations 3.31 and 3.32, 'plan' refers to the planned dose distribution, 'AC' refers to the threshold criteria, and D_V and V_D are the dose-volume metrics, where D_V is the dose in the volume V and V_D is the volume receiving at least an absorbed dose D [4]. In equations 3.33 and 3.34, w_{Group} and $w_{\mathrm{Structure}}$ are the group and structure weights, respectively.

$$\mathrm{Score}_{\mathrm{OAR}} = \begin{cases} \frac{(D_{\mathrm{V}})_{\mathrm{Plan}}}{D_{\mathrm{AC}}} \\ \frac{(V_{\mathrm{D}})_{\mathrm{Plan}}}{V_{\mathrm{AC}}} \end{cases} \tag{3.31}$$

$$\mathrm{Score}_{\mathrm{PTV}} = \begin{cases} \frac{D_{\mathrm{AC}}}{(D_{\mathrm{V}})_{\mathrm{Plan}}} \\ \frac{V_{\mathrm{AC}}}{(V_{\mathrm{D}})_{\mathrm{Plan}}} \end{cases} \tag{3.32}$$

$$\text{Group Score} = \sum_j w_{\mathrm{Structure(j)}}\ \mathrm{Score}_{\mathrm{Structure(j)}} \tag{3.33}$$

$$\text{Global Plan Score} = \sum_i w_{\mathrm{Group(i)}} \sum_j w_{\mathrm{Structure(i,j)}}\ \mathrm{Score}_{\mathrm{Structure(i,j)}} \tag{3.34}$$

In this work, PTV dose-volume objectives had their score formula reconceptualized to take advantage of the full scale between 0 (ideal) and 1 (threshold) values. In this range, scores were set to follow a linear trend. For metrics where values should be higher than the threshold criteria ($V_D \geq V_{\mathrm{AC}}$), volume objective scores are calculated using equation 3.35, and for metrics where values should be lower ($V_D \leq V_{\mathrm{AC}}$), equation 3.36 is used. Similarly, dose objective scores are determined by equations 3.37 and 3.38, depending on whether lower or higher values are desired. In the score formulas, V_{PTV} is the volume of the PTV and $D_{\mathrm{presc.}}$ is the prescribed dose for that target.

$$\mathrm{Score}(V_D \geq V_{\mathrm{AC}}) = \begin{cases} \frac{V_{\mathrm{PTV}} - (V_{\mathrm{D}})_{\mathrm{Plan}}}{V_{\mathrm{PTV}} - V_{\mathrm{AC}}} & (V_{\mathrm{D}})_{\mathrm{Plan}} \geq V_{\mathrm{AC}} \\ \frac{V_{\mathrm{AC}}}{(V_{\mathrm{D}})_{\mathrm{Plan}}} & (V_{\mathrm{D}})_{\mathrm{Plan}} < V_{\mathrm{AC}} \end{cases} \tag{3.35}$$

$$\mathrm{Score}(V_D \leq V_{\mathrm{AC}}) = \frac{(V_D)_{\mathrm{Plan}}}{V_{\mathrm{AC}}} \tag{3.36}$$

$$\mathrm{Score}(D_V \geq D_{\mathrm{AC}}) = \begin{cases} \frac{D_{\mathrm{Presc.}} - (D_V)_{\mathrm{Plan}}}{D_{\mathrm{Presc.}} - D_{\mathrm{AC}}} & (D_V)_{\mathrm{Plan}} \geq D_{\mathrm{AC}} \\ \frac{D_{\mathrm{AC}}}{D_{\mathrm{P}}} & (D_V)_{\mathrm{Plan}} < D_{\mathrm{AC}} \end{cases} \tag{3.37}$$

$$\mathrm{Score}(D_V \leq D_{\mathrm{AC}}) = \begin{cases} \frac{(D_V)_{\mathrm{Plan}}}{D_{\mathrm{AC}}} & (D_V)_{\mathrm{Plan}} \geq D_{\mathrm{AC}} \\ \frac{(D_V)_{\mathrm{Plan}} - D_{\mathrm{Presc.}}}{D_{\mathrm{TC}} - D_{\mathrm{Presc.}}} & (D_V)_{\mathrm{Plan}} < D_{\mathrm{AC}} \end{cases} \tag{3.38}$$

To increase flexibility in template creation, an updated version of SPIDERplan was also introduced, where each structure in the template can have multiple objectives with separate normalized weights within the group. The GPS for this version is given by the equation 3.39.

$$\text{Global Plan Score} = \sum_i w_{\mathrm{Group(i)}} \sum_j w_{\mathrm{Objective(i,j)}} \ \mathrm{Score}_{\mathrm{Objective(i,j)}} \tag{3.39}$$

To easily compare the score of objectives between plans, each plan is plotted as a polygon in one of three custom radar plots:

- Groups Plan Diagram (GPD): each group is represented as a section of the diagram with a central angle proportional to the group weight. For each section, the bisector represents the group score.

- Structures Plan Diagram (SPD): each group's section is further subdivided to represent the individual structures within that group. The central angle within the group section is proportional to the structure weight and the bisector represents the structure score.

- Structures Group Diagram (SGD): a selected group section of the SPD is projected on the entire plot.

When SPIDERplan diagrams have only one or two objectives, it is not possible to generate a polygon. A new representation was introduced in this work for such cases, where each plan score is represented by a line extending from the center of the diagram to the score value (Figure 3.8). These lines are arranged so that shorter lines are placed on top of longer ones, ensuring that each score is not obscured by scores of other plans.

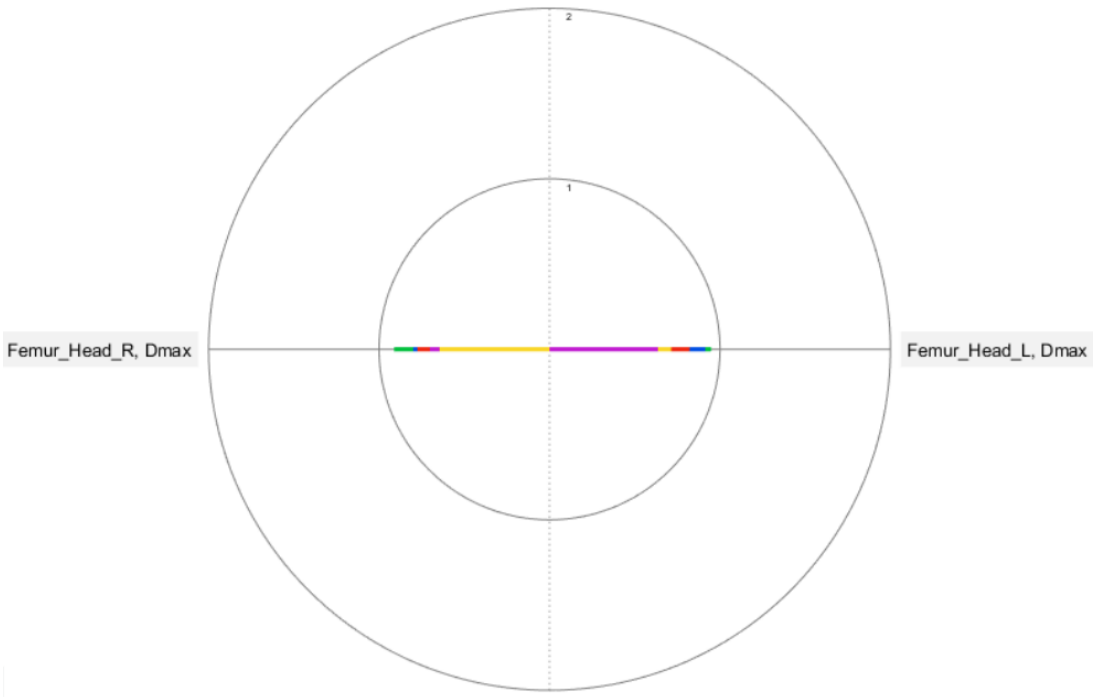


Figure 3.8: Example of a SGD with only two objectives where 5 plans are represented by different colors.

SPIDERplan GUI

Upon opening the interface, the user selects a treatment plan from the qualComp file system and chooses the desired SPIDERplan template (Figure 3.9). Once the template is selected, the SPIDERplan diagrams are generated. The GUI shown in 3.9 displays the templates already processed for each patient plan. To process with a new template, the user can select 'Add'.

A template can be applied even if the plan lacks some of the template structures. In such cases, only the objectives for the structures present in the plan are used, and the objective weights are renormalized accordingly.

The selected plan, referred as the main plan, can be compared with up to four additional plans that have been previously processed using the same template. These comparison plans are plotted as polygons on the main plan diagrams, and the GPS is calculated based on the objectives of the main plan. For comparison plans to be compatible, they must include at least the structures present in the main plan. This requirement is automatically managed by

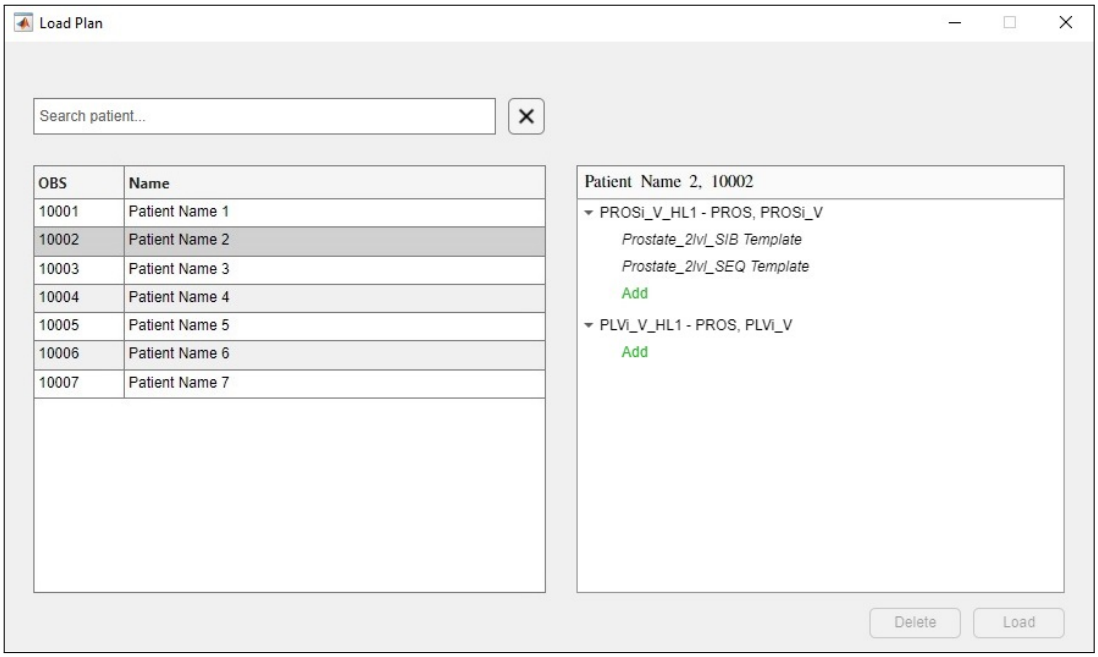


Figure 3.9: SPIDERplan GUI for selecting a template.

the software, which filters and displays only compatible comparison plans in the list shown in Figure 3.9. The main plan can be saved into the qualComp file system as a SPIDERplan file, after which it becomes available for future comparisons.

The SPIDERplan GUI is divided into three panels (Figure 3.10):

- Control panel: located on the left side of the screen (in blue), this panel includes various controls for managing SPIDERplan functionalities. In the middle, there are six buttons. The upper three buttons handle the display of the diagrams, allowing users to reset the view, drag, or zoom. The lower left button opens the interface shown in Figure 3.9 for changing the main plan, the lower center button will be used in the future for plan reporting, and the lower right button saves the main plan into the file system. At the bottom of the panel, a legend indicates the color coding of each plan, its name, and its GPS, where the main plan is highlighted in bold. Checkboxes enable users to toggle the visibility of individual plans in the diagrams, and a cross next to each comparison allows for its removal. Below the legend, there are two additional buttons, a plus sign to add a comparison plan by opening the GUI depicted in Figure 3.9, and a cross to remove all comparison plans.

- Diagram Panel: this central panel (in red) features four tabs dedicated to different SPIDERplan analyses: Group Plan Diagram, Structures Plan Diagram, Structures Group Diagram, and Statistics. The tab for the Structures Group Diagram also displays Dose-Volume Histograms (DVHs) for selected group structures in the bottom half of the panel.

- Information panel: it is placed at the right of the Diagram Panel (in green) and provides further information about a selected group or structure, such as group and objective

weights, template objectives, achieved metric values, and scores. Users can click on group or structure labels in the diagrams to update the information shown in this panel.

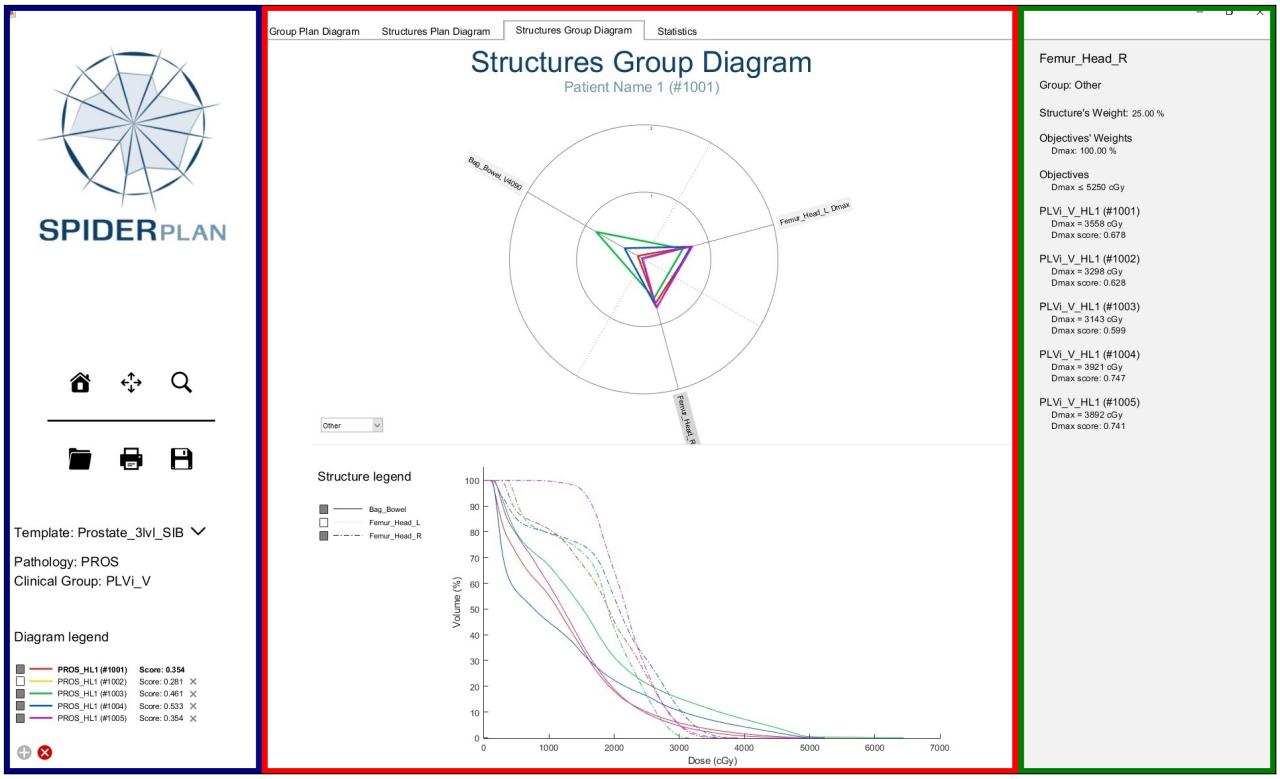


Figure 3.10: SPIDERplan GUI. The control panel is highlighted in blue, the diagram panel in red, and the information panel in green.

Examples for each of the four diagram panel tabs are provided in the following figures: Figure 3.11 for the Group Plan Diagram, Figure 3.12 for the Structures Plan Diagram, Figure 3.13 for the Structures Group Diagram, and Figure 3.14 for the Statistics tab. The same five prostate plans are compared in each tab.

To facilitate the comparison of the opened plans in the SPIDERplan GUI with the clinical practice, the Statistics tab features a violin plot of the SPIDERplan groups or objectives scores. Each 'violin' in the plot represents an estimate of the probability density function for a specific score, based on the observed score values within the file system for the same template. Score values for the main and comparison plans are also plotted as asterisks over the violins with the appropriate plan color scheme. The mean and median scores are represented with a dashed and a solid line, respectively.

The violins are generated using a Gaussian kernel with standard deviation set to half the standard deviation of the observed values. This specific bandwidth was chosen as it was found to effectively distinguish between different populations within the violin plots. The estimated probability density function is truncated at the minimum and maximum observed values for the purpose of providing the amplitude of observed data. The code implementation for this

functionality was based on a Zenodo repository [98].

As an application example, the violin plot for group scores shown in Figure 3.14 indicates that the Bladder group has the widest range of scores, reaching up to 2.5. However, the most frequent scores for this group are around the 0.5 mark.

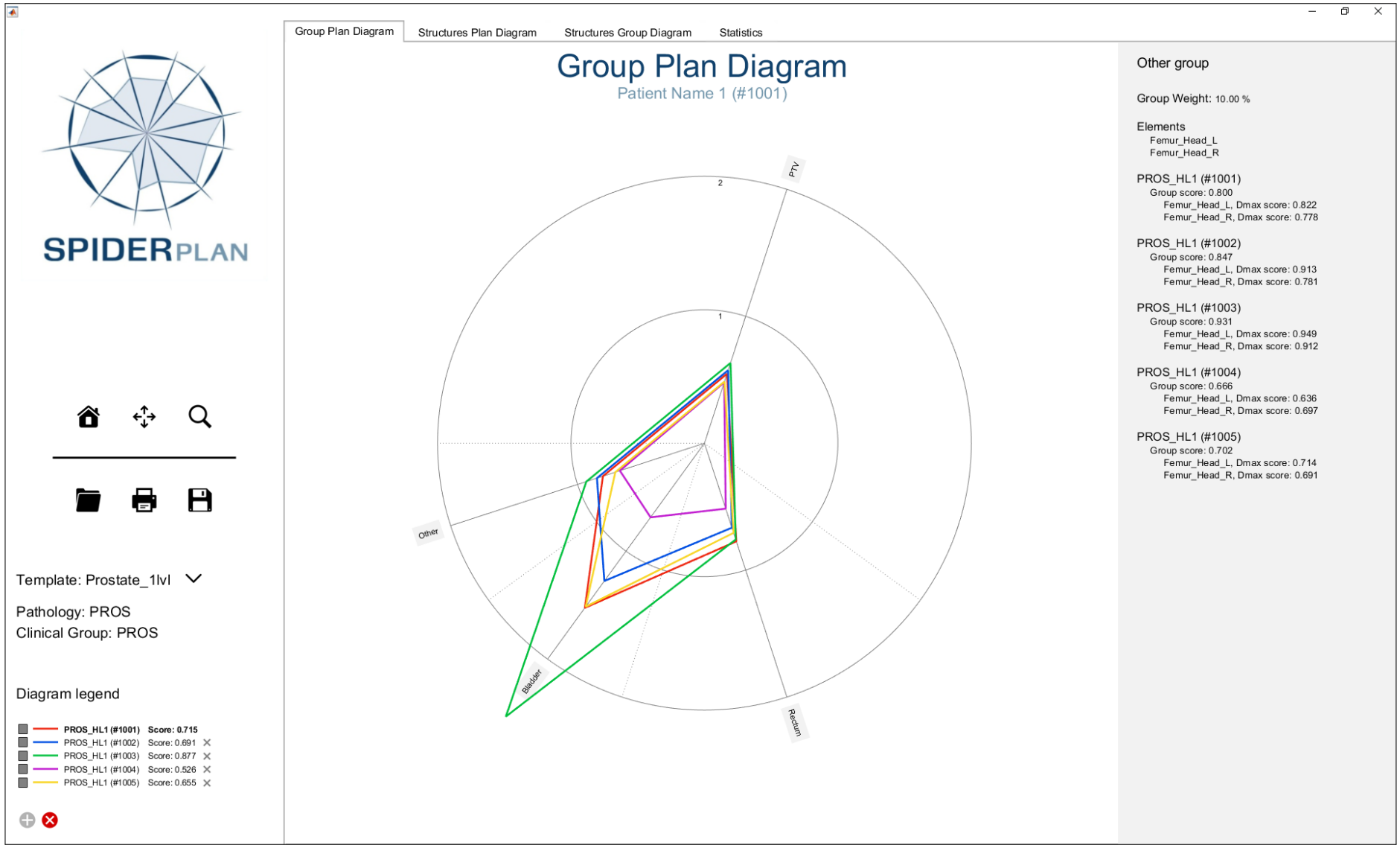


Figure 3.11: SPIDERplan GUI with five PROS plans being compared in the GPD diagram using the Prostate_1lvl template.

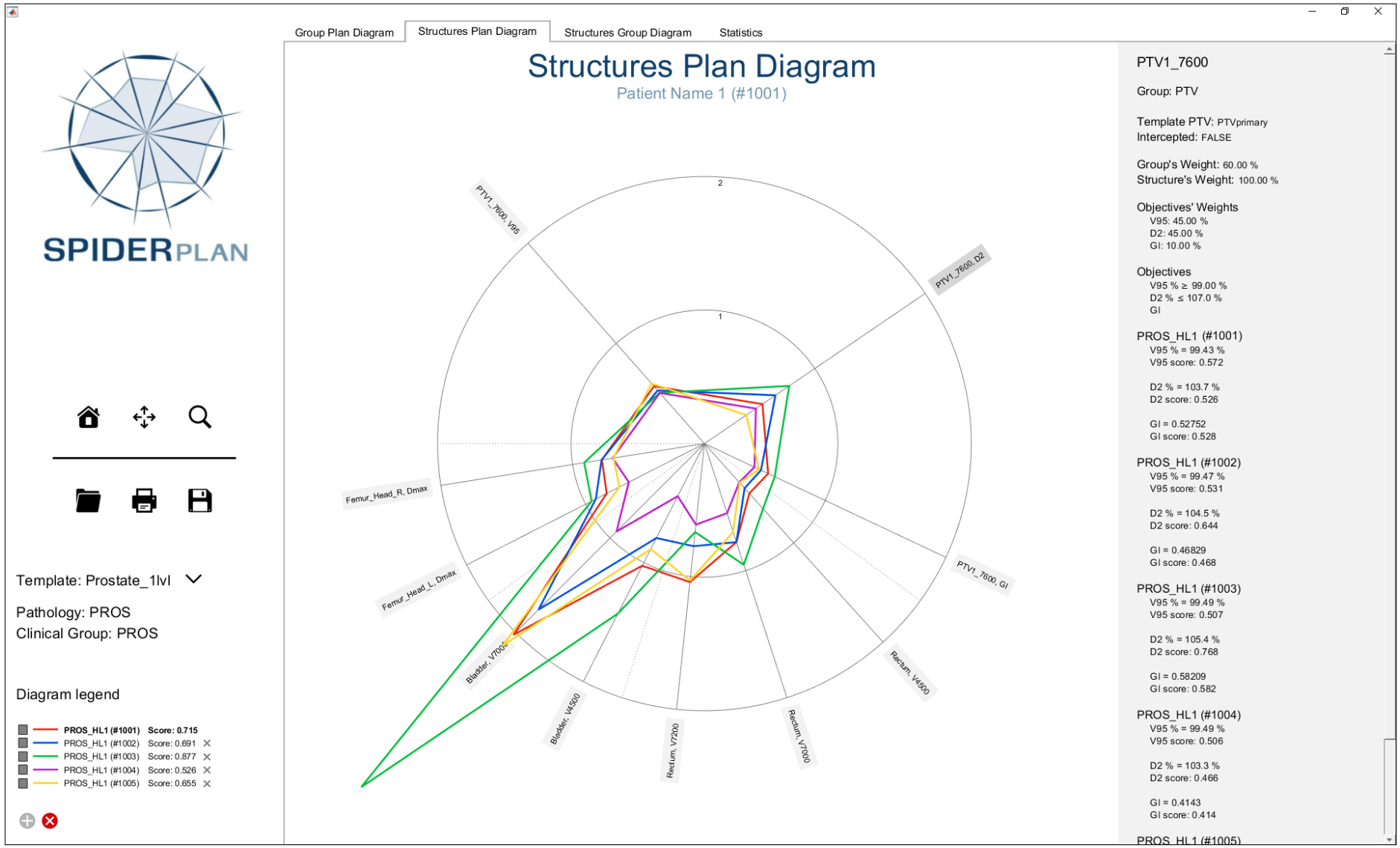


Figure 3.12: SPIDERplan GUI with five PROS plans being compared in the SPD diagram using the Prostate_1lvl template.

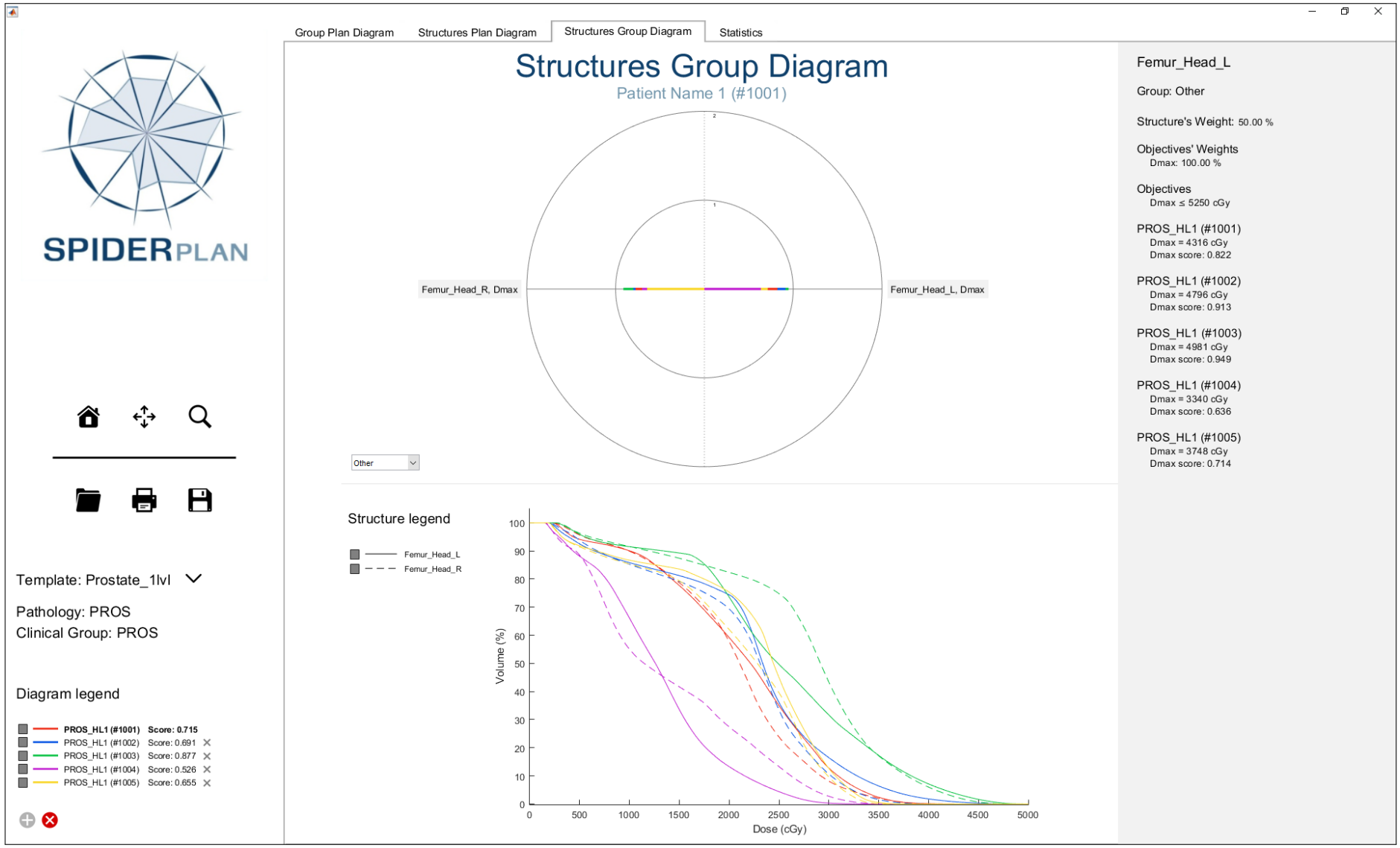


Figure 3.13: SPIDERplan GUI with five PROS plans being compared in the SGD diagram using the Prostate_1lvl template.

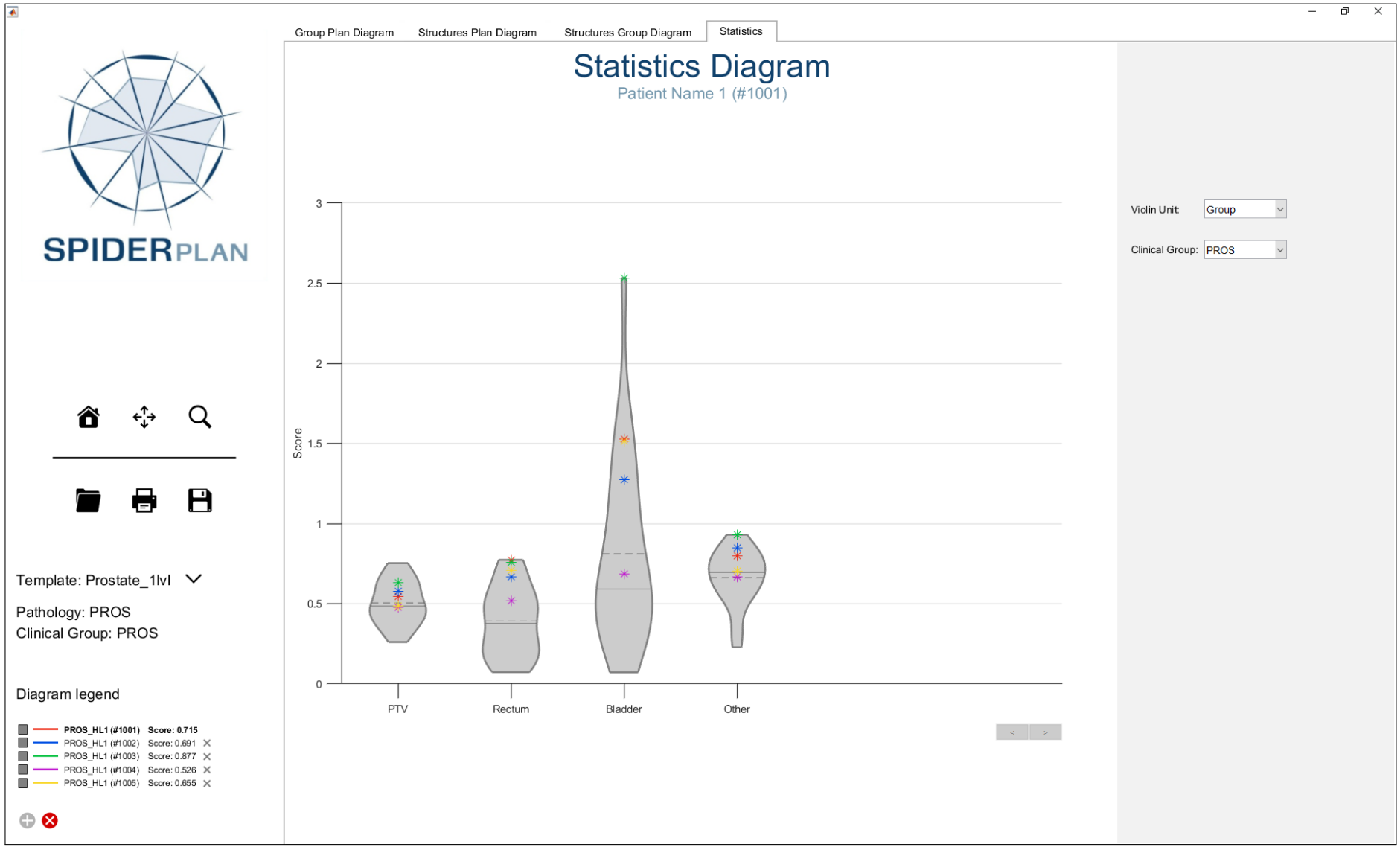


Figure 3.14: SPIDERplan GUI with the group scores of five PROS plans being compared in the Statistics diagram using the Prostate_1lvl template.

3.1.4 Search and Consulting Module

Templates

All templates for SPIDERplan are created and modified in a dedicated GUI (Figure 3.15), which not only facilitates the creation of templates but also ensures that all created templates are valid.

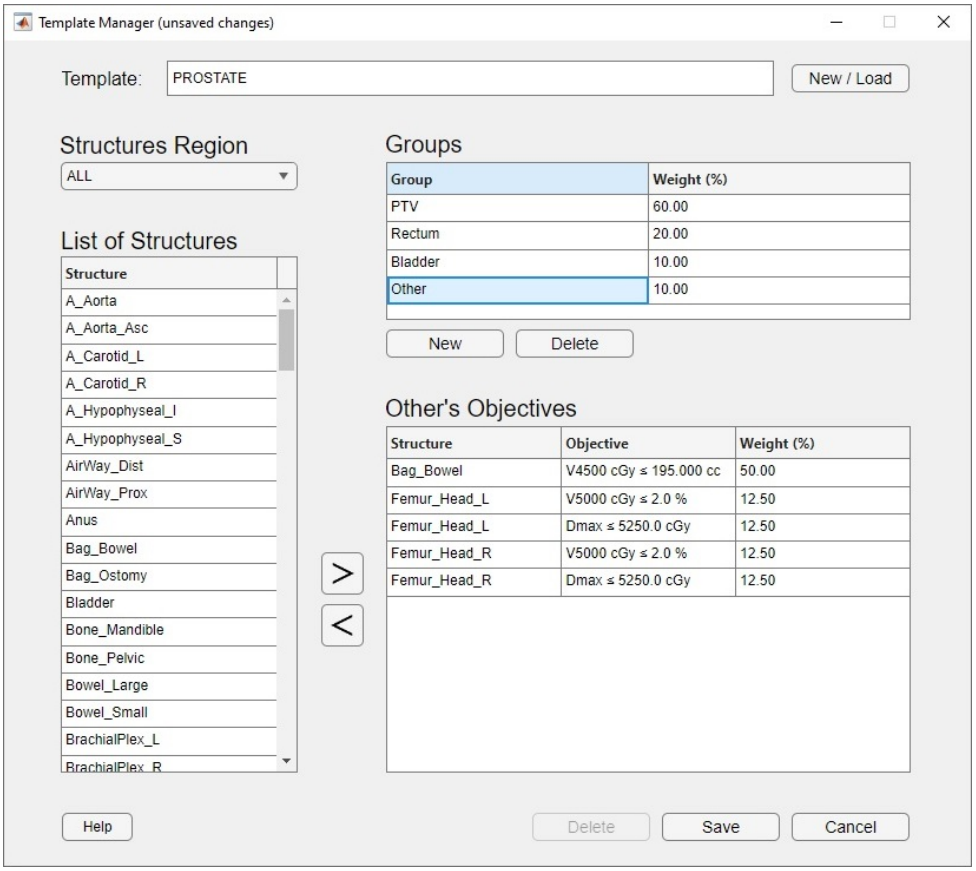


Figure 3.15: SPIDERplan template manager

In the GUI depicted in Figure 3.15, the user begins by clicking the 'New/Load' button in the upper right corner to either create a new template or load an existing one from the qualComp file system. Once a template is open, its name can be edited in the text field at the top and the three tables shown in the Figure become available for editing. In the upper right table, the user defines the template groups and assigns their weights. The objectives and corresponding structures and weights for the selected group are displayed in the lower right table. To add an objective, the user selects a structure from the left table and moves it to the objectives' table using the right arrow button ('>'). Similarly, objectives can be removed using the left arrow button ('<'). The left table lists standardized structure names the same protocoled names as those in the Structures Selection GUI (Figure 3.3). This list can be filtered using the region dropdown menu above it, which includes options like Pelvic and Thoracic regions. Once all template data is validated, the template can be saved in the qualComp file system, after which it becomes available for use in the SPIDERplan GUI.

Despite featuring a larger customization capability, the following rules are enforced by the pro-

gram to ensure good integration in the qualComp file system and to ensure the template is interpreted correctly by the SPIDERplan program:

- each template contains one PTV group, which cannot be removed or renamed.
- each objective in an OAR group must be associated with an OAR named according to the nomenclature protocol at IPO Coimbra.
- there cannot be groups sharing the same name nor two repeated objectives for the same structure.
- a structure may only be present in one of the groups.
- a structure for the body must be present.
- group weights and objective weights within a group must add up to 100%. Automatic organization can be prompted when saving.
- objectives must follow a pre-defined syntax for the specific metric and have a valid definition. Case differences, non-uniform spacing, and incorrect number of significant digits are automatically corrected.
- a template cannot be edited or deleted if there are SPIDERplan files in the qualComp file system referencing it.

Export DVHs

This sub-module exports preprocessed dose-volume histograms (DVHs) for all structures in a treatment plan as a spreadsheet file. In the exported file, the DVH of each structure is placed into a separate sheet. Additionally, the minimum, mean, and maximum doses for each structure are reported at the top of each sheet.

In the GUI (Figure 3.16), the DVHs are exported by hitting the 'export' button after selecting the desired patient and DVH file.

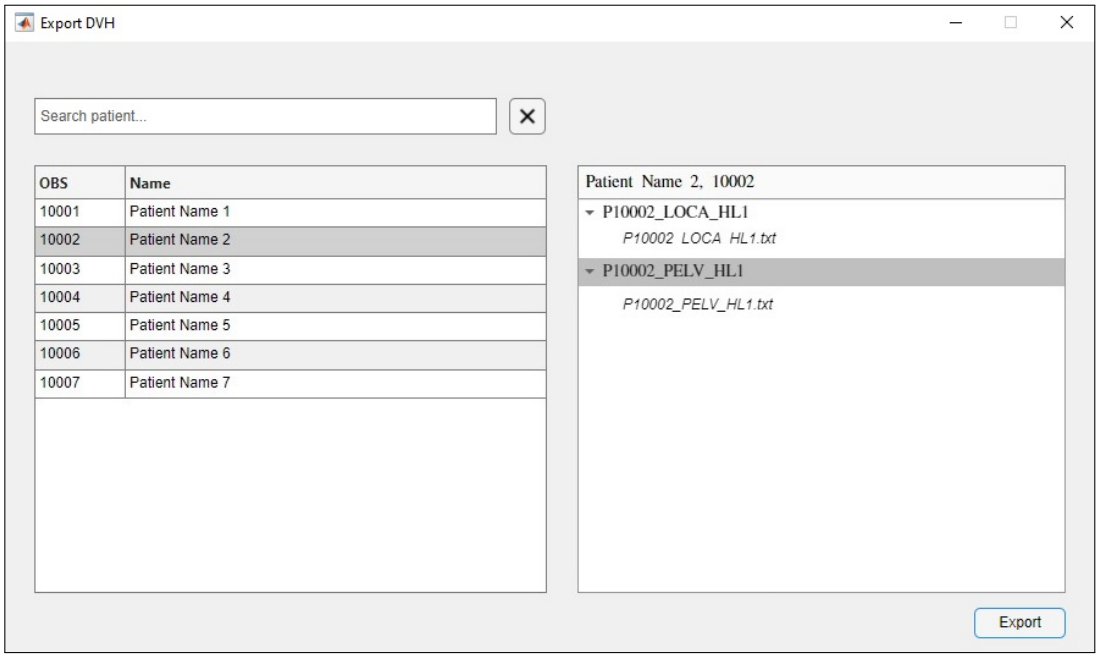


Figure 3.16: qualComp GUI to export plan DVHs.

3.2 Application to the Prostate Pathology

qualComp was employed to characterize the current clinical practice at IPO Coimbra concerning the prostate pathology. Plan complexity was assessed using the 15 complexity metrics described in section 3.1.2, while plan quality was evaluated using the SPIDERplan sub-module described in section 3.1.3.

A clinical protocol with treatment planning recommendations was issued at IPO Coimbra in April 2022, aiming to harmonize planning procedures among different planners. These recommendations included guidelines for the maximum number of monitor units (MUs), the configuration of the Aperture Shape Control (ASC) tool in the Eclipse TPS, and the number of arcs for each plan type. The impact of the implementation of this protocol on plan complexity and quality was assessed and considered while characterizing the current clinical practice.

A total of 578 VMAT plans for prostate cancer patients treated with a Halcyon linear accelerator were retrospectively selected. These plans were created between May 28, 2021 and April 26, 2024 using the Varian Eclipse TPS version 16.1 and the dose distribution was calculated with the anisotropic analytical algorithm (AAA). All structures were delineated following the local protocol at IPO Coimbra.

The plans are categorized into five distinct plan types based on their target regions and prescribed dose levels:

- PELV (N=133): irradiation of pelvic lymph nodes and prostate bed with a total dose of 4500 cGy and a fraction dose of 180 cGy. 129 plans used 2 arcs and 4 plans used 3 arcs.

- LOCA (N=133): Irradiation prostate bed with 2000 cGy (N=4), 2200 cGy (N=57), or 2400 cGy (N=72) with fraction doses of 200 cGy, respectively. All plans used 2 arcs.

- PROSi_V (N=131): Simultaneous irradiation of prostate with 7000 cGy and seminal vesicles with 6160 cGy. The fraction doses are 250 cGy for the prostate and 220 cGy for the seminal vesicles. 109 plans used 2 arcs and 22 plans used 3 arcs.

- PLVi_V (N=157): Simultaneous irradiation of prostate with 7000 cGy, pelvic lymph nodes with 5040 cGy, and seminal vesicles with 5040 cGy (N=47) or 6160 cGy (N=110). From highest to lowest, the three dose levels have fraction doses of 250 cGy, 220 cGy, and 180 cGy. 136 plans used 2 arcs, 13 plans used 3 arcs and 8 plans used 4 arcs.

- PROS (N=24): prostate with 7800 cGy (N=2), 7600 cGy (N=6), 7000 cGy (N=13), 6800 cGy (N=2), or 6000 cGy (N=1). Plans with 7000 cGy prescription doses have fraction doses of 250 cGy (N=10) or 200 cGy (N=3). Plans with 6800 cGy, 7800 cGy, or 7600 cGy prescription doses have fraction doses of 200 cGy, and the plan with a 6000 cGy prescription dose has a fraction dose of 300 cGy. All plans used 2 arcs.

The PELV and LOCA plans belong to the same sequential boost treatment courses, where an initial dose is delivered to the whole pelvic area (PELV), followed by a localized boost to the prostate bed (LOCA).

3.2.1 Plan Complexity

Plan Groups

The complexity of a treatment plan is highly influenced by the target's shape, size, and location [63, 99]. Given this, complexity assessment was performed considering two main groups of plans based on their anatomical similarities: Pelvic (N=290) and Prostatic (N=288). The Pelvic group consists of PELV (N=133) and PLVi_V (n=157) plans, while the Prostatic group consists of PROS (N=24), PROSi_V (N=131), and LOCA (N=133) plans.

Methodology

After computing the 15 complexity metrics implemented in qualComp for the 578 VMAT plans, the analysis was conducted in five main steps:

1. The relationship between the proximal and the distal layer implementations of MCS, PA, PI, and PM was evaluated using the Pearson correlation coefficient (r). Based on these correlations, it was assessed whether single layer metrics provide distinct information compared to the weighted metrics.

2. The relationship between the MCS, PA, PI, and PM metrics computed for the effective 5 mm apertures and their weighted implementations was evaluated using the Pearson correlation coefficient (r). Based on these correlations, it was assessed whether weighted metrics provide distinct information compared to the original formulations of the metrics

considering the effective 5 mm apertures.

3. Correlation matrices were generated for each group of plans (PELV and PROS) to evaluate the relationships between the retained metrics from steps 1. and 2. using the Pearson correlation coefficient.

4. The longitudinal evolution of the complexity metrics was analyzed using a moving average with a triangular kernel spanning 251 days. Three distinct time periods were considered on the observed longitudinal trend: before the protocol implementation (from May 2021 to April 2022), during protocol adoption (from May 2022 to April 2023), and practice stabilization (from May 2023 to April 2024). For each metric, the difference in mean complexity between the first and third periods was examined. The statistical significance of these differences was evaluated using the Welch's t-test at a 5% significance level.

5. Complexity values representative of the clinical practice at IPO Coimbra were summarized.

In steps 1. and 2., only the MCS, PA, PI, and PM were used for studying the usefulness of single layer and weighted metrics, as they are the metrics used by Tamura et al. [20] in their proposal of single-layer and weighted metrics.

Correlation strength was categorized into five levels: very weak ($0 \leq |r| < 0.2$), weak ($0.2 \leq |r| < 0.40$), moderate ($0.4 \leq |r| < 0.6$), strong ($0.6 \leq |r| < 0.8$), and very strong ($0.8 \leq |r| \leq 1$).

The Welch's t-test was chosen over the t-test for its better performance when the two samples compared differ in variance and size [100].

All statistical analyses were conducted using MATLAB R2023b.

3.2.2 Plan Quality

Plan Groups

SPIDERplan templates were configured considering the preferences of the radiation oncologists evaluating the created plans. At IPO Coimbra, the OAR constraints for a sequential boost scheme consider the dosimetric sum of the two plans. Therefore, PELV and LOCA plans were evaluated together (PELV_LOCA) using the dosimetric sum of the two dose distributions. All other plan types were evaluated separately, resulting in four groups: PELV_LOCA (N=133), PROSi_V (N=131), PLVi_V (N=157), and PROS (N=24).

The group definition used for plan complexity was not applied to plan evaluation due to differences in prescription doses and OAR constraints among plan types belonging to the same complexity group. Additionally, PELV_LOCA contains plan types from separate complexity groups.

GI

In this work, a gradient index (GI) based on the RTOG $R_{50\%}$ was proposed. This metric is normalized to a range of 0 to 1 and aims to minimize the dependency on the PTV volume, providing a consistent evaluation across various target sizes.

Firstly, $T(R, d)$ is defined as the function representing the $R_{50\%}$ for a spherical PTV of radius R, where the 50% prescription isodose volume is the isotropic expansion of the PTV by a distance d:

$$T(R,d) = \frac{\frac{4\pi(R+d)^3}{3}}{\frac{4\pi R^3}{3}} = \left(\frac{R+d}{R}\right)^3 \tag{3.40}$$

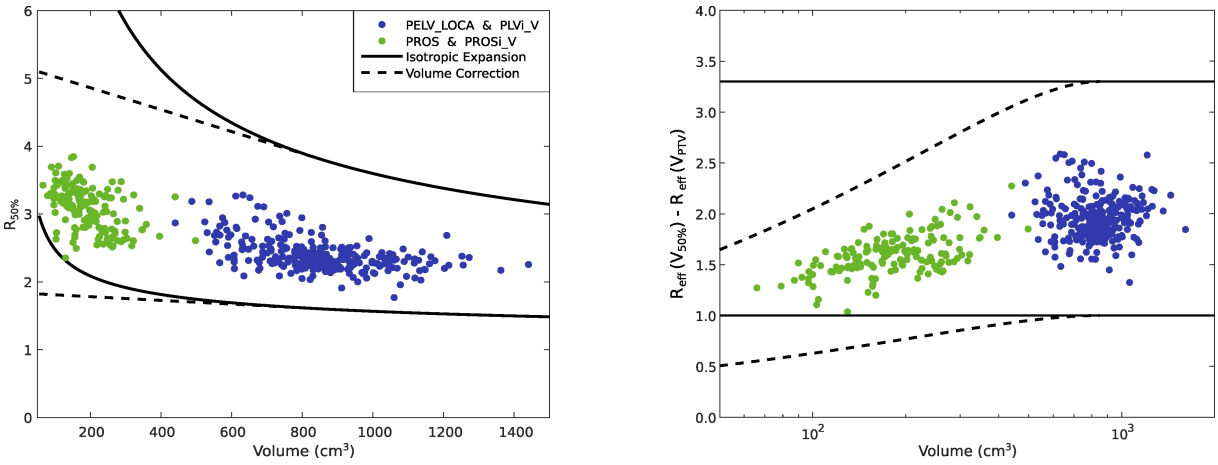
Then, GI is given by the relative position of the observed $R_{50\%}$ in relation to the two reference values, $T(R_{eff}, d_0)$ and $T(R_{eff}, d_1)$, where, R_{eff} is the radius of a sphere with the same volume as the PTV:

$$GI = \frac{R_{50\%} - T(R_{eff}, d_0)}{T(R_{eff}, d_1) - T(R_{eff}, d_0)} \tag{3.41}$$

For this work, $d_0 = 1\,\text{cm}$ and $d_1 = 3.3\,\text{cm}$ were used. The value of d_1 was selected as the midpoint between 2.3 cm and 4.3 cm, reflecting the typical range within which the planners at IPO Coimbra aim to contain the 50% isodose line relative to the PTV.

However, using the same isotropic expansions for all target volumes does not yield a volume-independent gradient metric for volumes smaller than the lymph node targets (Figure 3.17a). This issue arises because the difference between the effective radius of the 50% isodose volume and the effective radius of the PTV stabilizes only for reasonably large volumes (Figure 3.17b). To mitigate this effect, a correction was introduced for volumes less than $850\,\text{cm}^3$ (dashed lines), near the average volume for lymph node targets. In the graph of $R_{50\%}$ as a function of target volume (Figure 3.17a), two tangent lines to the isotropic expansion lines at this volume were used as substitutes for T in the GI calculation (equation 3.41).

With these new reference values, GI, the relative position of $R_{50\%}$ in relation to the two reference values, becomes significantly less dependent on the target volume, as shown in Figure 3.17a. This correction is equivalent to considering different isotropic expansions based on the target volume, as demonstrated in Figure 3.17b.



(a) (b)

Figure 3.17: (a) $R_{50\%}$ values and (b) differences between the effective radius of the 50% isodose and the effective radius of PTV ($R_{eff}(V_{50\%}) - R_{eff}(V_{PTV})$) for all plans. In both figures, the solid lines represent the reference values without correction, while the dashed lines represent values with the correction. In figure (b), The horizontal axis is presented on a logarithmic scale for better visualization of lower volumes.

SPIDERplan Templates

SPIDERplan was configured using four templates considering the different dose levels and boost schemes of each plan type. These templates include planning objectives set by the radiation oncologists at IPO Coimbra where each one is weighted within their group according to the relative importance given by the radiation oncologist during the plan evaluation. Four groups are considered in the templates: PTV (PTV objectives), Rectum (rectum objectives), Bladder (bladder objectives), and Other (objectives for femur heads and bowels).

Template objectives that consistently received null scores for all plans processed were removed. Specifically, these included both the objectives for the bowels and the $V_{5000\,\mathrm{cGy}} \leq 2\%$ objective for the femur heads in the Prostate_1lvl, Prostate_2lvl_SIB, and Prostate_3lvl_SIB templates.

The four different SPIDERplan templates used in this work are:

- Prostate_1lvl (table 3.2) for PROS plans
- Prostate_2lvl_SEQ (table 3.3) for PELV_LOCA plans
- Prostate_2lvl_SIB (table 3.4) for PLVi_V plans with two dose levels and PROSi_V plans
- Prostate_3lvl_SIB (table 3.5) for PLVi_V plans with three dose levels

Table 3.2: Group, objectives, and corresponding weights for template Prostate_1lvl

Group	w_{Group}	Structure	Objectives	$w_{\text{Objective}}$
PTV	60%	Prostate	$V_{95\%} \geq 99\%$	45%
			$D_{2\%} \leq 107\%$	45%
			GI	10%
Rectum	20%	Rectum	$V_{7200\,\text{cGy}} \leq 10\,\text{cm}^3$	33.3%
			$V_{7000\,\text{cGy}} \leq 15\%$	33.3%
			$V_{4500\,\text{cGy}} \leq 50\%$	33.3%
Bladder	10%	Bladder	$V_{7000\,\text{cGy}} \leq 15\%$	50%
			$V_{4500\,\text{cGy}} \leq 50\%$	50%
Other	10%	Femur_Head_L	$D_{\max} \leq 5250\,\text{cGy}$	50%
		Femur_Head_R	$D_{\max} \leq 5250\,\text{cGy}$	50%

Table 3.3: Group, objectives, and corresponding weights for template Prostate_2lvl_SEQ

Group	w_{Group}	Structure	Objectives	$w_{\text{Objective}}$
PTV	60%	Prostate bed	$V_{95\%} \geq 99\%$	35%
			$D_{2\%} \leq 107\%$	35%
		Lymph nodes	$V_{95\%} \geq 99\%$	20%
			GI	10%
Rectum	20%	Rectum	$V_{7000\,\text{cGy}} \leq 15\%$	50%
			$V_{4500\,\text{cGy}} \leq 50\%$	50%
Bladder	10%	Bladder	$V_{7000\,\text{cGy}} \leq 15\%$	50%
			$V_{4500\,\text{cGy}} \leq 50\%$	50%
Other	10%	Bag_Bowel	$V_{4500\,\text{cGy}} \leq 195\,\text{cm}^3$	50%
		Femur_Head_L	$D_{\max} \leq 5250\,\text{cGy}$	12.5%
			$V_{5000\,\text{cGy}} \leq 2\%$	12.5%
		Femur_Head_R	$D_{\max} \leq 5250\,\text{cGy}$	12.5%
			$V_{5000\,\text{cGy}} \leq 2\%$	12.5%

Table 3.4: Group, objectives, and corresponding weights for template Prostate_2lvl_SIB

Group	w_{Group}	Structure	Objectives	$w_{\text{Objective}}$
PTV	60%	Prostate	$V_{95\%} \geq 99\%$	35%
			$D_{2\%} \leq 107\%$	35%
		Vesicles/Lymph nodes	$V_{95\%} \geq 99\%$	20%
			GI	10%
Rectum	20%	Rectum	$V_{6500\,\text{cGy}} \leq 15\%$	33.3%
			$V_{5500\,\text{cGy}} \leq 25\%$	33.3%
			$V_{4500\,\text{cGy}} \leq 45\%$	33.3%
Bladder	10%	Bladder	$V_{6500\,\text{cGy}} \leq 15\%$	33.3%
			$V_{5500\,\text{cGy}} \leq 25\%$	33.3%
			$V_{4500\,\text{cGy}} \leq 45\%$	33.3%
Other	10%	Femur_Head_L	$D_{\max} \leq 5250\,\text{cGy}$	50%
		Femur_Head_R	$D_{\max} \leq 5250\,\text{cGy}$	50%

Table 3.5: Group, objectives, and corresponding weights for template Prostate_3lvl_SIB

Group	w_{Group}	Structure	Objectives	$w_{\text{Objective}}$
PTV	60%	Prostate	$V_{95\%} \geq 99\%$	22.5%
			$D_{2\%} \leq 107\%$	22.5%
		Vesicles	$V_{95\%} \geq 99\%$	25%
		Lymph nodes	$V_{95\%} \geq 99\%$	20%
			GI	10%
Rectum	20%	Rectum	$V_{6500\,\text{cGy}} \leq 15\%$	33.3%
			$V_{5500\,\text{cGy}} \leq 25\%$	33.3%
			$V_{4500\,\text{cGy}} \leq 45\%$	33.3%
Bladder	10%	Bladder	$V_{6500\,\text{cGy}} \leq 15\%$	33.3%
			$V_{5500\,\text{cGy}} \leq 25\%$	33.3%
			$V_{4500\,\text{cGy}} \leq 45\%$	33.3%
Other	10%	Femur_Head_L	$D_{\max} \leq 5250\,\text{cGy}$	25%
		Femur_Head_R	$D_{\max} \leq 5250\,\text{cGy}$	25%

Methodology

After processing all plans in SPIDERplan with the appropriate templates (section 3.2.2), the analysis was conducted in two main steps:

1. The longitudinal evolution of the plan quality was analyzed using a moving average with a triangular kernel spanning 251 days. The three time periods used in the complexity analysis (section 3.2.1) were considered in order to study the potential impact of complexity reduction on plan quality. The GPS and GI values between the first and third periods were compared for each group. The statistical significance of these differences was assessed

using Welch's t-test for the PELV_LOCA group and the Mann-Whitney U test for the PROSi_V groups, both at a 5% significance level. Due to scarcity of PLVi_V plans in the first period (N=5), the analysis for the PLVi_V group was not feasible. Similarly, for PROS plans, no statistical test was done due to the limited number of plans in both the first (N=7) and third periods (N=9).

2. SPIDERplan score values representative of the clinical practice at IPO Coimbra were summarized in violin plots. After a Kruskal-Wallis test at a 5% significance level to test for the presence of different distributions, the GPS and GI distributions for each group were compared pairwise using the Mann-Whitney U test at a 1.25% significance level.

The Welch's t-test was selected for the PELV_LOCA group given its better performance when the two samples compared differ in variance and size [100]. For the PLVi_V groups, the Mann-Whitney U test was used because the score values do not follow a normal distribution ($p < 0.05$ for the Kolmogorov-Smirnov test) and the central limit theorem does not apply for some of the sample sizes ($N < 30$).

The significance level in the pairwise comparisons in step 2 were adjusted for multiple testing using the Bonferroni correction.

All statistical analyses were conducted using MATLAB R2023b.

4 Results and Discussion

4.1 Complexity

4.1.1 Single Layer Complexity

Correlations between MCS, PA, PI, and PM for each layer are depicted in Figure 4.1. All four metrics exhibit very strong correlations ($|r| > 0.8$), with the exception of the PI metric (Figure 4.1c) for the Pelvic group, which shows only a strong correlation ($r = 0.704$) between layers.

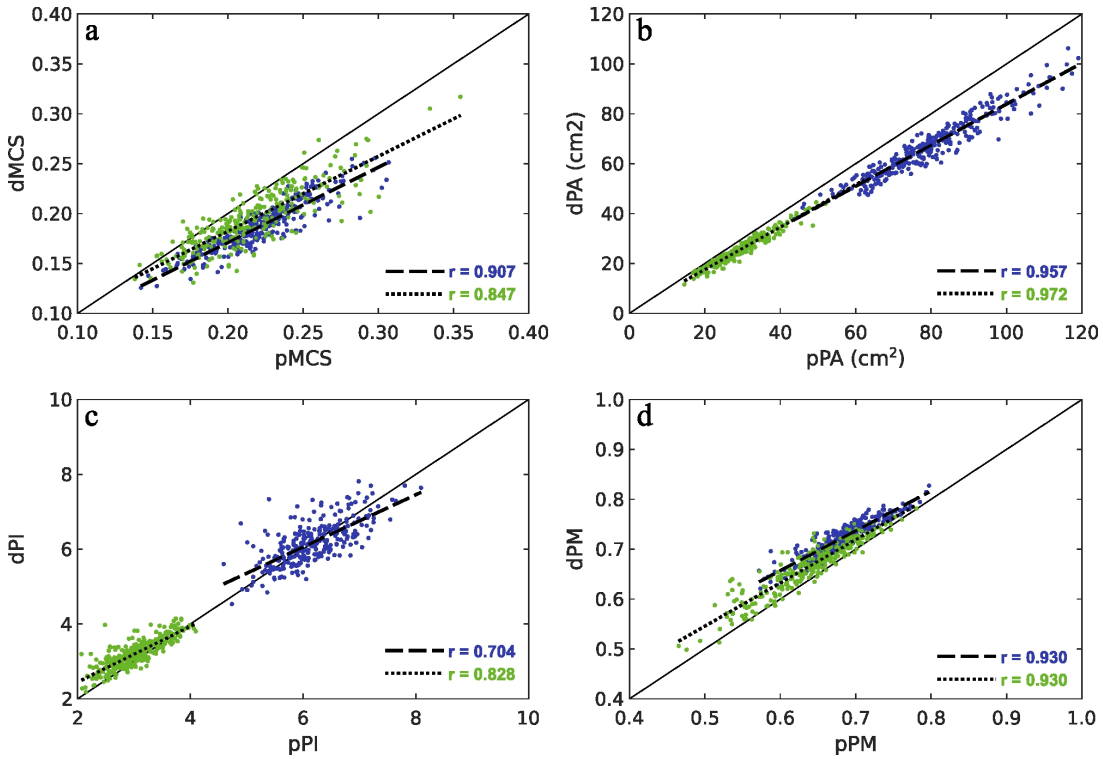


Figure 4.1: Comparison of complexity metrics values for MCS (a), PA (b), PI (c), and PM (d) between the proximal and distal layers for the Prostatic and Pelvic groups. Plans from the Prostatic group are depicted in green, while plans from the Pelvic group are depicted in blue. Each scatter plot shows the Pearson correlation coefficients (r) for the corresponding metrics in both groups. The dotted lines indicate the linear regressions for the Prostatic group, and the dashed lines for the Pelvic group.

The distal layer features smaller apertures (Figure 4.1b) and greater aperture variation (Figure 4.1d) over the beam, consistent with the findings of Tamura et al. [20] and Petroccia et al. [67].

PA values show bigger differences between layers for greater volumes, as seen in figure 4.1b. In fact, among the 227 out of 578 plans where every control point proximal aperture is larger than the distal aperture, only 18 out of these plans (8%) have pPA values less than $60\,\mathrm{cm}^2$. In the other hand, aperture irregularity, as measured by PI does not show systematic differences between layers.

The higher complexity of the distal layer is further evidenced by MCS (Figure 4.1a) and PM (Figure 4.1d), which is also in line with the findings reported by Tamura et al. [20].

Weighted metrics combine the single layer values of both layers. Therefore, given the overall very strong correlations observed between the proximal and distal layer values, weighted metrics can be employed instead of single layer metrics without losing significant information. From here onward, single layer complexity values will not be used for the characterization of clinical practice. The usefulness of weighted metrics will be assessed in the next subsection.

Although not directly applicable to the focus of this study, single-layer metrics can still serve a purpose in comparing different treatment planning systems and software versions.

4.1.2 Weighted and Effective 5 mm Metrics

The relationships between the weighted and the effective 5 mm implementations are depicted in Figure 4.2 for the MCS, PA, PI, and PM metrics.

PAw consistently showed higher values compared to PA across all plans (Figure 4.2b), which was to be expected as, for a single control point, the effective area (PA) is less than the smallest of the two layer areas (pPA and dPA), which in turn is less than the weighted (PAw) average of them: $\mathrm{PA} \leq \min\{\mathrm{pPA}, \mathrm{dPA}\} \leq \mathrm{PAw} \implies \mathrm{PA} \leq \mathrm{PAw}$.

PM and PI values were also consistently higher compared to PMw and PIw (Figures 4.2c and 4.2d), respectively, which can be explained by their computation involving the inverse of the aperture area. In contrast, MCS and MCSw (Figure 4.2a) did not show such a clear distinction. Tamura et al. [20] did not present the Pearson correlation coefficients for the PA, PI, and PM metrics. However, their reported mean values align with the observed trends in this study: higher mean values for PM and PI compared to PMw and PIw, and a higher PAw than PA. Regarding the relationship between MCS and MCSw, Tamura et al. reported a correlation coefficient of 0.6573, which is notably lower than the values observed in this work. This discrepancy might be due to variations in the computation of the weighting factors. In fact, the original paper [20] does not provide the mathematical formula for these weighting factors, allowing for different mathematical interpretations.

All four metrics demonstrated very strong correlations ($|r| > 0.8$), with the weakest correlation observed for the PI metric in the Pelvic group ($r = 0.953$). The strongest correlations were found between PA and PAw, with coefficients of $r = 0.995$ for the Prostatic group and $r = 0.994$ for the Pelvic group. These values suggest a strong inherent connection between PA and PAw. In actuality, PA already integrates the contribution of each layer through its computation, which

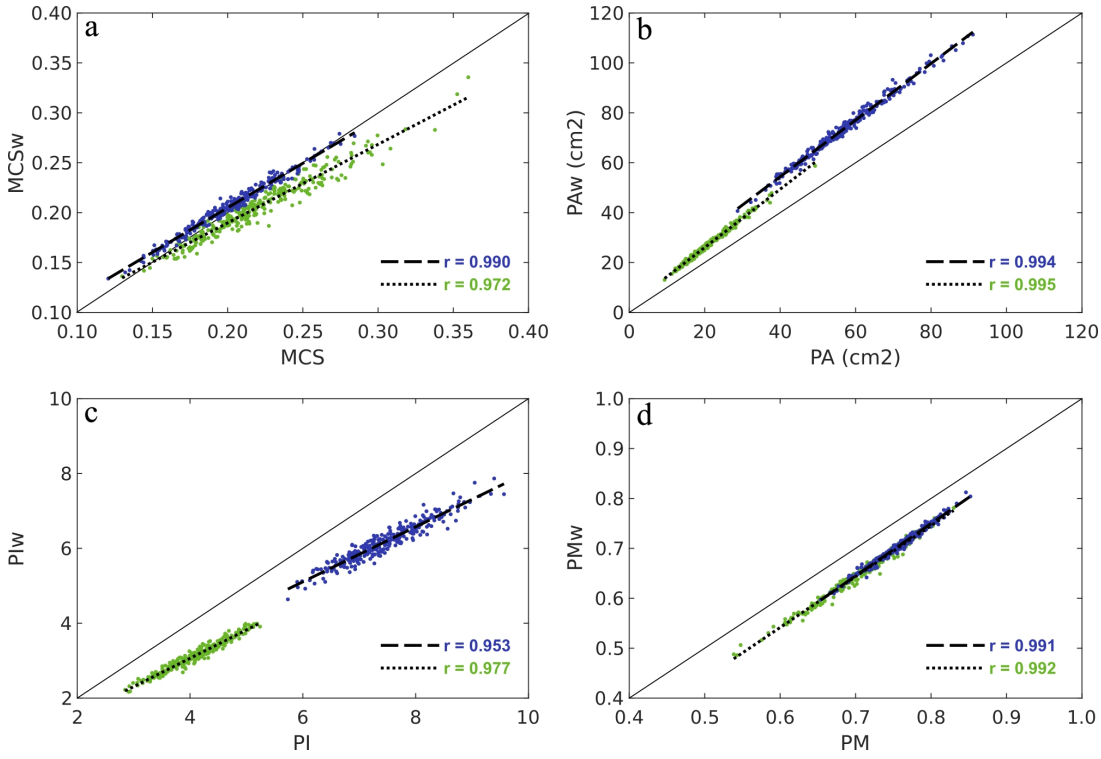


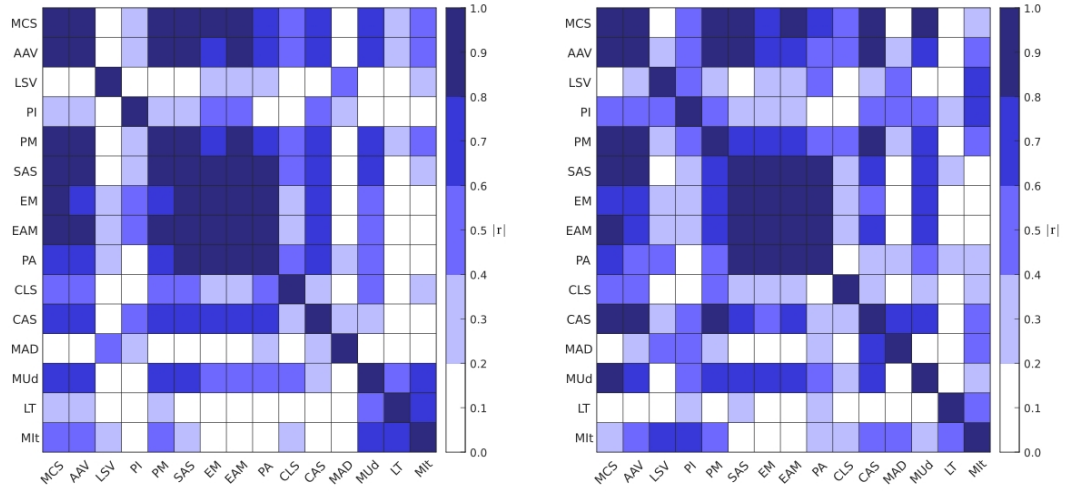
Figure 4.2: Comparison of the weighted metrics MCSw (a), PAw (b), PIw (c), and PMw (d) with their original formulation applied to the effective 5 mm apertures (MCS, PA, PI, PM). Plans from the Prostatic group are depicted in green, while plans from the Pelvic group are depicted in blue. Each scatter plot shows the Pearson correlation coefficients (r) for the corresponding metrics in both groups. The dotted lines indicate the linear regressions for the Prostatic group, and the dashed lines for the Pelvic group.

only considers the positions of leaf ends contributing to the effective aperture. Therefore, a greater contribution by the leaf ends of one layer results in a proportionally higher impact of that layer on the computed PA.

Given the lack of new significant information provided by the weighted metrics, they will not be employed hereafter.

4.1.3 Correlation Matrices

Correlations between the 15 complexity metrics for each group were studied using the Pearson correlation coefficient and are summarized in correlation matrices (Figure 4.3). Given the results of sections 4.1.1 and 4.1.2, only the computations of the metrics for the effective 5 mm apertures were considered. Overall, just four metrics (LSV, PI, CLS, and LT) did not show at least one strong correlation with other metric across the two groups.



(a) Pelvic group

(b) Prostatic group

Figure 4.3: Correlation matrices of complexity metrics for both plan groups. The color scheme was set to reflect the correlation strength categorization defined previously.

Some differences in metrics correlation was observed between groups. For instance, MIt correlates strongly with LSV and PI in the Prostatic group but only weakly or very weakly in the Pelvic group. On the contrary, MIt has strong correlations with MUd and LT in the Pelvic group, whereas these correlations were weak or moderate in the Prostatic group. Additionally, MCS, CAS, and MAD showed stronger correlations with other metrics in the Prostatic group compared to the Pelvic group.

Metrics that are sensitive to different sources of complexity and show very strong correlations with each other, such as SAS and AAV, were retained. This decision is based on the understanding that these correlations might be influenced by factors such as the TPS, linac, and treatment site, rather than indicating an inherent link between the metrics. Conversely, when metrics

sensitive to the same complexity sources showed very strong correlations, it was assumed that an inherent link was the most probable cause and a decision was made to retain only one of them. Despite the differences in correlation strengths, the same metrics were retained for both groups to simplify the analysis. Five different pairs of similar metrics were studied:

- LSV and PI: These metrics show a very low correlation in the Pelvic group and moderate correlation in the Prostatic group, despite both quantifying the aperture irregularity. This can be explained by their distinct perspectives on irregularity: while LSV focuses on the relative variation in the positions of adjacent leaf ends, PI evaluates how much the aperture shape deviates from an ideal circular shape. Both metrics are retained to provide a comprehensive characterization of aperture irregularity.

- EM and EAM: Both metrics are sensitive to irregular and smaller apertures and exhibit very strong correlations in both groups. EM will be chosen over EAM due to its much simpler computation

- MCS and AAV: MCS shows a very strong correlation with AAV across both groups, suggesting that most of the variance in MCS is attributable to its AAV component rather than its LSV component. Since AAV encapsulates much of the information provided by MCS while being sensitive to only a single complexity source, AAV will be retained alongside LSV, while the composite metric MCS will be excluded.

- AAV and PM: These metrics show very strong correlations across both groups, suggesting that the different reference area definitions were insufficient to distinguish them. AAV will be preferred due to its simpler computation.

- PA and SAS: These metrics show very strong correlations for both groups. However, they do not share the same correlation strengths in most of the remaining metrics (8 in the Pelvic group and 10 in the Prostatic). As such, both metrics were retained.

From the analysis conducted in this subsection, the EAM, MCS, and PM metrics were excluded, resulting in 12 out of the 15 metrics being retained for use in further analysis.

4.1.4 Longitudinal Evolution

The longitudinal evolution of complexity values for the 12 retained metrics is represented in Figure 4.4 for the Pelvic group and in Figure 4.5 for the Prostatic group, where the moving average is given by the solid line. For figure 4.4, it is observed that AAV, SAS, EM, PA, CAS, MUd, LT, and MIt follow the same longitudinal trend: higher complexity values before the protocol implementation, followed by a steady reduction and later stabilization. Based on this, the three distinct time periods used for analysis were defined: before the protocol implementation from May 2021 to April 2022 (first period), during protocol adoption fom May 2022 to April 2023 (second period), and clinical practice stabilization from May 2023 to April 2024 (third period), each spanning one year. In figures 4.4 and 4.5, the time periods are separated by vertical dashed lines.

Table 4.1 summarizes the mean complexity values for the first and third periods, and the differences between these means for the PELV plans. p-values were calculated using the Welch's t-test. PLVi_V plans were not included as these plans were scarce in the first period, with only 5 created, compared to 43 in the second period and 109 in the third period. This scarcity may result in exaggerated (e.g., for MIt in Figure 4.4l) or muted (e.g., for PA in Figure 4.4f) changes in mean complexity. The impact of the inclusion of PLVi_V is easily seen in Figure 4.4 comparing the whole group moving average (solid line) with the moving average for PELV plans (dashed line). The second period was not considered in this analysis as it represents a transitional period.

Table 4.2 is analogous to table 4.1 for the Prostatic group, where all plans of the group were included in the analysis.

Table 4.1: Comparison of mean metric values for PELV plans before protocol implementation and after practice stabilization. Values are reported as mean ± standard deviation. Except for AAV, LSV, and PA, all smaller values indicate less plan complexity. Statistically significant differences ($p < .05$) are in bold.

	Before Protocol (N=46)	After Stabilization (N=51)	Difference (%)	p-value
AAV	0.209 ± 0.027	0.266 ± 0.027	27.3	**0.00**
LSV	0.830 ± 0.013	0.827 ± 0.011	-0.4	0.23
PI	7.31 ± 0.48	7.06 ± 0.56	-0.3	**0.02**
SAS	0.453 ± 0.072	0.327 ± 0.046	-27.8	**0.00**
EM (cm^{-1})	1.08 ± 0.11	0.944 ± 0.068	-12.6	**0.00**
PA (cm^2)	49.3 ± 9.9	65.0 ± 8.7	31.8	**0.00**
CLS	0.0507 ± 0.0140	0.0380 ± 0.0168	-25.2	**0.00**
CAS	0.697 ± 0.046	0.637 ± 0.036	-8.6	**0.00**
MAD (cm)	27.9 ± 2.9	28.7 ± 2.2	2.9	0.16
MUd	4.97 ± 0.71	3.79 ± 0.28	-23.7	**0.00**
LT (m)	1.20 ± 0.05	1.14 ± 0.05	-5.0	**0.00**
MIt	123 ± 10	110 ± 10	-10.6	**0.00**

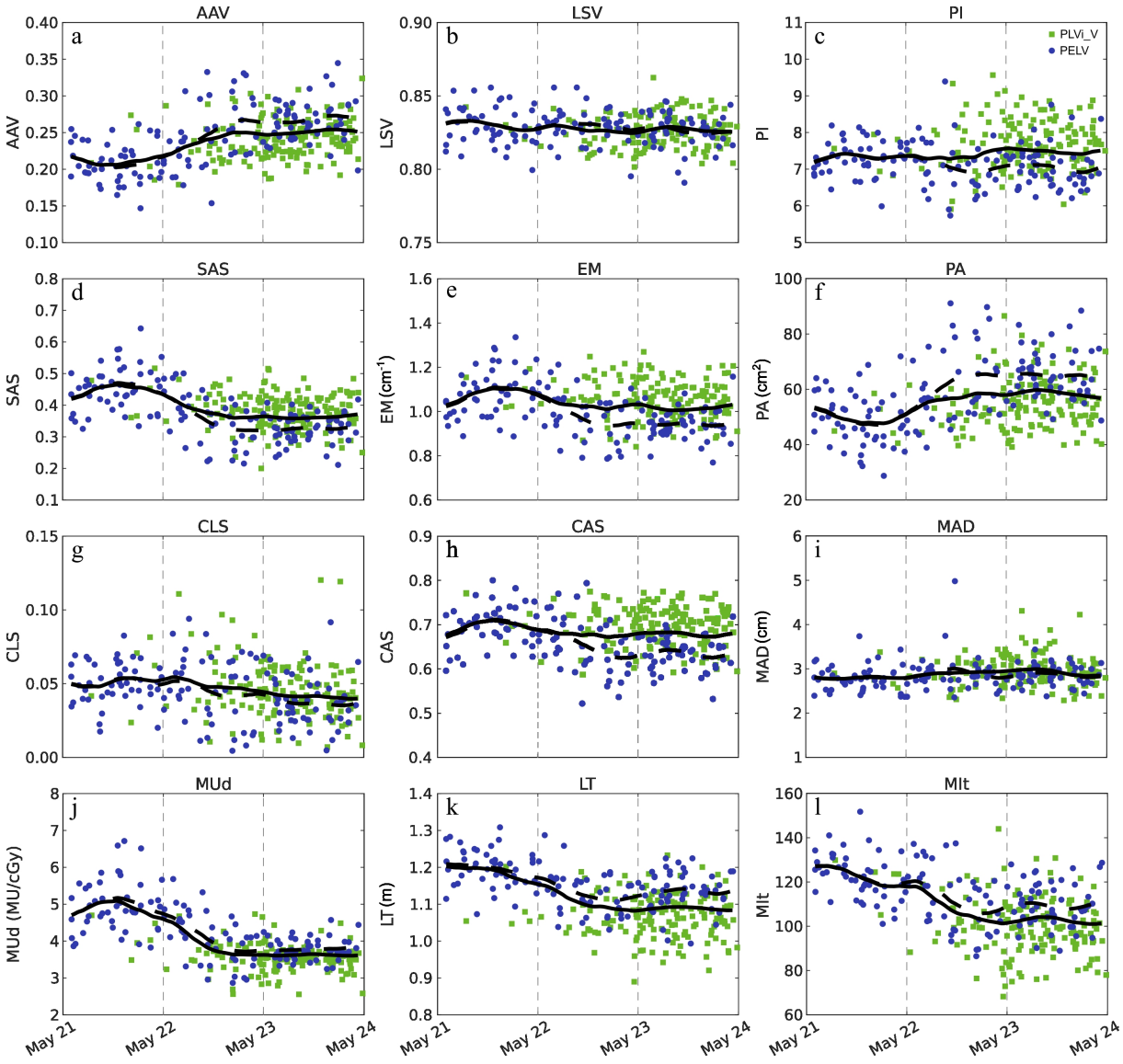


Figure 4.4: Longitudinal evolution of twelve complexity metrics for the Pelvic group from May 2021 to May 2024. PLVi_V plans are represented by green squares and PELV plans are represented by blue circles. The solid black line in each subplot is the moving average for the entire Pelvic group, while the dashed line is the moving average considering only the PELV plans.

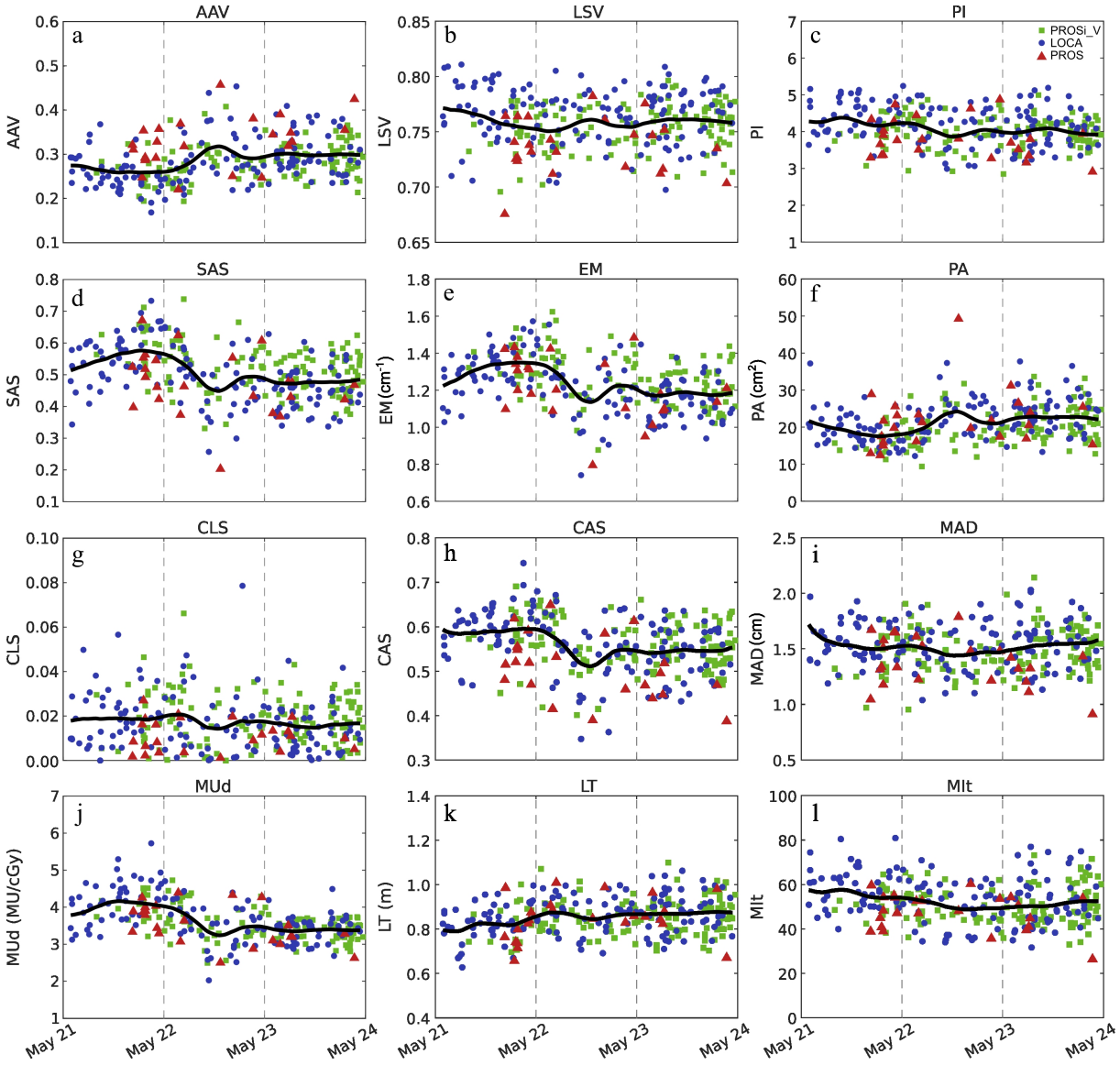


Figure 4.5: Longitudinal evolution of twelve complexity metrics for the Prostatic group from May 2021 to May 2024. PROSi_V plans are represented by green squares, LOCA plans are represented by blue circles, and PROS plans are represented by red triangles. The solid black line in each subplot is the moving average for the entire Prostatic group.

Table 4.2: Comparison of mean metric values for the Prostatic group before protocol implementation and after practice stabilization. Values are reported as mean ± standard deviation. Except for AAV, LSV, and PA, all smaller values indicate less plan complexity. Statistically significant differences ($p \leq 0.05$) are in bold.

	Before Protocol (N=75)	After Stabilization (N=125)	Difference (%)	p-value
AAV	0.262 ± 0.041	0.300 ± 0.043	14.5	**0.00**
LSV	0.760 ± 0.028	0.760 ± 0.023	0.0	0.95
PI	4.25 ± 0.49	3.98 ± 0.46	-6.4	**0.00**
SAS	0.561 ± 0.077	0.475 ± 0.067	-15.3	**0.00**
EM (cm^{-1})	1.32 ± 0.11	1.17 ± 0.11	-11.4	**0.00**
PA (cm^2)	18.6 ± 4.6	22.6 ± 5.0	21.5	**0.00**
CLS	0.0184 ± 0.0121	0.0160 ± 0.0096	-13.0	0.16
CAS	0.592 ± 0.064	0.546 ± 0.064	-7.8	**0.00**
MAD (cm)	15.4 ± 3.0	15.4 ± 3.1	-0.1	0.98
MUd	4.07 ± 0.51	3.37 ± 0.31	-17.2	**0.00**
LT (m)	0.820 ± 0.086	0.871 ± 0.087	6.2	**0.00**
MIt	55.6 ± 9.3	51.1 ± 10.2	-8.1	**0.01**

For the PELV plans, ten of the twelve metrics (AAV, PI, SAS, EM, PA, CLS, CAS, MUd, LT, and MIt) showed statistically significant differences, with all being reductions in complexity. The most notable differences were observed in the PA metric, with a 31.8% increase, followed by SAS with a 27.8% decrease, and AAV with a 27.3% decrease.

In the Prostatic group, nine out of twelve metrics (AAV, PI, SAS, EM, PA, CAS, MUd, LT, and MIt) exhibited statistically significant differences in complexity. Among these, only the LT metric showed a slight increase of 5.8%. The largest differences were observed in the PA metric with a 21.5% increase, followed by MUd with a 17.2% decrease, and SAS with a 15.3% decrease.

Both LSV and MAD metrics did not show statistically significant changes in either case. The MAD metric, being primarily dependent on the volume and location of the PTV, was anticipated to remain stable. The stability in the LSV metric indicates that the overall plan irregularity did not change sufficiently to cause variations in the leaf end positions between adjacent leaves.

Of the eight metrics (AAV, PI, SAS, EM, PA, CAS, MUd, and MIt) that demonstrated statistically significant reductions in complexity for both the PELV plans and the Prostatic group, all but the PI metric showed greater reductions in the PELV plans. This suggests that the Pelvic group plans benefited more substantially from the protocol implementation than plans from the Prostatic group.

Overall, after the implementation and adoption of the protocol, treatment plans feature larger aperture areas with less aperture variation, reduced monitor units, and lower mechanical strain on the linear accelerator. Despite these improvements, aperture irregularity and leaf movement

did not show substantial changes, with leaf movement increasing slightly (6.2%) in the Prostatic group.

Given the clear reduction in plan complexity, when possible, only plans from the stabilization period should be considered for the creation of RapidPlan models and for characterization, as their better reflect the current clinical practice at IPO Coimbra.

4.1.5 Clinical Practice Characterization

The typical complexity values for Halcyon VMAT plans at IPO Coimbra (third period) are reported in table 4.3 for each group. Unlike table 4.1, table 4.3 includes both plan types of the Pelvic Group. Complexity values reported for the Prostatic group are the same presented in table 4.2 for the third time periods.

Table 4.3: Typical complexity values for both groups.

	Pelvic Group	Prostatic Group
AAV	0.250 ± 0.028	0.300 ± 0.043
LSV	0.827 ± 0.011	0.760 ± 0.023
PI	7.49 ± 0.65	3.98 ± 0.46
SAS	0.363 ± 0.057	0.475 ± 0.067
EM (cm^{-1})	1.02 ± 0.09	1.17 ± 0.11
PA (cm^2)	58.5 ± 9.8	22.6 ± 5.0
CLS	0.0416 ± 0.0176	0.0160 ± 0.0096
CAS	0.680 ± 0.050	0.546 ± 0.064
MAD (cm)	29.3 ± 3.2	15.4 ± 3.1
MUd	3.62 ± 0.33	3.37 ± 0.31
LT (m)	1.09 ± 0.07	0.871 ± 0.087
MIt	103 ± 12	51.1 ± 10.2

4.2 Quality

4.2.1 Longitudinal Evolution

The longitudinal evolution of global plan score (GPS) and gradient index (GI) are represented for the four quality groups in figures 4.6 and 4.7, respectively, where the moving average is given by the solid line. The same three time periods identified in section 4.1 were used and are separated by vertical dashed lines. Visually, the quality values do not show an evident difference between the first and third time periods, except for the GPS values in PROS plans (Figure 4.6c), which improved over time. This, however, is likely attributable to differences in prescription doses rather than being a consequence of the decrease in plan complexity, given that, from the 9 plans created in the first period, 6 (66.6%) had prescription doses of 7600 cGy or 7800 cGy, whereas all plans created in the third period had a prescription dose of 7000 cGy. The possible influence of prescription dose on GPS will be further analyzed in section 4.2.2.

Table 4.4 summarizes the mean values of GPS and GI for the PELV_LOCA and PROSi_V groups in the first and third periods. p-values were calculated using the Welch's t-test for the PELV_LOCA group and the Mann–Whitney U test for PROSi_V. The PROS and PLVi_V plans were not evaluated given the low number of plans created during one of these periods ($N < 8$).

Table 4.4: Comparison of mean metric values for GI and GPS plans before protocol implementation and after clinical practice stabilization. Values are reported as mean ± standard deviation. The statistical significance of the differences in the means for PELV_LOCA group was assessed using the Welch's t-test. The statistical significance of the differences in the values' distributions for the PROSi_V group was assessed using the Mann–Whitney U test.

		Before Protocol	After Stabilization	Difference (%)	p-value
PELV_LOCA	GPS	0.395 ± 0.079	0.378 ± 0.092	-4.2	0.35
	GI	0.353 ± 0.088	0.348 ± 0.079	-3.5	0.47
PROSi_V	GPS	0.466 ± 0.107	0.479 ± 0.114	2.7	0.50
	GI	0.403 ± 0.095	0.393 ± 0.086	-2.5	0.72

As shown in Table 4.4, both PELV_LOCA and PROSi_V groups did not exhibited statistically significant differences between the first and third periods, despite the clear reduction in plan complexity observed in section 4.1.4. It is concluded that the planning guidelines at IPO Coimbra were effective in reducing complexity without deteriorating plan quality. This finding aligns with the results of Younge et al. [59] and Hui et al. [78], who maintained similar plan quality, as measured by a PQI, despite penalizing irregular apertures in the cost function.

To illustrate this result, two PROSi_V plans with prescription doses of 7000 cGy to the prostate and 6160 cGy to the seminal vesicles, created in the first and third time periods were selected and compared using the Structures Plan Diagram from SPIDERplan (Figure 4.8). The plan from the first period (red) exhibits higher complexity than the plan from the third period (green) across 10 of the 12 metrics (all metrics but LSV and LT), with an average complexity difference of 15.8% for these metrics. In spite of this, the green plan has better quality, achieving lower scores for every objective except two, as easily seen by its proximity to the origin on the diagram. This improved performance is further reflected in the GPS values, with the green plan scoring 0.318 compared to 0.444 for the red plan.

Although the longitudinal evolution of GPS and GI was not evaluated for the PLVi_V group due to the low number of plans in first period (N=5), it is likely, based on the trends observed in the PELV_LOCA and PROSi_V groups, that the reduction in plan complexity did not significantly impact the plan quality for this group. Therefore, all PLVi_V plans across the three time periods were included in the characterization analysis.

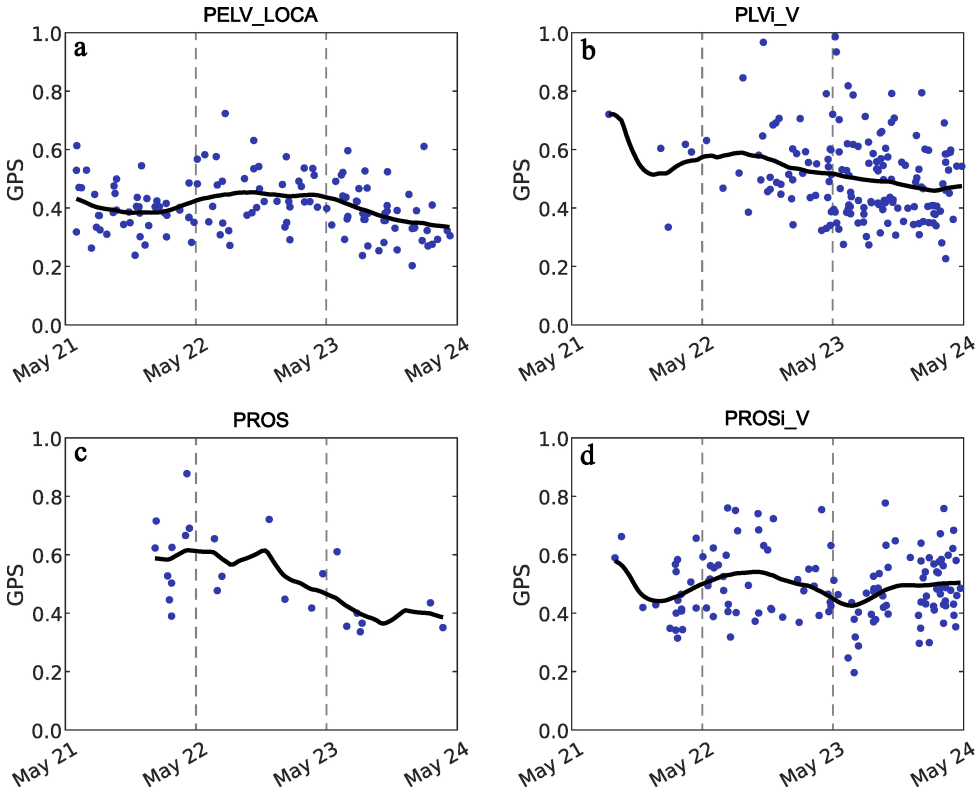


Figure 4.6: Longitudinal evolution of GPS

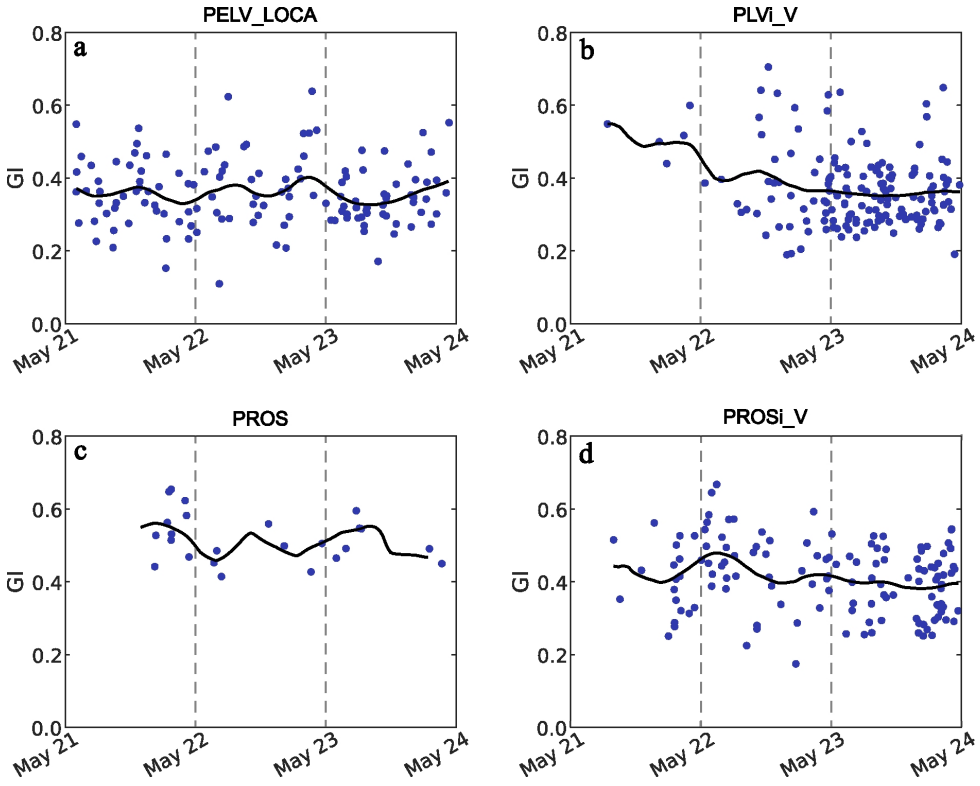


Figure 4.7: Longitudinal evolution of GI

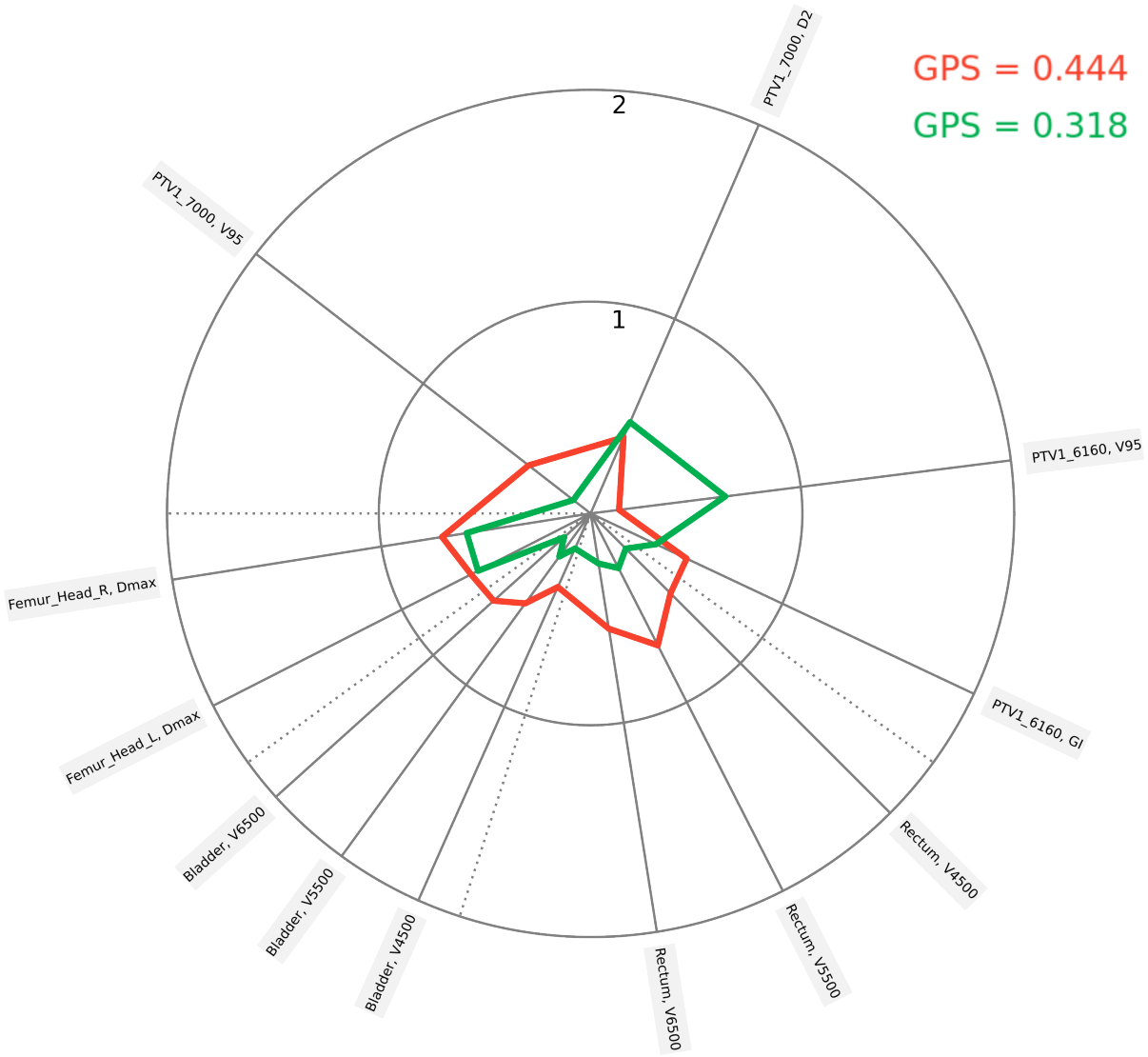


Figure 4.8: Structures Plan Diagram (SPD) for two PROSi_V plans created in the first (red) and the third (green) time periods. Diagram was created for the Prostate_2lvl_SIB template (Table 3.4).

4.2.2 Clinical Practice Characterization

Given that there is no clear change in quality between the first and third periods for PELV_LOCA, PROSi_V, and PLVi_V groups, all plans of these groups were considered for the clinical practice characterization. PROS plans exhibit high heterogeneity in terms of prescription dose, whose effect on the GPS values of PROS plans is assessed in this subsection.

The distributions of the SPIDERplan objectives' scores are illustrated through violin plots in figures 4.9, 4.10, 4.11, and 4.12 for the PELV_LOCA, PLVi_V, PROS, and PROSi_V plans, respectively.

Among the structures analyzed, the bladder exhibited the highest score values, particularly at the volume objectives for the higher doses ($V_{7000\,\mathrm{cGy}}$ or $V_{6500\,\mathrm{cGy}}$), with each group having at least one instance where the score for one of these bladder objectives exceeded 3. Following the bladder, the rectum also showed wide score amplitudes. This tendency of high scores for the bladder and rectum can be attributed to their anatomical proximity to the PTV, which often results in parts of these organs being encompassed by the PTV. Indeed, all 14 plans where at least one score for the bladder exceeded 2.5 involved scenarios where significant portions of these organs were within the PTV boundaries. Regarding the PTVs, the $D_{2\%}$ scores consistently stayed below threshold limits, meaning that the near-maximum dose was below 107% of the prescribed dose. On the other hand, 43 plans (9.7% of the cohort) had $V_{95\%}$ scores above the threshold of 1. This group includes 7 PELV_LOCA, 21 PLVi_V, 13 PROSi_V, and 2 PROS plans. Of these, most (7, 17, 11, and 2 cases, respectively), showed these values for the primary PTV (prostate or prostate bed). Only one PLVi_V plan failed the threshold criterion of $V_{95\%}$ for two targets. Additionally, 27 out of the 43 plans (62.8%) with $V_{95\%}$ scores above 1 also had scores exceeding 1 for at least one OAR objective. This could indicate that, in some of these cases, compromising the PTV coverage was necessary to mitigate the violation of OAR constraints.

The observable dependence of the GPS on patient anatomy poses challenges in the selection of plans for model training. In particular, higher GPS (lower quality) plans must be manually evaluated, as high GPS values may result from significant overlap of OARs with the PTV and arbitrarily excluding such plans could reduce RapidPlan's ability to generate optimal plans for similar anatomies [101].

The reconceptualized PTV score functions for the achieved objectives (score < 1) produced values covering the entire range, as demonstrated in the violin plots. If using the previously used score functions, the best achievable score for the $V_{95\%}$ objective, according to equation 3.32, would be $0.99/100 = 0.99$, a value that would not adequately differentiate between plans. In contrast, the highest value of $V_{95\%}$ for a PTV was 1.0479, which is equivalent to the 95% isodose covering 94.5% of the target volume, considerably less than the threshold value of 99%. This slight deviation in score is not evident in the violin plots. Moreover, even using the template with the highest objective weight (Prostate_1lvl), it only results in the small penalization of 0.015 on the GPS. With the current PTV score functions, plans with unmet PTV objectives but

good OAR sparing would be favored by GPS, despite not being clinically adequate. As such, future adjustments to PTV score functions are warranted.

The objective scores for the PROS plans were closely linked to the prescription dose, as to be expected. For instance, among the 9 plans where at least one bladder score exceeded 1, 6 plans (66.6%) had prescription doses of 7600 cGy or 7800 cGy, despite these plans constituting only 33.3% of the total PROS plans. A examination of the dose distribution of all plans with prescription doses higher than 7600 cGy did not provide evidence suggesting the observed differences are due to variations in the amount of bladder tissue within the PTV. However, the subjectivity of this examination is acknowledged.

Furthermore, all 7 plans with rectum $V_{7200\,\mathrm{cGy}}$ scores above 0.35 had prescription doses of 7600 cGy or 7800 cGy, accounting for 88.8% of these plans. In contrast, the 3 plans with lower prescription doses of 6000 cGy or 6800 cGy had $V_{7200\,\mathrm{cGy}}$ scores less than 0.1. Considering the maximum observed values for these objectives (around 1 for the rectum objective and 3.6 for the bladder) and the Prostate_1lvl template used for these plans (Table 3.2), differences of up to 0.2 (0.13 from the prostate and 0.07 from the rectum) can be explained by the differences in prescription dose.

These differences are similar to the observed amplitude of mean GPS values (Figure 4.7c), suggesting that the improvement in quality for PROS plans may have resulted from the unbalanced distribution of prescription doses between time periods. Hence, Figure 4.11 might be representative of the current clinical practice at IPO Coimbra. However, the reduced number of PROS plans hinders drawing conclusive results.

In the ideal scenario, only PROS plans from the third period would be use for model training. However, 7 plans are far from the minimum number of 20 plans recommended by the manufacturer [85]. Three solutions can be considered:

- Given that we do not have strong reasons to believe that plans from the first period were worse in quality, all 24 PROS plans can be used as a compromise, despite the different plan complexity levels. Regarding the widely different prescription doses, they would probably not pose a problem, as RapidPlan models are trained considering the relative doses on OARs compared to the prescription dose (considered the 100% level) [91]. Therefore, models trained on plans with different prescription doses should, in principle, not weaken the model, as observed by Fogliata et al. [91].

- The model trained on PROSi_V plans may be used if proven to successfully generalize to PROS plans, considering that the PROSi_V plans only differ in the inclusion of the seminal vesicles as a target, which is also a relatively small volume.

- Models provided by the manufacturer can be used, although it might be more difficult to achieve plans with similar quality to those in the database.

The PELV_LOCA group (mostly composed of 6700 cGy and 6900 cGy prescription doses) also demonstrated some dependency on the prescription dose. For instance, all plans where $V_{7000\,\mathrm{cGy}}$ scores for the bladder and rectum exceeded 0.1 had prescription doses of 6900 cGy, with maximum scores reaching 0.73 for the bladder and 4.1 for the rectum in these objectives. The different prescription doses account for up to 0.26 of variation in the GPS value, considering the objective weights for the Prostate_2lvl_SIB template (Table 3.3). However, the variation in GPS values for PELV_LOCA plans was less pronounced compared to the PROS plans. This is because the three objectives that contributed most to the GPS for these plans, $V_{4500\,\mathrm{cGy}}$ for the bladder and $D_{\max}$ for the femur heads (Figure 4.9), exhibited smaller discrepancies across different prescription doses. Specifically, 44% of the plans with a prescription dose of 6700 cGy had $V_{4500\,\mathrm{cGy}}$ scores higher than the mean score value, compared to 54% of the plans with a prescription dose of 6900 cGy. For the $D_{\max}$ scores of the left and right femur heads, the differences were 40% against 53%, and 49% against 47%, respectively.

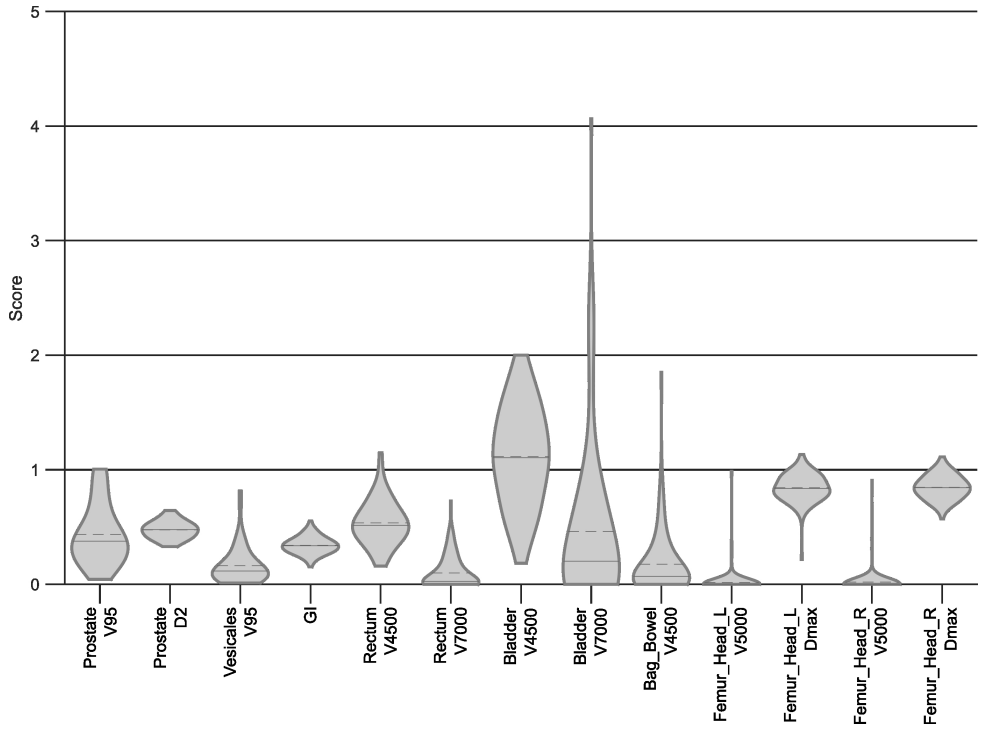


Figure 4.9: Violin Plot for PELV plans. In each violin, the dashed line represents the mean value and the solid line the median.

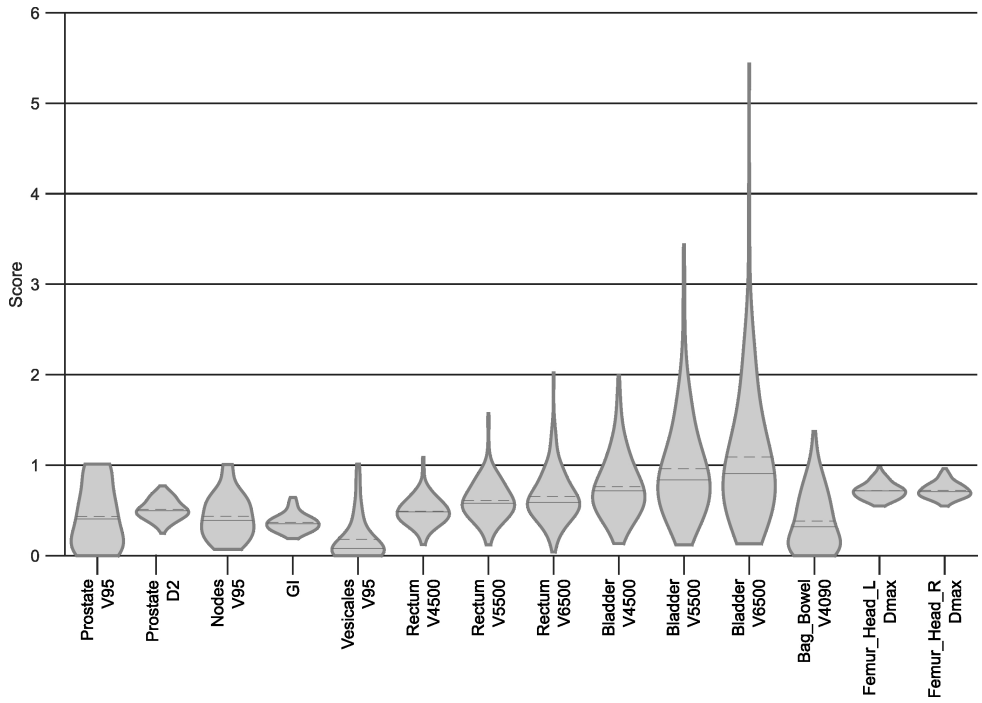


Figure 4.10: Violin Plot for PLVi_V plans. In each violin, the dashed line represents the mean value and the solid line the median.

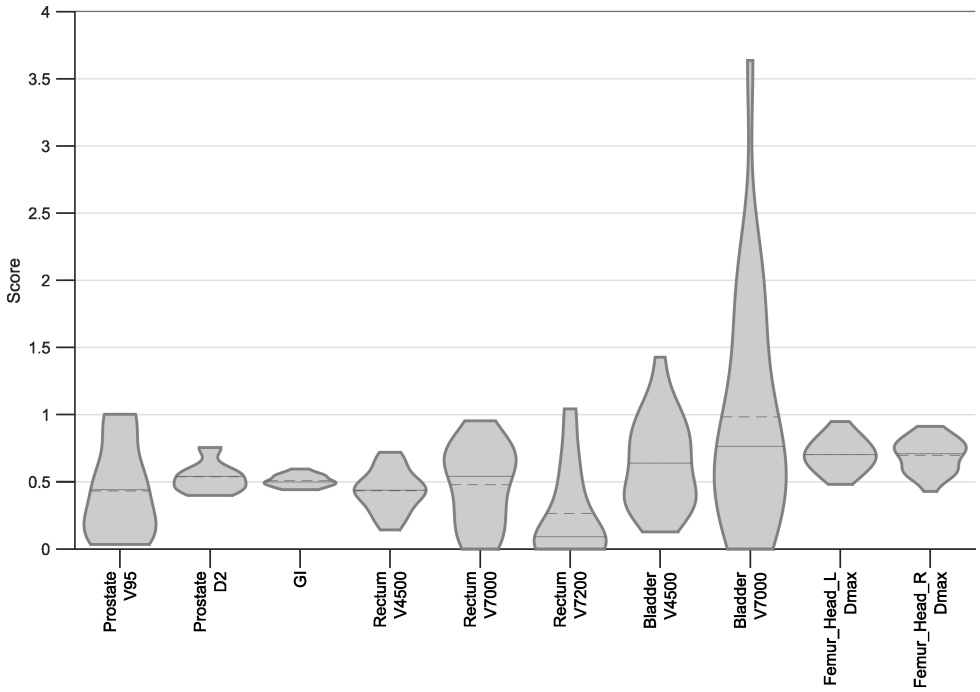


Figure 4.11: Violin Plot for PROS plans. In each violin, the dashed line represents the mean value and the solid line the median.

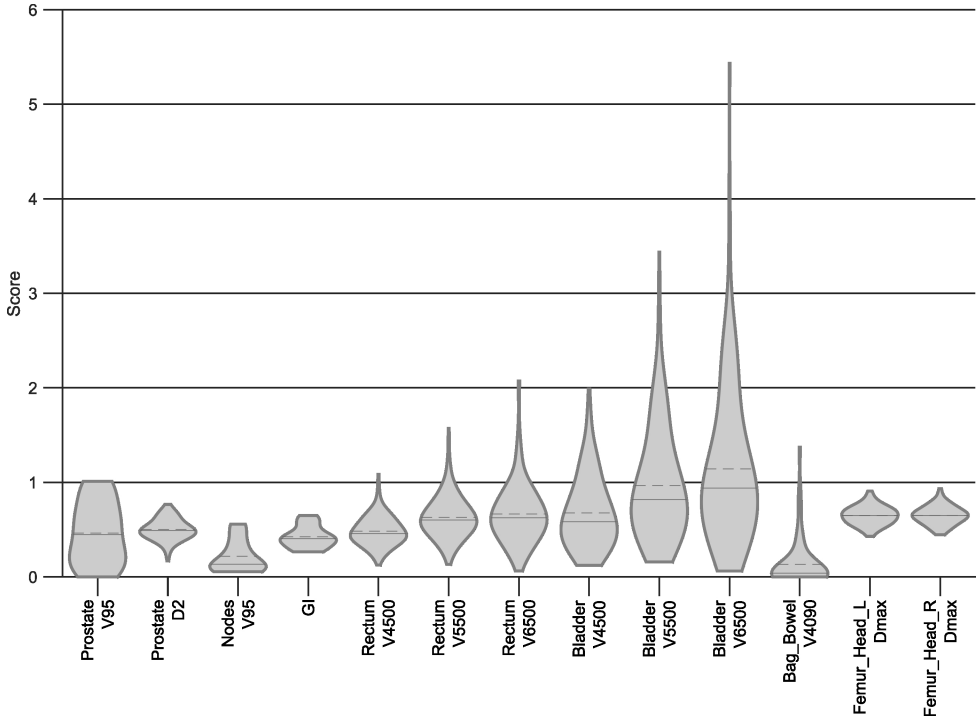


Figure 4.12: Violin Plot for PROSi_V plans. In each violin, the dashed line represents the mean value and the solid line the median.

The violin plots for the GPS and GI values across the groups are shown in figures 4.13 and 4.14, respectively. After both plots rejecting the null hypothesis of the Kruskal-Wallis test that all plan groups follow the same score distribution ($p < 0.05$), pairwise comparisons were conducted using the Mann-Whitney U test. A significance level of 1.25% was used after the Bonferroni correction for 4 comparisons.

For the GPS values, the distribution of PELV_LOCA was found to be significantly different from the distributions of the other plans (PLVi_V, PROS, and PROSi_V), as seen in Figure 4.13. Regarding the GI values, the test and the visual inspection of the violin plot suggest three distinct levels of distribution (Figure 4.14). From lowest to highest GI values, the levels would be composed by PELV_LOCA and PLVi_V, PROS, and PROSi_V plans. It is important to note that these results do not indicate what plan groups performed better dosimetrically, but rather the plans that better met the template objectives set by the radiation oncologist.

The GPS violin for the PROS group (Figure 4.13) reveals two distinct populations centered around two scores: approximately 0.48 for PROS plans with a prescription dose of 7000 cGy and approximately 0.67 for those with a prescription dose of 7600 cGy. This observation was further validated by the Mann-Whitney U test, which rejected the null hypothesis that there is no difference between these two sub-populations with different prescription doses. The result corroborates the previous assumption that the GPS values for the PROS group are dependent on the prescription dose.

Despite the Mann-Whitney U test also rejecting the null hypothesis for the GPS distributions for PELV_LOCA plans with prescription doses of 6700 cGy and 6900 cGy, the violin plot does not clearly depict separate populations. This lack of visual separation is due to the proximity of the mean GPS values (0.359 for 6700 cGy and 0.433 for 6900 cGy), which are relatively close compared to their standard deviations (0.088 and 0.093, respectively). For instance, if the scores followed normal distributions, and assuming that the violin plot accurately predicts the probability density function, their difference in mean value would need to exceed 0.180 to be noticeable [103].

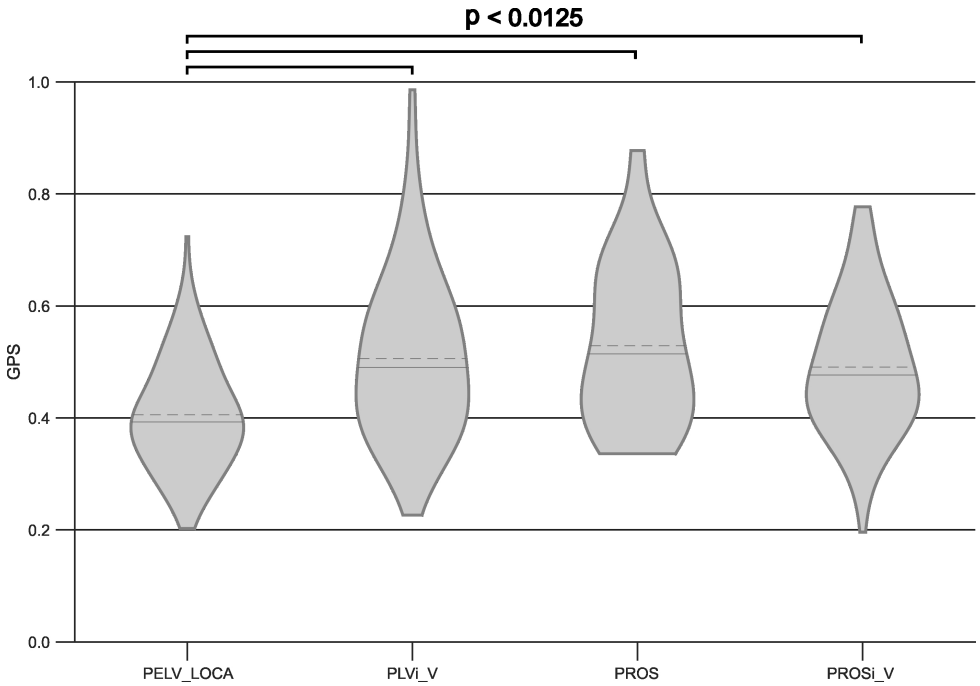


Figure 4.13: Violin Plot for SPIDERplan GPS. All statistical significant differences between distributions are highlighted

Figure 4.14: Violin Plot for GI metric. All statistical significant differences between distributions are highlighted

5 Conclusions

Treatment planning is time-intensive and highly dependent on the planner's skills and experience, prompting the implementation of automated planning solutions. One category of such solutions, termed knowledge-based planning (KBP), uses prior created treatment plans to generate planning objectives tailored to the patient anatomy. The present work aimed to establish a database of prostate radiotherapy plans created at IPO Coimbra for the Varian Halcyon linac and characterizing them in terms of quality and complexity for later validation of the Varian Eclipse KBP solution, RapidPlan.

To facilitate the characterization and creation of a plan database, an electronic platform called qualComp was developed for automatic calculation of the quality and complexity of each new plan, as well as the graphical comparison with the existing plan database. For plan quality, new score functions for the PTV objectives were introduced and a graphical interface was created for SPIDERplan with included statistical analysis. Regarding plan complexity, 15 metrics were adapted for the Halcyon MLC considering each layer separately, weighting the contribution of each layer to the effective aperture, and using the 5 mm effective apertures.

Plans from prostate patients treated with VMAT in a Halcyon linac over the span of three years since the linac installation at IPO Coimbra were retrospectively selected for characterization. To ensure the characterized plans are the representative of the current clinical practice at IPO Coimbra, the longitudinal evolution of plan quality and complexity was also performed, focusing on the impact of the introduction of guidelines aiming to standardize treatment planning and reduce plan complexity. Quality was measured using the SPIDERplan global plan score (GPS) with templates configured according to the clinical preferences of the radiation oncologists. A dose gradient index (GI) proposed in this work was used as one of the SPIDERplan template objectives. Single-layer and weighted complexity metrics were not found useful for plan characterization and thus excluded, while three metrics were omitted due to their strong correlation with similar metrics.

The longitudinal evaluation revealed a clear reduction in plan complexity followed by stabilization period, with newer plans featuring larger aperture areas, less aperture variation, reduced linac mechanical strain, and higher efficiency regarding monitor units, albeit with similar aperture irregularity. Conversely, plan quality, as measured by the SPIDERplan GPS, did not show statistically significant changes, indicating that the introduction of planning guidelines success-

fully reduced plan complexity without impacting plan quality.

Given these observed trends in the longitudinal evaluation, the characterization of prostate radiotherapy treatment plans created for the Halcyon linac at IPO Coimbra was based on the stabilization period for complexity and on all created plans for quality. These periods used for characterization will be use to test the model performance, as they are representative of current practices at IPO Coimbra.

To train RapidPlan models, it is advisable to select plans from the stabilization period, where both plan quality and complexity were found to be stable. An exception could be considered for plans involving only irradiation of the prostate (PROS), where the number of plans available for this period falls below the manufacturer's recommended minimum for model training. If a model for these plans is desired, a compromise could be made by selecting PROS plans from all periods, considering that, despite the reduction in complexity, there was no clear evidence that quality varied. Alternatively, models trained on simultaneous boost plans of the prostate and seminal vesicles (PROSi_V) were hypothesized to effectively generalize to PROS plans due to their similarities.

The SPIDERplan global plan score (GPS) was found to be ineffective for selecting individual plans for model training, as structures with substantial overlap with the target increase the GPS regardless of the plan's performance for the specific anatomy. In such scenarios, visual assessment of the dose distribution becomes necessary. The PTV score functions for SPIDERplan introduced in this study were found to better distinguished between plans that met the PTV objectives but were insensitive to unmet objectives, indicating a need for future adjustments.

Despite focusing on prostate cases, the methodology presented in this work can be easily applied to other pathologies. However, this may require extending qualComp to support radiotherapy plans for the intended linac, enabling it to calculate complexity metrics specific to that linac.

Overall, the development of the qualComp platform and the characterization of plan quality and complexity pave the way for reliable configuration, validation, and testing of automated planning solutions.

Bibliography

[1] F. Bray et al. “Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. In: *CA: A Cancer Journal for Clinicians* 74.3 (2024), pp. 229–263. DOI: 10.3322/caac.21834.

[2] Y. Lievens, J. Borras, and C. Grau. “Provision and use of radiotherapy in Europe”. In: *Molecular Oncology* 14 (May 2020). DOI: 10.1002/1878-0261.12690.

[3] S. O. Hunte et al. “Volumetric modulated arc therapy (VMAT): a review of clinical outcomes-what is the clinical evidence for the most effective implementation?” In: *Br. J. Radiol.* 95.1136 (Oct. 2022), p. 20201289. DOI: 10.1259/bjr.20201289.

[4] *ICRU Report 83 Prescribing, Recording, and Reporting Photon-beam Intensity-modulated Radiation Therapy (IMRT)*. Oxford University Press, 2010. DOI: 10.1093/jieru/ndq001.

[5] B. E. Nelms et al. “Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems”. In: *Pract. Radiat. Oncol.* 2.4 (Oct. 2012), pp. 296–305. DOI: 10.1016/j.prro.2011.11.012.

[6] P. Meyer et al. “Automation in radiotherapy treatment planning: Examples of use in clinical practice and future trends for a complete automated workflow”. In: *Cancer/Radiothérapie* 25 (June 2021). DOI: 10.1016/j.canrad.2021.06.006.

[7] A.-M. Fanou et al. “Implementation, Dosimetric Assessment, and Treatment Validation of Knowledge-Based Planning (KBP) Models in VMAT Head and Neck Radiation Oncology”. In: *Biomedicines* 11.3 (2023). ISSN: 2227-9059. DOI: 10.3390/biomedicines11030762.

[8] N. Lambri et al. “Evaluation of plan complexity and dosimetric plan quality of total marrow and lymphoid irradiation using volumetric modulated arc therapy”. In: *Journal of Applied Clinical Medical Physics* 24.6 (2023), e13931. DOI: 10.1002/acm2.13931.

[9] A. 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: 10.1016/j.ejmp.2023.102542.

[10] A Fogliata et al. “RapidPlan knowledge based planning: iterative learning process and model ability to steer planning strategies”. In: *Radiat. Oncol.* 14.1 (Oct. 2019), p. 187.

[11] V. Ahire et al. “Radiobiology of Combining Radiotherapy with Other Cancer Treatment Modalities”. In: *Radiobiology Textbook*. Ed. by S. Baatout. Cham: Springer International Publishing, 2023, pp. 311–386. ISBN: 978-3-031-18810-7. DOI: 10.1007/978-3-031-18810-7_6.

[12] L. Zhang et al. “Delivery of therapeutic radioisotopes using nanoparticle platforms: potential benefit in systemic radiation therapy”. In: *Nanotechnol. Sci. Appl.* 3 (Dec. 2010), pp. 159–170. DOI: 10.2147/NSA.S7462.

[13] E. Podgoršak. *Radiation Oncology Physics: A Handbook for Teachers and Students*. International Atomic Energy Agency, 2005. ISBN: 9789201073044. URL: https://www-pub.iaea.org/mtcd/publications/pdf/pub1196_web.pdf.

[14] A. Boyer et al. *Basic Applications of Multileaf Collimators*. AAPM Report No. 72. American Association of Physicists in Medicine, 2001. URL: https://aapm.org/pubs/reports/RPT_72.pdf.

[15] M. S. Huq et al. “A dosimetric comparison of various multileaf collimators”. In: *Phys. Med. Biol.* 47.12 (June 2002), N159–70. DOI: 10.1088/0031-9155/47/12/401.

[16] P. Symonds et al. *Walter and Miller’s Textbook of Radiotherapy*. Elsevier Health Sciences, 2012. ISBN: 9780443074868.

[17] F. Van den Heuvel, Q. Wu, and J. Cai. “In modern linacs monitor units should be defined in water at 10 cm depth rather than at dmax”. In: *Med. Phys.* 45.11 (Nov. 2018), pp. 4789–4792. DOI: 10.1002/mp.13015.

[18] A. Agarwal et al. “Investigating the electronic portal imaging device for small radiation field measurements”. In: *J. Med. Phys.* 42.2 (Apr. 2017), pp. 59–64. DOI: 10.4103/jmp.JMP_131_16.

[19] D. Pokhrel et al. “Feasibility of using ring-mounted Halcyon Linac for single-isocenter/two-lesion lung stereotactic body radiation therapy”. In: *Journal of Applied Clinical Medical Physics* 23.5 (2022), e13555. DOI: 10.1002/acm2.13555.

[20] M. Tamura et al. “Plan complexity quantification of dual-layer multi-leaf collimator for volumetric modulated arc therapy with Halcyon linac”. In: *Phys. Eng. Sci. Med.* 43.3 (Sept. 2020), pp. 947–957. DOI: 10.1007/s13246-020-00891-2.

[21] L. Begnozzi et al. “Quality assurance of 3D-CRT: Indications and difficulties in their applications”. In: *Critical Reviews in Oncology/Hematology* 70.1 (2009), pp. 24–38. ISSN: 1040-8428. DOI: 10.1016/j.critrevonc.2008.07.016.

[22] M Teoh et al. “Volumetric modulated arc therapy: a review of current literature and clinical use in practice”. In: *Br. J. Radiol.* 84.1007 (Nov. 2011), pp. 967–996. DOI: 10.1259/bjr/22373346.

[23] *Setting Up a Radiotherapy Programme*. Non-serial Publications. Vienna: International Atomic Energy Agency, 2008. ISBN: 92-0-101807-X. URL: https://www.iaea.org/publications/7694/setting-up-a-radiotherapy-programme.

[24] G. Chaput and L. Regnier. “Radiotherapy: Clinical pearls for primary care”. In: *Canadian family physician Medecin de famille canadien* 67 (Oct. 2021), pp. 753–757. DOI: 10.46747/cfp.6710753.

[25] D. S. Chang et al. “Advanced Treatment Planning for EBRT”. In: *Basic Radiotherapy Physics and Biology*. Cham: Springer International Publishing, 2014. ISBN: 978-3-319-06841-1. DOI: 10.1007/978-3-319-06841-1_12.

[26] I Rosenberg. “Radiation Oncology Physics: A Handbook for Teachers and Students”. In: *British Journal of Cancer* 98.5 (Mar. 2008), 1020–1020. ISSN: 1532-1827. DOI: 10.1038/sj.bjc.6604224. URL: http://dx.doi.org/10.1038/sj.bjc.6604224.

[27] L. Masi et al. “Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy”. In: *Medical physics* 40 (July 2013), p. 071718. DOI: 10.1118/1.4810969.

[28] M. Miften et al. “Tolerance Limits and Methodologies for IMRT Measurement-Based Verification QA: Recommendations of AAPM Task Group No. 218”. In: *Medical Physics* 45 (Feb. 2018). DOI: 10.1002/mp.12810.

[29] S. Chiavassa et al. “Complexity metrics for IMRT and VMAT plans: A review of current literature and applications”. In: *The British Journal of Radiology* 92 (July 2019), p. 20190270. DOI: 10.1259/bjr.20190270.

[30] A. Mans et al. “The NCS code of practice for the quality assurance and control for volumetric modulated arc therapy”. In: *Phys. Med. Biol.* 61.19 (Oct. 2016), pp. 7221–7235. DOI: 10.1088/0031-9155/61/19/7221.

[31] M. 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: 10.1016/j.ejmp.2019.05.024.

[32] D. A. Low et al. “A technique for the quantitative evaluation of dose distributions”. In: *Med. Phys.* 25.5 (May 1998), pp. 656–661.

[33] J. Tepper, R. Foote, and J. Michalski. *Gunderson & Tepper’s Clinical Radiation Oncology*. Elsevier, 2021. ISBN: 9780323672481.

[34] V. Hernandez et al. “What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans”. In: *Radiother. Oncol.* 153 (Dec. 2020), pp. 26–33. DOI: 10.1016/j.radonc.2020.09.038.

[35] L. P. Kaplan and S. S. Korreman. “A systematically compiled set of quantitative metrics to describe spatial characteristics of radiotherapy dose distributions and aid in treatment planning”. In: *Physica Medica* 90 (2021), pp. 164–175. ISSN: 1120-1797. DOI: doi.org/10.1016/j.ejmp.2021.09.014.

[36] E. Shaw et al. “Radiation therapy oncology group: Radiosurgery quality assurance guidelines”. In: *International Journal of Radiation Oncology*Biology*Physics* 27.5 (1993), pp. 1231–1239. ISSN: 0360-3016. DOI: 0.1016/0360-3016(93)90548-A.

[37] M. Yoon et al. “A new homogeneity index based on statistical analysis of the dose-volume histogram”. In: *J. Appl. Clin. Med. Phys.* 8.2 (Mar. 2007), pp. 9–17. DOI: 10.1120/jacmp.v8i2.2390.

[38] L. Yan et al. “A new homogeneity index definition for evaluation of radiotherapy plans”. In: *Journal of Applied Clinical Medical Physics* 20 (Oct. 2019). DOI: 10.1002/acm2.12739.

[39] L. Feuvret et al. “Conformity index: A review”. In: *International Journal of Radiation Oncology*Biology*Physics* 64.2 (2006), pp. 333–342. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2005.09.028.

[40] A. van’t Riet et al. “A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: Application to the prostate”. In: *International Journal of Radiation Oncology*Biology*Physics* 37.3 (1997), pp. 731–736. ISSN: 0360-3016. DOI: doi.org/10.1016/S0360-3016(96)00601-3.

[41] D. Baltas et al. “A conformal index (COIN) to evaluate implant quality and dose specification in brachytherapy”. In: *International Journal of Radiation Oncology*Biology*Physics* 40.2 (1998), pp. 515–524. ISSN: 0360-3016. DOI: 10.1016/S0360-3016(97)00732-3.

[42] G. M. M. Videtic et al. “A randomized phase 2 study comparing 2 stereotactic body radiation therapy schedules for medically inoperable patients with stage I peripheral non-small cell lung cancer: NRG oncology RTOG 0915 (NCCTG N0927)”. In: *Int. J. Radiat. Oncol. Biol. Phys.* 93.4 (Nov. 2015), pp. 757–764. DOI: 10.1016/j.ijrobp.2015.07.2260.

[43] I. Paddick and B. Lippitz. “A simple dose gradient measurement tool to complement the conformity index”. In: *Journal of neurosurgery* 105 Suppl (Dec. 2006), pp. 194–201. DOI: 10.3171/sup.2006.105.7.194.

[44] Q. Wu et al. “Quality of coverage: Conformity measures for stereotactic radiosurgery”. In: *Journal of applied clinical medical physics / American College of Medical Physics* 4 (Feb. 2003), pp. 374–81. DOI: 10.1120/1.1621372.

[45] K. Sung and Y. Choi. “Dose gradient curve: A new tool for evaluating dose gradient”. In: *PLOS ONE* 13 (Apr. 2018), e0196664. DOI: 10.1371/journal.pone.0196664.

[46] Y.-b. Cho et al. “A new conformity and dose gradient distance measure for stereotactic radiosurgery of brain metastasis”. In: *Journal of radiosurgery and SBRT* 8 (Jan. 2022), pp. 27–36.

[47] V. Y. Kuperman, Y. Altundal, and T. N. Kouskoulas. “Toward an improved assessment of dose conformity in radiotherapy”. In: *Medical Physics* 51.3 (2024), pp. 2210–2220. DOI: 10.1002/mp.16775.

[48] X Allen Li et al. “The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM”. In: *Med. Phys.* 39.3 (Mar. 2012), pp. 1386–1409.

[49] K. Rayn et al. “Scorecards: Quantifying Dosimetric Plan Quality in Pancreatic Ductal Adenocarcinoma Stereotactic Body Radiation Therapy”. In: *Advances in radiation oncology* 8 (June 2023), p. 101295. DOI: 10.1016/j.adro.2023.101295.

[50] T. Ventura et al. “SPIDERplan: A tool to support decision-making in radiation therapy treatment plan assessment”. In: *Rep. Pract. Oncol. Radiother.* 21.6 (Nov. 2016), pp. 508–516. DOI: 10.1016/j.rpor.2016.07.002.

[51] M. Esposito et al. “SBRT planning for spinal metastasis: indications from a large multicentric study”. In: *Strahlentherapie und Onkologie* 195 (Mar. 2019). DOI: 10.1007/s00066-018-1383-2.

[52] K. Balaji and V. Ramasubramanian. “Integrated scoring approach to assess radiotherapy plan quality for breast cancer treatment”. In: *Rep. Pract. Oncol. Radiother.* 27.4 (Sept. 2022), pp. 707–716. DOI: 10.5603/RPOR.a2022.0083.

[53] J. Krishnan, S. Rao, and S. Hegde. “Plan Quality Index - an integrated dosimetric approach for plan evaluation with consideration of quality of dose coverage to tumors and quality of organs’ sparing”. In: *Asian Pac. J. Cancer Prev.* 23.9 (Sept. 2022), pp. 3009–3017. DOI: 10.31557/APJCP.2022.23.9.3009.

[54] K. Kubo et al. “Inter-planner variation in treatment-plan quality of plans created with a knowledge-based treatment planning system”. In: *Physica Medica* 67 (Nov. 2019), pp. 132–140. DOI: 10.1016/j.ejmp.2019.10.032.

[55] T. Ventura et al. “Clinical validation of a graphical method for radiation therapy plan quality assessment”. In: *Radiation Oncology* 15 (Mar. 2020). DOI: 10.1186/s13014-020-01507-5.

[56] 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 & Biology* 60.6 (Feb. 2015), p. 2587. DOI: 10.1088/0031-9155/60/6/2587.

[57] J Götstedt, A. K. Hauer, and A Bäck. “Complexity metric as a complement to measurement based IMRT/VMAT patient-specific QA”. In: *Journal of Physics: Conference Series* 573.1 (Jan. 2015), p. 012016. DOI: 10.1088/1742-6596/573/1/012016.

[58] S. B. Crowe et al. “Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results”. In: *Australas. Phys. Eng. Sci. Med.* 37.3 (Sept. 2014), pp. 475–482. DOI: 10.1007/s13246-014-0274-9.

[59] K. Younge et al. “Penalization of aperture complexity in inversely planned volumetric modulated arc therapy”. In: *Medical physics* 39 (Nov. 2012), pp. 7160–70. DOI: 10.1118/1.4762566.

[60] J. M. Park et al. “Modulation indices for volumetric modulated arc therapy”. In: *Phys. Med. Biol.* 59.23 (Dec. 2014), pp. 7315–7340. DOI: 10.1088/0031-9155/59/23/7315.

[61] G. Nicolini et al. “What is an acceptably smoothed fluence? Dosimetric and delivery considerations for dynamic sliding window IMRT”. In: *Radiation oncology (London, England)* 2 (Nov. 2007), p. 42. DOI: 10.1186/1748-717X-2-42.

[62] W. Du et al. “Quantification of beam complexity in intensity-modulated radiation therapy treatment plans”. In: *Medical physics* 41 (Feb. 2014), p. 021716. DOI: 10.1118/1.4861821.

[63] A. L. McNiven, M. B. Sharpe, and T. G. Purdie. “A new metric for assessing IMRT modulation complexity and plan deliverability”. In: *Medical Physics* 37.2 (2010), pp. 505–515. DOI: 10.1118/1.3276775.

[64] J. Unkelbach et al. “Optimization approaches to volumetric modulated arc therapy planning”. In: *Med. Phys.* 42.3 (Mar. 2015), pp. 1367–1377. DOI: 10.1118/1.4908224.

[65] J. Götstedt and A. Bäck. “Edge area metric complexity scoring of volumetric modulated arc therapy plans”. In: *Physics and Imaging in Radiation Oncology* 17 (2021), pp. 124–129. ISSN: 2405-6316. DOI: 10.1016/j.phro.2021.02.002.

[66] J. Zheng, Y. Xia, and L. Sun. “A comprehensive evaluation of the application of the Halcyon(2.0) IMRT technique in long-course radiotherapy for rectal cancer”. In: *Technol. Cancer Res. Treat.* 21 (Jan. 2022), p. 15330338221074501. DOI: 10.1177/15330338221074501.

[67] H. M. Petroccia et al. “Spine SBRT with Halcyon™: Plan quality, modulation complexity, delivery accuracy, and speed”. In: *Front. Oncol.* 9 (Apr. 2019), p. 319. DOI: 10.3389/fonc.2019.00319.

[68] Y. Huang and Z. Liu. “Dosimetric performance evaluation of the Halcyon treatment platform for stereotactic radiotherapy: A pooled study”. In: *Medicine* 102 (Sept. 2023), e34933. DOI: 10.1097/MD.0000000000034933.

[69] H. Zhu et al. “Patient-specific quality assurance prediction models based on machine learning for novel dual-layered MLC linac”. In: *Med. Phys.* 50.2 (Feb. 2023), pp. 1205–1214. DOI: 10.1002/mp.16091.

[70] P. Quintero et al. “Effect of treatment planning system parameters on beam modulation complexity for treatment plans with single-layer multi-leaf collimator and dual-layer stacked multi-leaf collimator”. In: *Br. J. Radiol.* 94.1122 (June 2021), p. 20201011. DOI: 10.1259/bjr.20201011.

[71] K. Younge et al. “Predicting deliverability of volumetric-modulated arc therapy (VMAT) plans using aperture complexity analysis”. In: *Journal of Applied Clinical Medical Physics* 17 (July 2016), pp. 124–131. DOI: 10.1120/jacmp.v17i4.6241.

[72] T. Marsac et al. “An effective and optimized patient-specific QA workload reduction for VMAT plans after MLC-modelling optimization”. In: *Physica Medica* 107 (2023), p. 102548. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2023.102548.

[73] T Kairn et al. “Predicting the likelihood of QA failure using treatment plan accuracy metrics”. In: *Journal of Physics: Conference Series* 489.1 (Mar. 2014), p. 012051. DOI: 10.1088/1742-6596/489/1/012051.

[74] J. M. Park et al. “Reliability of the gamma index analysis as a verification method of volumetric modulated arc therapy plans”. In: *Radiation Oncology* 13 (Sept. 2018), p. 175. DOI: 10.1186/s13014-018-1123-x.

[75] J. 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: 10.1016/j.ijrobp.2019.07.049.

[76] N. Lambri et al. “PO-1617 Multicentric evaluation of a machine learning model to streamline the RT patient-specific QA process”. In: *Radiotherapy and Oncology* 182 (May 2023), S1311–S1312. DOI: 10.1016/S0167-8140(23)66532-6.

[77] A. Scaggion et al. “Limiting treatment plan complexity by applying a novel commercial tool”. In: *Journal of Applied Clinical Medical Physics* 21 (May 2020). DOI: 10.1002/acm2.12908.

[78] C. Hui, A. Pourmoghaddas, and Y. Mutaf. “The effects of flattening filter-free beams and aperture shape controller on the complexity of conventional large-field treatment plans”. In: *Journal of applied clinical medical physics* 24 (Aug. 2023), e14108. DOI: 10.1002/acm2.14108.

[79] V. Desai et al. “Multi-institution single geometry plan complexity characteristics based on IROC phantoms”. In: *Medical physics* (Apr. 2024). DOI: 10.1002/mp.17086.

[80] T. Santos, T. Ventura, and M. do Carmo Lopes. “Evaluation of the complexity of treatment plans from a national IMRT/VMAT audit – Towards a plan complexity score”. In: *Physica Medica* 70 (2020), pp. 75–84. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2020.01.015.

[81] C. McGarry et al. “PO-0991: The role of complexity metrics in a multi-institutional dosimetry audit of VMAT”. In: *Radiotherapy and Oncology* 115 (Apr. 2015), S530. DOI: 10.1016/S0167-8140(15)40983-1.

[82] V. 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: 10.1016/j.phro.2018.02.002.

[83] E. 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 (June 2019), pp. 73–82. DOI: 10.1016/j.ejmp.2019.05.003.

[84] E. J. Hall. “Intensity-modulated radiation therapy, protons, and the risk of second cancers”. In: *Int. J. Radiat. Oncol. Biol. Phys.* 65.1 (May 2006), pp. 1–7. DOI: 10.1016/j.ijrobp.2006.01.027.

[85] I. Ayuthaya et al. “Multi-planner validation of RapidPlan knowledge-based model for volumetric modulated arc therapy in prostate cancer”. In: *Journal of Applied Clinical Medical Physics* 25 (Nov. 2023). DOI: 10.1002/acm2.14223.

[86] S. Sinha et al. “Development and Validation of Single-Optimization Knowledge-Based Volumetric Modulated Arc Therapy Model Plan in Nasopharyngeal Carcinomas”. In: *Advances in Radiation Oncology* 9.1 (2024), p. 101311. ISSN: 2452-1094. DOI: 10.1016/j.adro.2023.101311.

[87] K. van Gysen et al. “Rolling out RapidPlan: What we’ve learnt”. In: *Journal of Medical Radiation Sciences* 67 (Sept. 2020). DOI: 10.1002/jmrs.420.

[88] L. M. H. Chua et al. “Dosimetric comparison of RapidPlan and manually optimised volumetric modulated arc therapy plans in prostate cancer”. In: *Journal of Radiotherapy in Practice* 20.3 (2021), 257–264. DOI: 10.1017/S1460396920000345.

[89] A. R. Delaney et al. “Effect of Dosimetric Outliers on the Performance of a Commercial Knowledge-Based Planning Solution”. In: *International Journal of Radiation Oncology*Biology*Physics* 94.3 (2016), pp. 469–477. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2015.11.011.

[90] M. Shepherd. “Developing knowledge-based planning for gynaecological cancer: A retrospective dosimetric validation study of Rapidplan”. In: (2017). DOI: 10.1594/RANZCR2017/R-0126.

[91] A. Fogliata et al. “RapidPlan head and neck model: The objectives and possible clinical benefit”. In: *Radiation Oncology* 12 (Apr. 2017). DOI: 10.1186/s13014-017-0808-x.

[92] A. 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 OncologyBiologyPhysics* 95 (July 2016), pp. 981–990. DOI: 10.1016/j.ijrobp.2016.02.017.

[93] M. Fusella et al. “Efficiently train and validate a RapidPlan model through APQM scoring”. In: *Medical Physics* 45.6 (2018), pp. 2611–2619. DOI: 10.1002/mp.12896.

[94] C. Mayo et al. “AAPM TG-263: Standardizing Nomenclatures in Radiation Oncology”. In: *International Journal of Radiation Oncology*Biology*Physics* 100 (Dec. 2017). DOI: 10.1016/j.ijrobp.2017.12.013.

[95] P. Gehring. “The Isoperimetric Inequality: Proofs by Convex and Differential Geometry”. In: *Rose-Hulman Undergraduate Mathematics Journal* 20.2 (2019). Available at: https://scholar.rose-hulman.edu/rhumj/vol20/iss2/4, Article 4.

[96] L. Masi et al. “Quality assurance of volumetric modulated arc therapy: Evaluation and comparison of different dosimetric systems”. In: *Medical Physics* 38.2 (2011), pp. 612–621. DOI: 10.1118/1.3533900.

[97] D. E. Knuth. “Two Notes on Notation”. In: *The American Mathematical Monthly* 99.5 (1992), pp. 403–422. DOI: doi.org/10.2307/2325085.

[98] M. H. Wong. *grpandplot: An open-source MATLAB tool for drawing box plot and violin plot with automatic multi-way data grouping*. Version 1.0.0. Nov. 2022. DOI: 10.5281/zenodo.7295877.

[99] E. Kamperis et al. “Complexity analysis of VMAT prostate plans: insights from dimensionality reduction and information theory techniques”. In: *Polish Journal of Medical Physics and Engineering* 29 (Sept. 2023). DOI: 10.2478/pjmpe-2023-0015.

[100] M. Ergin and O. Koskan. “Comparison of Student - t, Welch’s - t, and Mann - Whitney U Tests in Terms of Type I Error Rate and Test Power”. In: *Selcuk Journal of Agricultural and Food Sciences* 37 (Aug. 2023), pp. 223–231. DOI: 10.15316/SJAFS.2023.022.

[101] J. P. Tol et al. “Evaluation of a knowledge-based planning solution for head and neck cancer”. In: *Int. J. Radiat. Oncol. Biol. Phys.* 91.3 (Mar. 2015), pp. 612–620. DOI: 10.1016/j.ijrobp.2014.11.014.

[102] I. I. N. Ayuthaya, S. Suriyapee, and T. Sanghangthum. “Validation of RapidPlan knowledge-based model for volumetric-modulated arc therapy in prostate cancer”. In: *J. Med. Phys.* 47.3 (July 2022), pp. 250–255. DOI: 10.4103/jmp.jmp_138_21.

[103] A. E. W. Mark F Schilling and W. Watkins. “Is Human Height Bimodal?” In: *The American Statistician* 56.3 (2002), pp. 223–229. DOI: 10.1198/00031300265.