Medical Physics
The International Journal of Medical Physics Research and Practice

Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings

Julia Götstedt, Anna Karlsson Hauer, and Anna Bäck

Citation: Medical Physics **42**, 3911 (2015); doi: 10.1118/1.4921733
View online: http://dx.doi.org/10.1118/1.4921733
View Table of Contents: http://scitation.aip.org/content/aapm/journal/medphys/42/7?ver=pdfcov
Published by the American Association of Physicists in Medicine

Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings

Julia Götstedt
Department of Radiation Physics, University of Gothenburg, Göteborg 413 45, Sweden

Anna Karlsson Hauer and Anna Bäck[a)]
Department of Therapeutic Radiation Physics, Sahlgrenska University Hospital, Göteborg 413 45, Sweden

(Received 4 December 2014; revised 27 April 2015; accepted for publication 14 May 2015; published 10 June 2015)

Purpose: Complexity metrics have been suggested as a complement to measurement-based quality assurance for intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT). However, these metrics have not yet been sufficiently validated. This study develops and evaluates new aperture-based complexity metrics in the context of static multileaf collimator (MLC) openings and compares them to previously published metrics.
Methods: This study develops the *converted aperture metric* and the *edge area metric*. The *converted aperture metric* is based on small and irregular parts within the MLC opening that are quantified as measured distances between MLC leaves. The *edge area metric* is based on the relative size of the region around the edges defined by the MLC. Another metric suggested in this study is the *circumference/area* ratio. Earlier defined aperture-based complexity metrics—the *modulation complexity score*, the *edge metric*, the ratio monitor units (MU)/Gy, the *aperture area*, and the *aperture irregularity*—are compared to the newly proposed metrics. A set of small and irregular static MLC openings are created which simulate individual IMRT/VMAT control points of various complexities. These are measured with both an amorphous silicon electronic portal imaging device and EBT3 film. The differences between calculated and measured dose distributions are evaluated using a pixel-by-pixel comparison with two global dose difference criteria of 3% and 5%. The extent of the dose differences, expressed in terms of pass rate, is used as a measure of the complexity of the MLC openings and used for the evaluation of the metrics compared in this study. The different complexity scores are calculated for each created static MLC opening. The correlation between the calculated complexity scores and the extent of the dose differences (pass rate) are analyzed in scatter plots and using Pearson's r-values.
Results: The complexity scores calculated by the *edge area metric*, *converted aperture metric*, *circumference/area* ratio, *edge metric*, and MU/Gy ratio show good linear correlation to the complexity of the MLC openings, expressed as the 5% dose difference pass rate, with Pearson's r-values of −0.94, −0.88, −0.84, −0.89, and −0.82, respectively. The overall trends for the 3% and 5% dose difference evaluations are similar.
Conclusions: New complexity metrics are developed. The calculated scores correlate to the complexity of the created static MLC openings. The complexity of the MLC opening is dependent on the penumbra region relative to the area of the opening. The aperture-based complexity metrics that combined either the distances between the MLC leaves or the MLC opening circumference with the aperture area show the best correlation with the complexity of the static MLC openings.
 [http://dx.doi.org/10.1118/1.4921733]

Key words: complexity metric, quality assurance, MLC shape, aperture-based metric

1. INTRODUCTION

Intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) are treatment techniques developed to produce highly conformal treatment plans. Compared to conventional 3D conformal radiotherapy, these techniques may allow a higher dose to the target volume while simultaneously decreasing the toxicities in healthy tissue. Advanced dose optimization algorithms for IMRT and VMAT are designed to fulfill user-selected constraints and objectives. As a result, the treatment fields consist of the so-called control points with multileaf collimator (MLC) openings of various sizes and shapes. Irregularly shaped MLC openings with small subopening components are challenging from a dosimetric point of view, as the calculation, delivery, and measurement are affected, for example, due to lack of charged particle equilibrium (CPE).[1] Regarding the delivery, small MLC openings are considered complex due to the increased sensitivity of MLC positioning errors[2,3] and it has been shown that some dose calculation algorithms underestimate the delivered dose of small static MLC openings.[4] Treatment plans with fields composed of small MLC openings might therefore cause differences between the calculated and delivered dose distributions and can be considered as complex.

Due to the potential complexity of the IMRT and VMAT treatment fields, patient-specific pretreatment quality assurance (QA) is recommended.[5,6] One intention with QA is to detect treatment plans with clinically relevant differences between the planned and delivered doses. A common approach for IMRT/VMAT is to use measurement-based quality controls (QC), which compare a calculated and measured dose distribution to ensure that the planned absorbed dose distribution is delivered within predefined tolerances. Measurement-based QC is dependent on the measurement equipment and the choice of evaluation method. It has been shown that commonly used QC procedures are not always able to detect clinically relevant dose differences (dd).[7–10] Furthermore, measurement-based QC is a time consuming procedure that, at some radiation oncology departments, might limit the number of patients treated with advanced IMRT.

Complexity metrics have been suggested as an analytical and time-efficient independent alternative or complement to current QC methods, as they provide more information about the complexity of the treatment plan. This additional information can simplify the choice of plan for delivering robust treatment to the patient. A complexity metric could also be used as an optimization tool in the treatment planning system (TPS) to reduce the plan complexity at an early stage. The metrics are based on the different plan parameters that affect the complexity, i.e., parameters that might cause dosimetric differences between calculated and delivered dose distributions. The metrics can mainly be divided into two categories: fluence map-based metrics and aperture-based metrics.[11] A fluence map-based metric is a measure of intensity variations within the generated fluence distribution and these metrics have been developed for IMRT treatment plans with multiple static gantry entries.[12–15] Aperture-based metrics are based on the size and shape of MLC openings and can be valid for both IMRT and VMAT plans. Different approaches to quantify the complexity based on the geometry of the MLC opening have been described.[11,16,17] Some complexity metrics have been studied further and compared by investigating their correlation to the plan complexity.[18,19] These studies determine the plan complexity using measurement-based IMRT QC methods of treatment plans using gamma evaluation.[20] However, gamma evaluation does not always distinguish complex treatment plans.[7–10] Furthermore, the study of a treatment plan as a whole makes the analysis less clear and distinct because of variations in complexity between different control points in the treatment plan. Hence, a definitive validation of the proposed complexity metrics is still needed.

This study is a first step in our overall ambition to systematically develop and evaluate an aperture-based complexity metric suitable for both IMRT and VMAT treatment plans. New aperture-based metrics are designed and compared to other aperture-based metrics described in the literature. In this first step, their correlation to the complexity of created static MLC openings is investigated. The static MLC openings represent control points in an IMRT or VMAT treatment plan. The dynamic nature of the delivery is not addressed in this initial study.

2. MATERIALS AND METHODS

2.A. Created MLC openings of varying complexities

Small and irregular static MLC openings were created to simulate individual IMRT/VMAT control points with MLC openings of various complexities (i.e., with various sizes and shapes). These openings were created in the Eclipse treatment planning system (Eclipse™ version 11, Varian Medical Systems) by systematically decreasing the opening area or varying the MLC positions to create more irregular shapes. The Millennium™ 120 MLC (Varian Medical Systems) was used to create the various openings. The leaf width was 5 mm for the 40 central leaves and 10 mm for the 10 outer leaves on each side. Six series (A–F) of MLC openings were designed to study various complex situations based on, e.g., the MLC opening area, the MLC opening circumference, and the situation of several MLC subopenings. The MLC openings within each series were modified to gradually become more complex in their shape, size, or a combination of the two (Fig. 1).

In Series A, the MLC opening areas had decreasing size and circumference but maintained a constant square shape. For the MLC openings in Series B, the area was slightly decreased and the circumference of the areas was considerably increased. Series D and E had a constant total MLC opening area with increasing total circumference, and Series C and F had a constant total circumference and a decreasing total opening area. Series D also consisted of MLC openings with several small subopening components. The created static MLC openings challenge the accuracy of both the calculation and the delivery.

2.B. Measurement, dose calculation, and evaluation

All 30 MLC openings were delivered with a Clinac iX (Varian Medical Systems) linear accelerator and measured with an amorphous silicon electronic portal imaging device (EPID aSi1000, Varian Medical Systems) in integrated image mode as well as by using Gafchromic™ EBT3 film (Ashland). The EPID measurements were acquired for three occasions and the film measurements were only obtained once. The accelerator is calibrated to give 1 Gy at 10 cm depth in water for 120 monitor units (MU), 10×10 cm^2 field size, and source to surface distance (SSD) 90 cm. The number of MU for the MLC openings was adjusted to deliver 2 Gy, within ±0.2%, centrally in the most open part of the MLC opening at a depth of 10 cm in a solid water phantom, when calculated with the analytical anisotropic algorithm (AAA) algorithm in Eclipse™. The calculation grid size for the AAA calculations was set to 0.25×0.25 cm. The TPS commissioning, including assessment of the dosimetric parameters of the MLC, i.e., dosimetric leaf gap (DLG) and MLC transmission factor, was previously performed, and the dose calculation algorithm was carefully validated for clinical use. The regular machine QA at the hospital was based on the recommendations of the task group 142 report.[21] The DLG was determined according to the sweeping gap method described by the TPS manufacturer ("Dosimetric Leaf Gap Measurement," Varian Medical Systems, August 2011.) and entered into Eclipse. The value of the DLG was set to 0.2 cm and it was periodically validated

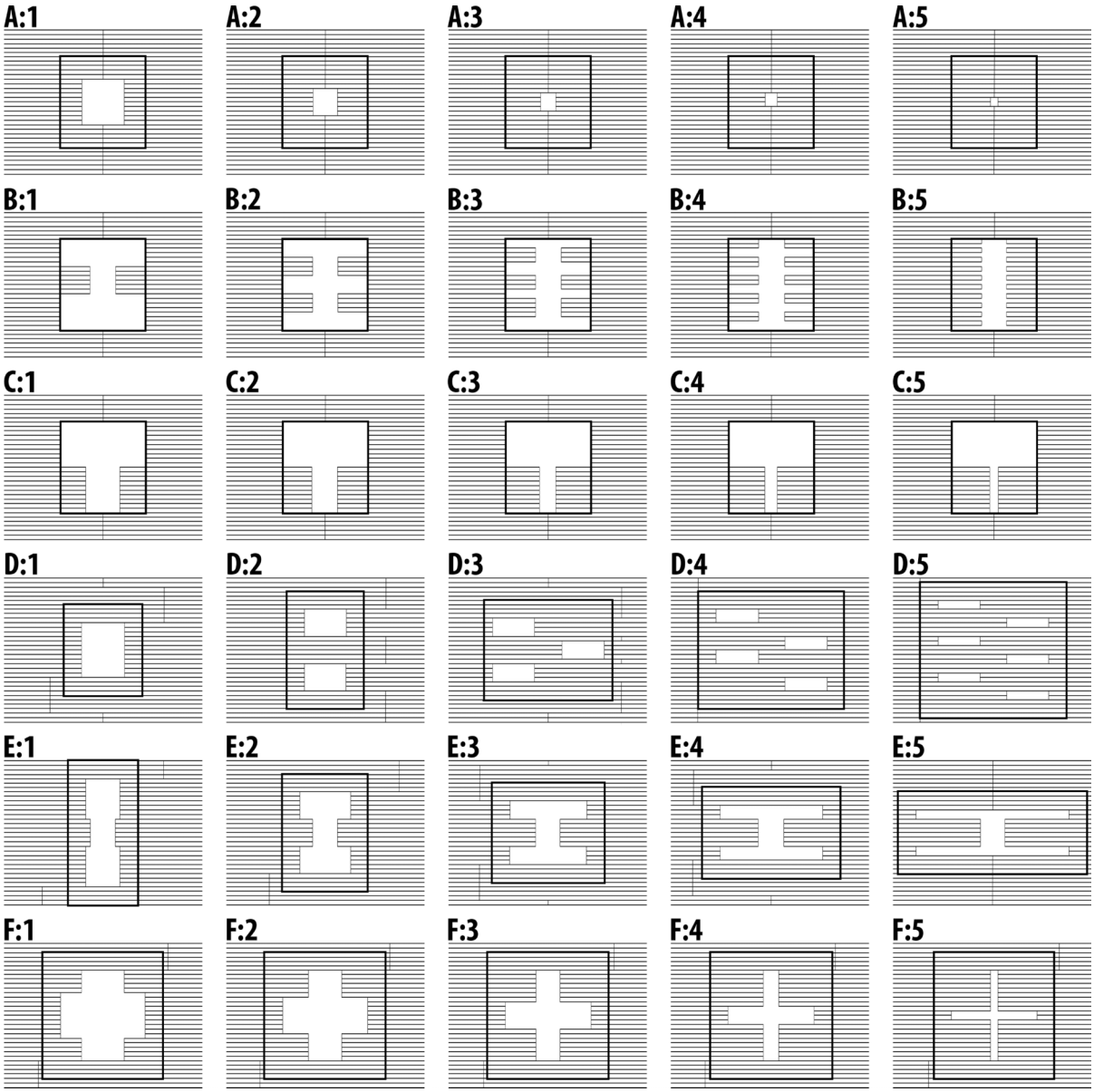

FIG. 1. Illustrations of the 30 created MLC openings (Millennium™ 120 MLC, Varian Medical Systems) grouped in Series A–F numbered from 1 to 5 with increasing number for more complex MLC openings. All MLC openings are to scale. For scale reference, the MLC leaves are 5 mm wide and the collimator jaws are positioned at 10×10 cm^2 for the MLC openings in Series A.

and verified according to the monthly QA program in the hospital. Eclipse currently allows only one value for the MLC transmission parameter and does not account for differences between inter- and intraleaf transmission or for changes in transmission with off-axis position, field size, and depth. The transmission was measured with an ionization chamber that was large relative to the MLC leaf width (Roos™ ionization chamber with electrode diameter 15.6 mm). The transmission factor that was entered into Eclipse, 0.015, was measured at a clinically relevant point (on the central axis at 10 cm depth in water with 10×10 cm^2 collimator setting), but the MLC transmission was also measured in other positions to ensure that there were only small variations from the value entered into Eclipse. The effect of the increased transmission through the drive screw slot as compared to the transmission closer to the leaf end was taken into account.

2.B.1. EPID

All of the static MLC openings were measured with EPID aSi1000 with the gantry at 0° and a source-to-measurement distance of 100 cm. A dosimetry calibration of the EPID was conducted according to the manufacturer's instructions (Portal Imaging and Portal Dosimetry Reference Guide, Varian Medical Systems, February 2008) before the measurement sessions. A 10×10 cm^2 calibration field shaped by the collimators was repeatedly measured during the sessions to verify the stability of the calibration, and a 10×10 cm^2 field shaped with the MLC was measured repeatedly for alignment. All of the measurements were analyzed within the Portal Dosimetry module in ARIA (Varian Medical Systems). The autoalign feature within Portal Dosimetry was applied to the MLC shaped 10×10 cm^2 fields, and the mean value of the

suggested corrections was used to align the measured static MLC openings. The metal support bar at the arm mount of the EPID contributed to a backscatter component that affected large fields closest to the gantry.[22] An algorithm that corrected for this effect was applied within the TPS when calculating EPID verification plans. Since the EPID was calibrated before each measurement session, no normalization was needed and the measurements were evaluated in calculated units (CU) in the Portal Dosimetry module.

2.B.2. Gafchromic EBT3 film

The MLC openings were measured with Gafchromic EBT3 film. Every pixel of the scanned film was calibrated based on its exposure to a known dose, using a double exposure method similar to the one described by Zhu *et al.*[23] Instead of assuming a linear-dose-to-density relation, a polynomial relation was established for every new batch of EBT3 films. In-house developed software in MatLab® (Mathworks, Natick, MA) was developed to facilitate the film calibration procedure. The film measurement was performed during three consecutive days. During Day 1, the unexposed films were scanned (with an Epson expression 1680 Pro) to correct for their initial nontransparency and the intrinsic irregularities within the film and the scanner. Later on that same day, the films were placed at a 10 cm depth in a solid water phantom and were exposed using a large (40×40 cm^2) open field. The uniform absorbed dose to the film was 2 Gy. In order to minimize the effect of any eventual beam asymmetry, the films were rotated 180° after the half exposure. Two films were placed in the phantom at the same time and were irradiated simultaneously. The exposed films were always scanned 18–24 h after irradiation to achieve a stable result. The sensitivity of each pixel was evaluated and a correction was applied based on the scan result of the open field irradiation. On Day 2, the 30 MLC openings were individually delivered to separate films. All irradiations were performed with the gantry at 0° and a SSD of 90 cm. The films were placed at a 10 cm depth in a $30\times30\times30$ cm^3 solid water phantom. A minimum of 3 cm margin was applied between the edge of the film and the MLC opening to avoid any uncertainties near the film edges. The films were scanned again during Day 3 within 18–24 h after the second irradiation. The measured dose distributions were corrected for any rotational errors and aligned to the calculated dose distribution. Vertical and horizontal dose profiles were used to correct for any positioning errors during the measurement. The median dose value within a centrally placed region of interest (ROI) of size 0.5×0.5 cm^2 was used to individually normalize the measured dose to the planned dose for the MLC openings of Series A, B, E, and F. For Series C, the ROI for normalization was positioned centrally in the upper part of the MLC opening; this area has a constant shape and size for all MLC openings in the series. Due to the small area of the MLC openings A:5, D:5, and F:5, a smaller square of size 0.3×0.3 cm^2 was used to exclude the surrounding dose gradients. For Series D, a mean value of the normalization values for each subopening was used. The image processing was performed using the evaluation software RIT113 (Radiological Imaging Technology ©).

2.B.3. Evaluation

The extent of dose differences between the calculated and measured dose distributions for the static MLC openings was used as a measure of field complexity. The planned and measured dose distributions were compared using a pixel-by-pixel comparison with two global dose difference criteria, 3% and 5%. Pixels that received less than 10% of the maximum dose in the calculated dose distribution were not considered in the evaluation. The relative number of pixels within 3% or 5% for each MLC opening, i.e., the pass rate, was used as a measure of complexity.

2.C. Development of new complexity metrics

An ideal complexity metric should consider calculation uncertainties and delivery issues that could lead to differences between the planned and delivered dose. Additionally, the metric should preferably be easy to apply and intuitive to understand. Simple plan parameters, such as the number of MUs and the circumference and area of the MLC opening, are examples of measures that are easy to interpret and relate to different complex situations. Metrics could also be defined as combinations of the various plan parameters to result in a more accurate measure of plan complexity. New aperture-based complexity metrics were developed in this study. An in-house developed MatLab® software was used to perform calculations of the metrics described below, i.e., converted aperture metric (CAM), edge area metric (EAM), and circumference/area metric. The calculations are based on the MLC positions stored in the RP-DICOM file.

Closed leaf pairs, which have a small physical opening of 0.05 cm to avoid eventual risk of collision, were not included in the calculation of any of the complexity scores calculated within this study.

2.C.1. Converted aperture metric

The *converted aperture metric* is an aperture-based metric based on the measured distances between the MLC leaves. This distance can be related to the size and shape of the MLC opening, where short distances indicate small and irregular parts. Within the *converted aperture metric*, distances both parallel and perpendicular to the direction of the MLC leaves were measured every 5 mm to quantify the MLC openings (Fig. 2). The 5 mm was chosen to cover the area of the MLC openings with a reasonable resolution and to give one measure for each MLC leaf pair in the MLC traveling direction.

A conversion function, $f(x) = 1 - e^{-x}$, was chosen to allow for a nonlinear relation between distances and their contribution to the complexity of the field. The idea of a conversion function was to penalize short distances more than larger ones. Measured distances, d_i, were converted to values ranging between 0 and 1, where noncomplex distances will be given a conversion value of 1. Distances larger than 4 cm were considered to be noncomplex due to the assumption that this distance is above the limit where the small field problematics end.[1] The *converted aperture metric* also includes the equivalent square field size, a_{Eq}, of the MLC opening. In this study, the

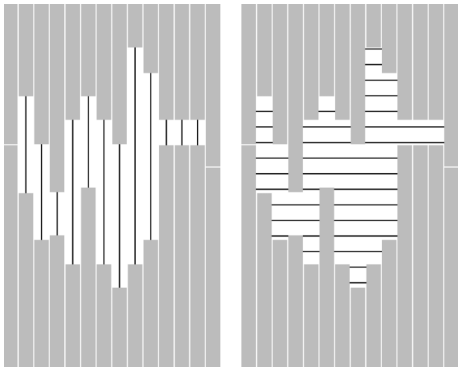

FIG. 2. Measured distances within the MLC opening, both parallel and perpendicular to the MLC traveling direction, are illustrated by the black lines within the white open part of the MLC opening. The MLC leaves are gray. For scale reference, the MLC leafs are 5 mm wide.

equivalent square field size was defined as the square root of the total area of the MLC opening. The equivalent square field size was converted according to the same conversion function as the measured distances d_i. The final complexity score is calculated in Eq. (1) as the mean value of all the conversion values of the measured distances multiplied with the conversion value of the equivalent square field size. To allow a more intuitive interpretation of the complexity score, the product of the conversion values is subtracted from one. This means that a more complex MLC opening will have a complexity score approaching unity,

$$\text{Converted aperture} = 1 - \overline{f(d_i)} \cdot f(a_{Eq}). \quad (1)$$

The *converted aperture metric* calculates a complexity score for each individual MLC opening of a control point and the final complexity score for a control point is defined as the mean of all calculated scores within the control point.

2.C.2. Edge area and circumference/area metrics

Another complexity metric was defined based on the relative amount of edge region for the MLC opening. For this complexity metric, the MLC openings were divided into two regions. One region, R_1, encloses the area 5 mm (at isocenter distance) on both sides of the MLC boundaries, inside and outside, of the opening. The other, region, R_2, includes the rest of the open area within the MLC opening. The 5 mm was chosen to fully enclose the penumbra region. The complexity metric, called the *edge area metric*, was then defined as the relative part of the MLC opening that is within the edge region. Thus, the complexity score was defined on a finite scale between 0 and 1, according to Eq. (2),

$$\text{Edge area} = \frac{R_1}{R_1 + R_2}. \quad (2)$$

In comparison, the ratio *circumference/area* was calculated for each MLC opening to be used as a more straight-forward, alternative complexity metric. *Circumference/area* was defined as the total circumference formed by the pattern of the MLC leaves expressed in cm, divided by the total area of the same MLC opening expressed in cm^2.

2.D. Previously defined complexity metrics for comparison studies

Five different previously defined aperture-based complexity metrics were included in this comparative study: the *modulation complexity score* (MCS),[16] the *edge metric* (EM),[17] MU/Gy,[15,18,19] the *aperture area* (AA) metric,[11] and the *aperture irregularity* (AI) metric.[11] The MCS was described by McNiven *et al.*[16] for the step-and-shoot IMRT, but the method can be modified for application to the sliding window IMRT or VMAT.[19] One part of the MCS is the aperture area variability (AAV), which considers area variations in the MLC opening during delivery of the treatment field. This part of the MCS was not considered, as the present study is based on static MLC openings and does not address dynamic delivery. The other part of MCS, the leaf sequence variability (LSV), is a measure of the irregularity between neighboring MLC leaves. The final score ranges from 0 to 1, where an unmodulated rectangular MLC opening is noncomplex according to definition and has a score of unity, regardless of the MLC opening size. The *edge metric*, described by Younge *et al.*,[17] is based on the ratio between the length of the MLC opening edge and the opening area. The contribution of the MLC edge length can be weighted differently for an edge that is along or opposed to the MLC leaf direction. Younge *et al.*[17] observed the strongest correlation between the metric and the extent of dose deviations between planned and measured VMAT plans when only the MLC edge along the MLC leaf direction with weight one was considered. This weighting was used within this study to calculate the *edge metric* scores. Another, more intuitive approach which has been used to score the field complexity is the ratio MU/Gy,[15,18,19] since complex fields often need to be delivered with more monitor units to deliver the same absorbed dose as noncomplex fields.[24] In this study, the absorbed dose value for the MU/Gy calculation was extracted from the AAA calculated dose distribution in the same region as used for normalization. Du *et al.*[11] suggested that a single metric is not enough to determine the complexity of a treatment plan; instead, several metrics should be used to quantify different aspects of the complexity. The AA and the AI, where AI is a measure of how the shape of the aperture differs from the shape of a circle, were two complexity metrics suggested.[11] As mentioned before, closed leaf pairs, which have a small physical opening of 0.05 cm to avoid eventual risk of collision, were not included in the calculation of any of the complexity scores calculated within this study.

3. RESULTS

3.A. Dose difference evaluation of static MLC openings of various complexities

The design of the new complexity metrics was based on the evaluation of the differences between the calculated and measured dose distributions for static MLC openings having different complexities (Fig. 1). The EPID was successfully calibrated, and all of the measurements with the 10×10 cm^2 field shaped by the collimator jaws matched the predicted

dose within 0.3%. The overall trends of the 3% and 5% dose difference evaluations were similar. The evaluated 5% dose differences are shown as percentage dose difference maps for the 30 MLC openings (Fig. 3). The dose differences contributing to a decreased pass rate were mainly from the pixels in the penumbra region. In the region closer to the EPID arm mount, the measurement was overestimated, probably due to an incorrect correction of the backscatter from the EPID arm mount. This was seen for the first MLC opening of Series E (E:1), where the effect is visible in the lower part of the dose difference map of E:1 (Fig. 3). Due to this effect, this MLC opening (E:1) was excluded from the complexity metric correlation analyses.

Figure 4 shows the pass rates for 3% and 5% dose difference evaluations for the EPID measurements and the 5% dose difference evaluation for the EBT3 film measurements for MLC openings of varying complexity. Data are shown for each of the series. The openings that were created to be noncomplex (low MLC opening numbering) resulted, with few exceptions, in higher pass rates, i.e., a larger fraction of the evaluated pixels were within the evaluation criteria. In comparison, MLC openings that were smaller or with a more irregular shape had relatively lower pass rates. The last MLC opening in Series A (A:5) had a higher pass rate than expected. This was probably due to the difficulties of measuring this very small MLC opening of 1×1 cm^2. The pass rate of the first MLC opening of Series E was lower than expected, likely due to an inaccurate correction of the backscatter from the EPID arm mount (as previously mentioned). The trend of the 3% and 5% dose difference evaluations was similar with an expected systematic shift toward lower pass rates for the more strict criterion of 3%. For some of the series, the film measurements showed more inconsequential results, with more variations in the pass rates within the series compared to the EPID measurements. This was probably due to the larger uncertainty incorporated in the film measurement procedure. The overall evaluation results were similar for EBT3 film and EPID measurements (Fig. 5).

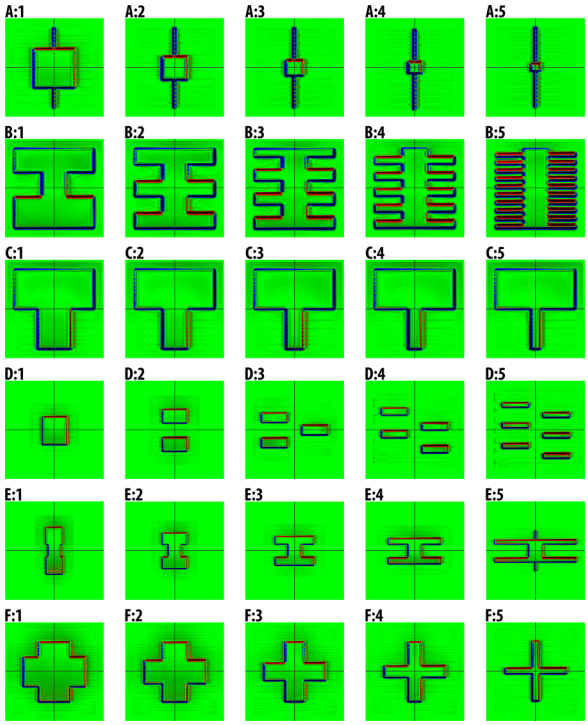

FIG. 3. Percentage dose difference maps for Series A–F from Portal Dosimetry (Varian Medical Systems) with window settings of ±5%, where pixels within the tolerance are colored green. Pixels where the measured data differ more than 5% from the calculated one are colored red (+5%) and blue (−5%). The evaluation region is defined by a dose cutoff at 10% of the maximum predicted dose. The isocenter for each MLC opening is marked with a cross hair. The figure is to scale within each series, but not in between series. For scale reference, please see Fig. 1.

The differences between the calculated and measured dose distributions for the 5% dose difference evaluations were generally smaller for the film measurements in which most of the MLC openings had pass rates between 70% and 90%. In comparison, the EPID measurements had pass rates in most cases between 40% and 85%. Even though there was an absolute difference between the EPID and EBT3 film pass rates, the overall linear correlation between the two measurement methods could be expressed with Pearson's r-values of 0.77 and 0.82 for 3% and 5% dose difference evaluation criteria, respectively. The complexity of the MLC openings was determined by the evaluated dose differences in a relative manner, and the absolute values of the pass rates were not of interest in this study. Correlation analyses between the dose difference pass rates and the complexity scores were conducted separately for the different types of measurement and evaluation methods.

3.B. Evaluation and comparison of complexity metrics

The complexity scores for all the metrics studied in this study were calculated for all the 30 static MLC openings and shown in Table I. The linear correlations between the field complexities (expressed as 5% dose difference pass rate for the EPID measurements) and the complexity scores for the different complexity metrics are illustrated in scatter plots (Fig. 6). The Pearson's r-values were calculated and listed in Table II including both 3% and 5% dose difference evaluations for both EPID and EBT3 film measurements.

The different complexity metrics studied here score the 30 static MLC openings in a similar way; however, the scores calculated by the MCS, AI, and AA differ from the other metrics and have a lower Pearson's r-value. The AI and AA are not designed to individually correlate with the dose difference pass rate, but rather to cover different aspects of field complexity.[11] The Pearson's r-values are similar for the 3% and 5% dose difference evaluations but are systematically lower for the EBT3 film measurements as compared to the EPID measurements. This is probably an effect of the larger uncertainty incorporated in the film measurement procedure as discussed before that caused a more inconsequent result for the film measurement presented in Fig. 4. The correlations between the different metrics in terms of Pearson's r-values are listed in Table III, but the AI and AA metrics are not included since they are not designed to be used individually to score field complexity.

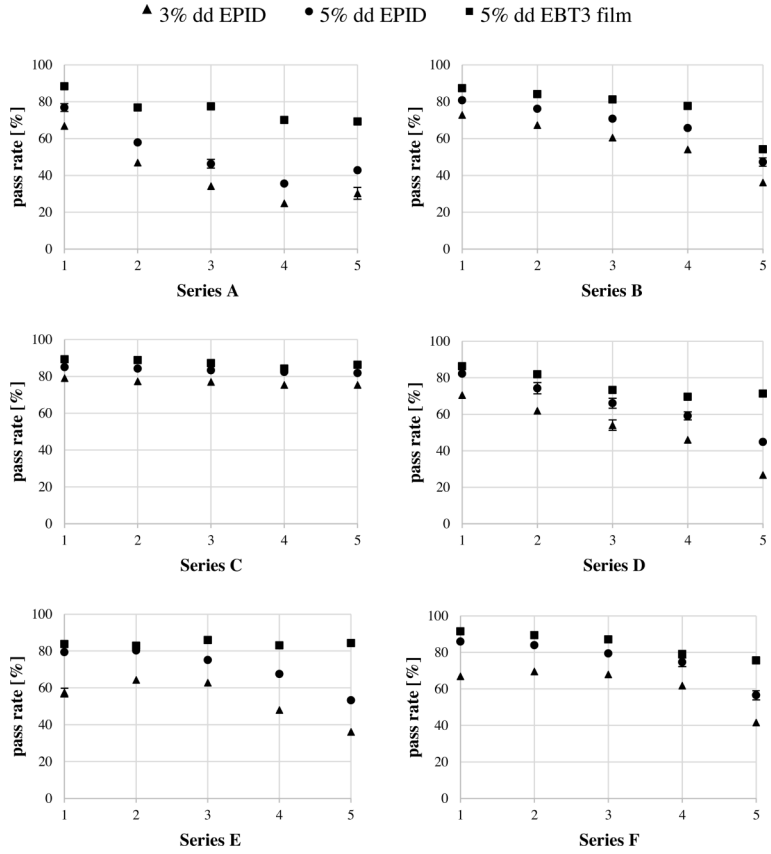

FIG. 4. Pass rates for 3% and 5% dose difference (dd) evaluations for the EPID measurements (triangles and circles, respectively) and for 5% dose dd evaluation for the film measurements (squares). The numbers 1–5 on the x-axis represent the different MLC openings within each series according to Fig. 1. Error bars (1 SD) for the variation of the three EPID measurements (triangles and circles) are shown in the case when the error bars are not within the size of the symbol.

4. DISCUSSION

Static MLC openings of various complexities were created to evaluate and compare different aperture-based complexity metrics. Varying degrees of complexity were obtained by modifying the size and shape of the MLC openings and assuming that smaller and more irregular shapes are more complex. Complexity has been defined in various ways by other authors. A frequently used definition is based on the extent of dosimetric differences between calculated and measured dose distributions, and that approach was used in this study.

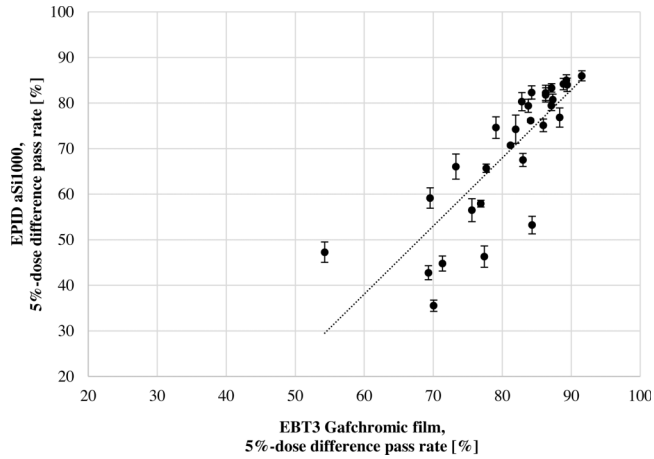

FIG. 5. Scatter plot illustrating the correlation between the 5% dose difference evaluation pass rate results for EPID and film measurements. The variation of the three EPID measurements is shown by the error bars (1 SD).

The results found in this study are dependent on the dosimetry system used for the measurements. The planar dosimetry systems, EPID and Gafchromic EBT3 film, were chosen for their outstanding spatial resolution which is advantageous in these types of measurements. All dosimetry systems have limitations and uncertainties and therefore two different dosimetry systems were chosen to increase the validity of the results. Furthermore, the EPID measurements were carried out at three occasions to partly take into account the precision of those measurements. Prior to the analysis, the film measurements were normalized centrally in the most open part of the MLC opening, and the EPID measurements were analyzed in CUs (not normalized). Some information will be lost due to the normalization; therefore, the pass rates for the film measurements can be expected to be higher compared with the evaluation results of the EPID measurements, as seen in Fig. 4. Despite the negative effect of lost information, normalization was still performed to avoid previously observed large uncertainties associated with non-normalized EBT3 Gafchromic film dosimetry. Another factor that distinguished the measurement methods was the alignment procedure. The alignment of the EPID measurements was in accordance with a correction based on an autoalignment feature in the evaluation module. This was done for a 10×10 cm^2 field shaped with the MLC that was repeatedly measured during the

TABLE I. The numerical complexity scores for the converted aperture metric (CAM), edge area metric (EAM), edge metric (EM), circumference/area (C/A), modulation complexity score (MCS), MU/Gy, aperture irregularity (AI), and aperture area (AA) for the 30 static MLC openings, Series A–F.

Series	CAM	EAM	EM (1/cm)	C/A (1/cm)	MCS	MU/Gy	AI	AA (cm^2)
A1	0.01	0.56	0.40	0.80	0.80	130	1.27	25.0
A2	0.10	0.75	0.67	1.00	0.80	138	1.27	16.0
A3	0.25	0.89	1.00	1.33	0.80	145	1.27	9.0
A4	0.40	0.96	1.33	2.00	0.80	154	1.27	4.0
A5	0.60	1.00	2.00	4.00	0.80	166	1.27	1.0
B1	0.02	0.50	0.43	0.68	0.80	124	2.94	79.0
B2	0.07	0.64	0.67	0.94	0.62	125	5.11	72.0
B3	0.13	0.74	0.91	1.20	0.47	126	7.82	68.5
B4	0.18	0.75	1.17	1.48	0.28	126	11.29	65.0
B5	0.31	0.85	2.25	2.55	0.00	126	33.75	65.0
C1	0.01	0.44	0.29	0.57	0.90	123	1.82	70.0
C2	0.02	0.47	0.31	0.62	0.90	124	1.96	65.0
C3	0.04	0.49	0.33	0.67	0.90	124	2.12	60.0
C4	0.06	0.51	0.35	0.70	0.90	125	2.22	57.5
C5	0.10	0.53	0.36	0.73	0.90	125	2.32	55.0
D1	0.01	0.52	0.33	0.73	0.90	129	1.28	30.0
D2	0.05	0.67	0.67	1.07	0.71	133	2.72	30.0
D3	0.11	0.77	1.00	1.40	0.61	137	4.68	30.0
D4	0.17	0.90	1.33	1.73	0.66	142	7.18	30.0
D5	0.27	1.00	2.00	2.40	0.52	154	13.76	30.0
E1	0.02	0.54	0.22	0.76	0.93	128	2.05	45.0
E2	0.03	0.56	0.40	0.80	0.90	128	2.29	45.0
E3	0.07	0.65	0.67	0.98	0.80	129	3.43	45.0
E4	0.11	0.74	0.93	1.20	0.77	130	5.16	45.0
E5	0.19	0.90	1.47	1.69	0.62	132	10.22	45.0
F1	0.00	0.42	0.27	0.53	0.90	121	1.70	75.0
F2	0.01	0.47	0.31	0.63	0.90	122	1.99	64.0
F3	0.04	0.56	0.39	0.78	0.89	125	2.50	51.0
F4	0.11	0.70	0.56	1.11	0.89	129	3.54	36.0
F5	0.35	1.00	1.05	2.11	0.88	139	6.70	19.0

EPID measurement sessions. With this alignment procedure, the effect of the tongue-and-groove design on the field edges along the MLC leaves partially disappeared at the 10×10 cm^2 field size. This slightly increased the pass rate results for field sizes close to 10 cm in the direction perpendicular to the MLC leaves. This method was selected to minimize the effect of a nonperfect MLC calibration and to apply a common alignment method to all MLC openings. For the film measurements, the measured MLC openings were aligned using dose profiles in the horizontal and vertical directions on an individual basis. The effect of the tongue-and-groove design was minimized for all openings, which again led to higher pass rates for the film measurements compared to the EPID measurements. Even though the evaluation of the differences between the calculated and measured dose distributions did not agree according to the definition of agreement by Bland and Altman (Ref. 25) for the EPID and film measurements, the reliability of the two different measurements methods was strengthened. This was because the trend of decreasing pass rate for decreasing area or increasing irregularity was similar and the evaluation results of EPID and film measurements had a good linear correlation, with a Pearson's r-value of 0.77 and 0.82 for 3% and 5% dose difference evaluation criterion, respectively. The dose difference pass rates are evaluated in a relative manner and the absolute value of the pass rates is not of interest in this study. The spread around the trend line (Fig. 5) increased with the increasing complexity of the MLC openings. This was probably because an increased complexity led to increased measurement and calculation uncertainty.

In this study, due to the careful alignment procedure for both the film and EPID measurements, a pixel-by-pixel dose difference comparison was used. A stricter criterion could theoretically better separate the pass rates of the various MLC openings, but the criterion should not be too strict so that noncomplex parts of the fields will fail because of dose difference criterion smaller than the uncertainty in the measurement. In this study, both 3% and 5% dose differences were used and the main results for the two criteria were similar.

The dose calculations in this study were performed with the AAA in Eclipse™ and a calculation grid size of 0.25×0.25 cm. Different dose calculation algorithms might have different limitations that would affect the deviations between

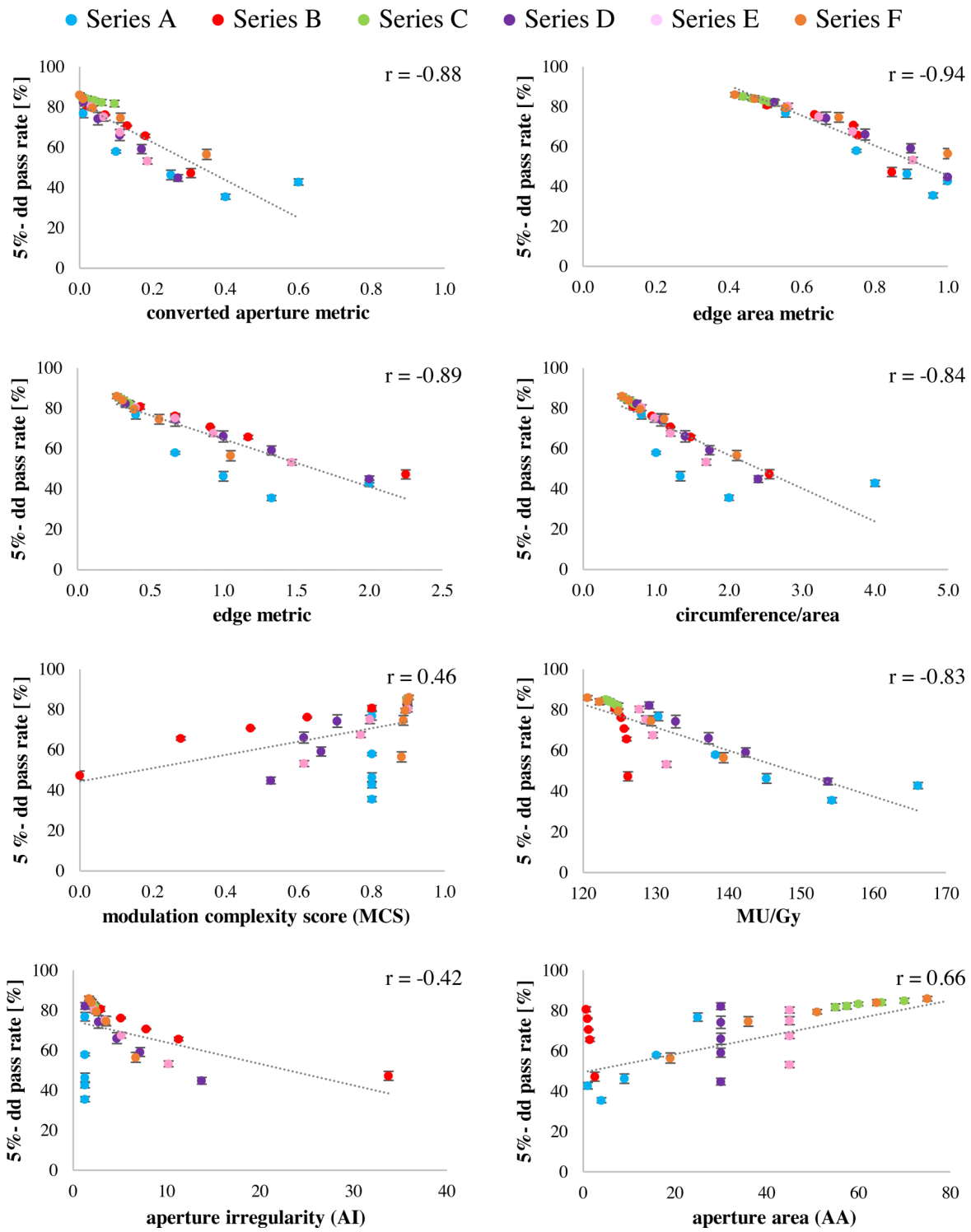

FIG. 6. Scatter plots illustrating the correlation between the 5% dose difference (dd) pass rates for the EPID measurements and each of the studied complexity metrics. Pearson's r-values are indicated in the upper right corner of each scatter plot.

the calculated and measured dose distribution. However, the principal limitations of modern dose calculation algorithms are similar. The results found in this study—based on the AAA—can be expected to be similar for other modern dose calculation algorithms, but this needs further investigation. The dose calculation grid size will also have an influence on the results and the importance of this also needs further investigations. Both the influence of different dose calculation algorithms and different calculation grid sizes will be investigated in future studies. The results are also dependent on the AAA beam modeling during commissioning, in particular, the choice of the dosimetric parameters of the MLC, i.e., DLG and MLC transmission factor, and the regular machine QA. The DLG in Eclipse accounts for the MLC's rounded leaf tips. The DLG influences the dose distribution around the field edge defined by the leaf tips of the MLC. The dose

TABLE II. The linear correlation, expressed in Pearson's r-values, between the 3% and 5% dd pass rates for both measurement methods (EPID and EBT3 film) and each of the studied complexity metrics.

	Pearson's r-value			
	EPID		EBT3 film	
Complexity metrics	3% dd	5% dd	3% dd	5% dd
Converted aperture metric	−0.85	−0.88	−0.78	−0.76
Edge area metric	−0.94	−0.94	−0.83	−0.79
Edge metric	−0.88	−0.89	−0.83	−0.87
Circumference/area	−0.83	−0.84	−0.78	−0.80
Modulation complexity score	0.44	0.46	0.59	0.67
MU/Gy	−0.81	−0.83	−0.66	−0.61
Aperture irregularity	−0.41	−0.42	−0.56	−0.68
Aperture area	0.66	0.66	0.49	0.45

differences contributing to a decreased pass rate are mainly from the pixels in the penumbra region and, therefore, different choices for the DLG influence the evaluated pass rates of this study. The main effect of different DLGs is probably seen in the absolute values of the evaluated pass rates, which are not of interest in this study. The effect on the relative relation of the pass rates due to different DLGs is probably smaller. In Series B, C, and F, the field edge defined by the leaf tips is constant and the effect of different DLGs on the relative relation of the pass rates is expected to be minimal. However, it would be of interest to study the robustness of the correlation between the calculated complexity scores and the evaluated pass rates due to differences in the DLG parameter. In the present study, the influence on the results due to different MLC transmission factors in the TPS was small—first, because only static MLC openings were included, and second, because the evaluation was limited to pixels that received more than 10% of the maximum dose in the calculated dose distribution. The machine status, e.g., MLC position calibration, also influenced the results. This was partly taken into account for the EPID measurements since the measurement was repeated for three occasions. The variation in machine status for the different measurements is included in the error bars in Figs. 4–6.

TABLE III. The linear correlation, expressed in Pearson's r-values, between the complexity scores calculated by the different complexity metrics studied.

Complexity metrics	Pearson's r-value
Converted aperture metric vs circumference/area	0.95
Converted aperture metric vs edge area metric	0.86
Converted aperture metric vs edge metric	0.84
Converted aperture metric vs MU/Gy	0.82
Converted aperture metric vs MCS	−0.33
Edge area metric vs edge metric	0.87
Edge area metric vs circumference/area	0.85
Edge area metric vs MU/Gy	0.81
Edge area metric vs MCS	0.45
Circumference/area vs edge metric	0.92
Circumference/area vs MU/Gy	0.81
Circumference/area vs MCS	−0.45
Edge metric vs MCS	−0.70
Edge metric vs MU/Gy	0.69
MCS vs MU/Gy	−0.09

In the dose difference evaluation of the static MLC openings, it was found that the dose differences contributing to a decreased pass rate were mainly from the pixels in the penumbra region (Fig. 3). The evaluation pass rates decreased for increasing circumference, for example, this occurred in Series D (Fig. 4) with constant total area of 30 cm^2 and a circumference varying from 22 to 72 cm. However, Series F, with a constant circumference of 40 cm and a varying area of 19–75 cm^2, also showed a decreasing dose difference pass rate. The complexity was dependent on the penumbra region relative to the area of the field. An almost constant dose difference pass rate was found for Series C. These MLC openings have a constant circumference of 40 cm and a very small variation in area, between 55 and 70 cm^2. The MLC openings were designed with a noncomplex rectangle of 5 $\times 10$ cm^2 and a complex part that was small relative to the total area of the field. The measurement results indicated that the smaller, complex part of the opening did not affect the overall pass rate of the total MLC opening to any considerable extent, even though the smaller part was becoming more and more complex within the series. However, this type of small opening can have clinical importance in patient cases.

The results of this study show that aperture-based complexity metrics based on the size and shape of the MLC openings can be appropriate for determining the complexity of static MLC openings. It is interesting to compare the *edge area metric*, the ratio of the *circumference/area*, and the *edge metric*, since they are all based on some measure of the field edge length in relation to the aperture area. These metrics correlate well to the complexities of the static MLC openings. The *converted aperture metric* also showed good correlation to the evaluated dose differences for the static MLC openings. The *converted aperture metric* is based on a conversion function that allows for a nonlinear relation between distances and their contribution to the complexity of the field. This function can be modified in order to achieve a better correlation between the calculated scores and the complexity of the MLC openings. For MLC designs with different leaf widths, the sampling resolution of the distances for the *converted aperture metric* may need to be altered to match the MLC leaf width. The sampling resolution should not be larger than the MLC leaf width. The correlation was not as good for the MCS, the MU/Gy ratio, or for the AA and AI. The MCS is not an appropriate metric for distinguishing static MLC openings of various complexities, since it has difficulties separating openings that have the same shape but various sizes (e.g., Series A and F). The metrics AA and AI were not designed to individually correlate with complexity since they are based on different aspects of the field complexity. For example, the MLC openings in Series A all have the same shape and therefore are equally scored using the AI, but they are well separated by size using the AA. The MCS, AA, AI, and MU/Gy ratio were not designed to be used on static MLC openings and they were previously evaluated only for IMRT treatment plans.

The MCS developed by McNiven *et al.*[16] defines the scores on a finite range between 0 and 1, where noncomplex apertures have a score of 1. The *converted aperture metric* and the *edge area metric* follow the same approach, but here, a higher complexity score indicates a more complex field and a more intuitive interpretation of the complexity metric. A complexity score that gives values between 0 and 1 makes it easier to interpret the absolute complexity of an individual treatment plan.

This study is a first step in our overall ambition to systematically develop and evaluate aperture-based complexity metrics. In this initial study, the aim is to start with the basics of IMRT/VMAT treatment plans and investigate the correlation of complexity metric scores to the complexity of static MLC openings representing control points of IMRT/VMAT treatment plans. The MLC openings used in this study were not based on MLC openings from clinical IMRT/VMAT treatment plans but were created in series with a controlled gradual increase in complexity within the series. Future work will focus on IMRT and VMAT treatment fields that include the dynamic delivery of the field. Next steps will also include further development of the complexity metrics to include a method for combining the control points into a complexity score for the whole treatment field so that the complexity score will have a good correlation with the complexity of the field. A more in depth investigation of the correlation of the complexity metric scores and the complexity of clinical IMRT/VMAT treatment plans will be included in future studies.

5. CONCLUSION

New aperture-based complexity metrics were developed in this work. The calculated scores correlated to the complexity of the created static MLC openings. The complexity of the MLC opening is dependent on the penumbra region relative to the area of the opening. Other aperture-based complexity metrics, previously described in the literature, were calculated for comparison. The aperture-based complexity metrics that combined either the distances between the MLC leaves or the MLC opening circumference with the aperture area show the best correlation with the complexity of the static MLC openings.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of this project by the Swedish Radiation Safety Authority. The authors are thankful for support from their colleagues at the Department of Therapeutic Radiation Physics at Sahlgrenska University Hospital. Special thanks also to Tom Bäck and Richard Hauer for help with the graphics.

There are no potential conflicts of interest.

[a)]Author to whom correspondence should be addressed. Electronic mail: anna.back@vgregion.se; Telephone: 0046-31-3427216.

[1]I. J. Das, G. X. Ding, and A. Ahnesjö, "Small fields: Nonequilibrium radiation dosimetry," Med. Phys. **35**, 206–215 (2008).

[2]T. LoSasso, C. S. Chui, and C. C. Ling, "Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy," Med. Phys. **25**, 1919–1927 (1998).

[3]M. Oliver, K. Bush, S. Zavgorodni, W. Ansbacher, and W. A. Beckham, "Understanding the impact of RapidArc therapy delivery errors for prostate cancer," J. Appl. Clin. Med. Phys. **12**, 32–43 (2011).

[4]L. S. Fog, J. F. B. Rasmussen, M. Aznar, F. Kjær-Kristoffersen, I. R. Vogelius, S. A. Engelholm, and J. P. Bangsgaard, "A closer look at RapidArc® radiosurgery plans using very small fields," Phys. Med. Biol. **56**, 1853–1863 (2011).

[5]*Guidelines for the Verification of IMRT*, edited by B. J. Mijnheer and D. Georg (ESTRO, Brussels, 2008).

[6]J. M. Moran, M. Dempsey, A. Eisbruch, B. A. Fraass, J. M. Galvin, G. S. Ibbott, and L. B. Marks, Safety considerations for IMRT (full report), 2011, available at: http://www.practicalradonc.org.

[7]J. Nilsson, A. Karlsson Hauer, and A. Bäck, "IMRT patient-specific QA using the delta4 dosimetry system and evaluation based on ICRU 83 recommendations," J. Phys.: Conf. Ser. **444**, 012048 (2013).

[8]B. E. Nelms, H. Zhen, and W. A. Tomé, "Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors," Med. Phys. **38**, 1037–1044 (2011).

[9]H. Zhen, B. E. Nelms, and W. A. Tomé, "Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA," Med. Phys. **38**, 5477–5489 (2011).

[10]J. J. Kruse, "On the insensitivity of single field planar dosimetry to IMRT inaccuracies," Med. Phys. **37**, 2516–2524 (2010).

[11]W. Du, S. Hyun Cho, X. Zhang, K. E. Hoffman, and R. J. Kudchadker, "Quantification of beam complexity in intensity-modulated radiation therapy treatment plans," Med. Phys. **41**, 021716 (9pp.) (2014).

[12]S. Webb, "Use of a quantitative index of beam modulation to characterize dose conformality: Illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy," Phys. Med. Biol. **48**, 2051–2062 (2003).

[13]M. M. Coselmon, J. M. Moran, J. D. Radawski, and B. A. Fraass, "Improving IMRT delivery efficiency using intensity limits during inverse planning," Med. Phys. **32**, 1234–1245 (2005).

[14]M. M. Matuszak, E. W. Larsen, and B. A. Fraass, "Reduction of IMRT beam complexity through the use of beam modulation penalties in the objective function," Med. Phys. **34**, 507–520 (2007).

[15]M. Nauta, J. E. Villarreal-Barajas, and M. Tambasco, "Fractal analysis for assessing the level of modulation of IMRT fields," Med. Phys. **38**, 5385–5393 (2011).

[16]A. L. McNiven, M. B. Sharpe, and T. G. Purdie, "A new metric for assessing IMRT modulation complexity and plan deliverability," Med. Phys. **37**, 505–515 (2010).

[17]K. C. Younge, M. M. Matuszak, J. M. Moran, D. L. McShan, B. A. Fraass, and D. A. Roberts, "Penalization of aperture complexity in inversely planned volumetric modulated arc therapy," Med. Phys. **39**, 7160–7170 (2012).

[18]C. K. McGarry, C. D. Chinneck, M. M. O'Toole, J. M. O'Sullivan, K. M. Prise, and A. R. Hounsell, "Assessing software upgrades, plan properties and patient geometry using intensity modulated radiation therapy (IMRT) complexity metrics," Med. Phys. **38**, 2027–2034 (2011).

[19]L. Masi, R. Doro, V. Favuzza, and S. Cipressi, "Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy," Med. Phys. **40**, 071718 (11pp.) (2013).

[20]D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," Med. Phys. **25**, 656–661 (1998).

[21]E. E. Klein, J. Hanley, J. Bayouth, F.-F. Yin, W. Simon, S. Dresser, C. Serago, F. Aguirre, L. Ma, B. Arjomandy, C. Liu, C. Sandin, and T. Holmes, "Task group 142 report: Quality assurance of medical accelerators," Med. Phys. **36**, 4197–4212 (2009).

[22]A. Van Esch, D. P. Huyskens, L. Hirschi, S. Scheib, and C. Baltes, "Optimized varian aSi portal dosimetry: Development of datasets for collective use," J. Appl. Clin. Med. Phys. **14**, 82–99 (2013).

[23]Y. Zhu, A. S. Kirov, V. Mishra, A. S. Meigooni, and J. F. Williamson, "Quantitative evaluation of radiochromic film response for two-dimensional dosimetry," Med. Phys. **24**, 223–231 (1997).

[24]R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," Med. Phys. **27**, 1226–1237 (2000).

[25]J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," Int. J. Nurs. Stud. **47**, 931–936 (2010).