Radiotherapy and Oncology 186 (2023) 109775

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Universal evaluation of MLC models in treatment planning systems based on a common set of dynamic tests

Jordi Saez [a,*], Raquel Bar-Deroma [b], Evelien Bogaert [c], Romain Cayez [d], Tom Chow [e], Catharine H. Clark [f,g,h], Marco Esposito [i,ae], Vladimir Feygelman [j], Angelo F. Monti [k], Julia Garcia-Miguel [l], Eduard Gershkevitsh [m], Jo Goossens [n], Carmen Herrero [o], Mohammad Hussein [f], Catherine Khamphan [p], Roel G.J. Kierkels [q], Wolfgang Lechner [r], Matthieu Lemire [s], Alexander Nevelsky [b], Daniel Nguyen [t], Lucia Paganini [u], Marlies Pasler [v,w], José Fernando Pérez Azorín [x,y], Luis Isaac Ramos Garcia [z], Serenella Russo [i], John Shakeshaft [aa], Laure Vieillevigne [ab], Victor Hernandez [ac,ad]

[a] *Hospital Clínic de Barcelona, Department of Radiation Oncology, Barcelona, Spain;* [b] *Rambam Health Care Campus, Department of Radiotherapy, Division of Oncology, Haifa, Israel;* [c] *Ghent University Hospital and Ghent University, Department of Radiation Oncology, Ghent, Belgium;* [d] *Oscar Lambret Center, Department of Medical Physics, Lille, France;* [e] *Juravinski Hospital and Cancer Centre at Hamilton Health Sciences, Department of Medical Physics, Ontario, Canada;* [f] *National Physical Laboratory, Metrology for Medical Physics Centre, London TW11 0PX, UK;* [g] *Radiotherapy Physics, University College London Hospital, 250 Euston Rd, London NW1 2PG, UK;* [h] *Dept Medical Physics and Bioengineering, University College London, Malet Place, London WC1 6BT, UK;* [i] *AUSL Toscana Centro, Medical Physics Unit, Florence, Italy;* [j] *Moffitt Cancer Center, Department of Radiation Oncology, Tampa, FL, USA;* [k] *ASST GOM Niguarda, Department of Medical Physics, Milano, Italy;* [l] *Consorci Sanitari de Terrassa, Department of Radiation Oncology, Terrassa, Spain;* [m] *North Estonia Medical Centre, Department of Medical Physics, Tallinn, Estonia;* [n] *Iridium Netwerk, Department of Medical Physics, Antwerp, Belgium;* [o] *Centro Médico de Asturias-IMOMA, Department of Medical Physics, Oviedo, Spain;* [p] *Institut du Cancer – Avignon Provence, Department of Medical Physics, Avignon, France;* [q] *Radiotherapiegroep, Department of Medical Physics, Arnhem/Deventer, the Netherlands;* [r] *Medical University of Vienna, Department of Radiation Oncology, Vienna, Austria;* [s] *CIUSSS de l'Est-de-l'Île-de-Montréal, Service de Radio-Physique, Montréal, Canada;* [t] *Centre de Radiothérapie de Mâcon, France;* [u] *Humanitas Clinical and Research Center, Radiotherapy and Radiosurgery Department, Rozzano, Italy;* [v] *Lake Constance Radiation Oncology Center, Department of Radiation Oncology, Singen, Friedrichshafen, Germany;* [w] *Radiotherapy Hirslanden, St. Gallen, Switzerland;* [x] *Medical Physics and Radiation Protection Department, Gurutzeta-Cruces University Hospital, Barakaldo, Spain;* [y] *Biocruces Health Research Institute, Barakaldo, Spain;* [z] *Servei de Radiofísica Hospitalària, Hospital Arnau de Vilanova, Lleida, Spain;* [aa] *Gold Coast University Hospital, ICON Cancer Centre, Gold Coast, Australia;* [ab] *Institut Claudius Regaud-Institut Universitaire du Cancer de Toulouse, Department of Medical Physics, Toulouse, France;* [ac] *Hospital Sant Joan de Reus, Department of Medical Physics, Reus, Spain;* [ad] *Universitat Rovira i Virgili, Tarragona, Spain;* [ae] *The Abdus Salam International Center for Theoretical, Trieste, Italy*

ARTICLE INFO

Article history:
Received 29 January 2023
Received in revised form 19 June 2023
Accepted 23 June 2023
Available online 28 June 2023

Keywords:
MLC
Comissioning
Audits
IMRT
VMAT
Standardization

ABSTRACT

Purpose: To demonstrate the feasibility of characterising MLCs and MLC models implemented in TPSs using a common set of dynamic beams.

Materials and methods: A set of tests containing synchronous (SG) and asynchronous sweeping gaps (aSG) was distributed among twenty-five participating centres. Doses were measured with a Farmer-type ion chamber and computed in TPSs, which provided a dosimetric characterisation of the leaf tip, tongue-and-groove, and MLC transmission of each MLC, as well as an assessment of the MLC model in each TPS. Five MLC types and four TPSs were evaluated, covering the most frequent combinations used in radiotherapy departments.

Results: Measured differences within each MLC type were minimal, while large differences were found between MLC models implemented in clinical TPSs. This resulted in some concerning discrepancies, especially for the HD120 and Agility MLCs, for which differences between measured and calculated doses for some MLC-TPS combinations exceeded 10%. These large differences were particularly evident for small gap sizes (5 and 10 mm), as well as for larger gaps in the presence of tongue-and-groove effects. A much better agreement was found for the Millennium120 and Halcyon MLCs, differences being within ± 5% and ± 2.5%, respectively.

Conclusions: The feasibility of using a common set of tests to assess MLC models in TPSs was demonstrated. Measurements within MLC types were very similar, but TPS dose calculations showed large variations. Standardisation of the MLC configuration in TPSs is necessary. The proposed

* Corresponding author at: Department of Radiation Oncology, Hospital Clínic de Barcelona, Villarroel 170, Barcelona 08036, Spain.
E-mail address: jordi.saez@gmail.com (J. Saez).

https://doi.org/10.1016/j.radonc.2023.109775

procedure can be readily applied in radiotherapy departments and can be a valuable tool in IMRT and credentialing audits.

The adoption of intensity-modulated radiation therapy (IMRT) techniques is widespread. Nevertheless, the dosimetric commissioning of IMRT treatment planning systems (TPS) poses substantial challenges. Current international recommendations and guidelines for IMRT commissioning include progressively complex tests that aim to identify beam modelling limitations in TPSs [1,2,3]. External end-to-end (E2E) audits with anthropomorphic phantoms are ultimately recommended to guarantee a good agreement between delivered and planned doses for clinical plans [4,5,6,7,8].

Several groups offer IMRT E2E audits as an external validation for quality assurance purposes and/or credentialing before participation in clinical trials [4,5,6,7,8]. In particular, results from the IROC-Houston auditing program with their head-and-neck anthropomorphic phantom show that, despite modest improvements over the last few years, 8–15% of participating institutions still fail the credentialing requirements, namely 7% for point dose differences and 7%-4 mm criteria for gamma index comparisons [9,10,11]. Significant efforts have been dedicated to unraveling the causes behind such failures. Carson [10] reported systematic differences between TPS calculations and measured doses. Kerns [12] recalculated 259 head and neck phantom irradiations using reference beam data and found considerable TPS errors in 68% of those irradiations not meeting the IROC-Houston criteria and in 17% of all the cases analysed. Remarkably, the reference recalculation system had, on average, better accuracy than the institutions' TPSs. A recent large-scale survey [13] revealed a high variability in the multileaf collimator (MLC) parameters used clinically, even amongst institutions with the same TPS, linear accelerator and MLC model. Furthermore, the dosimetric impact of such variations substantially affects calculated doses for IMRT and VMAT plans, with atypical MLC parameters associated with audit failures [11].

However, even as the mounting evidence underscores the influence of the MLC modelling in TPSs [10,11,12,13], to the best of the authors knowledge, no evaluation of the dosimetric MLC characteristics and their modelling in TPSs has been carried out in a systematic manner and no procedure has been demonstrated to be applicable to different MLCs and TPSs and to be sensitive to the major dosimetric properties of the MLC. Our work aims to fill this gap by providing a procedure and a set of tests that can be used experimentally across a range MLC/TPS combinations. Each TPS manufacturer recommends different tools and methods for adjusting the specific MLC parameters. As a result, medical physicists face a variety of methods of beam configuration, which can lead to different parameters being used in clinical practice.

Sweeping gap fields are commonly used to determine the dosimetric leaf gap (DLG) parameter in the Eclipse TPS [14,15]. Asynchronous sweeping gap tests (in which adjacent leaf tips are not aligned) have been proposed by Hernandez et al. [16] that are sensitive not only to the leaf positioning offset, but also to other MLC parameters such as the tongue-and-groove width and the leaf tip width. Saez et al. [17] recently provided a method based on those tests measured with a Farmer chamber to determine the MLC configuration parameters in RayStation. They showed that these parameters also improved the accuracy in clinical plans. Synchronous and asynchronous sweeping gap tests can therefore be used for configuring and fine-tuning the MLC models in TPSs [2,3] and could be helpful to standardise its configuration and evaluation, including in inter-departmental comparison or auditing of MLC/TPS model performance and in credentialing for trial participation or helping to resolve differences seen in E2E testing in these scenarios.

The purpose of the present study is to show the feasibility of characterising MLCs and MLC models implemented in different TPSs using a common set of dynamic synchronous and asynchronous sweeping gaps. Tests were prepared for the most widely used treatment units, MLC models, and TPSs and a large multicentric intercomparison and evaluation was conducted including the most common combinations of TPSs, machines and MLC models used in the radiotherapy community.

Materials and methods

Preparation and description of the tests

A DICOM work package containing images of a water phantom, a structure set containing evaluation and external contours, and a set of RT-plans with tests for a 6 MV beam (except for the Halcyon, which uses a 6FFF beam) was sent to twenty-five participating institutions.[1] The tests aimed to evaluate the MLC model in the TPS and contained the following beams:

a) Reference: Beam defining a 10 cm × 10 cm field.
b) Transmission: Reference beams with the MLC blocking the whole 10x10 aperture. In tertiary MLCs, closed leaf pairs were positioned outside the jaw-defining field [18]. In the case of the double-layer MLC, the same approach was followed for each layer, with the other layer used to define the 10 cm x10 cm field size [19]. In secondary MLCs that replace jaws, opposing leaves were positioned defining the minimum gap size as far as possible from the central axis [20].
c) Sweeping gaps (SG): Four different sliding window fields within which a given gap between opposing leaves (with size of 5, 10, 20 and 30 mm) moves at constant speed [14,17]. The centre of the gaps travelled from left to right over 12 cm. For linear accelerators with back-up or primary jaws, these were set symmetrically to define the reference field.
d) Asynchronous sweeping gaps (aSG): These tests were proposed for evaluating the tongue-and-groove dynamically [16]. The SG of 20 mm was modified by shifting adjacent leaves a certain distance '*s*' in each test, namely 0, 2, 5, 8, 10, 12, 15 and 20 mm, to expose a controlled amount of leaf sides. By design, the gap between opposing leaf pairs remains constant and the attenuation caused by the tongue-and-groove increases with the distance *s*.

The DICOM files were prepared specifically for each TPS and MLC combination, and some adaptations were required, depending on the TPS, to address some technical specifics, e.g., to adjust the nominal leaf positioning boundaries of the first and last pairs of leaves to the values expected by the system. For Pinnacle, beam sequences were introduced into the TPS using its scripting language.

[1] DICOM files will be provided upon request to the corresponding author.

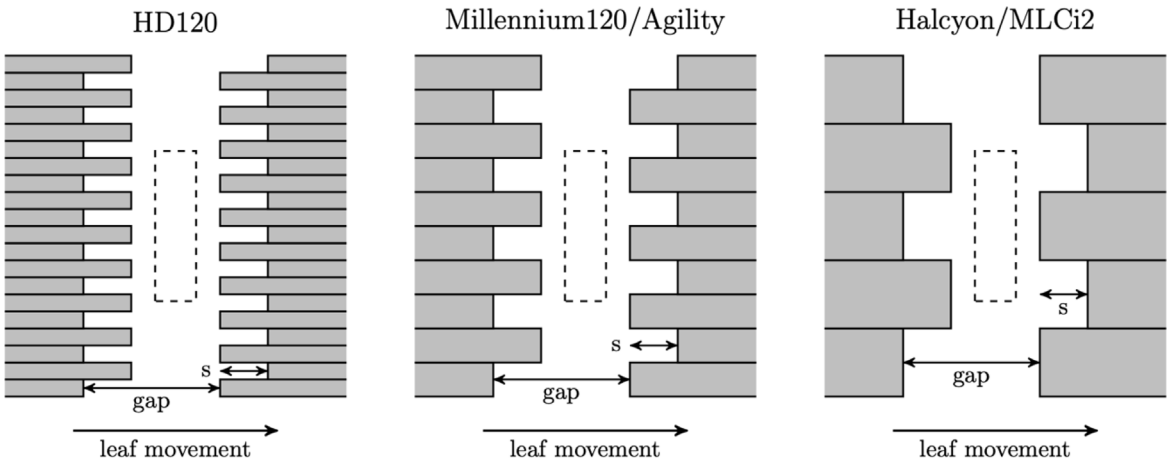

Fig. 1. Sketch of the asynchronous sweeping gap tests for different MLC models. The dashed rectangle represents the relative size of a Farmer-type ion chamber (22 mm length for PTW model 30013) with respect to the nominal leaf widths at the position of the ion chamber (2.5 mm for HD120, 5 mm for Millennium120 and Agility, 10 mm for Halcyon and MLCi2).

Data collection

All centres were instructed to carry out the measurements with a Farmer-type ion chamber in order to provide an appropriate sampling of several leaves as depicted in Fig. 1. The ion chamber was placed with its longitudinal axis perpendicular to the leaf motion in an isocentric set-up in water or solid water: depth 10 cm and source-to-surface distance of 90 cm. To better average the leaf effects, two different longitudinal positions were used separated by an odd multiple of half the leaf width, and the average of the two readings was taken. The longitudinal positions used for each MLC can be found in Table S1 in Supplementary Material.

TPS calculated doses matched the mean dose to the evaluation structure provided within the DICOM structure set, which mimicked the dimensions and position of the Farmer-type chamber active volume. Dose calculation algorithm and resolution reflected clinical practice at each centre. For Monte Carlo dose calculation algorithms, the evaluation structure was extended by 2 cm in the lateral direction in order to reduce the statistical uncertainty in the reported average doses.

Data analysis

The measured and calculated average doses obtained with the tests were analysed and compared. Measured doses were used to determine the dosimetric MLC characteristics, while calculated doses characterised the MLC models configured in TPSs. The comparison of measured and calculated average doses offered information about the accuracy of the MLC models configured in TPSs under different situations and for different TPSs, MLC types, and MLC configuration parameters.

First, the average transmission was computed from the measured doses. Second, all measured and calculated doses from the SG and aSG fields were normalised to their corresponding reference dose, effectively removing variations caused by dose calibration differences or daily beam output. Discrepancies between the renormalised TPS and measured doses were expressed as percentage differences. The SGs produce a homogenous dose at the position of the ion chamber, which depends on the fraction of MUs in which the chamber is exposed to the SG aperture, as well as on the transmission through the rounded leaf ends and through the full leaf thickness. The additional transmission through the rounded leaf ends can also be modelled by considering focused leaves with flat ends and by shifting each leaf's position by a quantity known as radiation field offset, which also absorbs the calibration offset introduced by the MLC controller [14,15]. The effect of the rounded leaf ends per leaf pair is known as DLG, which corresponds to twice the radiation field offset. To analyse the SG doses, offsets were computed from both measurement and TPS doses.

Next, the normalised doses for the aSG fields were used to assess the attenuation produced by the tongue-and-groove effect, which is caused by the increased transmission through the tongue-and-groove regions when the leaf sides are exposed because of the distance between adjacent leaves. Finally, the dose reduction observed for the normalised aSG doses were expressed as the dose percentage reduction with respect to the mean value for s = 0 mm and each MLC type. These dose percentages were used to compare and characterise the intrinsic tongue-and-groove effects of each MLC.

Results

The beams and dynamic MLC sequences were successfully imported into all TPSs. A summary of all the combinations of TPSs and MLC types evaluated is given in Table 1 (a detailed description of all TPSs, versions, calculation grid resolution, MLC parameters, and ion chamber used at each centre is provided as Table S2 and S3 in Supplementary Material). A few centres reported doses from different TPSs for the same MLC and/or doses for beam models configured differently for static-gantry IMRT or VMAT.

The measured transmission showed negligible variability within each MLC type: (1.23 ± 0.03)% for the HD120, (1.45 ± 0.03)% for the Millennium120, (0.44 ± 0.03)% for the Agility, (0.34 ± 0.01)% for the Halcyon distal layer, and (0.33 ± 0.01)% for the Halcyon proximal layer. These values indicate the *average* transmission, i.e., combining both intra- and inter-leaf transmissions, because of the Farmer-type ion chamber used.

The dose discrepancies between the experimental and TPS doses for the SGs are shown in Fig. 2a. As expected, discrepancies were larger for smaller gaps, but results depended on the MLC type. The largest discrepancies were found for the HD120 and Agility, with dose differences larger than ± 5% for the 10 mm gap and around 10% for the 5 mm gap. A much better agreement was obtained for the Millennium120 and the Halcyon's dual layer MLC, with most differences smaller than 5% for the 5 mm gap. A more detailed analysis showed that the discrepancies strongly depended on each individual TPS; for the HD120, for instance, two of the TPSs exhibited an agreement better than 1% for all SG sizes, even for the 5 mm gap.

Table 1
Combinations of TPSs and MLCs evaluated. The number of measured MLCs of each MLC type is indicated in brackets.

		TPS			
		Eclipse	Monaco	Pinnacle	RayStation
MLC type	Varian HD120 (6)	4	-	3	1
	Varian Millennium120 (6)	4	1	2	3
	Elekta Agility (10)	1	4	2	4
	Varian Halcyon (6)	6	-	-	3
	Elekta MLCi/MLCi2 (2)	-	1	1	1

The radiation field offset values shown in Fig. 2b facilitate the interpretation of the dose discrepancies in Fig. 2a. For the HD120, Agility and Halcyon, measured offsets had a very small variability, with all values within 0.1 mm from the mean. Even for the Millennium120, which had the largest spread in measured offsets, all values were within 0.2 mm of the mean (0.75 mm). Data from the TPSs, on the contrary, showed much less consistency. The offset values from TPS calculations for the Agility and HD120 MLCs had a spread of about 0.8 mm. For the Millennium120 and the Halcyon MLCs, the spread in the TPS offset was smaller, being ± 0.2 mm, and were within good agreement with measured offsets.

The measured and calculated doses for the aSG fields are given in Fig. 3a. All measured doses decreased when the distance between adjacent leaves '*s*' was increased. Individual measured doses for the same MLC type were almost coincident (as demonstrated by the thin shaded areas in Fig. 3a), which indicates that all the MLCs of the same type have very similar properties.

The TPS doses, on the contrary, showed an enormous variability for all cases except for the Halcyon MLCs, as seen with the varied dotted lines in Fig. 3a. The largest variations in TPS doses were also found for the HD120 and Agility MLCs, with differences rapidly rising as the distance between adjacent leaves increased.

The agreement between TPS and measured doses, shown in Fig. 3b, highlights the differences seen in Fig. 3a. Reasonable agreement was found for s = 0 mm, with all discrepancies < 5%, except for one case with the MLCi2, and most of them < 3%. However, discrepancies rapidly increased as the amount of leaf side exposed increased. The largest discrepancies were found for the HD120 and Agility, with differences exceeding 10% for many of the TPSs evaluated (3 out of 8 TPSs for the HD120 and for 4 out of 11 TPSs for the Agility). Dose discrepancies were smaller for the Millennium120, with all values roughly within ± 10% (and most of them within ± 5%) and for the Halcyon, with all differences within ± 5%.

To better explore the measured aSG doses, and how they varied between MLC types, the aSG measured doses shown in Fig. 3 were expressed as the percentage difference with respect to the average value at s = 0 mm for each MLC type. These curves (Fig. 4) reinforce that all measurements for the same MLC type were nearly indistinguishable. For Varian MLCs (HD120, Millennium120 and Halcyon), all the variations with respect to each MLC type average were within ± 0.5%. For Elekta MLCs (Agility and MLCi2) larger variations were found, but they were still within ± 1%. It is also clear that different MLC types have different characteristics. Therefore, each MLC type constitutes an "MLC class" with fixed and well-established dosimetric characteristics.

Discussion

The feasibility of a direct measurement-based assessment of the MLC dosimetric characteristics using a set of standardised tests, implemented into a wide range of different treatment units, was demonstrated. The presented methodology allows both the experimental characterisation of MLCs and the assessment of MLC models implemented in TPSs. The average dose produced in these tests is very sensitive to MLC characteristics such as the leaf tip, the tongue-and-groove, and the MLC transmission. Furthermore, these tests and methodology [17,16] allow isolation of the different MLC

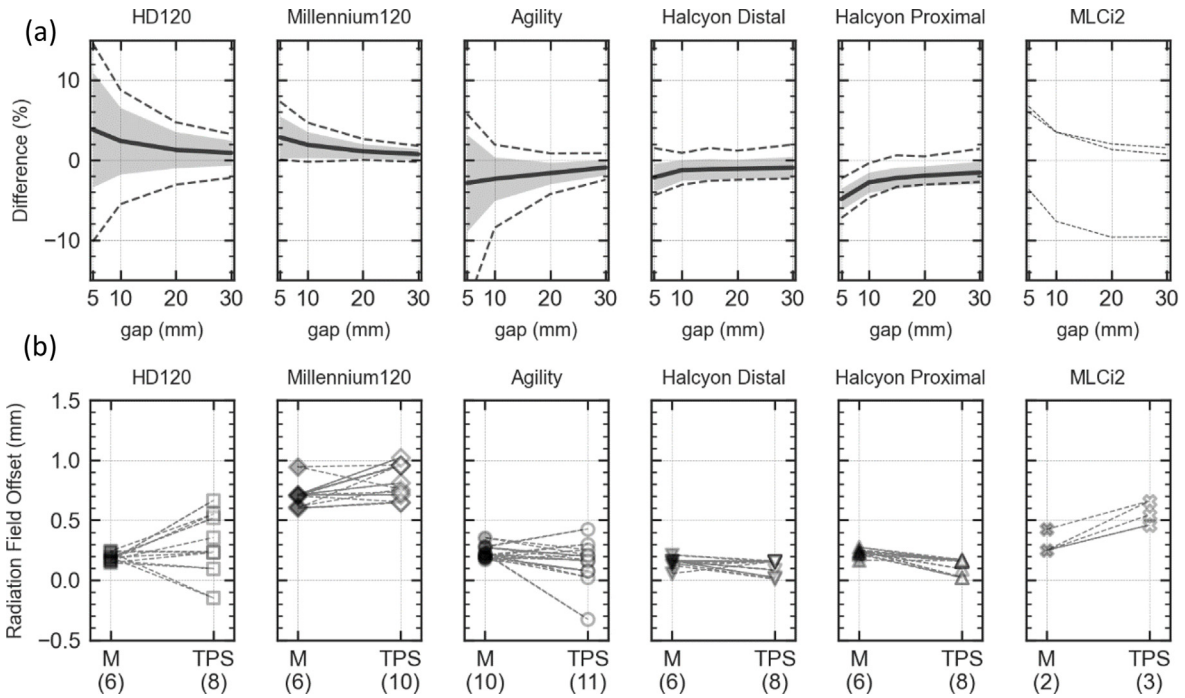

Fig. 2. (a) Dose differences between measured and TPS doses for each MLC type. The thick line represents the mean dose difference obtained for each gap, the grey region shows the range within one standard deviation from the mean, and the dashed lines represent the maximum differences. (b) Measured (M) and calculated (TPS) radiation field offsets for each MLC-TPS pair. The numbers under the x-axis labels indicate the number of MLCs and TPSs evaluated. MLCi2 data are shown for completeness but no statistical analysis was included due to the low number of systems evaluated (2).

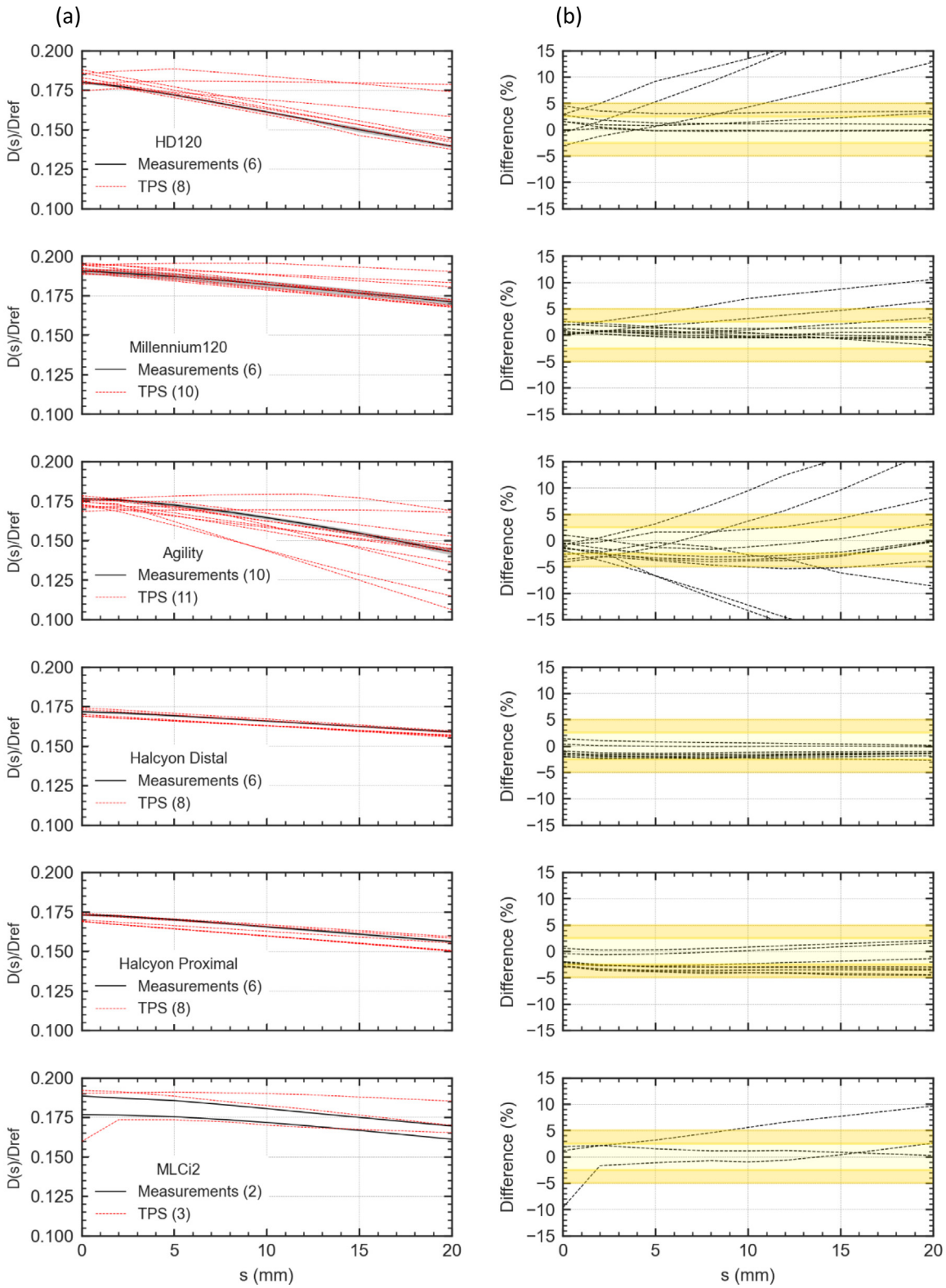

Fig. 3. (a) Doses for the aSG tests (20 mm gap size) as a function of the leaf side exposed s (see Fig. 1). Doses were normalised to the reference field dose $D(s)/D_{ref}$. The mean measured data is drawn as a thick black line with a light grey shaded area indicating one standard deviation (except for the MLCi2, where individual measured data are represented). Individual TPS doses are presented as dotted red lines. (b) Individual dose differences between each TPS-measurement pair. Light and dark shaded areas denote regions with dose differences of ± 2.5% and ± 5%, respectively.

characteristics, thus facilitating the assessment of MLC models in TPSs and the interpretation of the results.

Assessing the fine details of MLCs with a large Farmer-type ion chamber might seem counter-intuitive but has clear advantages. First, a large ion chamber is a robust detector with an uncompromised response for MV photon beams and is readily available in most departments. Second, a Farmer chamber integrates the dosimetric effects of the MLC across several leaves and, due to the sym-

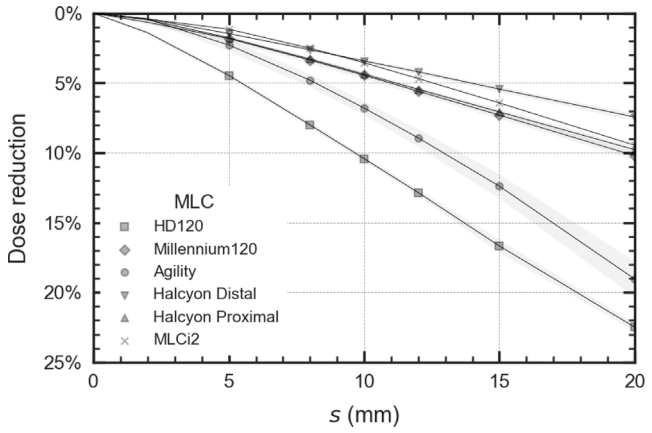

Fig. 4. Dose percentage reduction for the 20 mm sweeping gap with respect to the mean normalised dose for each MLC type. Measured doses for the aSG tests normalised to the dose of the 20 mm sweeping gap. For each MLC type, the line indicates the mean value and the grey shaded area one standard deviation (standard deviation not represented for MLCi2).

metry of the tests used, facilitates a good assessment of the average doses, which is the most clinically relevant quantity [21]. Average doses measured with a large ion chamber in dynamic beams have been used previously to fine-tune the MLC for Monte Carlo codes [18,22,23]. We followed a similar approach and incorporated dynamic measurements of the tongue-and-groove attenuation as a function of the distance between adjacent leaves and applied it systematically to different MLCs and TPSs. Although this study focused on 6 MV photon beams, the methodology is applicable for any energy.

Our results provide compelling evidence that all MLCs with the same design have almost identical properties and that very similar results are obtained provided that the same procedures are followed. The small variability found in the present study is compatible with data reported by several authors for Varian TrueBeams [24,25,26]. In particular, we attribute the low differences in the measured MLC transmission within each MLC type to the utilization of a common measurement set-up and to the robustness achieved by averaging readings from two longitudinal positions of the ion chamber. We found that the only significant differences between individual MLCs within a given class are due to slight differences in the MLC positioning calibration, which result in slightly different values of the measured offset. This high consistency also explains why an average MLC model can produce acceptable results in many cases [12]. However, adjusting the TPS parameter related to the MLC offset is important since it might produce large differences in calculated doses depending on the technique used [27,28,29].

On the other hand, our assessment of the MLC models in TPSs shows a concerning picture. We found a large variability between TPSs and some large discrepancies between measured and TPS doses. That indicates the use of suboptimal MLC parameters in the TPS and is in line with the large variability in the MLC parameters reported by Glenn et al [13] and with the challenges introduced by particular MLC designs [30,31]. The highest observed dose differences were found for the HD120 and Agility MLCs in the presence of tongue-and-groove effects. The impact in clinical plans is likely to be less as the effects will tend to be averaged across different gap sizes and *s* values, but several authors have reported tongue-and-groove effects to produce underdosages of up to 5%-7% in target volumes of clinical plans when not considered properly [32,33,34]. Variability and dose discrepancies were small only for the Halcyon system, probably because the only two TPSs available for this system provide either pre-commissioned and non-configurable beam and MLC models [19,35,36,37] or clear and robust recommendations and guidelines for beam and MLC configuration [19,38,39].

In our opinion, our results suggest that the observed TPS variability is due to the lack of standardised procedures for MLC configuration and commissioning [1], the insufficient sensitivity of the methodologies followed [40,41,42], the difficulty for fine-tuning multiple interrelated parameters [31,43], and potential compromises that might be required [19,30,31]. In case that important discrepancies between calculated and measured doses with these tests are found, a careful analysis of the MLC parameters used in the TPS is recommended. A comparison with MLC parameters from other institutions or with average results used by the community [13] may be useful to guide potential changes in the MLC parameters used. Additionally, it has also been shown that the tests used in the present study can facilitate an accurate and straightforward testing and determination of the MLC parameters in TPSs since the tests effectively identify and differentiate the various dosimetric effects of MLCs, such as transmission, rounded leaf end, and tongue-and-groove effects which could be helpful to standardise the process [17].

Given the small variability in measured data, a set of MLC parameters for each TPS-MLC combination could be derived from the average measurements for each MLC class by tuning the specific MLC model to the average measured doses, and these parameters could be directly used by the community to configure the MLC model in the TPS. Our results clearly show that only a slight fine-tuning of the MLC parameter related to the offset would be necessary to account for potential differences in the MLC positioning calibration. However, currently it is not straightforward to provide clear TPS-specific instructions to users on how to optimize the configuration parameters on their TPS to improve their MLC models. That demands specific analyses depending on the treatment unit and TPS and thus further work is required to derive sets of MLC parameters for each TPS and MLC models. In any case, a final verification should be conducted using plan-specific QA of a range of clinically representative plans or end-to-end tests.

The lack of a universal methodology for evaluating MLC models has prevented auditing groups, including trial credentialing groups, from directly assessing MLC models in TPSs [7,44,45,46]. This can be overcome using the presented procedure, which could be used as a troubleshooting tool both in on-site and remote audits to obtain complementary information on the MLC models used in clinical practice. Moreover, the proposed procedure is based on a set of dynamic beams and simple measurements taken with a Farmer chamber, which facilitate its implementation worldwide, including low- and middle-income countries.

One limitation of our work is that not every combination of TPSs and MLCs was evaluated; however, we covered the most widely used systems and showed that the presented methodology can be used as long as TPSs allow DICOM import or alternative procedures. Another limitation is that the accuracy of the different beam models in clinical plans was not investigated, although that was not the purpose of the present study as it has already been shown that MLC parameters obtained using these tests improve the calculation accuracy in clinical plans [17,47]. The impact of the differences found on clinical plans will depend on the specific characteristics of the plans generated by the TPS and used in clinical practice and must be carefully evaluated at each centre. For instance, the large dose differences observed for the SG and aSG tests for some TPSs will not necessarily produce unacceptable dose differences in clinical plans depending on the complexity of the treatment plans generated. However, MLC parameters should still be optimised as their dose impact can differ across different plans [48,49]. A better set of MLC parameters has the potential to improve the accuracy in some treatment plans, thus extending

the validity of TPS calculations to a wider range of plan characteristics.

In conclusion, we have presented a simple and robust methodology to assess MLC models in TPSs that can be universally applied and swiftly followed by medical physicists at local centres, in multicentric comparisons and specifically to aid trial credentialing audits in separating effects when differences are observed in E2E tests. We obtained very small variations in the measured characteristics for a given MLC type but concerning differences were found between measured average doses and TPS calculations. The proposed methodology can be helpful in tackling this problem and facilitating the standardisation of the MLC configuration, as well as the commissioning and assessment of MLC models in TPSs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

JS and VH would like to thank Stephen Kry from IROC, Agnes Angerud from RaySearch Labs, and Enrico Clementel from EORTC-Headquarters for their discussions and critical review of the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2023.109775.

References

[1] Smilowitz J.B, Das I.J, Feygelman V, Fraass B.A, Kry S.F, Marshall I.R, Mihailidis D.N, Ouhib Z, Ritter T, Snyder M.G, Fairobent L, AAPM Medical Physics Practice Guideline 5.a.: Commissioning and QA of Treatment Planning Dose Calculations - Megavoltage Photon and Electron Beams," *J Appl Clin Med Phys* 09 2015; 16:14–34.

[2] Geurts M.W, Jacqmin D.J, Jones L.E, Kry S.F, Mihailidis D.N, Ohrt J.D, Ritter T, Smilowitz J.B, Wingreen N.E. "AAPM MEDICAL PHYSICS PRACTICE GUIDELINE 5.b: Commissioning and QA of treatment planning dose calculations-Megavoltage photon and electron beams," *J Appl Clin Med Phys* Aug 2022; p. e13641.

[3] Schuring D, Westendorp H, van der Bijl E, Bol G, Crijns W, Delor A, Jourani Y, Loon Ong C, Penninkhof J, Kierkels R, Verbakel W, van de Water T, van de Kamer J. *NCS35: Quality assurance of Treatment Planning Systems,* Delft, The Netherlands: Netherlands Commission on Radiation Dosimetry, 2022.

[4] Molineu A, Followill DS, Balter PA, Hanson WF, Gillin MT, Huq MS, et al. Design and implementation of an anthropomorphic quality assurance phantom for intensity-modulated radiation therapy for the Radiation Therapy Oncology Group. Int J Radiat Oncol Biol Phys Oct 2005;63:577–83.

[5] Wesolowska P, Georg D, Lechner W, Kazantsev P, Bokulic T, Tedgren AC, et al. Testing the methodology for a dosimetric end-to-end audit of IMRT/VMAT: results of IAEA multicentre and national studies. Acta Oncol Dec 2019;58:1731–9.

[6] Clark CH, Aird EG, Bolton S, Miles EA, Nisbet A, Snaith JA, et al. Radiotherapy dosimetry audit: three decades of improving standards and accuracy in UK clinical practice and trials. Br J Radiol Nov 2015;88:20150251.

[7] Kazantsev P, Lechner W, Gershkevitsh E, Clark CH, Venencia D, Van Dyk J, et al. IAEA methodology for on-site end-to-end IMRT/VMAT audits: an international pilot study. Acta Oncol Feb 2020;59:141–8.

[8] Nakamura M, Minemura T, Ishikura S, Nishio T, Narita Y, Nishimura Y. An on-site audit system for dosimetry credentialing of intensity-modulated radiotherapy in Japanese Clinical Oncology Group (JCOG) clinical trials. Phys Med Aug 2016;32:987–91.

[9] Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. Med Phys Feb 2013;40:022101.

[10] Carson ME, Molineu A, Taylor PA, Followill DS, Stingo FC, Kry SF. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. Med Phys Dec 2016;43:6491.

[11] Glenn MC, Brooks F, Peterson CB, Howell RM, Followill DS, Pollard-Larkin JM, et al. Photon beam modeling variations predict errors in IMRT dosimetry audits. Radiother Oncol Jan 2022;166:8–14.

[12] Kerns JR, Stingo F, Followill DS, Howell RM, Melancon A, Kry SF. "Treatment planning system calculation errors are present in most imaging and radiation oncology core-Houston phantom failures". Int J Radiat Oncol Biol Phys 08 2017;98:1197–203.

[13] Glenn MC, Peterson CB, Followill DS, Howell RM, Pollard-Larkin JM, Kry SF. Reference dataset of users' photon beam modeling parameters for the Eclipse, Pinnacle, and RayStation treatment planning systems. Med Phys Jan 2020;47:282–8.

[14] LoSasso T, Chui CS, Ling CC. Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy. Med Phys Oct 1998;25:1919–27.

[15] Vial P, Oliver L, Greer PB, Baldock C. An experimental investigation into the radiation field offset of a dynamic multileaf collimator. Phys Med Biol Nov 2006;51:5517–38.

[16] Hernandez V, Vera-Sánchez JA, Vieillevigne L, Saez J. Commissioning of the tongue-and-groove modelling in treatment planning systems: from static fields to VMAT treatments. Phys Med Biol Aug 2017;62:6688–707.

[17] Saez J, Hernandez V, Goossens J, De Kerf G, Verellen D. "A novel procedure for determining the optimal MLC configuration parameters in treatment planning systems based on measurements with a Farmer chamber". Phys Med Biol 07 2020;65:155006.

[18] Arnfield MR, Siebers JV, Kim JO, Wu Q, Keall PJ, Mohan R. A method for determining multileaf collimator transmission and scatter for dynamic intensity modulated radiotherapy. Med Phys Oct 2000;27:2231–41.

[19] Hernandez V, Saez J, Angerud A, Cayez R, Khamphan C, Nguyen D, et al. Dosimetric leaf gap and leaf trailing effect in a double-stacked multileaf collimator. Med Phys Jul 2021;48:3413–24.

[20] Thompson CM, Weston SJ, Cosgrove VC, Thwaites DI. A dosimetric characterization of a novel linear accelerator collimator. Med Phys Feb 2014;41:031713.

[21] Hodapp N. The ICRU Report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). Strahlenther Onkol Jan 2012;188:97–9.

[22] Kim JO, Siebers JV, Keall PJ, Arnfield MR, Mohan R. A Monte Carlo study of radiation transport through multileaf collimators. Med Phys Dec 2001;28:2497–506.

[23] Siebers JV, Keall PJ, Kim JO, Mohan R. A method for photon beam Monte Carlo multileaf collimator particle transport. Phys Med Biol Sep 2002;47:3225–49.

[24] Glide-Hurst C, Bellon M, Foster R, Altunbas C, Speiser M, Altman M, et al. Commissioning of the Varian TrueBeam linear accelerator: a multi-institutional study. Med Phys Mar 2013;40:031719.

[25] Isono M, Akino Y, Mizuno H, Tanaka Y, Masai N, Yamamoto T. Inter-unit variability of multi-leaf collimator parameters for IMRT and VMAT treatment planning: a multi-institutional survey. J Radiat Res Mar 2020;61:307–13.

[26] Ghazal M, Södergren L, Westermark M, Söderström J, Pommer T. Dosimetric and mechanical equivalency of Varian TrueBeam linear accelerators. J Appl Clin Med Phys Dec 2020;21:43–53.

[27] Zygmanski P, Kung JH. Method of identifying dynamic multileaf collimator irradiation that is highly sensitive to a systematic MLC calibration error. Med Phys Nov 2001;28:2220–6.

[28] Rangel A, Dunscombe P. Tolerances on MLC leaf position accuracy for IMRT delivery with a dynamic MLC. Med Phys Jul 2009;36:3304–9.

[29] McVicker D, Yin F-F, Adamson JD. On the sensitivity of TG-119 and IROC credentialing to TPS commissioning errors. J Appl Clin Med Phys Jan 2016;17:34–48.

[30] Hernandez V, Angerud A, Bogaert E, Hussein M, Lemire M, García-Miguel J, et al. Challenges in modeling the Agility multileaf collimator in treatment planning systems and current needs for improvement. Med Phys Dec 2022;49:7404–16.

[31] Vieillevigne L, Khamphan C, Saez J, Hernandez V. On the need for tuning the dosimetric leaf gap for stereotactic treatment plans in the Eclipse treatment planning system. J Appl Clin Med Phys Jul 2019;20:68–77.

[32] Lorenz F, Nalichowski A, Rosca F, Killoran J, Wenz F, Zygmanski P. An independent dose calculation algorithm for MLC-based radiotherapy including the spatial dependence of MLC transmission. Phys Med Biol Feb 2008;53:557–73.

[33] Yang J, Tang G, Zhang P, Hunt M, Lim SB, LoSasso T, et al. Dose calculation for hypofractionated volumetric-modulated arc therapy: approximating continuous arc delivery and tongue-and-groove modeling. J Appl Clin Med Phys Mar 2016;17:3–13.

[34] Cadman P, Bassalow R, Sidhu NPS, Ibbott G, Nelson A. Dosimetric considerations for validation of a sequential IMRT process with a commercial treatment planning system. Phys Med Biol Jul 2002;47:3001–10.

[35] De Roover R, Crijns W, Poels K, Michiels S, Nulens A, Vanstraelen B, et al. Validation and IMRT/VMAT delivery quality of a preconfigured fast-rotating O-ring linac system. Med Phys Jan 2019;46:328–39.

[36] Hu Y, Byrne M, Archibald-Heeren B, Collett N, Liu G, Aland T. Validation of the preconfigured Varian Ethos Acuros XB Beam Model for treatment planning dose calculations: A dosimetric study. J Appl Clin Med Phys Dec 2020;21:27–42.

[37] Gao S, Netherton T, Chetvertkov MA, Li Y, Court LE, Simon WE, et al. Acceptance and verification of the Halcyon-Eclipse linear accelerator-treatment planning system without 3D water scanning system. J Appl Clin Med Phys Oct 2019;20:111–7.
[38] Saini A, Tichacek C, Johansson W, Redler G, Zhang G, Moros EG, et al. Unlocking a closed system: dosimetric commissioning of a ring gantry linear accelerator in a multivendor environment. J Appl Clin Med Phys Jan 2021.
[39] RaySearch Laboratories, RayStation10A ReferenceManual, Stockholm: RaySearch Labs, 2020.
[40] Nelms BE, Chan MF, Jarry G, Lemire M, Lowden J, Hampton C, et al. Evaluating IMRT and VMAT dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. Med Phys Nov 2013;40:111722.
[41] Kry SF, Molineu A, Kerns JR, Faught AM, Huang JY, Pulliam KB, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. Int J Radiat Oncol Biol Phys Dec 2014;90:1195–201.
[42] Koger B, Price R, Wang D, Toomeh D, Geneser S, Ford E. Impact of the MLC leaf-tip model in a commercial TPS: Dose calculation limitations and IROC-H phantom failures. J Appl Clin Med Phys Feb 2020;21:82–8.
[43] Snyder M, Halford R, Knill C, Adams JN, Bossenberger T, Nalichowski A, et al. Modeling the Agility MLC in the Monaco treatment planning system. J Appl Clin Med Phys May 2016;17:190–202.
[44] Kerns JR, Followill DS, Lowenstein J, Molineu A, Alvarez P, Taylor PA, Kry SF. "Agreement between institutional measurements and treatment planning system calculations for basic dosimetric parameters as measured by the imaging and radiation oncology core-Houston". Int J Radiat Oncol Biol Phys 08 2016;95:1527–34.
[45] Kerns JR, Followill DS, Lowenstein J, Molineu A, Alvarez P, Taylor PA, et al. Reference dosimetry data and modeling challenges for Elekta accelerators based on IROC-Houston site visit data. Med Phys May 2018;45:2337–44.
[46] Kry SF, Dromgoole L, Alvarez P, Leif J, Molineu A, Taylor P, Followill DS. "Radiation therapy deficiencies identified during on-site dosimetry visits by the imaging and radiation oncology core Houston Quality Assurance Center". Int J Radiat Oncol Biol Phys 12 2017;99:1094–100.
[47] Passal V, Barreau M, Tiplica T, Dufreneix S. Optimizing the effective spot size and the dosimetric leaf gap of the AcurosXB algorithm for VMAT treatment planning. J Appl Clin Med Phys May 2021.
[48] Middlebrook ND, Sutherland B, Kairn T. Optimization of the dosimetric leaf gap for use in planning VMAT treatments of spine SABR cases. J Appl Clin Med Phys Jul 2017;18:133–9.
[49] Kielar KN, Mok E, Hsu A, Wang L, Luxton G. Verification of dosimetric accuracy on the TrueBeam STx: rounded leaf effect of the high definition MLC. Med Phys Oct 2012;39:6360–71.