ELSEVIER

Radiotherapy and Oncology 62 (2002) 309–319

RADIOTHERAPY & ONCOLOGY
JOURNAL OF THE EUROPEAN SOCIETY FOR THERAPEUTIC RADIOLOGY AND ONCOLOGY

www.elsevier.com/locate/radonline

A quantitative evaluation of IMRT dose distributions: refinement and clinical assessment of the gamma evaluation

Tom Depuydt, Ann Van Esch*, Dominique Pierre Huyskens

Department of Oncology, Division Radiation Physics, University Hospital Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium

Received 16 July 2001; received in revised form 5 November 2001; accepted 30 November 2001

Abstract

Background and purpose: Although intensity modulated radiotherapy (IMRT) is a step forward in comparison to conventional, static beam delivery, quality assurance is more complex and labour intensive, demanding detailed two-dimensional dosimetric verification. Regardless of the technique used for measuring the dose distribution, what is essential to the implementation of routine verification of IMRT fields is the efficient and accurate comparison of the measured versus desired dose distribution. In order to achieve a fast, yet accurate quantitative measure of the correspondence between measured and calculated dose, the theoretical concept of the gamma evaluation method presented by Low et al. (Med. Phys., 25 (1998) 656) was converted into a calculation algorithm, taking into account practical considerations related to the discrete nature of the data.

Materials and methods: A filter cascade of multiple levels was designed to obtain fast and accurate comparison of the two dose distributions under evaluation. The actual comparison consists of classification into accepted or rejected datapoints with respect to user-defined acceptance criteria (dose difference and distance to agreement). The presented algorithm was tested on dosimetric images calculated and/or acquired by means of a liquid filled portal imaging device during the course of intensity modulated treatments of prostate cancer, including pre-treatment verification as well as verification during treatment. To assess its ability to intercept possible errors in dose delivery, clinically relevant errors were deliberately introduced into the dose distributions.

Results: The developed gamma filter method proves successful in the efficient comparison of calculated versus measured IMRT dose distribution. Secondly, intercomparison of dosimetric images acquired during different treatment sessions illustrate its potential to highlight variations in the dosimetric images. The simulated errors were unmistakably intercepted.

Conclusions: The readily obtained gamma evaluation images are an easy tool for quality control of IMRT fields. To reduce the artefacts related to the discrete nature and limited resolution of the data, a fast and accurate filter cascade was developed, offering the possibility to use the gamma method for day to day evaluation of patient dosimetric portal images with or without comparison to a predicted portal dose distribution.

Keywords: Gamma evaluation; Intensity modulated radiotherapy; Two-dimensional dose distributions

1. Introduction

Implementation of intensity modulated beams into the clinical routine of radiotherapy can offer substantial advantages to the patient either regarding dose distribution in the target volume or improved sparing of the surrounding normal tissue and critical organs (see e.g. Refs. [5,1]). In the University Hospital of Leuven, such intensity modulated fields are delivered by means of a dynamic multileaf collimator (dMLC). This technique is a step forward in comparison to conventional, static technique, but its quality assurance is more complex and demands two-dimensional verification.

Although ionisation chamber, TLD or diode array measurements provide accurate dosimetric data, they are incomplete for IMRT quality assurance because they only yield the dose in a single point or along one line. The two-dimensional dosimetric accuracy and reproducibility of the dMLC can e.g. be verified using films in a phantom [1,3,11]. Recently, it has been shown that accurate dosimetric images of IMRT beam profiles delivered with dMLC can also be obtained with a liquid filled electronic portal imaging device (EPID) (Portal Vision MarkII, Varian Medical Systems) [10], offering the advantage of on-line, digital images.

Regardless of the measurement technique, what is essential to the QA of the intensity modulated dose delivery is the efficient and accurate comparison of the measured versus calculated dose distribution.

A simple qualitative evaluation is made by superimpos-

* Corresponding author.

0167-8140/02/$ - see front matter
PII: S0167-8140(01)00497-2

ing the isodose distributions. Provided the relevant isodose lines have been chosen to plot, this evaluation can highlight areas of disagreement, but a more quantitative assessment for final approval is desirable.

The extraction and comparison of line profiles provides a more detailed print of the dose correspondence, but because of the limited selection important disagreements can be overlooked. In selecting the most critical and relevant line profiles, adequate experience of the physicist is imperative.

Furthermore, the above methods demand a lot of manual analysis and are therefore, time consuming. A higher level of automation and a more quantitative evaluation is desirable to accomplish full integration into daily clinical routine.

A first attempt to define a quantitative evaluation method was the use of the dose difference as acceptance criterion. This criterion can be used in low gradient areas but is inadequate to evaluate high gradient areas where a small spatial shift – of physical origin (e.g. an uncertainty in the positioning of the detector) or related to the calculation – will result in a large dose difference a priori.

Van Dyk et al. [9] subdivide the dose distribution comparisons into regions of high and low dose gradients, each with a different acceptance criterion. In low gradient regions, the doses are compared directly, with an acceptance tolerance placed on the difference between the measured and calculated doses. Visualisation of the dose difference distribution identifies regions of disagreement. Because the dose difference in high dose gradient regions may be misleading, Van Dyk et al. used the concept of distance-to-agreement (DTA). The DTA is the distance between a reference data point and the nearest point in the compared dose distribution that exhibits the same dose. The evaluation images displaying the dose difference and DTA are complementary in determining the acceptability of dose calculation versus delivery.

In order to merge both evaluation images into a single image, a composite analysis used by Harms et al. [2] uses a pass–fail criterion of both the dose difference and DTA: points that fail both criteria are identified on a composite distribution. The dose difference is displayed with the binary composite distribution highlighting regions of disagreement. A limitation of this technique is that the display of the dose difference may accentuate the impression of failure in high dose gradient regions. Also, it provides no quantitative measure of the magnitude of disagreement.

We have focused on the method presented by Low et al. [4] to simultaneously incorporate the dose and distance criteria. This method provides a numerical quality index – referred to as the gamma value – that serves as a measure of disagreement in the regions that fail the acceptance criteria and indicates the calculation quality in regions that pass.

1.1. Gamma evaluation: the theoretical concept

The gamma method, as presented by Low et al. [4], is designed for the comparison of two dose distributions: one is defined to be the reference information ($D_r(\boldsymbol{r})$) and the other is queried for evaluation ($D_c(\boldsymbol{r})$). Fig. 1 shows a schematic representation of the gamma analysis tool for two-dimensional dose distribution evaluations. The acceptance criteria are denoted by ΔD_M for the dose difference and Δd_M for the distance to agreement. For a reference point at position $\mathbf{r}_r$, receiving dose D_r, the surface representing these

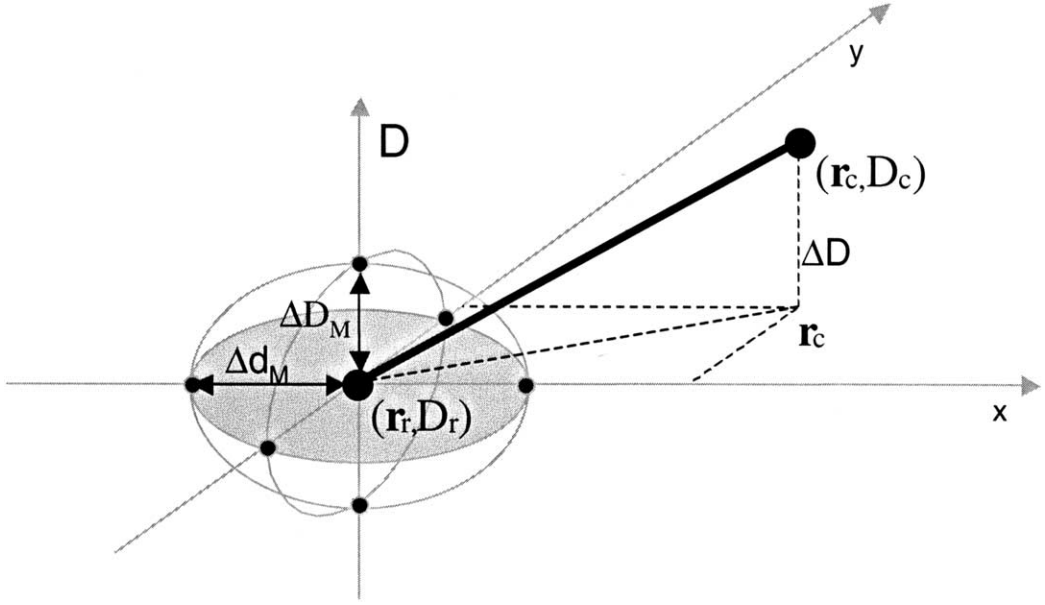

Fig. 1. Schematic representation of the theoretical concept of the gamma evaluation method. The reference and compared dose distributions are denoted by $(\boldsymbol{r}_r, D_r)$ and $(\boldsymbol{r}_c, D_c)$, respectively. The criteria defining the ellipsoid of acceptance are denoted by the dose difference tolerance ΔD_M and the maximal distance to agreement Δd_M.

acceptance criteria is an ellipsoid defined by:

$$1 = \sqrt{\frac{\Delta r^2}{\Delta d_M^2} + \frac{\Delta D^2}{\Delta D_M^2}}$$

where

$$\Delta r = |\boldsymbol{r}_r - \boldsymbol{r}_c|$$

is the distance between the reference and compared point and

$$\Delta D = D_c(\boldsymbol{r}_c) - D_r(\boldsymbol{r}_r)$$

is the dose difference at the position $\boldsymbol{r}_c$ relative to the reference dose D_r in $\boldsymbol{r}_r$. For the compared distribution to match the reference dose in $\boldsymbol{r}_r$, it needs to contain at least one point $(\boldsymbol{r}_c, D_c)$ lying within the ellipsoid of acceptance, i.e. one point for which:

$$\Gamma_r(\boldsymbol{r}_c, D_c) \equiv \sqrt{\frac{\Delta r^2}{\Delta d_M^2} + \frac{\Delta D^2}{\Delta D_M^2}} \leq 1$$

A quantitative measure of the accuracy of the correspondence is determined by the point with the smallest deviation from the reference point, i.e. the point for which $\Gamma_r(\boldsymbol{r}_c, D_c)$ is minimal. This minimal value is referred to as the quality index $\gamma(\boldsymbol{r}_r)$ of the reference point.

The pass–fail criterion therefore becomes

$\gamma(\boldsymbol{r}_r) \leq 1$, correspondence is within the specified acceptance criteria,
$\gamma(\boldsymbol{r}_r) > 1$, correspondence is not within specified acceptance criteria.

An implicit assumption is made that once the passing criteria are selected, the dose difference and DTA analyses have equivalent significance when determining calculation quality.

1.2. Gamma evaluation in clinical routine

Although in theory the gamma evaluation is a powerful concept for comparison of two-dimensional dose distributions, practical considerations need to be made when proceeding towards clinical implementation. The dose distributions under examination may originate from different sources such as treatment planning system calculations, film or EPID measurements, each having their own format and resolution. While Low et al. present a useful theoretical method in a continuous environment, in a discrete environment the limited sample rates and spatial resolutions need to be taken into account. When relying on a pixelwise calculation, the pixel size of the compared image $D_c(\boldsymbol{r})$ needs to be small with respect to the acceptance criteria. A straightforward solution to suppress calculation artefacts due to the finite sampling would be an interpolation algorithm. However, in regions of high dose gradients, even with a highly intensified calculation grid, gamma indices can be unjustly concluded to be larger than one. Hence, while interpolation considerably increases calculation time, it still does not assure an infallible report of the acceptability of the dose distribution. Furthermore, $\gamma(\boldsymbol{r}_r)$ remains dependent on the discretisation and in most cases will not be the analytical minimum one would obtain when both dose distributions were presented as continuous functions.

To overcome the above mentioned limitations, we present another strategy. Firstly, we are primarily interested in whether or not a point is within the specified acceptance criteria, without focussing on the exact numerical gamma value, since the latter is not considered to be of primary importance in routine evaluation and can even be misleading, as illustrated. It is therefore, only used as a pass/fail criterion. Secondly, we aim to restrict calculation time to a minimum and thirdly, we want to reduce the amount of falsely reported unaccepted points to a minimum. The objective of our approach is to produce an evaluation image indicating whether or not two dosimetric distributions are within a user-defined agreement. Areas of disagreement are highlighted and subsequently analysed further by means of line profiles.

2. Methods and materials

2.1. The algorithm: a level based approach

The input data of the algorithm consist of two matrices of dose values: the reference distribution D_r and the compared distribution D_c, both consisting of pixels each having a position $\boldsymbol{r}(x, y)$ and a dose value D. For every point of the reference dose distribution, we can construct an axial system with the origin at the reference position and reference dose, to allow a straightforward geometric representation of the gamma criterion as illustrated for a one-dimensional case in Fig. 2a, the largest symbol representing the reference point, the smallest symbols representing the datapoints of the compared distribution, open symbols indicate datapoints passing the acceptance criteria for the given reference point. Illustrated for a two-dimensional distribution in Fig. 2b, the 'district of the reference point' is defined to be the intersection of the ellipsoid with the xy-plane, i.e. a circle with radius Δd_M. The points for which the (x, y) coordinates lie within the district of the reference point are the ones focussed on for calculation. For every reference point, the gamma evaluation aims to answer the question whether or not the compared dose distribution surface (x, y, D_c) intersects the ellipsoid of acceptance. Reference points for which the compared distribution in its original format contains pixels lying within the ellipsoid allow straightforward acceptance, while reference points for which this is not the case require further investigation. We have therefore designed a filter cascade to categorise all reference points as accurately as possible while avoiding excessive calculation time. Schematically illustrated in Fig. 3, it consists of three successive filters.

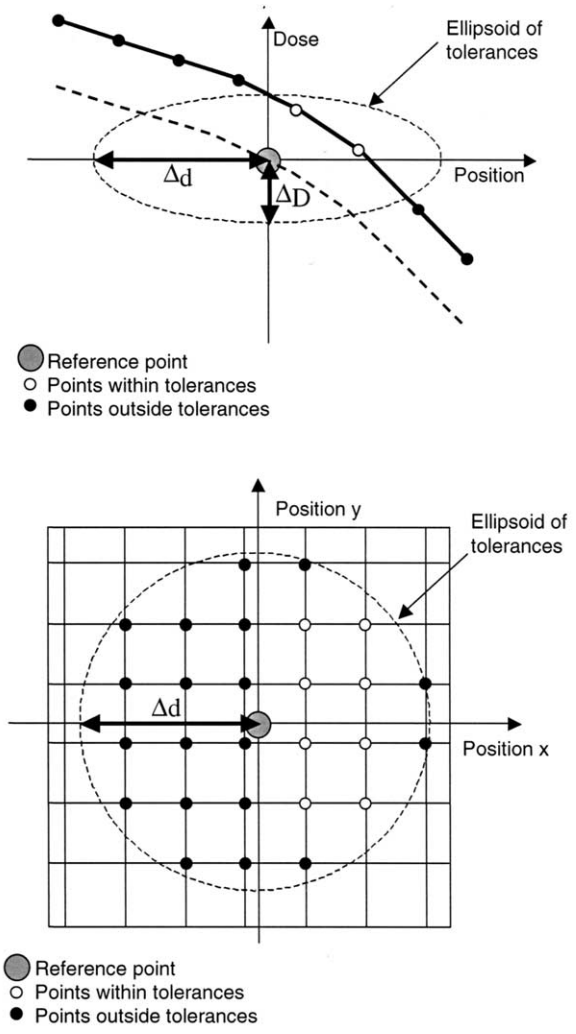

Fig. 2. Geometric representation of the gamma criterion for discrete dose distributions: (a) Illustrated for a one-dimensional case: the origin of the axial system is placed at the reference point (large symbol), the smaller, open symbols represent datapoints of the compared distribution conform with the reference point within the specified acceptance criteria. Closed symbols are characterised by a $\Gamma(\boldsymbol{r}_c, D_c)$ exceeding unity. (b) The concept of the 'district of the reference point' in the *xy*-plane is illustrated for a two-dimensional case.

2.1.1. Level 1

The first level is directly derived from the theoretical gamma criterion. However, rather than calculating $\Gamma(\boldsymbol{r}, D_c)$ for every pixel in the compared image and defining the smallest of these values to be the gamma index of the reference point, the calculation is restricted to the points in the district of the reference point. Furthermore, as soon as one pixel is found for which $\Gamma(\boldsymbol{r}_c, D_c)$ is smaller than unity, calculation is stopped and the reference point is classified as accepted. When no such $\Gamma(\boldsymbol{r}, D_c)$ is found, the point under investigation is passed onto the next filter.

2.1.2. Level 2

All points subjected to the filter of the second level have in their district only points with $\Gamma(\boldsymbol{r}, D_c)$ exceeding unity. In

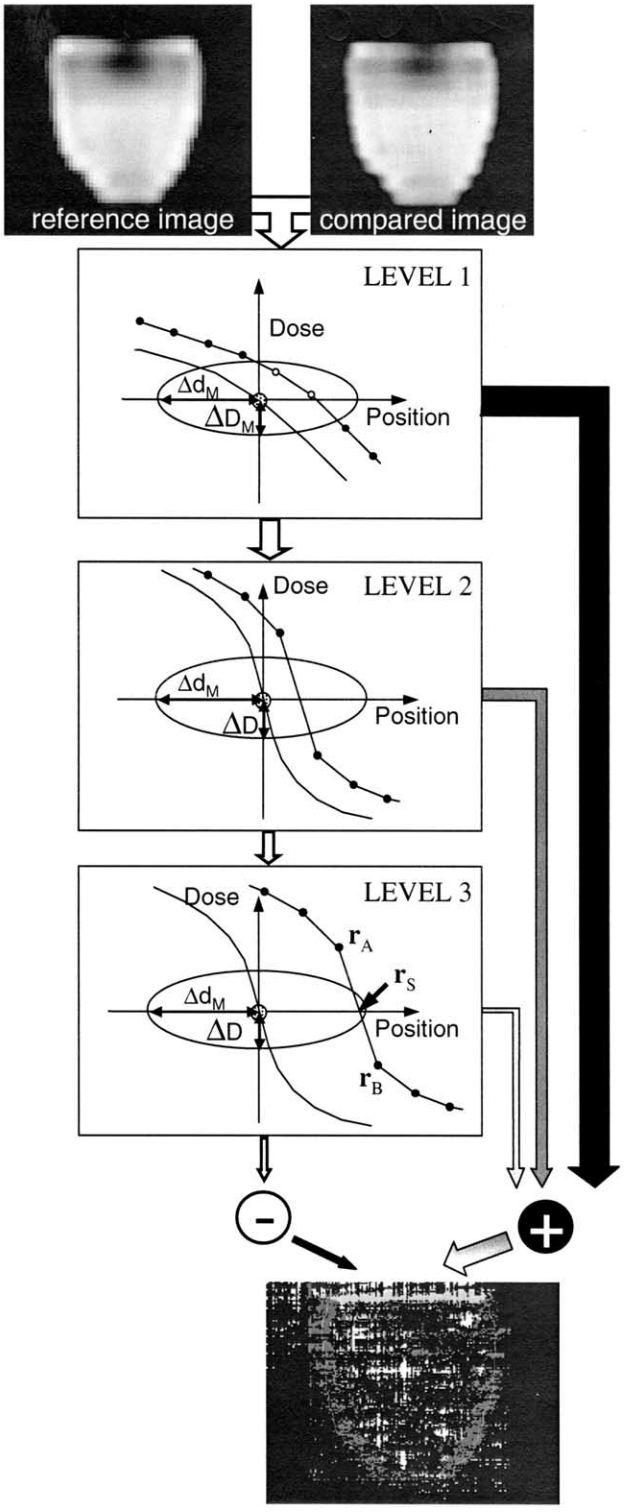

Fig. 3. Schematic outline of the filter cascade designed to categorise all reference points as accurately as possible according to the theoretical concept of the gamma evaluation. For simplicity, the illustrations representing each filter level are shown for a one-dimensional case. Input into the gamma filter are the reference and compared distributions, resulting in an output distribution of black, dark grey and light grey pixels, corresponding to reference points being accepted after passing the first, second and third level of the algorithm, respectively. The white pixels indicate points that are found to be not within the specified tolerance limits.

regions with large dose gradients, it is possible that the compared dose distribution intersects the ellipsoid but the datapoints sampling the distribution are situated outside the ellipsoid of acceptance. Hence, straightforward calculation of the γ index would falsely reject the investigated reference point. Although linear interpolation could yield a datapoint for which $\Gamma(\boldsymbol{r}, D_c)$ is smaller than one, it is readily seen that in this case excessive interpolation may be required. However, since the actual value of the γ index does not concern us, an alternative approach is possible. When ΔD is of opposite sign for at least two sample points in the district of a reference point, the compared distribution must intersect the ellipsoid at some point and the reference point should be classified as accepted. Hence, the identification of such points can be restricted to a quest for two datapoints in their district with dose differences of opposite sign.

Reference points rejected by the second level filter are submitted to a third test.

2.1.3. Level 3

Of reference points being subjected to the third test the following is known: geometrically speaking, all datapoints $(\boldsymbol{r}_c, D_c)$ in their district lie either above or below the ellipsoid of acceptance. However, in rare cases, some reference points rejected by the second filter should still be classified as accepted. As illustrated in Fig. 3 (Level 3), intersection of the compared distribution surface with the ellipsoid is still possible when datapoints at the outside boundary of the district have ΔD of opposite sign compared to the datapoints within the district. The third filter is therefore designed to take these datapoints at the outside boundary of the district into account. Fig. 3 (Level 3) shows the intersection point $\boldsymbol{r}_S(D_c = 0)$ of the line through $D_c(\boldsymbol{r}_A)$ and $D_c(\boldsymbol{r}_B)$, calculated through linear interpolation. Provided this intersection point is situated inside the ellipsoid (i.e. $\boldsymbol{r}_S < \Delta d_M$), the reference point is accepted in the comparison procedure.

The third filter concludes the algorithm and reference points not exhibiting the above mentioned intersection point in their ellipsoid of acceptance are finally classified as being in disagreement with the compared distribution.

Although a small number of situations can still be imagined for which classification by the above filter cascade would lead to wrongful rejection of a reference point, the clinical relevance of these situations becomes minor. Even so, if required, additional levels could be added to the algorithm.

2.2. Clinical implementation/illustrations

To explore the gamma evaluation algorithm – as described in the above paragraph – as a verification tool, we have studied the corresponding gamma image at different steps of the quality assurance, following the chronology of clinical routine. Additionally, to probe its sensitivity in error detection we have deliberately introduced possible discrepancies into compared images.

As the treatment of prostate cancer was selected for the initiation of IMRT into the University Hospital Leuven, most of the gamma evaluation tests have been performed on these patient plans. In Leuven, IMRT is delivered by means of the sliding window technique on a 2100C Linac (Varian Medical Systems), equipped with a dynamic MLC (40 leaf pairs). The irradiation of the prostate and seminal vesicles by means of IMRT is routinely planned according to a fixed plan set-up, consisting of two static open beam fields used for portal imaging and five intensity modulated fields. Patients are irradiated in a supine position. The five intensity modulated fields are set at predefined gantry angles (RPO: 255°, RAO: 315°, ANT: 0°, LAO: 45° and LPO: 105°) optimised by means of the Helios inverse planning modules in the Cadplan treatment planning system (Varian Medical System). A dose of 70 Gy is delivered to the prostate in daily fractions of 2 Gy, seminal vesicles receive a total dose of 50 Gy (2 Gy/fraction). During the optimisation of the dose delivery to the target volumes, constraints on the dose to the rectum are enforced to reduce rectal toxicity of the treatment. The two orthogonal static fields are used to check the positioning of the patient daily: prior to the delivery of the intensity modulated fields, the bony structures are verified with respect to the field outline through comparison of a reference digitally reconstructed radiograph (DRR) and a portal image. These fields are only delivered with a minimum amount of dose to minimise their input on the intensity modulation.

2.2.1. Pre-teatment verification

Once the optimal treatment plan has been decided on, all treatment fields are dosimetrically verified by means of EPID dosimetry prior to treatment initiation. Pre-treatment verification of intensity modulated fields involves evaluation of a dosimetric image – acquired with the liquid filled portal imager – through comparison with a predicted image [10]. The intensity modulated fields are delivered to the portal imager without a patient or phantom in between. For this test, the complete treatment plan is exported from the TPS to the patient record in the record and verify system and delivered as such at the linear accelerator (permitting only overrides on the table parameters and the gantry angle). Hence, simultaneously with the dosimetric check, the complete data-transfer to the record and verify system and the dMLC controller at the linear accelerator is verified. The data are acquired with a dosimetric acquisition mode especially designed for this purpose, as described in Ref. [10]. All fields are delivered at zero gantry angle, allowing a polystyrene build-up plate of 2-cm thickness to be placed on top of the detector housing. Although gravitational effects on the positioning of the leaves are thus not included, the verification becomes not only faster, but also more accurate since bulging of the liquid need not be corrected for. The portal imager is placed at a fixed position of 145 cm

from the source (this being the pre-defined position used for imaging in clinical routine). The acquired images are subsequently compared to predicted portal dose images. The calculation of the portal dose in the absence of a patient is based on the algorithm developed by Storchi et al. [7,8] using pencil beam kernels. The prediction is always calculated at isocentre distance and stored in a 128×128 matrix (pixel size 0.25 cm). The measured dose distribution consists of a 256×256 matrix with a pixel size of 0.127 cm. Since these data are obtained at a distance of 145 cm from the source, the resolution recalculated back to isocentre distance is improved to 0.088 cm. Because of its lowest resolution, the predicted image is set to be the reference image in the gamma evaluation algorithm.

2.2.2. Verification during treatment

The introduction of IMRT into the treatment of prostate cancer at the radiotherapy department of the U.Z. Gasthuisberg has allowed acquisition of dosimetric data during the actual delivery. Analysis of the data thus obtained can be performed from different angles of incidence.

Firstly, as in pre-treatment verification, the acquired data were compared to a predicted image. The format of both images is identical to the format described in the above paragraph. The acquisition, however, is now performed with the gantry angle – and hence the inclination of the portal imager – at the correct angle for delivery. Because of this rotation, no build-up plate can be placed on top of the detector housing. With a patient in the beam, patient anatomy as described by the planning CT-scan is taken into account in the portal dose image prediction algorithm [6]. It is worthwhile to mention, however, that the contribution of electron contamination and the presence of the treatment table at the linear accelerator are not taken into account in the prediction algorithm. Only comparison or the relative dose distributions, i.e. comparison of the images both normalised to their datapoint on the beam axis, was evaluated.

Secondly, the day to day reproducibility was assessed by intercomparison of the acquired portal images. For every IMRT patient, dosimetric EPID images are acquired daily for all treatment fields, providing us with an extended dosimetric database for every patient over the whole course of the treatment. All originating from the same source, these images have identical dimensions (256×256) and since the portal imager is automatically placed at a fixed, pre-programmed position, they also have the same scale. In the analysis of the gamma matrix, we aimed to evaluate the absolute reproducibility of the dosimetric image during use in clinical routine, bearing in mind the bulging of the detector liquid, inaccuracies in positional reproducibility of the detector, as well as the lack of build-up. Relative (i.e. with both images normalised to their datapoint on the beam axis) and absolute (i.e. with both images in their original form) gamma evaluations were performed simultaneously for intercomparison. As the reference image in the gamma algorithm, we used the image acquired on the first day of treatment as well as the image obtained during the previous treatment session.

Dosimetric images acquired during treatment inherently contain information on patient anatomy. Although it is not the primary goal of the gamma evaluation to verify the position of the patient – since the orthogonal images acquired prior to IMRT delivery are more suitable for this purpose – we have examined the capability of the gamma evaluation to extract such information from image intercomparison. In extracting the appropriate images for these purposes, we made use of the anatomical information contained in the two orthogonal images acquired daily to verify patient positioning prior to IMRT dose delivery. We have selected dosimetric images corresponding to a positional shift of $\sim$5 mm in the lateral direction. These positional shifts are determined as a shift in the bony structures of the patient (pelvic bone and femoral heads) with respect to the isocentre of the linear accelerator; reflecting whole body motion. Internal motion of the prostate, on the contrary, cannot directly be evaluated from the portal images. Indirectly, the position of the prostate is closely linked to the position (and filling) of the rectum; e.g. excessive rectal filling will push the prostate to a more anterior position. Hence, in addition to examining dosimetric images in which a positional shift of the whole patient had been noted, we have extracted dosimetric images for which the anterior portal image pointed out the presence of air in the rectum and compared them to reference images without air.

2.2.3. Error detection

To assess the capability of the gamma index to detect possible errors in dose delivery, we have deliberately introduced clinically credible mistakes into compared images.

As a first example, dose was delivered to the EPID using an erroneous energy on the linear accelerator (6 MV instead of 18 MV, in this case). Both relative and absolute comparisons were made with an image acquired with the correct energy selection. Different tolerance settings were tested ($\Delta D_M = 1$–3.3% and $\Delta d_M = 3$ mm).

Treatment delivery with an erroneous field choice was simulated by comparing the EPID images of non-corresponding fields. The error can consist of either the wrong field in the correct treatment plan (e.g. RAO instead of RPO), or the correct field in the wrong treatment plan (e.g. corresponding to another patient).

Thirdly, dose discrepancy due to malfunctioning of the MLC was mimicked by blocking the motion of one leaf during the complete delivery.

3. Results

In the resulting gamma evaluation images, white pixels indicate rejected reference points, black pixels were accepted by the filter of the first level while dark and light grey colours indicate acceptance at the second and third level filter,

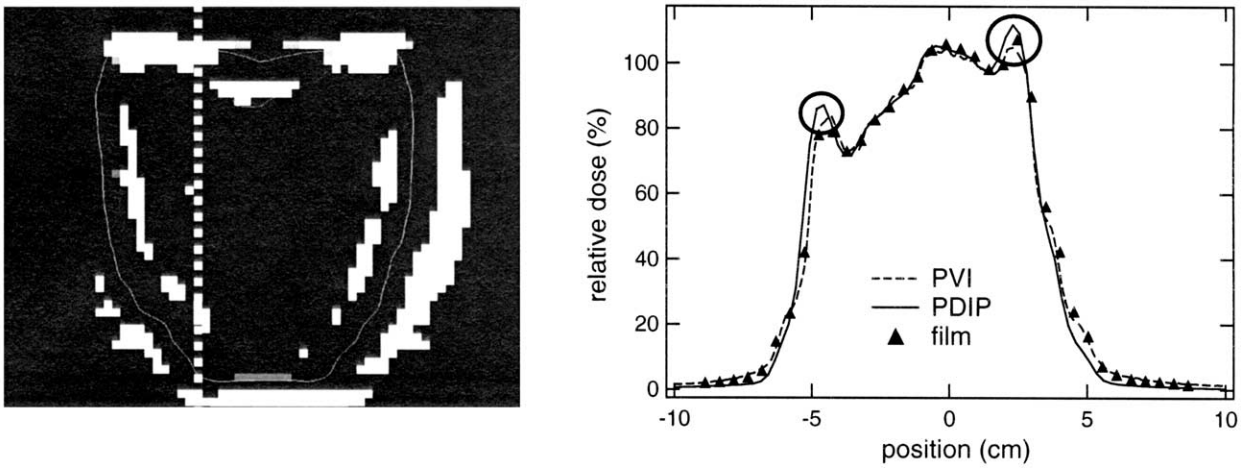

Fig. 4. (a) Gamma evaluation chart for pre-treatment verification of an anterior IMRT prostate field obtained with tolerance settings of $\Delta D_M = 5\%$ and $\Delta d_M = 4.5$ mm. The calculated portal dose distribution is defined to be the reference image. It is compared to a dosimetric distribution measured by means of an EPID (PortalVision MarkII). (b) Line profiles extracted from the dosimetric distributions (calculated and measured by means of the EPID). The regions where points are rejected are indicated in the figure. A line profile measured with film at the plane of the EPID excludes errors related to the dosimetric data acquisition.

respectively. The 50% isodose line (relative to the datapoint on the beam axis) of the reference dose distribution is routinely plotted (in white) as a visual guide to the field outline.

3.1. Pre-treatment verification

In Fig. 4 the gamma evaluation in the pre-treatment verification process (i.e. without patient) on the anterior field of a treatment plan is shown. Different sets of tolerance variables can be used to illustrate the actual error size. Although some discrepancy zones occur in the central part of the field for $\Delta D_M = 3\%$ and $\Delta d_M = 3$ mm, the largest and most persistent areas of disagreement are situated at the field edge, as situated by the display of the 50% isodose line. Mainly two zones of particularly large deviations, exceeding tolerance levels of $\Delta D_M = 5\%$ and $\Delta d_M = 4.5$ mm, are recognised in Fig. 4a; this is illustrated in Fig. 4b by the line profiles through the original dosimetric distributions. Added to Fig. 4b is a profile measured with film, excluding possible EPID measurement errors.

Although not displayed, similar behaviour is recognised in the other treatment fields.

3.2. Verification during treatment

3.2.1. Comparison between delivered and predicted dose distributions

The results of the relative gamma evaluation of the anterior field delivered to a prostate patient are displayed in Fig. 5. Agreement between predicted and measured dose is satisfying inside the target volume: acceptance criteria of $\Delta D_M = 4.5\%$ and $\Delta d_M = 3$ mm are not violated inside the 50% isodose area, except for the small discrepancy zone at the edge of the field, already spotted in the pre-treatment verification. An extended area of disagreement, even for tolerance levels beyond $\Delta D_M = 7\%$ and $\Delta d_M = 3$ mm persists in the zone outside the target volume. Conformity between calculation and measurement is of lesser quality for the oblique fields (not displayed).

3.2.2. Intercomparison between delivered dose distributions

Fig. 6 illustrates the absolute day to day reproducibility of the EPID images (as well as the reproducibility of the actual delivery): over the course of the treatment, images are compared to the image acquired during the first day of treatment. Tolerance levels can be reduced to $\Delta D_M = 1\%$ and $\Delta d_M = 3$ mm without introducing significant areas of discrepancy in the gamma display. The comparisons between the dosimetric images acquired during the first and sixth treatment session are displayed for the anterior (Fig. 6a) and right anterior oblique (Fig. 6b) field. Similar

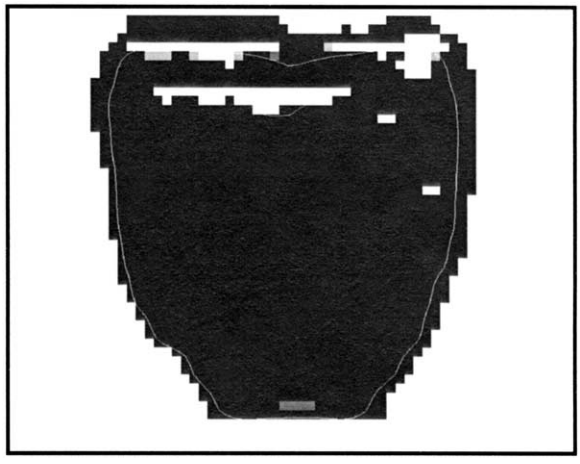

Fig. 5. Gamma comparison between delivered and predicted dose distribution during patient treatment (anterior prostate field); acceptance criteria were defined to be $\Delta D_M = 4.5\%$ and $\Delta d_M = 3$ mm.

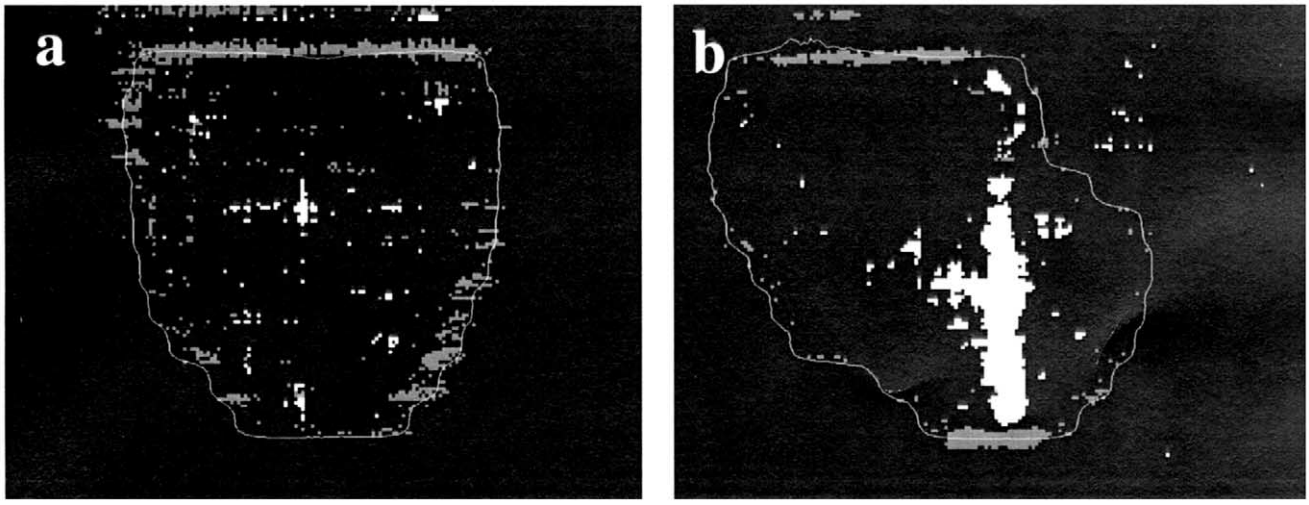

Fig. 6. Intercomparison between dosimetric images acquired during the first and sixth treatment session: the gamma evaluation chart of the (a) anterior and (b) RAO field is calculated with $\Delta D_M = 1\%$ and $\Delta d_M = 3$ mm.

results are obtained when images of successive treatment days are intercompared. In contrast to the anterior field, results on the oblique fields are expected to be subject to bulging in the detector. In the gamma evaluation of the RAO field overall reproducibility of the data is of the same quality as for anterior fields. However, a consistent, elongated zone of discrepancy occurs in several gamma evaluations (e.g. Fig. 6b), corresponding to a changing position of the moveable table support bar. Since the table is behind the patient with respect to the accelerator head for all fields, its exact position is of no clinical relevance to the treatment and hence not verified by the radiographers.

The effect of changes of the patients anatomy and its corresponding effect on the dosimetric acquisition is illustrated in Fig. 7, where the rectal filling of the patient differs from session to session (as shown by the localisation images in Fig. 7). The effect in terms of transit dosimetry with the EPID is evaluated by means of the gamma chart: with tolerances of $\Delta D_M = 1\%$ and $\Delta d_M = 3$ mm, a blotch in the centre of the image – corresponding to the position of the air cavity in the rectum – reflects the disagreement between both dosimetric acquisitions.

When comparing dosimetric images of two treatment sessions with a lateral shift of 5 mm in patient positioning,

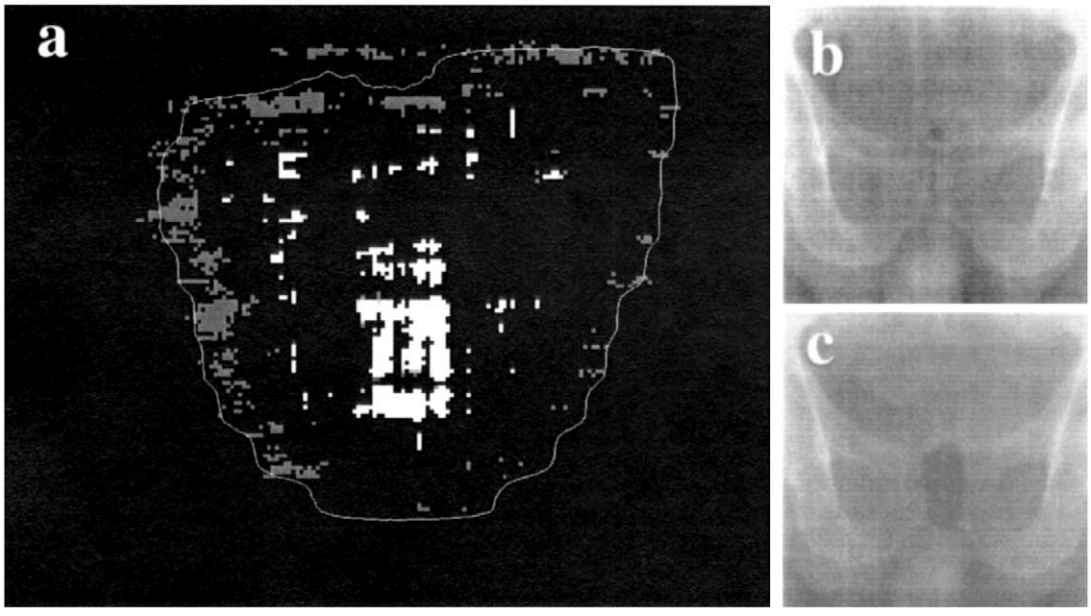

Fig. 7. Gamma comparison (a) of two dosimetric images with different rectal filling; the corresponding localisation images obtained with the liquid filled EPID reveal the absence (b) and presence (c) of an air bubble.

the gamma image ($\Delta D_M = 1\%$ and $\Delta d_M = 3$ mm) reports correspondence of the same quality as for two dosimetric images acquired with near identical patient positioning.

3.3. Error detection

Fig. 8 shows gamma evaluations of identical treatment delivery (i.e. with identical leaf motion) except for the energy set-up at the accelerator: for the compared image 6 MV was selected instead of the 18 MV used for the reference image. Both the absolute and relative gamma evaluation detect a definite disagreement when the standard tolerances of $\Delta D_M = 1\%$ and $\Delta d_M = 3$ mm are used. Regarding the absolute comparison the detection of a wrong delivery energy is unmistakable, even with the more relaxed dose difference acceptance level of $\Delta D_M = 3\%$ (Fig. 8a1). The disagreement detected with the relative comparison is less striking: discrepancies are primarily observed in the background area when an accepted dose difference of 1% is used. Relaxed tolerance settings ($\Delta D_M = 3\%$) entirely fail to detect the error (Fig. 8a2).

An example of a resulting gamma evaluation when incorrect fields are delivered, is depicted in Fig. 8b where delivery of a field with the correct field name and gantry angle, but originating from the wrong patient plan, is simulated. Only a small strap of coincidental agreement is present. Near total disagreement is also observed when the incorrect field of the correct patient plan is selected (not shown).

Failure of one leaf pair to move during the actual irradiation causes a well defined streak of underdosage in the acquired image, unmistakably showing up in the gamma evaluation as well, even if extremely large dose discrepancies are accepted. Allowing $\Delta D_M = 10\%$ ($\Delta d_M = 3$ mm), Fig. 8c provides an illustration of such an artificially created mechanical failure.

4. Discussion

4.1. Pre-treatment verification

Fig. 4 illustrates the potential and the usefulness of the

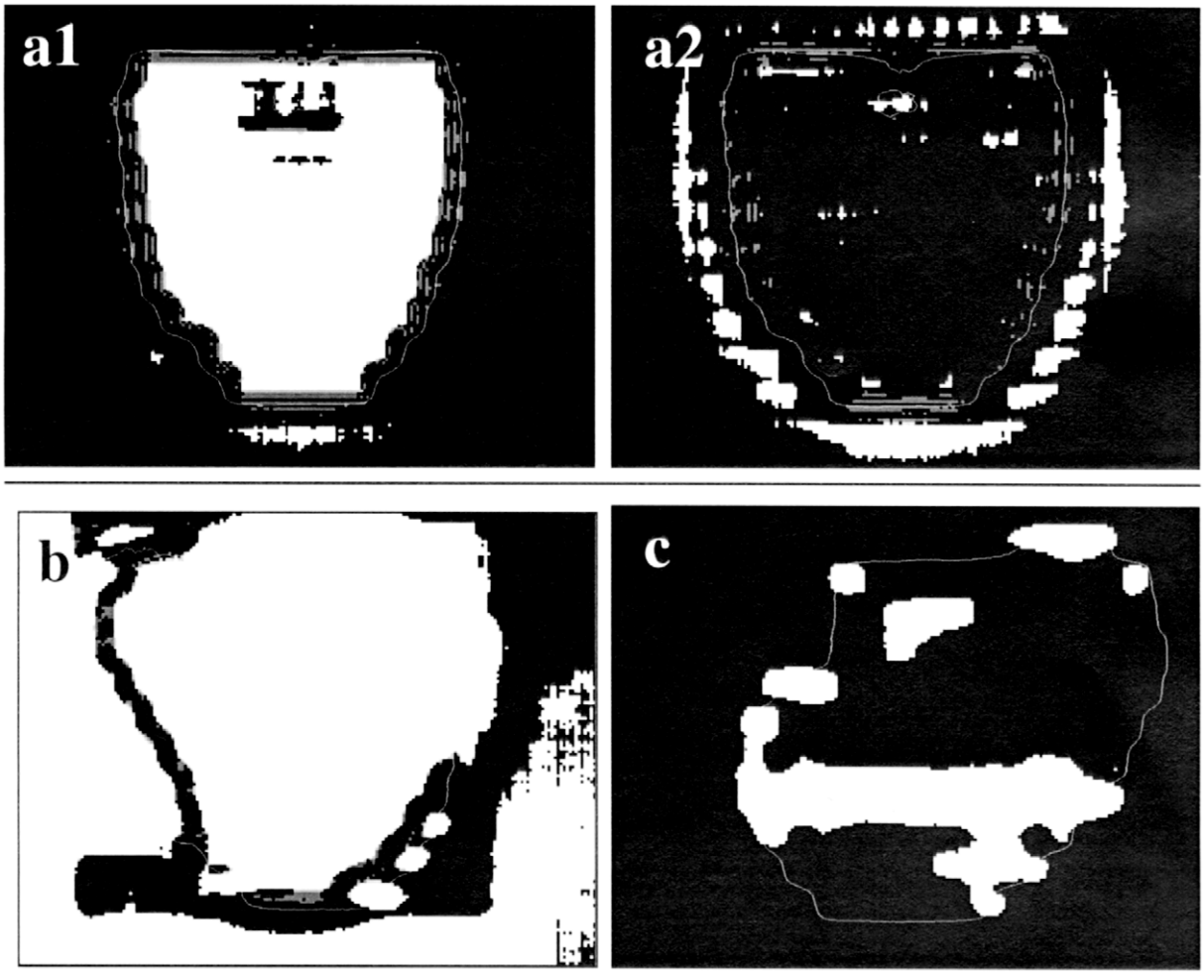

Fig. 8. Error detection by means of the gamma evaluation for (a) erroneous energy selection at the accelerator: (a1) relative and (a2) absolute comparison ($\Delta D_M = 3\%$ and $\Delta d_M = 3$ mm), for (b) erroneous field choice (correct field name, but from the wrong patient plan) ($\Delta D_M = 3\%$ and $\Delta d_M = 3$ mm) and (c) mechanical MLC failure (blocked leaf) ($\Delta D_M = 15\%$ and $\Delta d_M = 3$ mm).

gamma method to quantitatively evaluate the delivered dose distribution versus the calculated dose distribution. Moreover, this kind of comparison highlights some imperfections in the calculation algorithm. As the image prediction is based on a convolution of the actual fluence distribution with scatter kernels, the origin of these discrepancies can be two-fold: the kernels describing the scatter behaviour of the high energy photons are inadequate and/or the calculation algorithm for the fluence distribution is lacking. Although these hypotheses need further investigation, especially since it is not straightforward to separate both causes, some conclusions can already be drawn. Regarding the modelling of the actual fluence distribution, the following comments can be made: as the extended zones of unaccepted points outside the 50% isodose lines in Fig. 4 indicate, leaf leakage is not accurately modelled. In fact, when calculating the actual fluence distribution, the Helios software applies a single average value (2.0%) of leaf leakage, not taking into account the variations between intra- and interleaf transmission (ranging from 1.5 to 2.2%). Although this averaging is clinically justifiable, it contributes to the increased number of rejected points at the outer surface of the field. Also contributing to this number of points is the apparent underestimation of the outscattered dose, this being related to the tails of the scatter kernels. These inadequacies are nicely visualised in two dimensions in the gamma evaluation. The optimal shape of the scatter kernels will therefore definitely be subject to further investigation. As dose calculations become more accurate (e.g. through Monte Carlo calculations), artefacts due to the dose calculation will be reduced.

4.2. Verification during treatment

Depending on the philosophy of the department for quality assurance of IMRT, different approaches exist. Firstly, measured dose distributions can be compared to the portal dose distributions calculated by the TPS. This comparison can be restricted to the first treatment day or can be performed during each treatment session. Alternatively, daily acquired images can be intercompared, more specifically all images can be compared to the measurement acquired and approved on the first treatment day.

Following the comparison of measured and calculated images during treatment, the same conclusions as for pretreatment evaluation can be drawn regarding the image prediction. Furthermore, as can be seen in Fig. 5, the entire background calculation is off by more than $\Delta D_M = 5\%$. This is due to the scatter produced by the patient's body but not modelled in the prediction algorithm. Should the prediction calculation be used for automatic error detection during treatment, the electron contamination must be taken into account in the prediction algorithm. Alternatively, sufficient build-up could be permanently mounted on top of the detector housing to absorb the scattered electrons and prevent them from reaching the actual imaging plane. Also, a correction should be applied for the bulging of the detector. As the daily reproducibility of the images – including the bulging – appears to be excellent, this should be possible [10].

In general, the gamma evaluation readily shows that reproducibility of the detector for absolute dosimetry is better than 1%, taking into account a positional inaccuracy of 3 mm. As an alternative to treatment verification by means of a calculated image, this allows verification of daily dosimetric images by means of a measured reference image.

Although the image acquired and approved on the first treatment day is a suitable option, it does present some disadvantages since it inherently contains characteristics specific to one single treatment day. As illustrated, for example, the shifting position of the table support bar as well as the possibility of air in the rectum, can result in significant areas of disagreement in the gamma image. An alternative method would consist of comparing daily dosimetric image with a mean dosimetric image, i.e. with the arithmetic mean of all preceding, accepted images. Not only does this provide a more general reference image, it simultaneously yields an overview of the overall delivery at the end of a treatment. The a posteriori comparison of the averaged dosimetric image with all individual images would provide information on temporary, daily characteristics (e.g. rectal air) as well as on gradual changes (e.g. weight loss, …). Although the sensitivity of the gamma evaluation algorithm does not suffice for patient positioning verification for prostate treatments, this is a site dependent characteristic and has to be further investigated for other anatomical regions.

Also, in general, we would like to point out that, although the majority of the points going through the level based gamma algorithm, are decided upon during the first filter (i.e. are classified as 'good' points), a significant number of points would have unjustly been rejected without the implementation of the second and third level filters. Notwithstanding the use of interpolation, data classification by mere calculation of the gamma index as presented in the paper of Low et al. [4], would have misjudged all points represented by the grey colours in the displayed figures.

4.3. Error detection

Concerning the investigated treatment errors, the gamma evaluation proves very efficient in automatic error detection. Especially since the simulated errors, although of primary importance, could go undetected if not automatically verified. When only visual verification is used, for example, all presented errors – except the blocked leaf – would lead to clinically familiar portal images. The highly unacceptable gamma matrix, however, provides an alarming warning. To reduce the clinical workload to an absolute minimum, a threshold value could be defined for a maximum acceptable number of rejected points. Better still, a maximum density

of rejected points would be a more efficient criterion since a concentrated cluster of rejected points will be significant, while the same amount of rejected points, spread out over the entire image may be clinically irrelevant. A signal requesting closer investigation should be provided when this threshold value is surpassed.

5. Conclusions

The presented algorithm to implement the theoretical concept of the gamma evaluation method into clinical routine was assessed on images acquired during the course of an intensity modulated treatment of prostate cancer. The readily obtained gamma evaluation image is easy to interpret. Furthermore, the method is much faster than (and even superior to) a classical two-dimensional interpolation algorithm. In different stages of the treatment planning, dose distributions of different sources were compared and these comparisons clearly pointed out areas of disagreement. Efficient error detection was also illustrated.

Acknowledgements

This work was supported by Varian Medical Systems.

References

[1] Burman C, Chui C, Kutcher G, et al. Planning, delivery and quality assurance of intensity modulated radiotherapy using dynamic multileaf collimator: a strategy for large-scale implementation for the treatment of carcinoma of the prostate. Int J Radiat Oncol Biol Phys 1997;39:863–873.

[2] Harms WB, Low DA, Purdy JA, Wong JW. A software tool to quantitatively evaluate 3D dose calculation algorithms. Int J Radiat Oncol Biol Phys 1994;30:187.

[3] LoSasso T, Chui C, Ling CC. Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy. Med Phys 1998;25:1919–1927.

[4] Low DA, Harms WB, Sasa Mutic, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25:656–661.

[5] Mohan R, Wang X, Jackson A, et al. The potential and limitations of the inverse radiotherapy technique. Radiother Oncol 1994;32:232–248.

[6] Pasma KL, Heijmen BJM, Kroonwijk M, Visser AG. Portal dose image prediction for dosimetric treatment verification in radiotherapy: an algorithm for open beams. Med Phys 1998;25(6):830–840.

[7] Storchi P, Woudstra E. Calculation models for determining the absorbed dose in water phantoms in off-axis planes of rectangular fields of open and wedged photon beams. Phys Med Biol 1995;40:511–527.

[8] Storchi P, Woudstra E. Calculation or the absorbed dose distribution due to irregularly shaped photon beams using pencil beam kernels derived from basic input data. Phys Med Biol 1996;41:637–656.

[9] Van Dyk J, Barnett RB, Cygler JE, Shragge PC. Commissioning and quality assurance of treatment planning computers. Int J Radiat Oncol Biol Phys 1993;26:261–273.

[10] Van Esch A, Vanstraelen B, Verstraete J, Kutcher G, Huyskens D. Pre-treatment dosimetric verification by means of a liquid-filled electronic portal imaging device during dynamic delivery of intensity modulated treatment fields. Radiother Oncol 2001;60(2):181–190.

[11] Wang X, Spiridon S, LoSasso T, Stein J, Chui C, Mohan R. Dosimetric verification of intensity modulated fields. Med Phys 1996;23:317–327.