Manuscript Number: EJMP-D-16-00472R1

Title: Challenges in calculation of the gamma index in radiotherapy -
towards good practice

Article Type: Review article

Keywords: Gamma index; IMRT; VMAT; Quality Assurance; detector arrays;
MATLAB

Corresponding Author: Dr. Mohammad Hussein, PhD

Corresponding Author's Institution: Royal Surrey County Hospital NHS
Foundation Trust

First Author: Mohammad Hussein, PhD

Order of Authors: Mohammad Hussein, PhD; Catharine H Clark, PhD; Andrew
Nisbet, PhD

Abstract: The gamma index (γ) is one of the most commonly used metrics for the verification of complex modulated radiotherapy. The mathematical definition of the γ is computationally expensive and various techniques have been reported to speed up the calculation either by mathematically refining the γ or employing various computational techniques. These techniques can cause variation in output with different software implementations. The γ has traditionally been used to compare a 2D measured plane against a 2D or 3D dose distribution. Recently, software algorithm and hardware improvements have led to the possibility of using measured 2D data from commercial detector arrays to reconstruct a 3D-dose distribution and perform a volumetric comparison against the treatment planning system (TPS). A limitation in this approach is that commercial detector arrays have so far been limited by their spatial resolution which may affect the accuracy of the reconstructed 3D volume and subsequently the γ calculation. Additionally, 2D versus 3D γ comparison adds a layer of complication in the calculation of the γ given the increase in the number of calculation points and the result cannot be as easily interpreted in the same way as 2D comparison. This review summarises and highlights the computational challenges of the γ calculation and sheds light on some of these issues by means of a bespoke MATLAB software to demonstrate the impact of interpolation, γ search distance, resolution and 2D and 3D calculations. Finally, a recommendation is made on the minimum information that should be reported when publishing γ results.

Royal Surrey County Hospital
NHS Foundation Trust

Dr Mohammad Hussein
Department of Medical Physics
St. Luke's Wing
Royal Surrey County Hospital
Guildford
Surrey
GU2 7XX
Tel: +44 1483 406660
Email: m.hussein@nhs.net

03 January 2017

Professor Paolo Russo
Universita' di Napoli Federico II
Dipartimento di Fisica
Compl. univ. Monte S. Angelo
Via Cintia
I-80126 Naples
Italy

Title: 'Challenges in calculation of the gamma index in radiotherapy – towards good practice'

Dear Professor Russo,

We have revised our above paper based on the feedback we received from the reviewers. We have carried out major revisions according to the recommendations. The changes made are highlighted in blue in the manuscript and we have attached an itemised response to the reviewer comments.

Yours sincerely,

Dr Mohammad Hussein, Dr Catharine Clark and Prof Andrew Nisbet

Response to reviewer comments

The authors wish to thank the reviewers and editors for their helpful and constructive comments. We have highlighted major changes within the manuscript in blue. An itemized response to specific comments is provided below.

Reviewers' comments:

AE:

This is a well-written review of critical aspects associated with using the gamma index in radiotherapy. Though largely discussed in the past, the topic is of interest to readers of Physica Medica who are likely to benefit from this review. However, the referees raised several questions with which I concur. I would encourage the authors to submit a revision that addresses the issues raised in the review process before this manuscript is considered for publication.

Thank you for your comments, we have addressed the issues raised.

Reviewer #1:

The authors discuss the pitfalls and good practice in gamma index calculation. This is an interesting topic: the gamma index is widely used, yet sometimes a bit indiscriminately - often, details about the settings (global/local, threshold dose, etc.) are lacking in the scientific literature, and it is important to draw attention to these and other critical issues. A number of additional - less widely known - issues are addressed by calculations using an in-house software and some commercial systems are also evaluated. I find this paper very useful and interesting and would like to see it published. Although the structure does not adhere to the typical standards (methods, results etc.), I find it adequate and suitable for the present purpose, the text itself is very well written and enjoyable to read.

Thank you for your constructive comments. We have made changes as you have recommended in the specific comments.

I have one suggestion:

With the Octavius 4D, I have the impression that the best agreement is always found for the coronal, worst for the sagittal plane. This can be explained if the depth dose curve is not quite correct or the extrapolation is slightly off. It might be interesting to include this issue, if it can be reproduced.

Thank you for this interesting suggestion. We haven't seen this and attempted to investigate this issue but could not reproduce it. For the purpose of this review we looked at the 3D volume gamma from the Octavius4D. The coronal measurement was using the OctaviusII octagonal phantom, not the coronal plane reconstruction in the Octavius4D. Further investigation at this time is slightly outside the scope of this review, there are a couple of potentially confounding factors that need to be considered to prove this. The number of points being evaluated would need to be controlled somehow. The dose gradients in the coronal plane would also need to be controlled in both planes. Possibly this could be achieved by creating a number of symmetric test volumes where the surface area seen in the coronal plane is the same as in the sagittal plane.

Some minor points:

Abstract, l.27: shedS light on
Corrected, thank you.

p.8 l.22: 2D plane vs 3D volume (leave out gamma analysis here as you do in the bullet before and after this one)

Done as suggested, thank you.

p.11, heading: Comparison against two (lower case) Commercial (capital) Software - I guess software is also plural, but most non-native readers will think this is strange - why not say "Software Systems" or "Implementation"?

Corrected as suggested and have gone with "Software Systems", thank you.

Reviewer #3:

General comments:

The gamma index is commonly used for specific QA patient. The authors present the advantages and the disadvantages of this metrics. This review is of interest for the community as they give some recommendations for good practice. However I think this review could be improved before publication because some parameter studies are incomplete.

Thank you for your constructive comments. We have made changes as you have recommended in the specific comments.

Detailed comments:

Page 3, line 16: Several point measurements are needed to derive an 1D measurement. In the context, I guess the authors mean point measurement instead of 1D.

Corrected to say "single point (e.g. ionization chamber measurement, 1D (e.g. line profile), ..."

Page 3, line 26: Please add (rr) and (re) to define these quantities.

These have now been defined.

Page 4, line 11: I suggest to define the dose difference earlier (page 3, line 35).

This has been defined earlier as suggested.

Figure 1: I think it would be clearer to name differently the three DE points (DE1, rE1); (DE2, rE2); (DE3, rE3). And <GAMMA> could be expressed for those points. In addition <DELTA>D and DTA should be defined in the caption. And it is not clear if they represent the criteria or the value for points DE(rR), rR and DE(rE1).

Page 4, line 26: The numbering could be 2.1 Formalism of the gamma index (page 3, line 24) and this section would become 2.2 Page 4, line 58: This paragraph should not be in the global and local section. I suggest to move this paragraph page 3, line 8.

Modified as suggested.

Page 5, line 49: In the second level, the authors wrote that for the 2 points in the vicinity of the reference point, Δr should be of opposite signs. Δr is a distance and is always positive. I think this is an error: ΔD should be of opposite signs.

Thank you, this mistake has been corrected.

Page 7, line 10: I would appreciate the authors refer the following paper:

Blanpain B, Mercier D. The delta envelope: a technique for dose distribution comparison. Med Phys. 2009 Mar;36(3):797-808.

This method is interesting and should be discussed in this review, here or in section 6.

We have added a short discussion as requested.

Page 7, lines 14-47: This paragraph is quite long and I am not convinced it provides additional useful information for the review. It could be shortened.

We agree and this has been significantly shortened.

Page 7, line 52: Ref 50 does not exist, I guess it should be 5.

Thank you, it should be 5. We used Mendeley for referencing and this appears to have been an error with the citation for the paper by Wendling et al. All references have been carefully re-checked, however.

Page 9, line 19: Do the authors present results for the 2.5 or 5mm interpolated calculation grids? Please, specify at least in the table 1 caption. ? Why do you use a 1mm DTA? Compared to the pixel spacing, this choice is not recommended (as you state later in the review). In addition, why do you choose gamma analysis instead of dose difference to compare two calculated dose distributions?

The results are presented for the 2.5mm grid. The reference to the 5mm grid is a typing error.

We agree that dose difference would be appropriate; however we found the gamma index passing rate was the easiest metric in comparing the difference between the different interpolation techniques. For dose difference, we found using something like the mean difference or median difference was not sensitive and we needed to rely on visual inspection of the maps. We agree that the use of 1mm DTA contradicts later recommendations; we re-run the analysis with 0.1mm instead (this made negligible difference to the results).

Page 10, line 1: Please, in equation 7, define $x_l = x_0/2l$.

Added as suggested

Page 10, line 10: Equation 8 is now $\Delta r/x_l$.

Agree, this has been corrected.

Page 10, line 20: Ref 51 does not exist. Again I guess it should be 5.

Corrected, same issue as mentioned above.

Figure 4: Again what is the evaluated dose distribution? It looks like the 3%/1mm was plotted using $x_0 = 2.5\text{mm}$ and the 3%/3mm using 5mm. Please precise at least in the caption.

To clarify x_0 was 2.5mm in all cases, so the ratio distance criterion/pixel spacing (according to equation 8) will be different for 3mm and 1mm distance criterion.

To make this figure clearer we have updated the caption. We hope that this has clarified this point for the reviewer. **Please note that this is now Figure 6 in the revision.**

Page 10, line 55: The was copied. A word, probably "plan" is missing.

Correct, the word "plan" has now been added

Page 10, line 58: would exists. I think it should be written "would exist".

Corrected, thank you.

Figure 6: What is the grid resolution? What is the impact of grid resolution?

The evaluated grid resolution was 2.5mm, however also added an additional graph side-by-side showing the similar trend if the evaluated distribution was 1.25mm resolution. Please note, this is now Figure 5 in the revision.

Page 12, line 15: This comparison could be illustrated with results (for example % rates with or without interpolation for each solution: in-house and commercial).

The results for this are indeed illustrated in Figure 8.

Section 4: The study is performed for a single plan (data interpolation, search distance, comparison). Why do the authors use only one plan? A larger number of plans could have been included to state results.

We agree with the reviewer that a single case is insufficient. The example we had was a fairly extreme error in a head & neck plan, we decided to also test a simple prostate plan (with very minor realistic errors included). We have found very consistent results between these two very contrasting cases (see updated Figure 4, 5, 6 and 8), for this reason we argue that these examples are sufficient for the purpose of the review.

Page 13, line 25: What are the deliberate errors for those two plans?

Description has been added in the test. These were single leaf MLC errors of 1, 2 and 5mm.

Page 14, line 24: I am surprised by this statement. I would prefer suggest to use a detector with a good spatial resolution! Please, comment.

This line wasn't recommending a specific spatial resolution. Rather it was pointing out that if one has to use a detector array to reconstruct a 3D distribution then it should be done using the spacing of the detector array to avoid introducing sampling issues by trying to reconstruct a finer resolution distribution. We have added a line suggesting detector arrays with finer resolution would be more ideal but also being pragmatic in highlighting that such systems are not widely available, currently have limited field size and fairly expensive.

Page 15, line 45: disadvantages instead of disadvantage.

Corrected, thank you.

EDITOR'S COMMENTS

We will publish this good paper in EJMP as a review, after the major revision required by the Associate Editor: the two reviewers indicate some points where a better description is needed.

Highlights

- Topical review of the issues affecting gamma index computation
- Demonstrates the factors affecting 2D and 3D gamma index computation using an independent software
- Recommendations for good practice in the calculation and reporting of the gamma index

Challenges in calculation of the gamma index in radiotherapy – towards good practice

Running head: Challenges in calculation of the gamma index

M Hussein^{1,2*}, CH Clark^{1,2,3}, A Nisbet^{1,2}

¹Department of Medical Physics, Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK

²Centre for Nuclear and Radiation Physics, University of Surrey, Guildford, UK

³National Physical Laboratory, Teddington, UK.

***Corresponding author:**

Dr Mohammad Hussein

Department of Medical Physics,

St Luke’s Cancer Centre, Royal Surrey County Hospital, Egerton Road, Guildford, Surrey, UK, GU2 7XX

E-mail: mo1310@gmail.com / m.hussein@nhs.net

Tel: +44 1483 406660

Keywords: Gamma index, IMRT, VMAT, Quality Assurance, detector arrays, MATLAB

Abstract

The gamma index (γ) is one of the most commonly used metrics for the verification of complex modulated radiotherapy. The mathematical definition of the γ is computationally expensive and various techniques have been reported to speed up the calculation either by mathematically refining the γ or employing various computational techniques. These techniques can cause variation in output with different software implementations. The γ has traditionally been used to compare a 2D measured plane against a 2D or 3D dose distribution. Recently, software algorithm and hardware improvements have led to the possibility of using measured 2D data from commercial detector arrays to reconstruct a 3D-dose distribution and perform a volumetric comparison against the treatment planning system (TPS). A limitation in this approach is that commercial detector arrays have so far been limited by their spatial resolution which may affect the accuracy of the reconstructed 3D volume and subsequently the γ calculation. Additionally, 3D versus 3D γ comparison adds a layer of complication in the calculation of the γ given the increase in the number of calculation points and the result cannot be as easily interpreted in the same way as 2D comparison. This review summarises and highlights the computational challenges of the γ calculation and sheds light on some of these issues by means of a bespoke MATLAB software to demonstrate the impact of interpolation, γ search distance, resolution and 2D and 3D calculations. Finally, a recommendation is made on the minimum information that should be reported when publishing γ results.

1 Introduction

The gamma index (γ) is one of the most commonly used metrics for the verification of complex radiotherapy deliveries such as intensity modulated radiotherapy (IMRT) and volumetric modulated arc radiotherapy (VMAT) [1]. The metric has been widely accepted and is implemented into most commercial verification analysis software. By combining dose difference and distance-to-agreement, the γ provides the means for an efficient analysis which is particularly important within a busy clinical environment [2,3]. Its popularity can be seen in the number of times that it has been used in the scientific peer reviewed literature. In the Elsevier Scopus citation database it was found that the original γ paper [1] has been cited 1088 times in the literature since it was published, as of January 2017. Of these, there were 978 original research articles; the remainder were composed of 81 conference proceedings, 20 review papers and the remainder as book chapters, letters, or Editorials.

The mathematical definition of the γ is computationally expensive and a full calculation can take a significant amount of time depending on the number of data points and the processing speed of the computer being used [4,5]. This has led to computational challenges where there have been various reports in the literature focused on either mathematically refining the γ or employing various computational techniques to speed up the process [4–8]. These various techniques can potentially cause variation in output with different software implementations. Often the exact technique employed to calculate the γ in commercial software is not well defined, with manufacturers typically referencing the original paper by Low et al [1], but the implementation having subtle variations.

The γ has traditionally been used to compare a 2D measured plane against a 2D or 3D dose distribution. There have been quasi-3D commercial systems available [2,9–11]; however these have not constructed a true 3-dimensional dose distribution. Recently, software algorithm and hardware improvements have led to the possibility of using measured 2D data from commercial detector arrays to reconstruct a 3D-dose distribution and perform a volumetric comparison against the treatment planning system (TPS). A limitation in this approach is that commercial detector arrays have so far been limited by their spatial resolution which may affect the accuracy of the reconstructed 3D volume and subsequently the γ calculation due to under-sampling [12]. Additionally 3D versus 3D γ comparison adds an extra layer of complication in the calculation of the γ given the increase in the number of calculation points and therefore the limited speed of calculation and the result cannot be as easily interpreted in the same way as a 2D comparison.

This review article seeks to summarise and highlight the computational challenges of the gamma index calculation and shed light on some of these issues by means of bespoke in-house written MATLAB software to demonstrate the impact of interpolation, gamma index search distance, resolution and 2D and 3D calculations.

2 Definition of the gamma index

The gamma index combines dose difference and distance difference to calculate a dimensionless metric for each point in the evaluated distribution. The **reference dose distribution** is generally taken as the 'gold standard', e.g. it could be the dose distribution that has been measured. In theory the distribution could be a single point (e.g. ionisation chamber measurement), 1D (e.g. a line profile), 2D (e.g. film measurement) or 3D (e.g. gel dosimetry, Monte Carlo simulation). The **evaluated dose distribution** is what is being compared. In most cases this will be the predicted TPS dose distribution that is being checked for accuracy in modelling the delivered dose.

The most common passing criteria used is 3%/3mm which was originally recommended in the work by Low et al [1]. The γ was originally developed to compare measured water tank beam data against a treatment planning system algorithm. The criteria of 3%/3mm were used due to the limitations of TPS algorithms at the time, where particularly penumbra modelling was a source of uncertainty [1]. Because the γ takes into account dose difference and distance difference it was well-suited to the modulated fields in IMRT, however the criteria of 3%/3mm has persisted.

2.1 Formalism of the gamma index

The γ is calculated based on finding the minimum Euclidean distance for each reference point, see Figure 1 in conjunction with the following description. For each reference point in the dose distribution, calculate against each point in the evaluated distribution:

1. the distance between reference to evaluated point: $\Delta r(\mathbf{r}_R, \mathbf{r}_E)$
2. the difference between the reference and evaluated dose; $\Delta D(\mathbf{r}_R, \mathbf{r}_E)$

Where $\mathbf{r}_R$ is the reference point, $\mathbf{r}_E$ is the evaluated point. The dose difference is calculated using Equation 1:

$$\Delta D(r_R, r_E) = D_E(r_E) - D_R(r_R) \quad (1)$$

Where $D_E(r_E)$ is the dose at a point in the evaluated dose distribution, r_E , and $D_R(r_R)$ is reference point dose.

Then for each point in the evaluated distribution, calculate the γ using Equation 2:

$$\Gamma(\mathbf{r}_R, \mathbf{r}_E) = \sqrt{\frac{\Delta r^2(\mathbf{r}_R, \mathbf{r}_E)}{\delta r^2} + \frac{\Delta D^2(\mathbf{r}_R, \mathbf{r}_E)}{\delta D^2}} \quad (2)$$

Where δr is the distance difference criterion and δD is the dose difference criterion.

The γ is then taken as the minimum value calculated over all evaluated points as shown in Equation 3:

$$\gamma(\mathbf{r}_R) = \min\{\Gamma(\mathbf{r}_R, \mathbf{r}_E)\} \forall \{\mathbf{r}_E\} \quad (3)$$

The δr and δD criterion form an ellipsoid around the reference point as shown in Figure 1. If an evaluated point is located within this then the reference point will pass since γ will be < 1 .

For nomenclature it is standard to report the passing criteria in the format $\delta D(\%)/\delta r(mm)$; e.g. 3%/3mm. This standard nomenclature is used throughout this review. In order to eliminate dose in the out-of-field region where a large relative dose difference can be calculated and hence skew the γ result, it is typical to set a lower dose threshold below which the γ result is ignored. Therefore, it is common to limit the γ calculation to all points that are ≥ 10 -20% of the maximum dose value within the dose distribution [13].

FIGURE 1.

2.2 Global & local γ calculations

Typically the γ calculations are categorised into two different types; local and global. The contrast between the two types is the way the dose difference is calculated. For a local γ , Equation 1 gives the definition for a local dose difference. For global gamma, Equation 1 has to be modified to become Equation 4:

$$\Delta D(r_R, r_E) = \frac{D_E(r_E) - D_R(r_R)}{D_{norm}} \quad (4)$$

Where D_{norm} is a normalisation dose value which can be defined as any value; for example, as the maximum dose within the reference dose distribution or a point selected in a high dose low gradient region. The two types of γ have advantages and disadvantages. The local γ will tend to highlight failures in high dose gradient regions and in low dose regions, whereas the global γ will tend to mask these errors but show the errors within the higher dose regions within the dose distribution. The choice of the γ calculation will depend on the needs of the test. The majority of published works within the reference list report global γ , although care should be taken in interpreting results, as it is often not defined which is used.

3 The computing challenge of the gamma index

The γ is a computationally expensive process due to the need to search all points in the evaluated distribution. This becomes more complex when comparing two 3D dose distributions. Ideally the γ would be calculated quickly to give a result within a reasonable time. The computer hardware used will have an impact on the speed of the γ calculation. Given previous limitations with computer technology, a number of studies in the literature have focussed on ways to mathematically decrease the calculation time of the γ . A common feature with manufacturer manuals is there is seldom detailed information on the way the γ calculation has been performed; with often the ‘go-to’ solution being to simply cite the original paper by Low et al. A true γ calculation can take from a few minutes up to potentially days for complex 3D dose distributions, which is clearly unacceptable from a clinical point of view [7].

3.1 Mathematical techniques to refine/speed up the gamma index calculation

A refinement of the γ was proposed by Depuydt et al [6]. In their work, the authors introduced a filter cascade of multiple levels that were designed to speed up the comparison. The focus of this algorithm is whether or not a point passes or fails the γ criteria rather than calculating an absolute γ value for each point in the reference distribution. The algorithm employs 3 levels in the calculation. In the first level, the minimum γ is searched for within a limited search distance to reduce the calculation time. As soon as an evaluated point is found where $\Gamma(\mathbf{r}_R, \mathbf{r}_E) < 1$ then the calculation is stopped and the algorithm moves on to the next reference point, starting at Level 1. If there are no

evaluated points where $\Gamma(\mathbf{r}_R, \mathbf{r}_E) < 1$ then the algorithm moves on to Level 2. In the second level, the algorithm searches for at least 2 evaluated points in the vicinity of the reference point where the $\Delta D(\mathbf{r}_R, \mathbf{r}_E)$ for the two points is of opposite signs. In this scenario it is assumed that the evaluated distribution must intersect the region defined by the passing criteria and therefore the reference point is classified as passed. Failing this, the algorithm moves to the third and final level. In most cases, reference points rejected in Level 2 are because the evaluated points are truly outside the pseudo-space defined by the passing criteria. However, there may be rare occasions when the reference points should not have failed because of the discrete nature of the evaluated distribution. This is because it is possible for two discrete evaluated points at the outside edge of the passing region to produce a failed result, but interpolating between them means that there is an intersection with the passing region. Therefore Level 3 is designed to take these possibilities into account. If the reference point fails level 3 then it is classified as having failed the test.

Chen et al [7] consider the possibility to speed up the search distance algorithm by using a fast Euclidean distance transform and predict a speedup of the order of tens of thousands for 3D γ calculations. Bakai et al [4] published a revision of the original γ formalism which considered gradient-dependent local acceptance thresholds. The actual number calculated by the Bakai method is called the χ index and is defined using Equation 5:

$$\chi = \frac{D_E(r_E) - D_R(r_R)}{\sqrt{\delta r^2 + \delta D^2 \cdot \nabla D_R(r_R)}} \quad (5)$$

In this method, one begins by calculating the local gradient for each point in the reference distribution, $\nabla D_R(r_R)$, to build a gradient map or cube. Thus the comparison is performed in the dose domain. The numerator part of equation 5 means that this method only works when the reference and the evaluated dose distributions have the same array sizes. This formalism means that a χ distribution can be calculated efficiently using simple matrix operations which are optimised for speed in numerical analysis coding software such as MATLAB, or could be programmed in other languages using existing libraries and utilising multiple threads in the computer processing unit (CPU). The χ index retains the negative or positive sign, unlike the γ . The value of $|\chi|$ is also approximately equivalent to the γ [4].

Another similar technique to the χ index is the delta envelope proposed by Blanpain and Mercier [14]. The delta envelope is based on the union of the γ index ellipsoids centred on each reference

point. The envelope is effectively a maximum and minimum dose difference tolerance at each reference point. The 'delta index', δ , is calculated for each evaluated point checking if the dose difference to the reference point is within the max-min tolerance given by the δ envelope. Similar to the χ index, the δ can take a value of -1 to +1. Because the delta envelope method reduces the analysis to a dose difference, the calculation is potentially quicker; 10s compared to 2 minutes γ index for 3D dose distributions [14]. One advantage of the delta envelope method is it avoids the uncertainty introduced by the resolution of the evaluated dose distribution where the gamma index can report different results whether this is interpolated or not (as discussed in more detail in section 4).

Wendling et al [5] developed a fast algorithm through speeding up the search routine by pre-sorting the distance from the reference point to evaluated points within a fixed search circle or sphere for 2D and 3D respectively. The theory is that the search loop should be stopped when there is a low chance that evaluated data points will reach the minimal Γ . In their approach, the calculation starts at the reference point and increases outwards, terminating when the condition in Equation 6 is met [5]:

$$\frac{\Delta r^2(\mathbf{r}_R, \mathbf{r}_E)}{\delta r^2} > \min(\Gamma(\mathbf{r}_R, \mathbf{r}_E)) \quad (6)$$

The potential limitation of this technique is that there is potential for overestimation of the γ when dose differences are very large within the search region and then sharply drop off just outside this region. If an exhaustive search was performed in this scenario it would likely find the minimum γ in this region where the calculation has stopped. Ju et al [8] proposed a re-interpretation of the γ to avoid the need to interpolate the evaluated distribution to a finer grid, thus reducing the calculation time. For 1D, 2D, and 3D distributions the evaluated distribution is divided into line segments, triangles and tetrahedral respectively. The closest distance between any reference point and these simplexes is calculated using matrix multiplication and inversion. The finding is that the method is as accurate as 16 times finer linear interpolation of the evaluated dose distribution with an order of ~20 times speedup in calculation time.

3.2 Future trends in computational techniques

As computer hardware continues to progress, the speed of the γ calculation will naturally continue to become quicker. In the last few years there have been innovative approaches to utilising the processing power provided by graphical processing units (GPU) which are commercially available with multiple cores that are in the region of 1000+. This allows for significant potential for parallel computing in areas such as Monte Carlo calculations or in this case, fast γ calculations. For the γ , it would be possible to utilise the GPU to perform the otherwise computer intensive task of the minimum distance search by searching and calculating γ for each reference in parallel. Gu et al [15] studied accelerated γ calculations by combining the geometric technique proposed by Ju et al [8] and the pre-sorting technique mentioned above by Wendling et al [5] and by implementing the calculation onto the GPU. They found a 45 – 70 times speedup of the calculation compared to the traditional implementation on the CPU. Persoon et al [16] found a speedup of 57 ± 15 for patient cases when using the GPU against the CPU.

4 Evaluation of the software effect on the gamma index calculation using a Bespoke MATLAB® Software

In order to investigate the different approaches that can be used to calculate γ , a software tool was written and implemented in MATLAB v2012a – 2014a (Mathworks Inc.). MATLAB uses some in-built functions for manipulation of matrices that can simplify the coding of the γ algorithm, but require the two compared datasets to have the same matrix size. This software was written so that it can accept two datasets with different resolution and matrix size, and by default is set up to do no interpolations on the reference dataset. This tool was implemented as a graphical user interface (GUI) for user-friendliness as shown in Figure 2. Some of the current features of the software are:

- Able to handle 2D or 3D DICOM dose distributions
- Datasets do not have to have the same resolution/matrix size as each other
- Able to perform γ analysis in 2D plane vs 2D plane, 2D plane vs 3D volume, and 3D volume vs 3D volume
- Specify whether to search the whole evaluated distribution, or limit the search to a user-defined distance from each reference point
- Perform global or local γ calculation, with ability to set dose difference and distance criterion.
- Allows the user to set lower and upper dose thresholds (a yellow outline is given to visualise this; can be used e.g. to focus on a high dose region)
- Allows the user to specify interpolation factors for either dataset 1 or dataset 2

- Allows the calculation of different γ metrics; % of points passing with $\gamma < 1$, mean γ , median γ , maximum γ , or the minimum γ in the top X% pixels (e.g. minimum γ in the top 1% pixels; $\gamma_{1\%}$)

FIGURE 2.

By default, the dataset 1 was designated as the evaluated distribution, and dataset 2 was designated as the reference distribution. In the software, the in-built 2D and 3D interpolation functions (called *interp2* and *interp3* respectively) in MATLAB were used. The software was designed so that it is possible to interpolate either dataset 1 or 2 for investigative purposes (dataset 2 was always defaulted to have no interpolation unless the user specified otherwise).

4.1 Which type of interpolation to use for the evaluated distribution?

The interpolation options available in MATLAB were to use (1) a linear interpolation, (2) nearest neighbour, (3) cubic method or (4) spline method. In order to test the different interpolation techniques available, a double 360° arc Head & Neck RapidArc™ treatment plan was calculated in the Varian Eclipse Treatment Planning System (TPS) using the Analytical Anisotropic Algorithm (AAA) v11 with a dose grid spacing of 1.25mm. It was then exported as a 2D DICOM plane with 1.25mm and 2.5mm. The 2.5mm grid calculation was then interpolated to a grid size of 1.25mm using the different interpolation techniques and compared directly against the 1.25mm dose calculation using the global γ and 1% / 0.1mm passing criteria. The % points passing with $\gamma < 1$, and mean γ were quantified as well as visual inspection of the gamma map for each interpolation technique. No lower threshold was used for the γ calculation.

Figure 3 shows the calculated gamma index maps for the four different interpolation techniques and Table 1 shows the % points passing $\gamma < 1$ and mean γ results. It was found that the spline algorithm gave the best agreement to the non-interpolated data. For the purposes of this software the spline algorithm was therefore deemed to be optimal.

FIGURE 3.

TABLE 1.

4.2 Impact of limiting the search distance on calculation time

The gamma index calculation is a time consuming computational process. For example, performing a full gamma index comparison between two planes with a grid spacing of 2.5mm and 81 x 81 points

on a PC desktop with a quad-core Intel i7 4GHz CPU, and 16GB of RAM took ~380s to complete. A simple way to speed up the γ calculation significantly is to limit the search in the evaluated dose distribution to a set distance around each reference point. An interesting observation made in the study by Wendling et al [5] is that by setting a limited maximum search distance, it is only necessary to calculate the distance between a reference point and all the evaluated points bound by the search distance once; this can be defined as a 2D or 3D array, $\mathbf{R}$. Similarly, it would then be possible to calculate the dose difference between the reference point and all the evaluated points which can be defined as an array $\mathbf{D}$ which has the same size as $\mathbf{R}$. This significantly reduces computing overhead and makes it possible to perform matrix operations which are optimised for speed in programmes such as MATLAB by calculating all elements using parallelisation.

To evaluate the impact of using a limited search distance, two very different plan scenarios were used. The first was to take the same Head & Neck RapidArc treatment plan described above was used with a dose grid of 2.5mm. The plan was copied and a collimator angle error of +5 degrees was introduced to the two arcs, such that a large range of failed and passing points with varying levels of dose gradient would exist, as shown in Figure 4. The second case was a simple prostate plan with subtle errors introduced to slightly disrupt the dose distribution. These were achieved by applying a gantry sag to the MLC positions in DICOM plan file, as described Carver et al [17], using Equation 7.

$$MLC_{mod} = MLC_{orig} + A \sin(\theta) \quad (7)$$

Where MLC_{mod} is the modified MLC position, MLC_{orig} is the original position, and A is the specified maximum MLC position change (in this case we used 5mm), and θ is gantry angle. [17]. This led to realistic small deviations as shown in Figure 4.

The error and normal plans were exported in DICOM format, as 2D planes with 81 x 81 points at 2.5mm spacing, to be evaluated against each other in the MATLAB software.

FIGURE 4.

For global γ passing criteria of 3%/3mm, 3%/2mm, 3%/1mm, the search distance was set in increments until Equation 8 reached at least 5:

$$\psi = \frac{L}{\delta r}$$

Where L was the search distance. As a baseline, full γ comparisons were calculated. The most informative parameter is the maximum calculated γ value as this will be affected by limited search distances. The calculation time was also recorded for each permutation. As δr is varied, the physical search distance will have a different impact depending on the value chosen for δr . Therefore the maximum γ was plotted against ψ ; see Figure 5. In graph (a), it can be seen that for different passing criteria there was a consistent trend towards the maximum γ having no variation once ψ became ≥ 3.0 . This was consistent for both the simple plan and the extreme case. At this ratio, the maximum γ was the same as that where the entire evaluated distribution was searched. Even up to a higher $\psi = 5$, the calculation time was still small at ~ 0.2 s. This was performed for the evaluation distribution having a grid spacing of 2.5mm. To check if the trend is consistent for different resolution, the evaluated distribution of the prostate plan was re-calculated at 1.25mm spacing and the analysis repeated for 3%/2mm. As can be seen in Figure 5 (b) the trend is consistent for different evaluated dose resolution.

FIGURE 5.

4.3 The impact of the extent of data Interpolation

A fundamental issue with the γ calculation is that the result will be influenced by the data point spacing of the evaluated dose distribution. There can be inaccuracies when the pixel spacing is equivalent to the distance criterion which can lead to overestimation of the γ . Previously Low *et. al.* [1] recommended that the pixel spacing should be $\frac{1}{3} \delta r$. More recently, Wendling *et. al.* concluded that the spacing should be of the order of $\frac{1}{10} \delta r$ [5]. In order to achieve smaller point spacing, it is necessary to interpolate the evaluated dose distribution to a finer grid size.

After confirming the spline algorithm for interpolation and speeding up the calculation by limiting the search distance, it was necessary to benchmark how much interpolation is needed. The same plans described in 4.2 were employed again. The normal plan was again set as the evaluated distribution and varying levels of interpolation were applied to this dataset. In MATLAB an integer interpolation factor, I , is specified. This changes the original spacing, x_0 to x_I according to Equation 9:

$$x_I = \frac{x_0}{2^I} \quad (9)$$

Where x_0 was 2.5mm. For passing criteria of 3%/1mm, 3%/2mm, 3%/3mm, I was varied from 0 to 7 in increments of 1. For each passing criteria the ratio was calculated according to Equation 10:

$$\lambda = \frac{\delta r}{x_I} \quad (10)$$

For each calculation, the mean γ was calculated and the time taken for the calculation was recorded. The results of this are shown in Figure 6 where it can clearly be seen that the mean gamma varies sharply when the ratio of δr and x_0 are close to 1 and begins to stabilise by the time the ratio has become 10. This is in keeping with the recommendations by Wendling et al [5]. Up to this ratio, the calculation time was only of the order of 1 second and therefore as the default for further evaluations using the software the interpolation factor was set so λ was ≥ 10 .

FIGURE 6.

4.4 Comparison against two Commercial Software Systems

The bespoke software was tested against two other established commercially available software; being the PTW Verisoft v5.1 package and IBA OmniPro v7.0. The same combination of the normal and error head & neck RapidArc plan was used. In order to test the software, the error plan was calculated and exported as 2.5mm, 5mm and 10mm dose grids to compare against the 2.5mm grid normal plan. Global γ comparisons were made using 3%/2mm passing criteria and no lower dose threshold. The mean gamma and % passing rates were compared. Figure 7 shows γ maps for the 10mm grid spacing error plan against the normal plan.

FIGURE 7.

The top two images are from OmniPro and Verisoft respectively. It is clear that the OmniPro software prefers to plot the γ map as an intensity image where each pixel is given a discrete colour based on the γ . The Verisoft software uses a colour contour approach which is visually easier to

interpret. The two styles were replicated in MATLAB and the maps are given for the bespoke software below the respective commercial software maps. The figure shows good visual agreement in the γ maps between the bespoke software and the commercial packages. The comparison between the γ passing rate in the MATLAB software, OmniPro and Verisoft is shown in Figure 8. This shows very good agreement between the MATLAB and Verisoft calculations for all the different grid spacing of the error plan.

FIGURE 8.

Interestingly, if the MATLAB software is re-run with no interpolation setting, there is excellent agreement against OmniPro. This implies that Verisoft interpolates the evaluated dose distribution into a finer grid spacing, whereas OmniPro does not do this. This is a good example of how there can be variability in the implementation of the γ calculation. The above demonstration has focussed purely on the software side, however most commercial systems are designed to take into account the associated detector array configuration and therefore a more robust analysis would include an evaluation of the combined hardware and software. This aspect has been reported by Hussein et al where it has been shown that different commercial systems can give different gamma index results [2]. Agnew and McGarry have developed a tool to allow the quality assurance of gamma index software [18].

5 Gamma index calculation in 2D and 3D

Traditionally the γ has been used to compare a 2D measured plane against a 2D or 3D dose distribution. There is a growing trend in using 3D back-projected EPID based in-vivo systems and independent calculations using Monte Carlo [19] to perform 3D versus 3D γ comparison. DVH-based analysis techniques have been proposed, however the γ continues to be prevalent. Recently, software algorithm and hardware improvements have led to the possibility of using measured 2D data from commercial detector arrays to reconstruct a 3D-dose distribution and perform a volumetric comparison against the TPS. There are some issues that need to be considered when moving to a 3D volume γ . As there are more degrees of freedom and more data points, it is possible that 3D volume γ could give systematically better passing rates than 2D, and therefore the passing criteria need to be carefully reconsidered and whether they should be tightened. Secondly, if using a detector array to reconstruct a 3D dose distribution and compare against the TPS, there are potential issues by the limited spatial resolution which may affect the accuracy of the reconstructed

3D volume depending on the type of interpolation algorithm and subsequently the γ calculation [12].

5.1 Experimental demonstration

These issues are demonstrated in an experiment conducted using the PTW OCTAVIUS® 4D , which is a commercial system that can use 2D measured data to reconstruct a 3D dose cube with 2.5mm grid spacing [20–22]. Measurements were performed using the OCTAVIUS® Detector 729 and the high resolution OCTAVIUS® Detector 1000^{SRS} inserted into the OCTAVIUS 4D. The Detector 729 consists of a matrix of 729 cubic vented ionization chambers with 5 mm x 5 mm cross-section, spaced 10 mm centre-to-centre, giving a total area of 270 mm x 270 mm [23]. The Detector 1000^{SRS} consists of a matrix of 977 liquid-filled ion chambers with 2.3 mm x 2.3 mm x 0.5 mm volume and 2.5 mm centre-to-centre detector spacing in the central 55 mm x 55 mm area and 5 mm in the outer 110 mm x 110 mm area. Traditional 2D coronal plane measurements were also separately performed using the OCTAVIUS Detector 729 within the OCTAVIUS II phantom for comparison. VERISOFT software version 6.0 was used for all measurement acquisitions and analysis.

Deliberate error tests for a prostate cancer 5-field IMRT and a single 360 RapidArc case were used following the same methodology as described by Hussein et al [2,24]. Both cases had single leaf MLC errors of 1, 2 and 5mm errors, respectively, added to the plan. In each case the plan with error was compared against the original unperturbed plan. For the 1000^{SRS} array, the 3D reconstructed dose cylinder is limited to having 110 mm as its length and 55 mm radius. In order to maintain a consistent analysis, the same cylindrical volume was reconstructed using the Detector 729 measurement. For the 2D coronal plane measurements using the OCTAVIUS 729 in the OCTAVIUS II phantom, the region of interest for the analysis was set to 110 mm x 110 mm.

Predicted 2D and 3D γ pass rates were simulated in the bespoke MATLAB software by exporting the normal plan and error plan predicted dose distributions in DICOM format. The same cylinder and plane dimensions, above, were applied to the simulated 3D and 2D dose distributions respectively. For all evaluations, the global γ was used with passing criteria of 3%/2mm, and a 20% threshold. The γ results based on measurements were compared against the 3D volume predicted analysis. The concordance correlation coefficient, ρ_c , was used to assess agreement between the analyses.

5.2 Experimental analysis

The impact of the 2D and 3D γ calculation can be significant depending on the spacing of detectors within an array and the way that it is used. A summary of mean percentage of points passing with $\gamma < 1$ is given in Table 2. There was good agreement between the predicted γ for the 2D coronal plane and predicted 3D volume analysis ($\rho_c > 0.90$). There was statistically strong agreement between the 3D 1000^{SRS} measured pass-rate and the 3D predicted pass-rate ($\rho_c = 0.93$). In the case of a planar measurement using the sparsely arranged 2D-ARRAY 729 in the OCTAVIUS II (as the originally intended way for this array to be used for measurement) the γ passing rate was in reasonable agreement with the predicted 3D volumetric passing rate ($\rho_c = 0.81$). However when the 2D-ARRAY 729 was used in the OCTAVIUS 4D phantom to reconstruct a 3D dose distribution it gave the worst overall agreement at $\rho_c = 0.34$; the γ distributions in Figure 9 show that this is due to bands of failed points caused by artefacts. The reason for these is given Figure 10, which shows a comparison of 2D dose profiles (plotted as discrete points) for the RapidArc prostate normal plan measurement from the 3D reconstructed dose distributions using the 729 and 1000^{SRS} array. The profile from the TPS predicted dose distribution is plotted for comparison. It is shown that, for the 729 array, there has been linear interpolation between points due to the low resolution which resulted in the dose points in those regions appearing to have an artificially lower dose than the TPS dose. This problem disappears when using the higher resolution 1000^{SRS} detector array where the dose reconstruction is more accurate, leading to a statistically stronger agreement. Interestingly, reconstructing the Detector 729 measurement into a 10mm grid spacing dose distribution improved the agreement. This would suggest that a 3D distribution should be reconstructed at the native resolution of the detector array, rather than interpolating which can introduce sampling uncertainties. Ideally a finer resolution detector array, like the 1000^{SRS}, should be used but there are currently limitations such as field size limitations, lack of wide availability, and cost of equipment. This experiment demonstrates that sparse detector arrays may increase measurement uncertainty when used to reconstruct 3D dose distributions and subsequently lead to some uncertainty with the gamma index calculation. This effect has also been reported by Xing et al [25] who compared the 3D γ between the OCTAVIUS 4D and Sun Nuclear 3DVH™ and found average passing rate differences of -1.3% (max of 11.7%) due to the different interpolations by the two systems [25].

TABLE 2.

FIGURE 9.

FIGURE 10.

6 Further considerations of the gamma index computation

1 It is common to report the results of a γ analysis as the percentage of points that achieved $\gamma < 1$; i.e.
2 the γ passing rate. The main characteristic of this metric is that it can condense a verification
3 measurement into a single value; this is both the advantage and pitfall of this metric. If implemented
4 carefully, it can be used to streamline QA by making it possible to choose decision thresholds for a
5 passing rate, thereby reducing the analysis time [26–30]. The disadvantage is that the passing rate
6 does not provide any details of where failed points are. Another disadvantage is that the γ itself is
7 inherently an absolute metric, i.e. it provides no information on whether a failed point is due to
8 positive or negative dose or distance fluctuations. For example, it is possible for there to be a failed
9 point in an OAR region where the measurement is lower than the predicted dose by more than the
10 dose difference criterion. In this case, the failed point is clinically acceptable as the OAR is receiving a
11 lower than expected dose and the aim of radiotherapy is to keep dose to an OAR region as low as
12 possible. Conversely, it is possible to have a failed point in a PTV region where the measured dose is
13 higher which may also be acceptable.

24 The use of the γ has been shown to have a weak correlation against clinically relevant metrics [31–
25 34]. However, recently Sumida et al [35] proposed a radiobiological γ which integrates tumour
26 control probability (TCP) and normal tissue complication probability (NTCP). Essentially this works by
27 perturbing a planned dose distribution at a voxel-by-voxel basis using a relative dose map measured
28 by a detector array. A 3D γ is calculated on the normal versus perturbed dose distribution and those
29 voxels with a $\gamma > 1$ are weighted based on NTCP or TCP depending on whether they are in a target or
30 normal tissue structure to give a radiobiological γ . The aim is that the weighting can help avoid the
31 scenario above whereby the dose difference can be systematically at the level of the dose difference
32 criterion but still pass.

41 For the same passing criteria, different devices and software combinations exhibit varying levels of
42 agreement with the MATLAB predicted γ analysis. As shown in this study, passing criteria of say
43 3%/3mm may not give the same results for measurements by different QA systems [2]. Crowe et al
44 report a method for calculating different tolerance thresholds for different passing criteria by
45 comparing the sensitivity of different passing criteria against 3%/3mm [36]. It is also important to be
46 aware that between different versions of the same software subtle changes to the gamma index
47 calculation can cause a difference in the result and users should take care to re-perform
48 benchmarking of the gamma index when any new software upgrade is released.

57 The two types of γ calculation reported have advantages and disadvantages. The local γ will tend to
58 highlight failures in high dose gradient regions and in low dose regions, whereas the global γ will

tend to mask these errors but show the errors within the higher dose regions within the dose distribution. The choice of the type of γ will depend on the need of the test. Most published works within the reference list report global γ . As Stojadinovic et al [37] point out, there is an infinite number of choices for the normalisation points for a global comparison whereas this choice is removed with a local comparison. Hussein et al previously investigated the impact of normalisation point choice by comparing analysis for a range of prostate and head & neck plans with deliberately introduced changes. They investigated the impact of selecting a point in a uniform region, maximum dose, or the mean dose within the PTV region and found no significant difference in the global γ passing rate, hence confirming the suitability of the global γ [24].

The suitability of the γ evaluation method in detecting clinically significant deviations has previously been questioned [31,38,39] and alternatives have been suggested [31,40–43]. However, the γ has been widely accepted and is implemented into most commercial software. The γ provides the means for an efficient analysis which is particularly important within a busy clinical environment [3]. It has also been used effectively within dosimetry audits of complex radiotherapy [44–48]. If one is performing a retrospective analysis of patient-specific QA in order to streamline the process, the γ provides a suitable means to explore trends over a period of time [49,50].

7 Conclusions

A common factor in most modern commercial radiotherapy verification systems is the reliance on the γ method to provide the quantitative evaluation of the measured dose distribution against the TPS calculated dose distribution. The mathematical definition of the γ is straightforward; however it has presented computing challenges in terms of the speed of the calculation of the metric. The different computational approaches that are possible can produce variability in the calculation of the γ between different software. Two of these parameters are whether or not to interpolate the evaluated dose distribution such that the pixel spacing is sufficiently less than the distance criterion to avoid uncertainties in the gamma calculation, and whether to search the entire evaluated distribution or limit the search. Furthermore the impact of the 2D and 3D γ can be significant depending on the spacing of detectors within a commercial detector array and the way that it is used. Given that various factors affect the gamma index calculation, as shown in this review, it is recommended that authors should report the following in published literature; the type of normalisation used (for example: to the maximum dose, a user selected point in a high dose low gradient region), whether the analysis was absolute or relative, whether local or global calculation

has been used, what (if any) low dose cut-off has been used, passing criteria as X%/Xmm, and the software used and version.

Acknowledgements

The authors would like to thank PTW (Freiburg, Germany) for kindly loaning the PTW OCTAVIUS 4D and 1000^{SRS} detector array.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys* 1998;25:656–61.
- [2] Hussein M, Rowshanfarzad P, Ebert MA, Nisbet A, Clark CH. A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems. *Radiother Oncol* 2013;109:370–6.
- [3] Van de Werf E, Lievens Y, Verstraete J, Pauwels K, Van den Bogaert W. Time and motion study of radiotherapy delivery: Economic burden of increased quality assurance and IMRT. *Radiother Oncol* 2009;93:137–40.
- [4] Bakai A, Alber M, Nüsslin F. A revision of the γ -evaluation concept for the comparison of dose distributions. *Phys Med Biol* 2003;48:3543.
- [5] Wendling M, Zijp LJ, McDermott LN, Smit EJ, Sonke J-J, Mijnheer BJ, et al. A fast algorithm for gamma evaluation in 3D. *Med Phys* 2007;34:1647.
- [6] Depuydt T, Van Esch A, Huyskens DP. A quantitative evaluation of IMRT dose distributions: refinement and clinical assessment of the gamma evaluation. *Radiother Oncol* 2002;62:309–19.
- [7] Chen M, Lu W, Chen Q, Ruchala K, Olivera G. Efficient gamma index calculation using fast Euclidean distance transform. *Phys Med Biol* 2009;54:2037.
- [8] Ju T, Simpson T, Deasy JO, Low DA. Geometric interpretation of the γ dose distribution comparison technique: Interpolation-free calculation. *Med Phys* 2008;35:879.
- [9] Li G, Zhang Y, Jiang X, Bai S, Peng G, Wu K, et al. Evaluation of the ArcCHECK QA system for IMRT and VMAT verification. *Phys Medica* 2013;29:295–303.
- [10] Bedford JL, Lee YK, Wai P, South CP, Warrington AP. Evaluation of the Delta4 phantom for IMRT and VMAT verification. *Phys Med Biol* 2009;54:N167-76.

- [11] Feygelman V, Opp D, Javedan K, Saini AJ, Zhang G. Evaluation of a 3D diode array dosimeter for helical tomotherapy delivery QA. *Med Dosim* 2010;35:324–9.
- [12] Bailey DW, Nelms BE, Attwood K, Kumaraswamy L, Podgorsak MB. Statistical variability and confidence intervals for planar dose QA pass rates. *Med Phys* 2011;38:6053–64.
- [13] Abolaban F, Zaman S, Cashmore J, Nisbet A, Clark CH. Changes in Patterns of Intensity-modulated Radiotherapy Verification and Quality Assurance in the UK. *Clin Oncol (R Coll Radiol)* 2016;28:28–34.
- [14] Blanpain B, Mercier D. The delta envelope: a technique for dose distribution comparison. *Med Phys* 2009;36:797–808.
- [15] Gu X, Jia X, Jiang SB. GPU-based fast gamma index calculation. *Phys Med Biol* 2011;56:1431–41.
- [16] Persoon LCGG, Podesta M, van Elmpt WJC, Nijsten SMJJG, Verhaegen F. A fast three-dimensional gamma evaluation using a GPU utilizing texture memory for on-the-fly interpolations. *Med Phys* 2011;38:4032.
- [17] Carver A, Gilmore M, Riley S, Uzan J, Mayles P. An analytical approach to acceptance criteria for quality assurance of Intensity Modulated Radiotherapy. *Radiother Oncol* 2011;100:453–5.
- [18] Agnew CE, McGarry CK. A tool to include gamma analysis software into a quality assurance program. *Radiother Oncol* 2016;118:568–73.
- [19] Boylan CJ, Aitkenhead AH, Rowbottom CG, Mackay RI. Simulation of realistic linac motion improves the accuracy of a Monte Carlo based VMAT plan QA system. *Radiother Oncol* 2013;109:377–83.
- [20] McGarry CK, O’Connell BF, Grattan MWD, Agnew CE, Irvine DM, Hounsell AR. Octavius 4D characterization for flattened and flattening filter free rotational deliveries. *Med Phys* 2013;40:91707.
- [21] Stathakis S, Myers P, Esquivel C, Mavroidis P, Papanikolaou N. Characterization of a novel 2D array dosimeter for patient-specific quality assurance with volumetric arc therapy. *Med Phys* 2013;40:71731.
- [22] Allgaier B, Schüle E, Würfel J. Dose reconstruction in the OCTAVIUS 4D phantom and in the patient without using dose information from the TPS. *PTW White Pap* 2013;D913.200.0:0–7.
- [23] Poppe B, Blehschmidt A, Djouguela A, Kollhoff R, Rubach A, Willborn KC, et al. Two-dimensional ionization chamber arrays for IMRT plan verification. *Med Phys* 2006;33:1005–15.
- [24] Hussein M, Adams EJ, Jordan TJ, Clark CH, Nisbet A. A critical evaluation of the PTW 2D-Array seven29 and Octavius II phantom for IMRT and VMAT verification. *J Appl Clin Med Phys* 2013;14:4460.
- [25] Xing A, Arumugam S, Deshpande S, George A, Vial P, Holloway L, et al. Evaluation of 3D Gamma index calculation implemented in two commercial dosimetry systems. *J Phys Conf Ser* 2015;573:12054.
- [26] Park JI, Park JM, Kim J-I, Park S-Y, Ye S-J. Gamma-index method sensitivity for gauging plan delivery accuracy of volumetric modulated arc therapy. *Phys Medica* 2015;31:1118–22.
- [27] Vieilleveigne L, Molinier J, Brun T, Ferrand R. Gamma index comparison of three VMAT QA systems and evaluation of their sensitivity to delivery errors. *Phys Medica* 2015;31:720–5.
- [28] Sdrolia A, Brownsword KM, Marsden JE, Alty KT, Moore CS, Beavis AW. Retrospective review

- of locally set tolerances for VMAT prostate patient specific QA using the COMPASS® system. *Phys Medica* 2015;31:792–7.
- [29] Rinaldin G, Perna L, Agnello G, Pallazzi G, Cattaneo GM, Fiorino C, et al. Quality assurance of Rapid Arc treatments: Performances and pre-clinical verifications of a planar detector (MapCHECK2). *Phys Medica* 2014;30:184–90.
 - [30] Heo T, Ye S-J, Carlson J, Park JM. The effect of beam interruption during VMAT delivery on the delivered dose distribution. *Phys Medica* 2015;31:297–300.
 - [31] Zhen H, Nelms BE, Tome W a. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys* 2011;38:5477–89.
 - [32] Stasi M, Giordanengo S, Cirio R, Boriani A, Bourhaleb F, Cornelius I, et al. D-IMRT verification with a 2D pixel ionization chamber: dosimetric and clinical results in head and neck cancer. *Phys Med Biol* 2005;50:4681–94.
 - [33] Jin X, Yan H, Han C, Zhou Y, Yi J, Xie C. Correlation between gamma index passing rate and clinical dosimetric difference for pre-treatment 2D and 3D volumetric modulated arc therapy dosimetric verification. *Br J Radiol* 2015;88:20140577.
 - [34] Fredh A, Scherman JB, Fog LS, Rosenschold PM af. Patient QA systems for rotational radiation therapy: A comparative experimental study with intentional errors. *Med Phys* 2013;40:31716–9.
 - [35] Sumida I, Yamaguchi H, Kizaki H, Aboshi K, Tsujii M, Yoshikawa N, et al. Novel radiobiological gamma index for the evaluation of three-dimensional predicted dose distribution. *Int J Radiat Oncol Biol Phys* 2015;92:779–86.
 - [36] Crowe SB, Sutherland B, Wilks R, Seshadri V, Sylvander S, Trapp J V, et al. Technical Note: Relationships between gamma criteria and action levels: Results of a multicenter audit of gamma agreement index results. *Med Phys* 2016;43:1501.
 - [37] Stojadinovic S, Ouyang L, Gu X, Pompoš A, Bao Q, Solberg TD. Breaking bad IMRT QA practice. *J Appl Clin Med Phys* 2015;16:154–65.
 - [38] Nelms BE, Zhen H, Tomé W a. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys* 2011;38:1037.
 - [39] Cozzolino M, Oliviero C, Califano G, Clemente S, Pedicini P, Caivano R, et al. Clinically relevant quality assurance (QA) for prostate RapidArc plans: Gamma maps and DVH-based evaluation. *Phys Medica* 2014;30:462–72.
 - [40] Agazaryan N, Solberg TD, DeMarco JJ. Patient specific quality assurance for the delivery of intensity modulated radiotherapy. *J Appl Clin Med Phys* 2003;4:40–50.
 - [41] Hu W, Graff P, Boettger T, Pouliot J. A spatially encoded dose difference maximal intensity projection map for patient dose evaluation: A new first line patient quality assurance tool. *Med Phys* 2011;38:1748.
 - [42] Li H, Dong L, Zhang L, Yang JN, Gillin MT, Zhu XR. Toward a better understanding of the gamma index: Investigation of parameters with a surface-based distance method. *Med Phys* 2011;38:6730–41.
 - [43] Park S-Y, Park JM, Kim J-I, Kim H, Kim IH, Ye S-J. Textural feature calculated from segmental fluences as a modulation index for VMAT. *Phys Medica* 2015;31:981–90.
 - [44] Clark CH, Hansen VN, Chantler H, Edwards C, James H V, Webster G, et al. Dosimetry audit for a multi-centre IMRT head and neck trial. *Radiother Oncol* 2009;93:102–8.

- [45] Budgell G, Berresford J, Trainer M, Bradshaw E, Sharpe P, Williams P. A national dosimetric audit of IMRT. *Radiother Oncol* 2011;99:246–52.
- [46] Hussein M, Tsang Y, Thomas R, Gouldstone C, Maughan D, Snaith J, et al. A methodology for dosimetry audit of rotational radiotherapy using a commercial detector array. *Radiother Oncol* 2013;108:78–85.
- [47] Clark CH, Hussein M, Tsang Y, Thomas R, Wilkinson D, Bass G, et al. A multi-institutional dosimetry audit of rotational intensity-modulated radiotherapy. *Radiother Oncol* 2014;113:272–8.
- [48] Lafond C, Chiavassa S, Bertaut C, Boussion N, Chapel N, Chapron L, et al. DEMAT: A multi-institutional dosimetry audit of rotational and static intensity-modulated radiotherapy. *Phys Medica* 2016;32:664–70.
- [49] Childress NL, White RA, Bloch C, Salehpour M, Dong L, Rosen II. Retrospective analysis of 2D patient-specific IMRT verifications. *Med Phys* 2005;32:838.
- [50] De Martin E, Fiorino C, Broggi S, Longobardi B, Pierelli A, Perna L, et al. Agreement criteria between expected and measured field fluences in IMRT of head and neck cancer: The importance and use of the γ histograms statistical analysis. *Radiother Oncol* 2007;85:399–406.

Table 1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1 Comparison of a head & neck dose distribution calculated at 2.5mm dose grid spacing in the TPS and interpolated to 1.25 mm spacing against the TPS dataset calculated with 1.25mm dose grid spacing. Gamma index analysis of different interpolation techniques using global 1%/0.1mm, 0% threshold.

Interpolation type	Passing rate (%)	Mean γ
Spline	99.2	0.06
Linear	97.9	0.14
Cubic	99.0	0.08
Nearest neighbour	79.0	0.73

Table 2 Summary of mean and minimum γ passing criteria for each system for the range of errors introduced. The concordance correlation coefficient, ρ_c , is also given assessing agreement with the predicted 3D volume γ pass-rate.

% detectors/pixels passing with $\gamma < 1$ and ρ_c				
	Mean	min	ρ_c	Pearson ρ
Predicted				
3D volume	98.8	92.4	N/A	N/A
2D coronal plane	98.6	91.8	0.98	0.99
Measured				
OCTAVIUS729 2D coronal	97.8	86.9	0.81	0.99
OCTAVIUS 1000 SRS 3D volume	98.1	91.5	0.93	0.98
OCTAVIUS729 3D volume; 2.5mm grid	97.2	94.6	0.34	0.47
OCTAVIUS729 3D volume; 10mm grid	99.2	95.4	0.80	0.95

Figure 1 Schematic representation of the gamma index method in 1D. Adapted from Low et al (1998). The y-axis is Dose, D , and the x-axis is distance, r . The cross is the reference point and the blue line represents the evaluated dose distribution with the solid circles being discrete points along the line. The δr and δD criterion create an acceptance ellipse around the reference point. In this schematic and using Equation 2, point $D_E(r_R), r_R$ would have $\Gamma > 1$, $D_E(r_E)$ would be $\Gamma < 1$, and $D_E(r_{E1})$ would be $\Gamma = 1$, $D_E(r_E)$ would be $\Gamma < 1$ (since it is inside the acceptance ellipse). Therefore the result of Equation 3 would be $\gamma < 1$ for the reference point.

Figure 2 Screenshot of the gamma index calculation software implemented in MATLAB.

Figure 3 Comparison of different Matlab interpolation algorithms; (a) spline, (b) linear, (c) cubic and (d) nearest neighbour.

Figure 4 Gamma index map for the deliberate error distribution against the normal distribution, showing failed points expected in the 2.5mm pixel spacing reference distribution for (a) the head & neck VMAT plan with 5° collimator error (Dataset1), and (b) prostate IMRT with MLC sag due to gravity (Dataset2).

Figure 5 (a) Comparison of the maximum γ as a function of varying the search distance for global gamma passing criteria 3%/3mm, 3%/2mm, 3%/1mm for Dataset 1 and Dataset 2. The evaluated dose grid spacing was 2.5mm. The average calculation time taken is plotted against the right axis. (b) Comparison repeated with evaluated dose grid spacing at 1.25mm for Dataset 2, 3%/2mm.

Figure 6 Comparison of the mean γ as a function of varying the interpolation factor for global gamma passing criteria 3%/3mm, 3%/2mm, 3%/1mm for Dataset 1 and Dataset 2. The x-axis shows the ratio of the distance criterion divided by the pixel spacing (where $x_0=2.5\text{mm}$ in Equation 8). The calculation time is plotted against the right axis for each passing criteria.

Figure 7 Gamma index map for the 10mm resolution reference distribution for Dataset 1 from (a) OmniPro, (b) Verisoft, (c) MATLAB plotted using the *imagesc* function, and (d) MATLAB plotted using the *contourf* function.

Figure 8 Comparison of % points passing with $\gamma < 1$ between MATLAB, Verisoft and OmniPro analysis for reference Dataset 1 and 2 distributions with 2.5mm, 5mm and 10mm pixel spacing. In all cases the evaluated dose distribution had $x_0 = 2.5\text{mm}$ and analysis used global γ 3%/2mm with no lower dose threshold. (a) Dataset 1, where evaluated dose distribution was not interpolated in MATLAB; i.e. $x_l = 2.5\text{mm}$, (b) Dataset 1, where evaluated dose distribution was interpolated in MATLAB; i.e. $x_l = 1.25\text{mm}$, and similar respectively for Dataset 2 in figure (c) and (d).

Figure 9 2D coronal planes. **Top row:** TPS, 1000^{SRS} in OCTAVIUS 4D, OCTAVIUS 729 in OCTAVIUS 4D, and 2D-ARRAY 729 in OCTAVIUS II dose distribution for the plan with 5mm MLC error. **Bottom row:** 3%/3mm γ distribution for the plan with 5mm MLC error in MATLAB predicted, OCTAVIUS 1000^{SRS}, OCTAVIUS 729, and coronal 2D-plane 729 array.

Figure 10 2D dose profile for the 3D reconstructed measured dose from the 729 and SRS-1000 arrays compared against the TPS dose profile.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10