Fractal analysis for assessing the level of modulation of IMRT fields

Marcel Nauta
Department of Physics and Astronomy, University of Calgary, Calgary, Alberta T2N 1N4, Canada

J. Eduardo Villarreal-Barajas and Mauro Tambasco[a)]
Department of Physics and Astronomy, University of Calgary, Calgary, Alberta T2N 1N4, Canada; Department of Oncology, University of Calgary, Calgary, Alberta T2N 1N4, Canada; and Tom Baker Cancer Centre, Calgary, Alberta T2N 4N2, Canada

(Received 29 March 2011; revised 18 July 2011; accepted for publication 15 August 2011; published 14 September 2011)

Purpose: To investigate the potential of three fractal dimension (FD) analysis methods (i.e., the variation, power spectrum, and variogram methods) as metrics for quantifying the degree of modulation in planned intensity modulated radiation therapy (IMRT) treatment fields, and compare the most suitable FD method to the number of monitor units (MUs), the average leaf gap, and the 2D modulation index (2D MI) for assessing modulation.
Methods: The authors implemented, validated, and compared the variation, power spectrum, and variogram methods for computing the FD. Validation of the methods was done using mathematical fractional Brownian surfaces of known FD that ranged in size from 128×128 to 512×512. The authors used a test set consisting of seven head and neck carcinoma plans (50 prescribed treatment fields) to choose an FD cut-point that ensures no false positives (100% specificity) in distinguishing between moderate and high degrees of field modulation. The degree of field modulation was controlled by adjusting the fluence smoothing parameters in the Eclipse^{TM} treatment planning system (Varian Medical Systems, Palo Alto, CA). The moderate modulation fields were representative of the degree of modulation used clinically at the authors' institution. The authors performed IMRT quality assurance (QA) on the 50 test fields using the MapCHECK^{TM} device. The FD cut-point was applied to a validation set consisting of four head and neck plans (28 fields). The area under the curve (AUC) from receiver operating characteristic (ROC) analysis was used to compare the ability of FD, number of MUs, average leaf gap, and the 2D MI for distinguishing between the moderate and high modulation fields.
Results: The authors found the variogram FD method to be the most suitable for assessing the modulation complexity of IMRT fields for head and neck carcinomas. Pass rates as measured by the gamma criterion for the MapCHECK^{TM} IMRT field measurements were higher for the moderately modulated fields, and a gamma criterion with 1 mm distance-to-agreement and 1% dose difference showed a clear separation between the 94% pass rates of the moderate and high modulation groups. From the ROC analysis of the test set, the authors found the AUC of the variogram FD, number of MUs, average leaf gap, and 2D MI methods to be 0.99 (almost perfect), 0.91 (excellent), 0.91 (excellent), and 0.92 (excellent), respectively. A cut-point of $FD > 2.25$ correctly identified 92.8% of the high modulation fields and 100% of the moderately modulated fields in the validation set, satisfying the condition of no false positives.
Conclusions: Of the three FD methods investigated, the variogram method is the most accurate and precise metric for identifying high modulation treatment fields. It is also more accurate and precise than the number of MUs, the average leaf gap, and the 2D MI. Although MapCHECK^{TM} IMRT QA does a reasonable job at identifying high modulation fields, the variogram FD method provides one with the opportunity to quantitatively and accurately assess modulation and adjust overly modulated fields at the treatment planning stage before they are sent to the treatment machine for QA or patient treatment. [DOI: 10.1118/1.3633912]

Key words: IMRT, complexity, quality assurance, fractal dimension, treatment planning, variogram method

I. INTRODUCTION

Intensity modulated radiation therapy (IMRT) is a radiotherapy method of treating cancer that includes a set of typically five to nine nonparallel opposed intensity modulated radiation beams that conform to the planning target volume (PTV). The modulation is designed to optimize the effectiveness of a treatment plan by delivering the intended prescribed dose to the PTV while minimizing dose to the organs at risk ("normal critical structures"). However, this optimization process could lead to overly modulated fields that increase beam-on time and are mechanically more

challenging to deliver.[1–3] Longer beam-on time increases the amount of radiation dose to organs at risk due to inter- and intra-leaf transmission, leakage, and scatter. It can also lead to patient discomfort, which increases the risk of movement during treatment. Furthermore, increased stress on the mechanical components of the linear accelerator can lead to increasing costs associated with machine downtime and repair. Overly modulated fields may also be more likely to suffer from tongue-and-groove effect,[4] which may not be modeled well by the treatment planning system, and may not even show up in the IMRT quality assurance (QA), depending on the QA detector and/or pass criteria. Finally, overly modulated fields may be more likely to suffer from the interplay effect when used to treat moving tumors.[5]

A relatively recent AAPM Task Group report[6] investigating a reasonable and achievable standard for IMRT commissioning pointed out the need to have measures of beam modulation to ensure that plans are comparable with regards to their complexity. This need stems from the fact that the level of complexity of individual plans is related to the delivery accuracy and associated quality assurance metrics.[7–9] Also, a recent study by Kruse[8] reported that highly modulated fields are not detected by patient-specific IMRT QA based on the single field planar dosimetry performed using an electronic portal imaging device (EPID). Hence, there is a need for accurate quantification and assessment of the level of beam modulation to ensure that treatment plans containing overly modulated fields (i.e., fields containing more modulation than necessary to satisfy the clinical objectives) are modified to an acceptable level of modulation.

Mohan *et al.*[3] used the term "complexity" to describe the degree of modulation as characterized by frequency and amplitude variations of a beam. Webb[10] proposed the modulation index (MI) concept for assessing the degree of modulation. Webb defined MI as the integral of a spectrum comprised of the number of adjacent-element intensity changes along the x-direction that exceed a certain fraction of the standard deviation in the beam. More recently, Giorgia *et al.*[1] generalized this definition to changes along the x-direction, y-direction, and x-y diagonal direction and called it the 2D MI. In this study, we investigate the usefulness of three fractal dimension (FD) approaches for quantifying the degree of modulation of radiation fields, and we test the hypothesis that FD classifies moderate versus highly modulated fields more accurately than the number of monitor units (MUs), the average leaf gap, and the generalized 2D MI. The motivation for investigating FD is that it is by definition a quantitative measure of complexity as characterized by amplitude variations at different scales or frequencies, and as such, it lends itself naturally as a quantitative index of modulation.

To test our hypothesis, we compared and validated our implementation of the variation, power spectrum, and variogram FD methods using mathematical objects (fractional Brownian surfaces) of known FD. From this comparison, we found that the variogram method was the most suitable fractal analysis method for assessing modulation complexity. Next, we used a commercial treatment planning system (Eclipse™, Varian Medical Systems, Palo Alto, CA) to generate a test set and a validation set that contained fields of both moderate and high degrees of modulation. We performed MapCHECK™ IMRT field measurements and assessed pass rates for the test set using the gamma criterion. In addition, we performed a receiver operating characteristics (ROC) analysis to investigate the classification performance of the variogram FD method, the number of MUs, the average leaf gap, and the 2D MI. Using this analysis, we chose an FD cutoff that maximized the sensitivity for identifying a highly modulated field while maintaining 100% specificity. Finally, we used the validation set to independently test the performance of this cutoff.

II. METHODS AND MATERIALS

A fractal surface can be thought of as a surface that has fine structure at arbitrarily small scales, which makes it too irregular and complex to be described by traditional Euclidean geometry.[11] Such surfaces can be better described by a quantity known as FD, which is an indicator of the variation in the level of structural detail of an object as the object is viewed at different scales.[12] The FD of a surface can be any real number that is ≥ 2 and ≤ 3, and the greater its' FD the more structural complexity it possesses.

In this study, we applied the concept of FD as a practical tool to characterize the complexity of IMRT fluence fields. In practice, the FD measurement is distorted by the finite size and digitization of an image.[13] We began our investigation by evaluating the variation,[14] the variogram,[15] and the radial power spectrum[16] methods for calculating the fractal dimension of finite digital representations of mathematical fractal surfaces. All of these methods are based on examining changes in structural characteristics as a function of scale, and they were implemented in MATLAB® (MathWorks, Natick, MA).

II.A. Variation method

This method involves characterizing the way that a volume that covers a fractal surface changes as neighborhoods of size ε in the surface domain change. That is, the fractal surface $\{S(x,y); x,y \in [0,1]\}$ is divided into neighborhoods (uniform squares of side length ε in our implementation), and for the i^{th} square S_i, the minimum volume $v_i(\varepsilon) = \varepsilon^2 \times [S_{i,\max} - S_{i,\min}]$ covering S_i is computed, and the total volume covering $S(x,y)$ is given by $V(\varepsilon) = \sum_i v_i(\varepsilon)$. This process is repeated for different sizes of ε. The boundaries of the fractal surface are taken into account by adding the volume of the partial boxes to the total volume. $V(\varepsilon)$ and ε are related to the fractal dimension by the following equation:

$$FD = \lim_{\varepsilon \to 0} \left(3 - \frac{\log V(\varepsilon)}{\log \varepsilon} \right). \quad (1)$$

Assuming the log-log plot of $V(\varepsilon)$ and ε has an intercept of zero the FD is found from the slope (m) for small ε by

$$FD = 3 - m. \quad (2)$$

II.B. Power spectrum method

In this method, a fractal surface $S(x,y)$ is Fourier transformed by means of a fast Fourier transform (FFT) and the power spectrum is found by taking the modulus square of the FFT. Using the same approach as Veenland *et al.*,[16] we reduced the two-dimensional (2D) power spectrum to a one-dimensional (1D) power spectrum $P(f)$ by averaging the logarithm of the 2D power spectrum over concentric circles of logarithmically decreasing frequency size. The resulting 1D power spectrum has the form $P(f) \propto f^{-\beta}$, where β is the spectral exponent that is related to FD. To find β, we computed the slope of log $P(f) = |FFT(f)|^2$ versus log f ignoring the lowest and highest frequency components. For a fractal surface, FD is then found by the following equation:[17]

$$FD = \frac{(8-\beta)}{2}. \tag{3}$$

II.C. Variogram method

The variogram method is based on the statistical Gaussian modeling of images.[18] Given an image, the FD is estimated by assuming that it is derived from a fractional Brownian surface.[15] The FD of a profile is related to the semivariogram function $\gamma(h)$ used in spatial statistics, and it is defined as follows:[19]

$$\gamma(h) = \frac{1}{2N}\sum_{i=1}^{N}(F(x_i) - F(x_i+h))^2, \tag{4}$$

where N is the number of pairs of data points whose lag is h. $\gamma(h)$ is related to FD by a power law of the lag given by $\gamma(h) = ch^{4-2D}$ for some constant c.[19] Taking the log of both sides it is clear that FD is related to the slope m of a plot of $\log(\gamma(h))$ versus $\log(h)$ by the equation

$$FD = \frac{(4-m)}{2}. \tag{5}$$

This method was defined for profiles and was extended to surfaces (represented by a matrix of intensity values) by calculating the FD of each row and column, averaging them all together, and adding one. Due to the finite size of an image, the range of lag for the straight line appearing on the $\log(\gamma(h))$ versus $\log(h)$ is less than 10% of the profile length.[19]

II.D. Comparison of the fractal dimension methods

The scale ranges used to find the slopes from which the FD was computed for the different methods (Eqs. (2), (3), and (5)) is shown in Fig. 1. The methods were tested for accuracy and precision across a range of FDs and image sizes. More specifically, the performance of each FD method was evaluated with a set of digital image representations of mathematical fractal surfaces known as fractional Brownian surfaces.[12] These digital surfaces were generated to have known FDs ranging from 2.05 to 2.95 in steps of 0.05. At each of these steps, 80 surface images were generated for each image size (128×128, 256×256, and 512×512 pixels). These image sizes cover the range of the head and neck IMRT fluence field sizes used in this study: 129×195 to 301×319 pixels, with an average size of 226×261 pixels and pixel dimension of 0.786×0.786 mm^2.

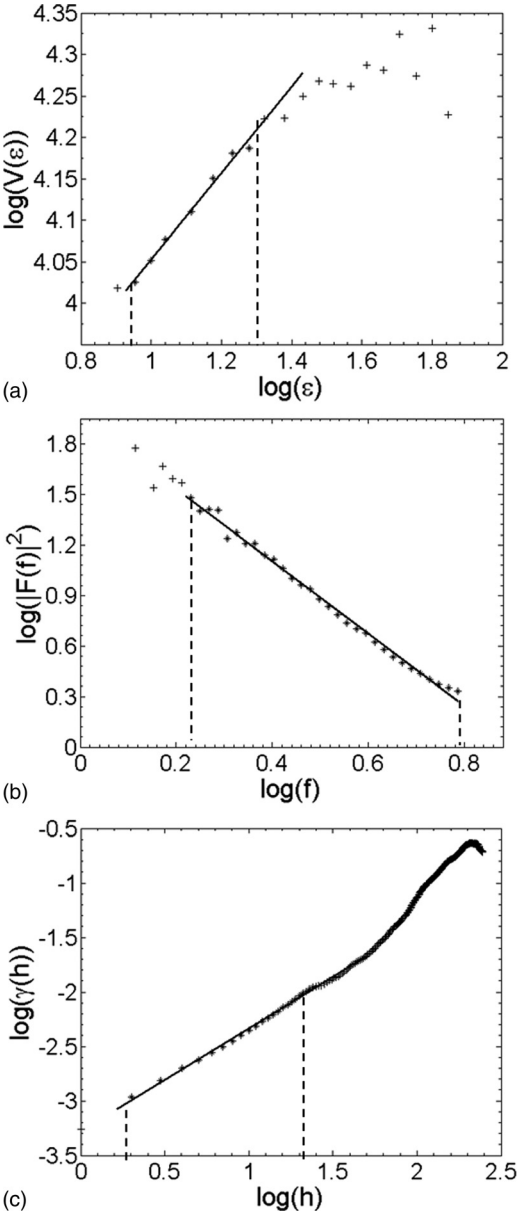

FIG. 1. Representative plots showing the scale ranges (x-axes) corresponding to the linear regions that were used to compute the slope of the graphs from which FD is computed using the (a) variation, (b) power spectrum, and (c) variogram methods. The scale ranges are in terms of the logarithm of the number of pixels for (a) and (c) and the logarithm of the pixel frequency for (b), where a pixel side length = 0.786 mm.

II.E. The 2D modulation index

The modulation index (MI) was first introduced for 1D beams by Webb[10] and extended to 2D by Giorgia *et al.*[1] The general idea of MI involves computing the magnitude of the difference between adjacent pixel intensity values and generating a spectrum of the number of differences that are less than some percentage of the standard deviation of the pixel intensities of the whole image. The spectrum is then normalized and integrated to give the MI.

We set σ to be the standard deviation of an image (I), n, and m the size of the image in the x and y directions, respectively, and

$$\Delta_x(i,j) = |I(i+1,j) - I(i,j)|$$
$$\Delta_y(i,j) = |I(i,j+1) - I(i,j)|$$
$$\Delta_{xy}(i,j) = |I(i+1,j+1) - I(i,j)|$$

for i from 1 to $(n-1)$ and j from 1 to $(m-1)$. Also, for f from 0 to 1 in steps of 0.01 the difference spectra $N_x(f)$, $N_y(f)$ and $N_{xy}(f)$ are defined as as the number of changes in pixel intensities Δ_x,Δ_y, and Δ_{xy}, respectively, that are greater than $f\sigma$. Hence, with these definitions the MI is expressed in terms of the integral of the average normalized difference spectra. That is,

$$2D\ \ MI = \int_0^1 z(f)df, \tag{6}$$

where

$$Z(f) = \frac{(Z_x(f) + Z_y(f) + Z_{xy}(f)}{3},$$

and

$$Z_x(f) = \frac{1}{(n-1)m} N_x(f),$$
$$Z_y(f) = \frac{1}{(m-1)n} N_y(f),$$
$$Z_{xy}(f) = \frac{1}{(n-1)(m-1)} N_{xy}(f).$$

II.F. Test and validation sets

All of the IMRT plans for the head and neck carcinoma patients used in this study were generated with a commercial treatment planning system (Eclipse™, Varian Medical Systems, Palo Alto, CA). The plans were chosen randomly from the set of nasopharynx and oropharynx cases at the authors' institution. The treatment fields were originally optimized using x and y fluence smoothing parameters of 55 and 45, respectively, and we defined these as moderately modulated. The IMRT actual fluencies of these moderately modulated fields passed patient-specific QA done with Portal Dosimetry (Varian Medical Systems, Palo Alto, CA) at the authors' institution, and were delivered clinically to treat the patients. This patient-specific QA was performed with the portal dosimeter at 100 cm from the source, and the passing criteria for each field was 95% of the points passing the gamma analysis based on a percent dose difference (PDD) and distance to agreement (DTA) of 3% and 3 mm, respectively. We defined the corresponding highly modulated fields as fields that satisfied the same clinical criteria as the moderately modulated fields in terms of PTV coverage and avoidance of organs at risk, as per Radiation Therapy Oncology Group (RTOG) protocols 0022 and 0225,[20,21] but with greater modulation. In order to produce highly modulated fluence fields, the moderately modulated plans were re-optimized using the same optimization objectives as the moderately modulated plans but had the fluence smoothing parameters set to 20 and 10 for the x and y directions. The smoothing parameters were used because they provide a controlled and objective way of modifying the degree of modulation. Figure 2 illustrates how the fluence smoothing parameters change the fluence modulation of a field. It is should be noted that the highly modulated plans did not exhibit a clinically significant improvement for either

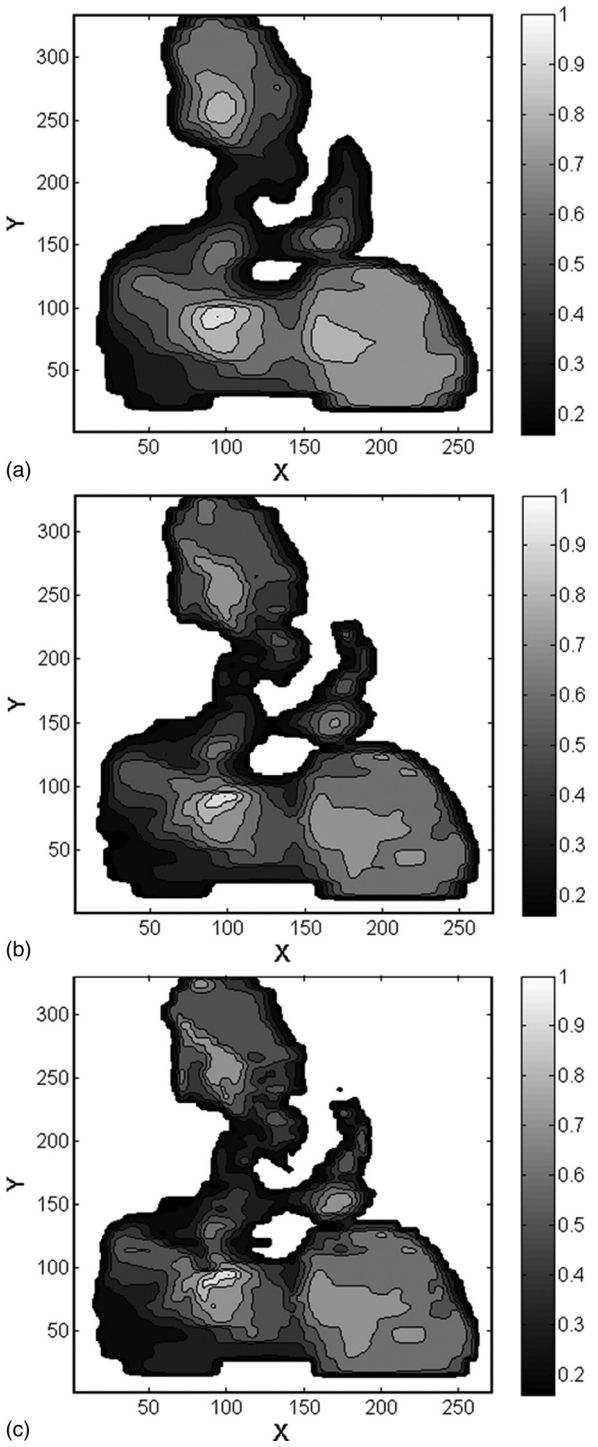

FIG. 2. Three representative contour plots of the same IMRT actual fluence field that was planned in Eclipse IMRT using x-y fluence smoothing of (a) 50-40, (b) 30-20, and (c) 20-10 to illustrate the change in modulation with the fluence smoothing parameters. The corresponding variogram fractal dimensions for the plots are: (a) 2.16, (b) 2.22, and (c) 2.27.

PTV coverage or sparing of organs at risk; however, the highly modulated plans required 50% more total MUs per plan on average than the moderately modulated plans. It is also worth noting that all of the treatment fields (moderate and highly modulated) were delivered at a dose rate of 300 MU/min.

The test set was composed of 50 moderate and 50 highly modulated IMRT actual fluence fields derived from seven patients. This set was used to compare the ability of FD, number of MUs, average leaf gap, and 2D MI to classify the level of modulation, and to determine a suitable cutoff value for FD. We also constructed an independent head and neck carcinoma validation set in the same way using four additional IMRT patient plans containing a total of 28 moderate and 28 high modulation fields to validate the performance of FD with the cutoff value that was determined from the seven patient test set.

The IMRT actual fluence fields were exported from Eclipse as standard digital imaging and communications in medicine (DICOM) files, and subsequently loaded into MATLAB® for analysis. Prior to the FD and 2D MI analyses all split fields were combined to give the number of whole fields stated above, i.e., the number of split field pairs for the test and validation sets. These whole fields were cropped from each direction to the first row or column that had an average of 2% of the maximum fluence intensity. This was done to effectively remove the noisy pixels surrounding the fluence field pattern. Also, the number of MUs and the average leaf gap obtained from the Eclipse™ treatment planning system were for the whole fields. That is, we added the number of MUs from a pair of split fields to give the total MUs for the whole field, and we computed a weighted average of the average leaf gaps for a pair of split fields. The weightings for the average leaf gap of a split field was calculated as the number of MUs given for the split field divided by the total MUs for the split field pair.

II.G. Comparison of FD, number of MUs, average leaf gap, and 2D MI

We applied ROC analysis[22] to the test set to compare the classification performance of FD, the number of MUs, average leaf gap, and the 2D MI as indicators of modulation complexity. We generated the sensitivity (fraction of true positives) and 1 – specificity (fraction of false positives) values for the ROC plots by varying the threshold values used to classify the fields as moderately or highly modulated. To measure classification performance, we computed the area under each ROC curve (AUC). Values of AUC can range from 0.5 (chance accuracy) to 1.0 (perfect accuracy), with the following intermediate benchmarks: 0.6 (poor), 0.7 (fair), 0.8 (good), 0.9 (excellent), >0.95 (almost perfect).

The test set was used to identify overly modulated IMRT fields (as defined by the $x = 20$, $y = 10$ fluence smoothing parameters) and ensure there are no false positives (i.e., 100% specificity). To do this, the FD cutoff was set to be greater than the maximum FD found for the moderately modulated fields. Although the choice of a cutoff point is somewhat arbitrary, the reason for choosing the cut-point in this way was to err on the side of having sufficient modulation to achieve the clinical objectives of the treatment plan. To assess classification performance, the cut-point chosen from the test set was applied to the validation set.

II.H. Dose measurements

We used the MapCHECK™ 2D diode detector array device (Sun Nuclear Corporation, Melbourne, FL) to perform 2D dose measurements for the 50 moderate and 50 high modulation fields comprising the test set. The measurements were performed at a 0° gantry and collimator setting. A $5 \times 30 \times 30$ cm^3 standard grade solid water slab (Gammex, Inc., Middleton, WI) was placed on top of the MapCHECK™ device to measure the planar dose at a water equivalent depth of 7 cm. Planar dose calculations were performed in the Eclipse™ treatment planning system using a dose matrix of 22×22 cm^2 with a spatial resolution of 0.2 cm (110×110 points). Measurements were compared to the planar dose calculations using the MapCHECK™ gamma analysis software tool (Version 3.04). For the gamma analysis, the PDD and DTA were set to (PDD, DTA) = (3%, 3 mm), (2%, 2 mm), and (1%, 1 mm), and a threshold of 10%. For the (3%, 3 mm) setting, the analysis was performed with and without the Van Dyke gamma criteria and measurement uncertainty, and for the (2%, 2 mm) and (1%, 1mm) gamma analysis the dose difference threshold was set to 2 cGy without the Van Dyke gamma criteria and measurement uncertainty.

III. RESULTS

III.A. Comparison of the different fractal dimension methods

Ideally, the fractal analysis method used to assess the complexity of an IMRT field should possess all of the following characteristics:

1. It should be independent of image size.
2. The computed FD should strictly increase as the theoretically known FD increases.
3. It should be precise.

We chose the variogram FD method to assess the complexity of the IMRT fields because of the three methods investigated with the fractional Brownian surfaces of known FD, the variogram method satisfied all three conditions to a greater degree than the other methods (Fig. 3). Figure 4 shows an example of the changes in the variogram FD for different levels of fluence smoothing for the 7 fields of a typical the head and neck case. It also shows that the fluence smoothing parameters effectively produce a systematic shift in the degree of field modulation as measured by the variogram FD method (cf. Fig. 2). That is, the greater the field modulation the higher the FD, as expected and the standard deviation in going from one level of modulation to the next for the data in Fig. 4 is only ~15%. It should be noted that some overlap in the FD values for the different levels of fluence smoothing (cf. Fig. 4) is expected, as the degree of modulation needed to achieve the clinical objectives defined

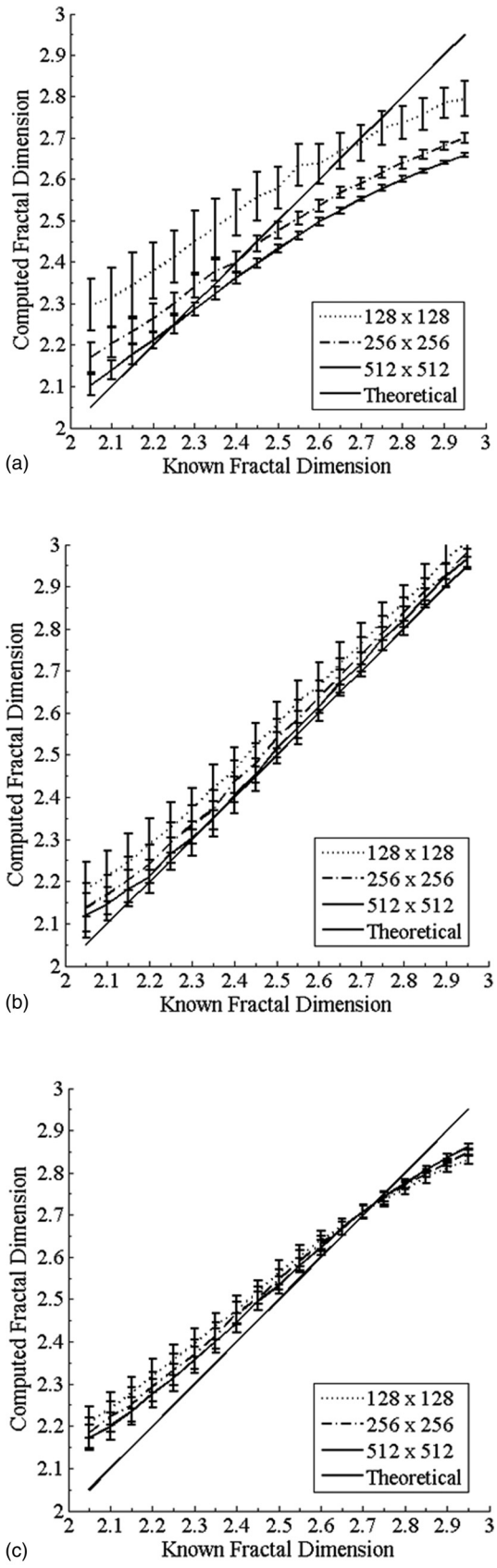

FIG. 3. Each point on the graphs represents the FD and its associated uncertainty as estimated from the standard deviation of 80 fractional Brownian surfaces of known fractal dimension. Results are shown for: (a) the variation method, (b) the power spectrum method, and (c) the variogram method.

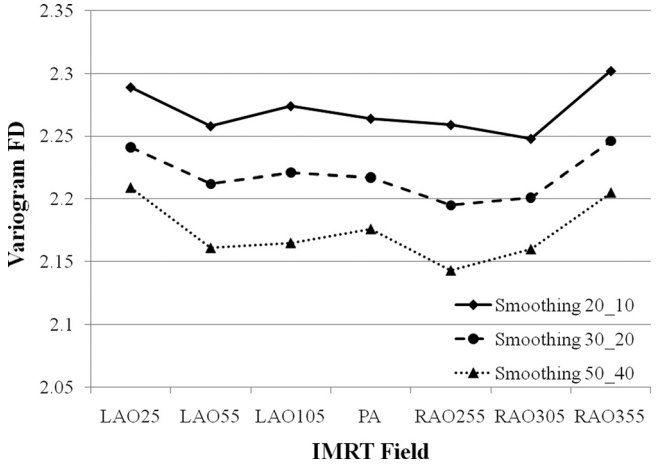

FIG. 4. Variogram FD analysis of 7 IMRT fields of the same head and neck case planned using three *x*-*y* fluence smoothing parameters. The LAO105 field corresponds to the contour images shown in Fig. 2.

by the RTOG protocols[20,21] also depends on the coupling between patient-specific attributes and the beam angle.

III.B. Comparison of FD, number of MUs, average leaf gap, and 2D MI

Figure 5 displays the distribution of the values of variogram FD, number of MUs, average leaf gap, and 2D MI for the moderate and high modulation IMRT fluence fields in the form of boxplots. It is obvious from Fig. 5 that FD is the only method of the three that has no overlap between the extreme points of the moderate and high modulation data and their inter-quartile ranges (IQR), shown as the regions within the boxes.

From the ROC analysis (Fig. 6 and Table I), we observe that the classification performance of the variogram FD is almost perfect, and the number of MUs, average leaf gap and 2D MI indicators are excellent. Hence, the FD indicator is the best of the three methods at discriminating the two modulation levels.

A minimal FD cut-point of $FD > 2.25$ was necessary to reduce the likelihood of false positives (i.e., achieve 100% specificity) in the identification of overly modulated IMRT fields (see Fig. 5). The corresponding sensitivity for this classification threshold is 80% in the test set. With this cut-point, the variogram FD analysis of the validation set correctly identified 26 of the 28 high modulation fields (92.8% sensitivity), and correctly identified all 28 of the moderately modulated fields (100% specificity), as expected.

III.C. Dose measurements

For each of the moderate and corresponding high modulation test fields, the pass rate for the different gamma analyses settings of the moderate modulation field was always greater than or equal to the corresponding high modulation field. The boxplots in Fig. 7 show that although there is overlap between the moderate and high modulation pass rates for different fields (due to patient-specific attributes), the distributions of pass rates are shifted lower for the high modulation (20-10

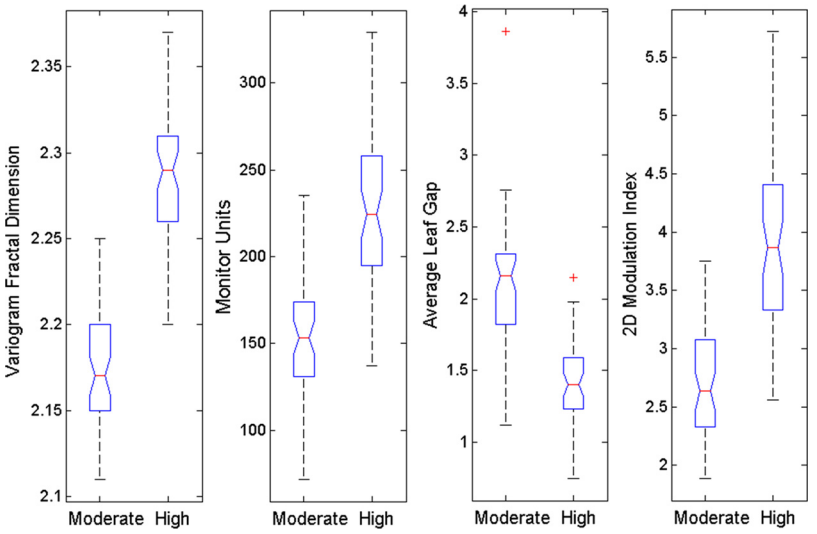

FIG. 5. Boxplot displays of the distributions of the variogram FD, number of MUs, average leaf gap, and 2D MI of the moderate and high modulation IMRT fields.

smoothing) group. Hence, Fig. 7 clearly demonstrates that MapCHECK™ planar dose measurements can detect differences in field modulation. It is also worth noting that the separation between the moderate and high modulation test groups is particularly evident for the (PDD, DTA) = (1%, 1 mm) and 2 cGy dose threshold criteria, in which the inter-quartile ranges between the two groups are well separated.

IV. DISCUSSION AND CONCLUSIONS

This study was motivated by the need to ensure an optimal balance between the level of modulation necessary for IMRT plans to achieve the clinical objectives and the level needed to avoid excessive modulation, which leads to higher doses to organs at risk, longer treatment times, increased stress on machine components,[1–3] and inaccuracies due to tongue-and-groove effects[4] and interplay effects[5] for moving targets.

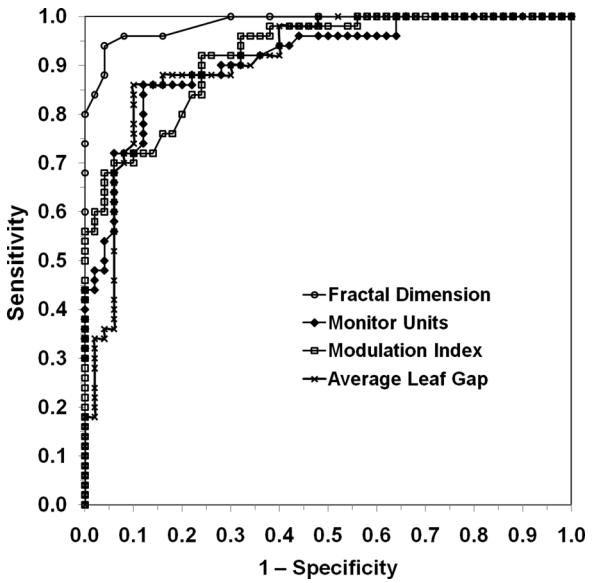

FIG. 6. ROC curves representing the classification performance of the variogram FD, number of MUs, average leaf gap, and 2D MI as indicators of moderate versus high modulation IMRT fields.

Also, for the sake of consistency it is desirable to generate plans that are comparable with regards to the degree of modulation for a given treatment site.[6] To this end, we investigated the ability of fractal analysis, a mathematical approach that lends itself naturally as a measure of complexity, to accurately quantify the level of modulation (spatial complexity) of IMRT fluence fields. We evaluated three fractal dimension analysis methods (variation, power spectrum, and variogram) and found the variogram method to be the most suitable (i.e., most precise and least sensitive to image size) for assessing the complexity of head and neck IMRT plans. Comparing the variogram FD method with three other previously used indicators of IMRT fluence field complexity, namely, the number of MUs, the average leaf gap, and the 2D MI, we found that the variogram FD is a better index for quantitatively identifying over-modulated IMRT fluence fields. This is clearly demonstrated by the ROC analysis, which showed the classification performance of the variogram FD to be almost perfect, and the number of MUs, average leaf gap, and 2D MI indicators to be excellent.

A possible reason that FD performed better than the 2D MI method is that FD is a measure of the variation in complexity across a range of spatial scales (c.f., Fig. 1), whereas the 2D MI method measures variations only for the small range defined by the size of adjacent image pixels that exceed a predefined change in intensity. In other words, this study suggests that it is better to have a metric for modulation that quantifies variations in modulation across a relevant range of spatial scales (or frequencies) and not just adjacent elements. The drawback for using the number of MUs and

TABLE I. AUC from the ROC analysis.

Method	AUC	Standard error
Variogram FD	0.99	0.01
MU	0.91	0.03
Average leaf gap	0.91	0.03
2D MI	0.92	0.03

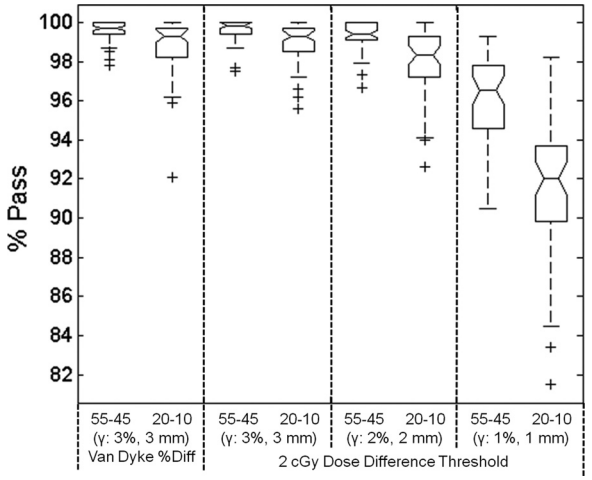

FIG. 7. Boxplot displays of the distributions of the gamma analysis of the planar dose measurements made with the MapCHECK™ 2D diode detector array device. The gamma analyses are shown for different settings of PDD, DTA, Van Dyke gamma criteria, and dose difference threshold.

the average leaf gap as indicators of modulation is that aside from modulation, they are also affected by the prescribed dose and patient-specific attributes such as the size and depth of the target volume. Hence, these confounding factors reduce the effectiveness of using the number of MUs and the average leaf gap as indicators of modulation.

It is worth noting that both the moderate and high modulation plans used in this study satisfied the RTOG 0022 and 0225 clinical plan objectives (verified by dose volume histograms), and they are typical head and neck carcinoma plans at the authors' institution. Hence, the moderate modulation fields defined by the 55 and 45 fluence smoothing parameters for the x- and y-directions, respectively, in Eclipse™ were both clinically adequate and avoided excessive complexity in modulation. However, this does not mean that these particular fluence smoothing parameters are necessary or sufficient to achieve the same degree of modulation for all head and neck treatment cases. At the authors' institution, the fluence smoothing parameters used to achieve the desired clinical objectives and degree of modulation for head and neck carcinoma plans typically range from ($x=50$, $y=40$) to ($x=60$, $y=50$) depending on patient-specific attributes such as the size of the patient, and the tumor type, size, and location relative to the organs at risk. Hence, in addition to the fluence smoothing parameters used in Eclipse™ to define and create the moderate and high modulation fields, the degree of modulation in a plan is also affected by the prescribed dose constraints dictated by patient-specific attributes.

The moderate and high modulation fluence fields were also compared to planar dose measurements made with MapCHECK™, and the gamma analyses showed that high modulation fields have lower pass rate distributions than moderately modulated fields (Fig. 7). However, the distributions in Fig. 7 show that if one uses the common clinical pass rate threshold of 95% and a 2 cGy dose passing threshold, then one would need to employ the 1% PDD and 1 mm DTA criteria to effectively detect over modulation and discriminate between the moderate and high modulation fields. Since both the moderate and high modulation fields satisfied the same RTOG clinical objectives, it is possible to use the MapCHECK™ planar dose measurements to detect and avoid the less desirable most highly modulated fields without compromising the clinical objectives. However, the advantage of using the variogram FD method to identify overly modulated fields is that it can be used to quickly and easily flag and modify high modulation fields before IMRT QA is performed giving medical physicists and dosimetrists more time to make the appropriate modifications to overly modulated fields.

In this study, we focused on assessing the level of modulation for head and neck carcinoma plans constructed using the dynamic IMRT approach of the Eclipse™ treatment planning system. Keeping in mind that the first aim of this study was to compare the most common fractal analysis methods to find the most suitable one, the reason we decided to focus on the head and neck site is because we wanted to investigate the potential of the most suitable fractal method using treatment fields that are generally more complex than other typical IMRT planning sites such as prostate or breast. The greater complexity of the head and neck IMRT plans is due to the many organs at risk concentrated in the region, which generally leads to higher field modulation than is commonly observed at other treatment sites.[3,7] This more extreme variation in modulation was useful for deciding on a reasonable upper FD classification cut-point that can be used as a guideline for deciding on whether a treatment field is overly modulated and whether further reduction in modulation is achievable. For example, if the modulation of treatment fields for less complex sites such as the prostate bed exceed the FD cut-point chosen for head and neck fields, then it would be obvious that the treatment plan is not optimal, and that a further reduction in field modulation is likely achievable without compromising the clinical objectives and constraints. Having said this, an area of future work will be to find FD cut-points for other common IMRT treatment sites, and investigate the relationship between exceeding these cut-points and dosimetric deliverability as assessed by patient-specific quality control of the IMRT plans.

Although this study was performed using dynamic IMRT fields, fractal analysis of IMRT fields generated using the static (step and shoot) approach can also be performed. The only parameter that may change is the FD cutoff point (i.e., $FD > 2.25$) used to ensure high specificity in distinguishing the moderate and high modulation fields. This is because the static method's finite number of discreet beam segments may lead to greater intensity variations that result in a systematic shift toward slightly higher modulation.

Since the general rule of thumb for IMRT practice is to avoid excessive complexity, the aim of this study was to develop a quantitative QA tool for assessing plan complexity that is easy to implement, rapid, and precise. The tool we have developed has direct clinical implications for workflow because it can be used to quantitatively assess and classify the degree of modulation using an FD threshold (*e.g.*, $FD > 2.25$ for head and neck IMRT) to identify overly modulated planned IMRT fluence fields before an IMRT QA

is performed. Hence, the feedback received from the fractal analysis can be used to try to modify an overly modulated field before it is measured, and ensure that the degree of modulation in treatment plans is more consistent within and across institutions, and remains consistent with new planning system upgrades. Furthermore, an FD threshold could also aid in setting treatment planning guidelines for dose objectives for the targets and organs at risk, and parameters such as fluence smoothing in Eclipse™ that reduce the risk of producing overly modulated fields for a given treatment site. Finally, the variogram FD method may also be incorporated as a cost function in an IMRT optimization algorithm to help avoid the pitfalls of excessive modulation.

[a]Author to whom correspondence should be addressed. Electronic mail: mtambasc@ucalgary.ca. Telephone: (403) 521-3749; Fax: (403) 521-3327.

[1]N. Giorgia, F. Antonella, V. Eugenio, C. Alessandro, A. Filippo, and C. Luca, "What is an acceptably smoothed fluence? Dosimetric and delivery considerations for dynamic sliding window IMRT," Radiat. Oncol. **2**, 42 (2007).
[2]S. Pandya and J. Burmeister, "Effect of fluence smoothing on plan quality and delivery accuracy in intensity modulated radiotherapy [Abstract]," Med. Phys. **35**, 2755 (2008).
[3]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).
[4]J. S. Li, T. Lin, L. Chen, R. A. Price, Jr., and C. M. Ma, "Uncertainties in IMRT dosimetry," Med. Phys. **37**, 2491–2500 (2010).
[5]L. E. Court, M. Wagar, D. Ionascu, R. Berbeco, and L. Chin, "Management of the interplay effect when using dynamic MLC sequences to treat moving targets," Med. Phys. **35**, 1926–1931 (2008).
[6]G. A. Ezzell, J. W. Burmeister, N. Dogan, T. J. LoSasso, J. G. Mechalakos, D. Mihailidis, A. Molineu, J. R. Palta, C. R. Ramsey, B. J. Salter, J. Shi, P. Xia, N. J. Yue, and Y. Xiao, "IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119," Med. Phys. **36**, 5359–5373 (2009).
[7]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).
[8]J. J. Kruse, "On the insensitivity of single field planar dosimetry to IMRT inaccuracies," Med. Phys. **37**, 2516–2524 (2010).
[9]L. Dong, J. Antolak, M. Salehpour, K. Forster, L. O'Neill, R. Kendall, and I. Rosen, "Patient-specific point dose measurement for IMRT monitor unit verification," Int. J. Radiat. Oncol. Biol. Phys. **56**, 867–877 (2003).
[10]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).
[11]K. J. Falconer, "*Fractal Geometry Mathematical Foundations and Applications* (Wiley, Chichester, England, 2003).
[12]H. O. Peitgen, D. Saupe, and M. F. Barnsley, *The Science of Fractal Images* (Springer-Verlag, New York, 1988).
[13]Q. Huang, R. Lorch, and R. Dubes, "Can the fractal dimension of images be measured," Pattern Recognit. **27**, 339–349 (1994).
[14]B. Dubuc, S. W. Zucker, C. Ticot, J. F. Quiniou, and D. Wehbi, "Evaluating the fractal dimension of surfaces," Proc. R. Soc. London A **425**, 113–127 (1989).
[15]M. F. Goodchild, "Fractals and the accuracy of geographical measures," Math. Geol. **12**, 85–98 (1980).
[16]J. F. Veenland, J. L. Grashius, M. F. van der, A. L. Beckers, and E. S. Gelsema, "Estimation of fractal dimension in radiographs," Med. Phys. **23**, 585–594 (1996).
[17]M. J. Turner, J. M. Blackledge, and P. R. Andrews, *Fractal Geometry in Digital Imaging* (Academic Press, San Diego, 1998).
[18]R. Lopes and N. Betrouni, "Fractal and multifractal analysis: a review," Med. Image Anal. **13**, 634–649 (2009).
[19]S. Murata and T. Saito, "The variogram method for a fractal model of a rock joint surface," Geotech. Geologic. Eng. **17**, 197–210 (1999).
[20]RTOG Protocol 0022, Phase I/II study of conformal and intensity modulated irradiation for oropharyngeal cancer [Electronic Citation], http://208.251.169.72/members/protocols/h0022/h0022.pdf.
[21]RTOG Protocol 0225, A phase II study of intensity modulated radiation therapy (imrt) +/− chemotherapy for nasopharyngeal cancer [Electronic Citation], http://208.251.169.72/members/protocols/0225/0225.pdf.
[22]C. E. Metz, "Basic principles of ROC analysis," Semin. Nucl. Med. **8**, 283–298, (1978).