

A method to verify sections of arc during intrafraction portal dosimetry for prostate VMAT

James L Bedford and Ian M Hanson

Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London SM2 5PT, UK.

Email James.Bedford@icr.ac.uk

Abstract

This study investigates the use of a running sum of images during segment-resolved intrafraction portal dosimetry for volumetric modulated arc therapy (VMAT), so as to alert the operator to an error before it becomes irremediable. At the time of treatment planning, predicted portal images were created for each segment of the VMAT arc, and at the time of delivery, intrafraction monitoring software polled the portal imager to read new images as they became available. The predicted and measured images were compared and displayed on a segment basis. In particular, a running sum of images from 10 segments (a “section”) was investigated, with mean absolute difference between predicted and measured images being quantified. Images for 13 prostate patients were used to identify appropriate tolerance values for this statistic. Errors in monitor units of 2-10%, field size of 2-10 mm, field position of 2-10 mm and path length of 10-50 mm were deliberately introduced into the treatment plans and delivered to a water-equivalent phantom and the sensitivity of the method to these errors was investigated. Gross errors were also considered for one case. The patient images show considerable variability from segment to segment, but when using a section of the arc the variability is reduced, so that the maximum value of mean absolute difference between predicted and measured images is reduced to below 12%, after excluding the first 10% of segments. This tolerance level is also found to be applicable for delivery of the plans to a water-equivalent phantom. Using this as a tolerance level for the error plans, a 10% increase in monitor units is detected, 4 mm increase or shift in multileaf collimator settings can be detected, and an air gap of dimensions 40 mm × 50 mm is detected. Gross errors can also be detected instantly after the first 10% of segments. The running difference between predicted and measured images over 10 segments is able to identify errors at specific regions of the arc, as well as in the overall treatment.

Short title: Intrafraction portal dosimetry for VMAT

PACS: 87.53.Bn, 87.55Qr

1. Introduction

Portal dosimetry using an electronic portal imaging device (EPID) is widely used to verify the fidelity of a radiotherapy delivery to the treatment plan in both the pre-treatment context and also in the *in vivo* context. The methods are comprehensively reviewed by van Elmpt *et al* (2008), Mijnheer *et al* (2013) and McCurdy *et al* (2018). Both forward-projection and back-projection methods for EPID dosimetry at the time of treatment are in use for intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT). The forward-projection method predicts the expected images from the treatment plan, accounting for attenuation and scatter within the patient, and compares these images with measured images. Meanwhile, the back-projection method uses the measured images and projects the fluence back into the patient model to calculate the dose within the patient.

Most of these methods operate on a per-fraction basis. In other words, the total dose or image intensity delivered during a single fraction of treatment is verified. This is a valuable step in satisfying the requirements of safety in modern radiotherapy (British Institute of Radiology 2008). However, this approach suffers from the drawback that if an error in the delivery is present, a substantial incorrect dose can be delivered before the fraction data are analysed. As treatments become increasingly hypofractionated, this is more of a limitation. Another approach is therefore to analyse the acquired images as they are obtained, so that an error can be immediately identified and the treatment paused or terminated. Such real-time methods are described by Fuangrod *et al* (2013) for dynamic IMRT, by Woodruff *et al* (2013) and Podesta *et al* (2014) for pre-treatment verification of VMAT, and by Fidanzio *et al* (2014), Woodruff *et al* (2015) and Spreeuw *et al* (2016) for *in vivo* monitoring of VMAT.

Apart from the advantage of promptly alerting to errors, this approach has the benefit of allowing time-resolved image analysis. Analysing images on a segment by segment basis allows the method to have greater sensitivity to changes in patient anatomy between planning and treatment. This is demonstrated by Persoon *et al* (2016) for VMAT treatments with reconstruction of the measured dose using cone-beam CT images. Schyns *et al* (2016) also show the superiority of time-resolved images for identifying anatomical changes that are located separately from the isocentre.

The principal objectives of this paper are threefold: firstly, to present a novel method of carrying out intrafraction portal dosimetry for VMAT using a commercial portal imager and the forward-projection method; secondly: to introduce the concept of analysing each VMAT delivery in sections, so as to obtain reliable information with a reasonably fine time resolution; and thirdly: to demonstrate these methods retrospectively for a cohort of prostate patients. The retrospective analysis is accomplished by firstly reviewing the images for these patients to identify appropriate tolerance levels in the comparison metrics. Both patient images and images acquired on a water-equivalent phantom are used, the latter to investigate both correct treatment deliveries and for deliberately

introduced errors in monitor units, aperture size, aperture position and path length. The speed with which catastrophic errors can be detected is then demonstrated.

2. Methods

2.1. Portal image prediction

The AutoBeam (v5.8) in-house inverse treatment planning system was used to generate VMAT treatment plans. After inverse planning, AutoBeam was used to generate a predicted portal image for each segment in the arc (Bedford *et al* 2018). Dose was firstly recalculated with 100 interpolated apertures between the apertures at the control points themselves, to model the MLC leaf motion between control points. This was necessary to provide the correct predicted portal image accumulated between each pair of control points, rather than simply the portal image predicted from the specific apertures at the control points themselves. The dose distribution in the isocentre plane was then calculated for a water-equivalent environment by deconvolving scatter and reconvolving under conditions of unit density. This step was comparable to that of Wendling *et al* (2012). The predicted images were 512×512 pixels in size, with resolution $0.8 \text{ mm} \times 0.8 \text{ mm}$ at the detector plane, which was located 1600 mm from the source.

The predicted portal image for each segment, together with the total summed image for the full arc and a small amount of header information, was stored as a single binary file. For 180 segments, this was 190 MB in size. Portal image prediction was carried out for the patient planning CT scan itself, and additionally for the same monitor units as the treatment plan but with the isocentre located at the centre of a water-equivalent phantom of dimensions 300 mm long \times 300 mm wide \times 200 mm high.

2.2. Intrafraction portal dosimetry system

The operation of the intrafraction portal dosimetry system was as shown in Figure 1. The basic means of operation was that the iViewGT system (Elekta AB, Stockholm, Sweden) acquired images into an acquisition directory and the in-house intrafraction portal dosimetry software (AutoDose v1.0) read these images and compared them against the predicted images. During normal image acquisition in the iViewGT system, with control via the iCom interface, a new directory was created, containing a frame summary file (.xml) and a series of images (.jpg) referenced by the frame summary file. The frame acquisition rate was around 2 Hz, so image files appeared at this rate, but the frame summary file was updated at a lower frequency, typically about 0.25 Hz. Consequently, several image files appeared before they were referenced by the frame summary. The software was designed to facilitate several attempts to read the frame summary file, since at the time of reading, it could be in use by the iViewGT system for updating. The AutoDose software was designed to complete the basic cycle of image reading, assignment to segments and generation of statistics within approximately 100 ms using a single thread of execution, so that it would successfully complete while the next image

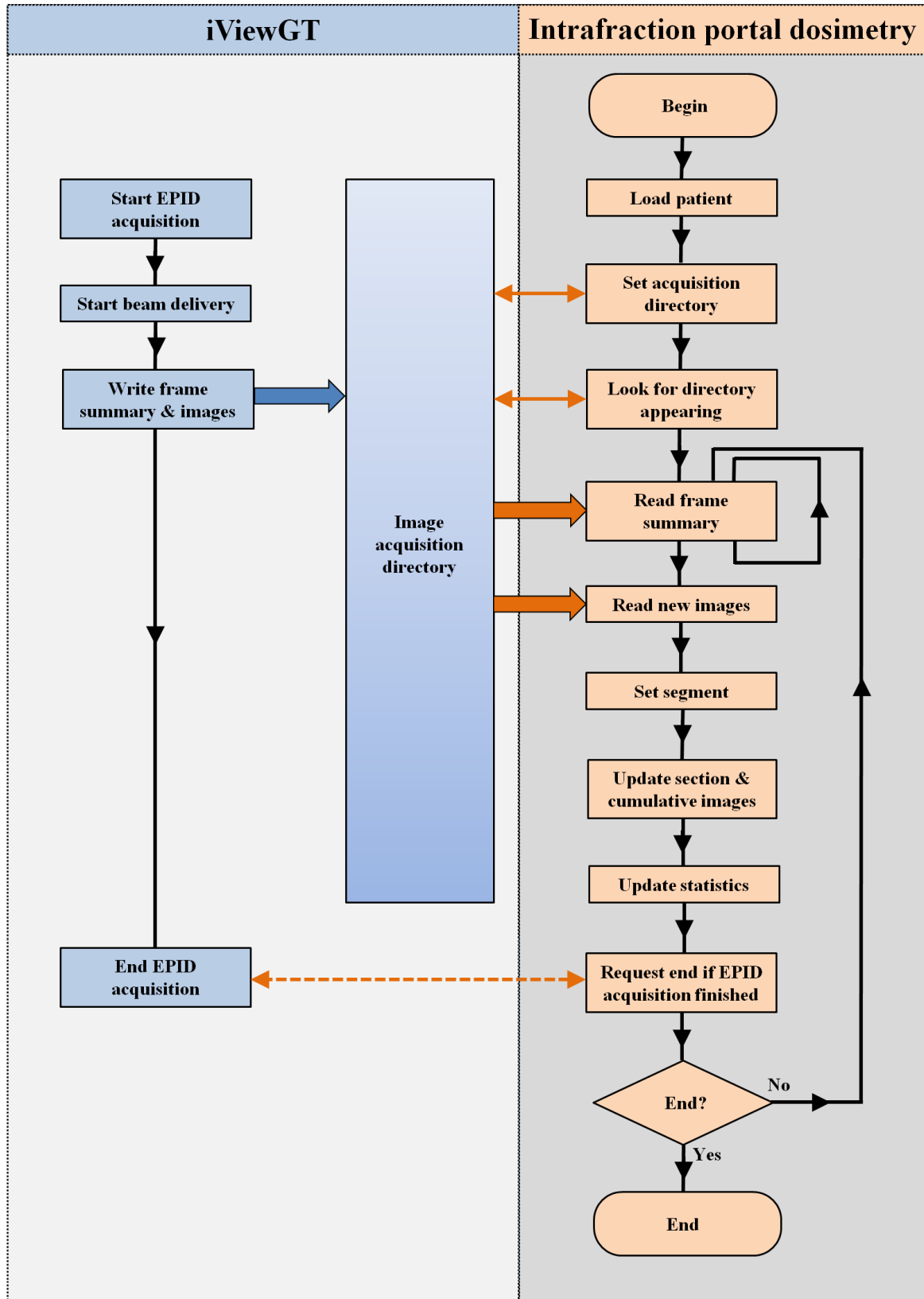


Figure 1. Flow diagram for intrafraction portal dosimetry. The normal image acquisition arrangement is shown on the left and the additional software on the right.

was being acquired by the iViewGT system. Use of a single thread was to avoid allowing the AutoDose program from consuming system resources required by iViewGT for normal operation. The system ran successfully on the iViewGT hardware, which consisted of a 4-core 2.90 GHz i5-4570S processor (Intel, Santa Clara, CA) with 4 GB of memory.

The gantry angle at which each image was acquired was stored in the frame summary file. To establish the impact of any time lag (Mans *et al* 2010), the range of gantry angles in relation to the stated gantry angle was determined empirically by acquiring images of a horizontally positioned steel rule while delivering a VMAT arc consisting of a 200 mm \times 200 mm aperture. The arc was delivered in both clockwise and counterclockwise directions and the actual image angle was taken to be 90° or 270° when the projection of the rule was minimized. Using this method, it was found that the stated gantry angle in the frame summary file corresponded most closely to the end of the image acquisition period. Consequently, the arrangement shown in figure 2 was used for the assignment of images to segments. Each image was taken as extending in gantry angle from the angle of the previous image to its own angle. This range of angles was then assigned to the segments of the arc according to the proportion of the image angle range lying in the gantry angle space of each segment. The total image intensity for a given segment was the total of all individual image intensities assigned to that segment.

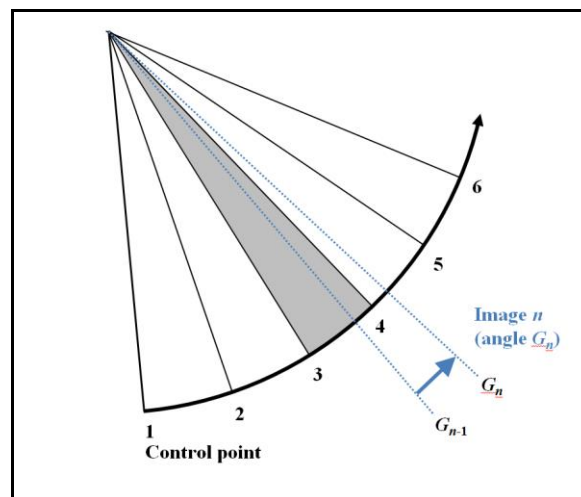


Figure 2. Assignment of an image with gantry angle G_n to the appropriate segment. Segment s spans from control point $s-1$ to control point s . The stated angle G_n is taken to be at the end of the image angle range, while the gantry angle, G_{n-1} , of the previous image is taken to be at the start of the image angle range. In this case, 50% of the image intensity associated with image n is assigned to segment 4, shaded, and the remaining 50% is assigned to segment 5.

2.3. Image statistics

Four types of statistic were used for analysis of the segment-resolved images. These were the central pixel value, the mean image intensity, the mean absolute difference between the predicted and actual images, taken as a percentage of the maximum predicted image value, and the mean absolute difference between the predicted and actual images, taken as a percentage of the local image intensity. For the latter three types of analysis, a threshold of 10% of the predicted image intensity was used.

Three types of temporal analysis were used for the segment-resolved images. The first was the instantaneous value of the image statistics. These segment by segment statistics were found to be very noisy, so the second type of analysis was a running sum of the images relating to 10 segments, referred to as the *section value* as it contained information about a section of the VMAT arc. The section image for segment s was a sum of images from segments $s-9$ to s . The final type of analysis was to view the images cumulatively, *i.e.* the cumulative image for segment s was the sum of all images from segments 1 to s . Due to the large fluctuations in all three types of temporal analysis at the beginning of the arc, the first 10% of segments were excluded from the generation of statistics. The choice of 10% of segments was chosen empirically and was found to give good results in all cases.

2.4. System operation

Figure 3 shows the main user interface for the AutoDose software. It was operated by loading the predicted image file using the file menu and then selecting whether to display the instantaneous, section or cumulative representation of the images. The specific statistic to be used was also set, and the image threshold for the analysis. The caution and error levels were set, the caution level indicating that there was potentially an error, and the error level indicating a definite error. The user selected the image acquisition directory to begin image acquisition. During image acquisition, the traffic light display showed the status of the measurement, according to the maximum error encountered between the predicted and measured signals after the initial 10% of segments. This tricolour scheme was also used in the background to the displayed graph to show exactly at which segments the errors occurred. There were further tabs comparing either instantaneous, section or cumulative images (see inset to figure 3).

2.5. Patients and treatment plans

Thirteen patients who gave consent for their images to be used for research purposes were retrospectively studied. Three gold seed fiducials were implanted into nine of the patients. All patients were CT scanned with a 1.5 mm slice thickness using a Brilliance Big Bore CT scanner (Philips Healthcare, Cleveland, OH), and regions of interest were delineated using Pinnacle³ (v9.10, Philips Radiation Oncology Systems, Madison, WI). Eleven of the patients received 60 Gy in 20 fractions to the most central of three planning target volumes, and the remaining two patients

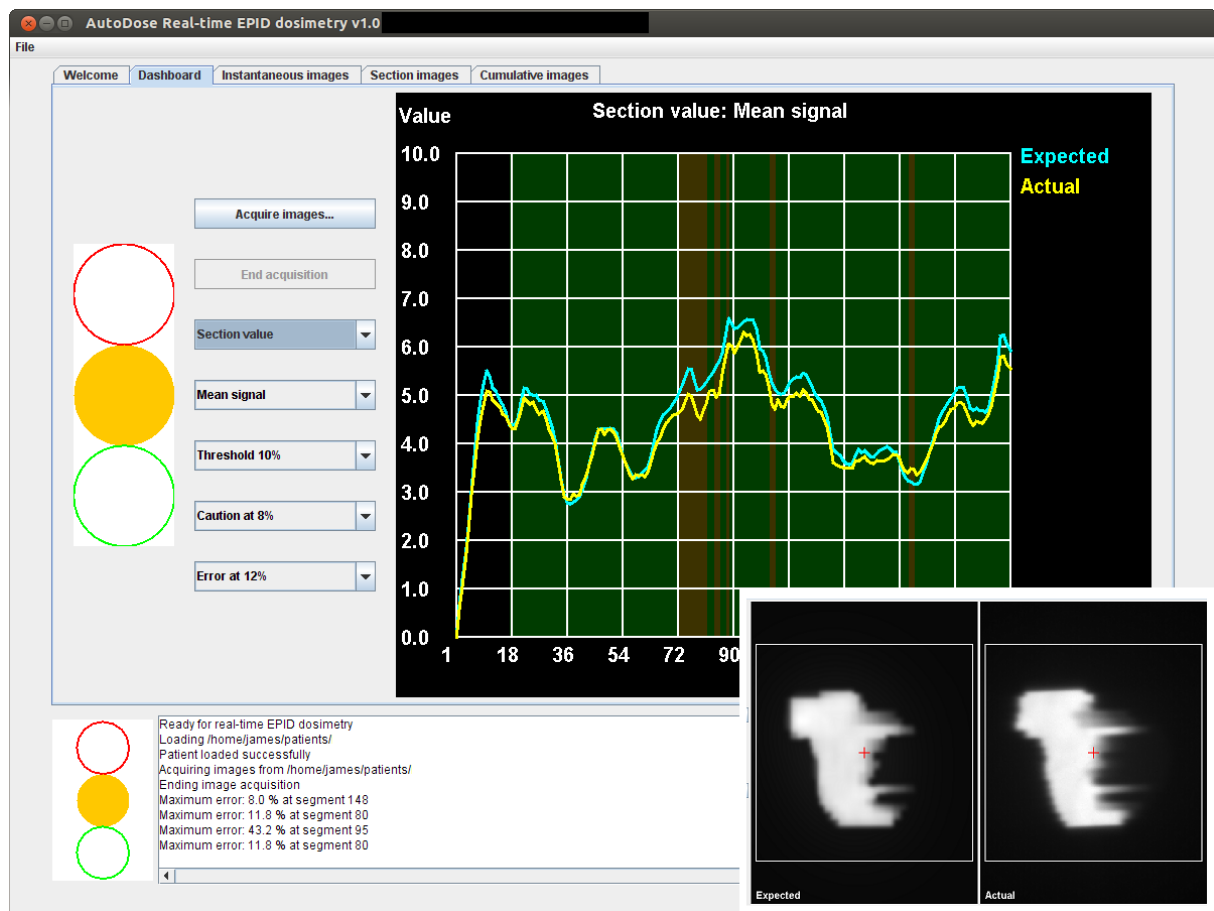


Figure 3. Graphical user interface for the AutoDose software, showing the difference between the predicted and measured images. The inset shows one of the further windows for comparison between predicted and measured images.

received 36 Gy in six fractions. Details of the planning target margins and prescribed dose levels are contained in a previous report (Bedford *et al* 2018).

The in-house treatment planning system AutoBeam was used to create treatment plans for the 6 MV beam of a Versa HD linear accelerator with Agility multileaf collimator (MLC). Plans consisted of a single counterclockwise VMAT arc, from gantry angle 179° to 181° in 11 of the patients, and from 120° to 240° in the remaining two, with control points spaced at intervals of 2° in gantry angle. Collimator angle was 2° in all cases. Absorbed dose was calculated using a fast convolution algorithm on a $4.0\text{ mm} \times 4.0\text{ mm} \times 4.0\text{ mm}$ dose grid and predicted images were then calculated as described above. The plan for each case was then transferred to Pinnacle³ for final dose calculation on a $2.5\text{ mm} \times 2.5\text{ mm} \times 2.5\text{ mm}$ grid. This last step involved a renormalization of up to 1.6% in order to base the plan on the more accurate dose calculation in the clinical treatment planning system. This renormalization was not included in the predicted images, so formed a small systematic error in each case.

2.6. Measured images

The iViewGT system was used for acquisition of all images (Hanson *et al* 2014). This consisted of an amorphous silicon EPID (Perkin Elmer, Santa Clara, CA), which was configured to provide images of size 512×512 pixels over $410 \text{ mm} \times 410 \text{ mm}$ at the detector plane, which was located 1600 mm from the source. A total of 46 image sequences were acquired and analysed for the 13 patients. These images were used to validate the time-dependent analysis in AutoDose and to establish tolerance levels. The plans were also delivered to a stack of Solid Water (Radiation Measurements, Inc., Middleton, WI), with dimensions 300 mm long (superior-inferior direction) by 300 mm wide (lateral direction) by 200 mm high (anterior-posterior direction). The isocentre was located at the centre of the phantom. Further understanding of the sensitivity of the method to different types of error was also gained by deliberately introducing errors into the treatment plans of three of the patients, delivering these plans to the Solid Water phantom and then analysing the images from these. The errors consisted of:

1. An increase in monitor units at all control points of the arc, from 2% to 10% in 2% steps.
2. A retraction of all MLC leaves at all control points of the arc, from 2 mm to 10 mm in 2 mm steps.
3. A shift in MLC opening on all leaves at all control points of the arc, from 2 mm to 10 mm in 2 mm steps.
4. Replacement of a 50 mm slab of the phantom with two 300 mm long (superior-inferior direction) by 150 mm wide (lateral direction) by 50 mm high (anterior-posterior direction) slabs of polymethylmethacrylate, which were moved apart laterally from 10 mm to 50 mm in 10 mm steps, so as to create a medially located air gap. The replacement slab and air gap were situated 20 mm above the couch top, so that the uppermost (most anterior) edge was 30 mm below (posterior) to the isocentre.

The forward-projection EPID dosimetry model used in this study was previously compared with a commercial back-projection method for this set of images (Bedford *et al* 2018) so that the time-integrated performance was well-understood. Finally, the performance of the error detection system in identifying major errors was demonstrated by introducing the following errors into the treatment plan for one patient, delivering the plan to the Solid Water phantom and analysing the images:

1. A major error in which the monitor units for the control point at the centre of the arc were increased by a factor of 10.
2. A major error in which the monitor units for all control points in the last half of the arc were increased by a factor of 2.
3. A major error in which the aperture was changed to $200 \text{ mm} \times 200 \text{ mm}$ for all control points in the last half of the arc.

3. Results

3.1. System performance

Times for reading and processing images for a typical delivery are shown in figure 4. The variability in time is mostly due to the different number of images being acquired and processed for each segment. For this case, the number of images is 375, so there are approximately two images per segment. In practice, the iViewGT image acquisition software writes out image files as soon as the images are acquired, but the frame summary file is not updated so frequently. It is not until the frame summary file has been updated that the acquired images can be processed, since the frame summary contains the gantry angle and scaling information. Consequently, when the frame summary is updated, a number of new images are available and these are referred to in the update. The images typically span around five segments, so the AutoDose software updates around five segments at a time. This cannot be referred to as “real time” but it is sufficiently frequent for practical purposes of treatment monitoring.

3.2. Patient images

Examples of the mean signal viewed instantaneously, as sections (running cumulative images for 10 segments), and cumulatively, are shown in figure 5. The instantaneous images contain considerable variability in the agreement between predicted and measured images and are consequently not found to be very useful. The variability results from the binning of the acquired images into the relevant segments. Although each image is apportioned into one or more segments according to figure 2, some variability remains. The gantry speed is very variable on the Versa HD system, with the gantry occasionally reversing direction for a small part of the arc, and this is likely

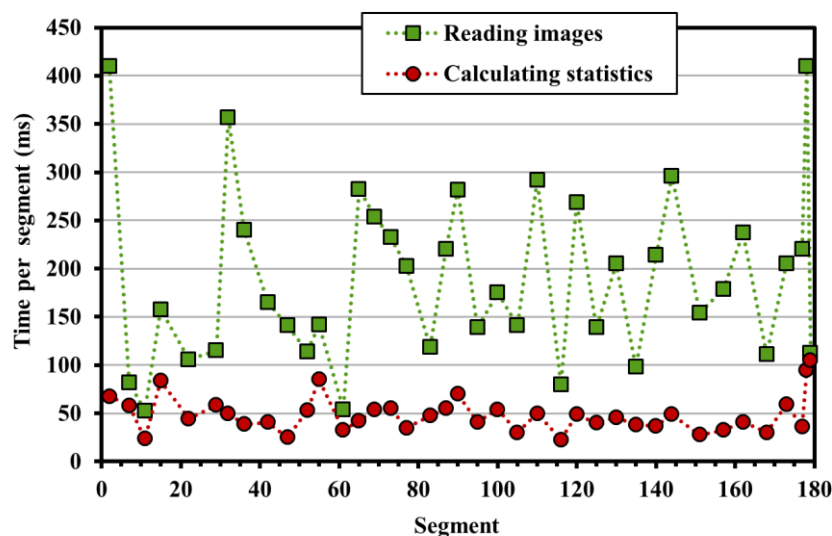


Figure 4. Times taken for reading and processing images during delivery of one of the treatment plans.

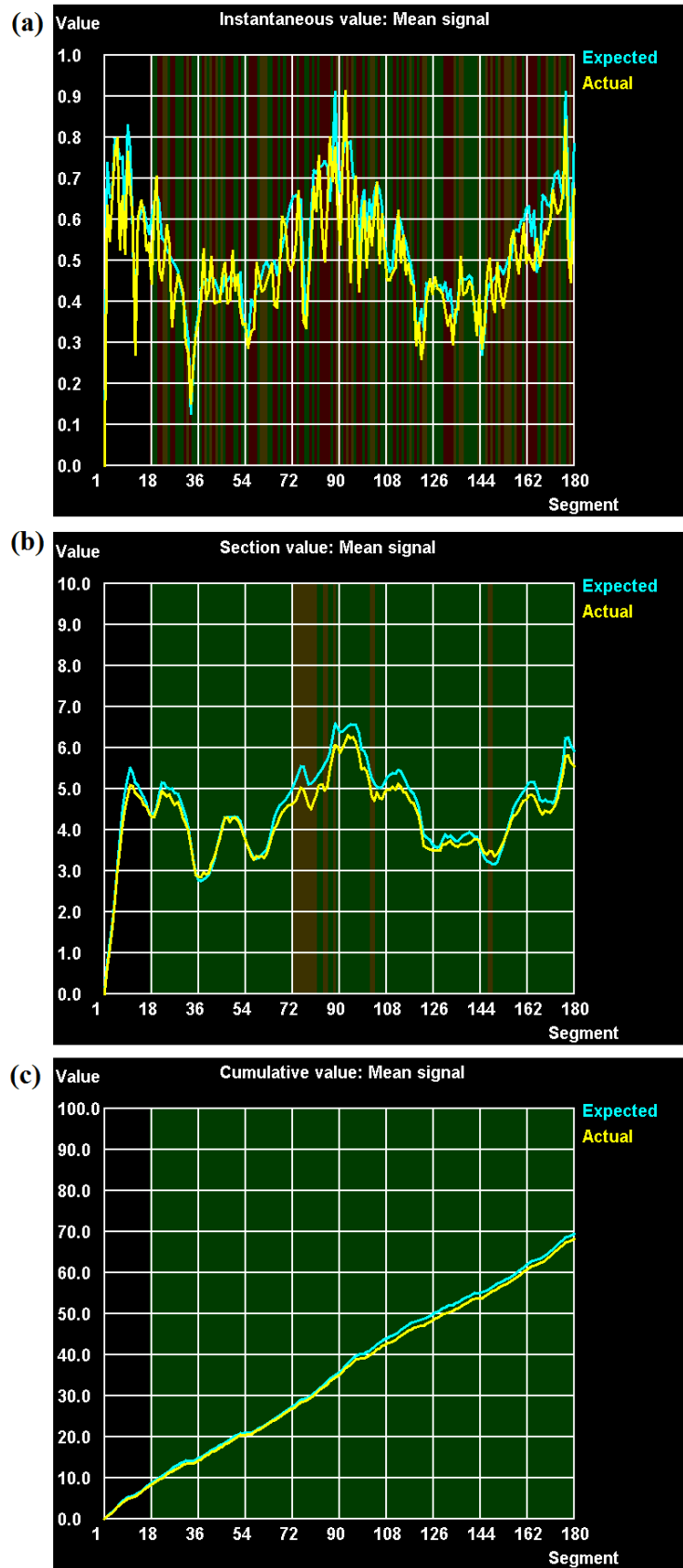


Figure 5. A typical (non-error) patient trace for mean signal with a threshold of 10%. (a) Instantaneous images, (b) section images, and (c) cumulative images.

to contribute to the fluctuation in instantaneous image intensity between segments. Conversely, the cumulative images are mainly useful to indicate whether the overall treatment is accurate, without giving any indication of errors which may occur at specific segments. The work which follows therefore focusses mostly on the section images, which give a relatively stable indication of what is taking place at each part of the treatment fraction.

The mean difference between predicted and measured images, as a percentage of the image maximum, is shown for each of the 13 patients in figure 6. The instantaneous values shown in figure 6a contain too much variability to be useful, due to the difficulty of binning the images into the correct segments. The section images are much more regular from control point to control point (note that the scale differs from that of figure 6a). The smoothest signal is obtained from the integrated images, with progressively improving agreement between measured and predicted images as the delivery proceeds, due to the longer period of accumulation, which tends to cancel noise relating to variability in segment binning. However, this process tends to obscure any true segment-specific errors. There is an observable trend in the results for the different patients, which is due to the use of standardized planning parameters in most of the cases. In particular, during the inverse planning, the control points are collected into groups of 20° gantry angle, so that each set of 10 control points accounts for one sweep of the MLC leaves across the PTV. Two groups of control points, i.e. 20 control points, therefore form a cycle of the MLC from one side of the PTV to the other. This periodicity can be seen in figure 6.

The complete results for all of the patient images are summarized in table 1. The large errors seen for the central signal are due to the predicted intensity being very low at some segments, so that the relative error is large. The instantaneous errors are also large in general. Taking the patient cases to be correctly delivered without error, which is a reasonable assumption given the analysis of the complete data set using both forward- and back-projection (Bedford *et al* 2018) a suitable caution level is 10% and an error level is 12% for the section values of the global mean difference. It is important in the operation of this software that patient treatments are not interrupted unnecessarily, so these levels are designed so that none of the patient cases trigger the error level of 12%.

3.3. Phantom images

The mean difference between predicted and measured images as a percentage of the maximum predicted intensity is shown on a section basis for one of the patient prescriptions delivered to the patient himself and to a water-equivalent phantom in figure 7. A weak correlation between the two deliveries is observable. The complete results for the 13 prescriptions delivered to the water-equivalent phantom are similar to those of figure 6, with the same periodicity being present. Again, the results suggest that a caution level of 10% and an error level of 12% are appropriate.

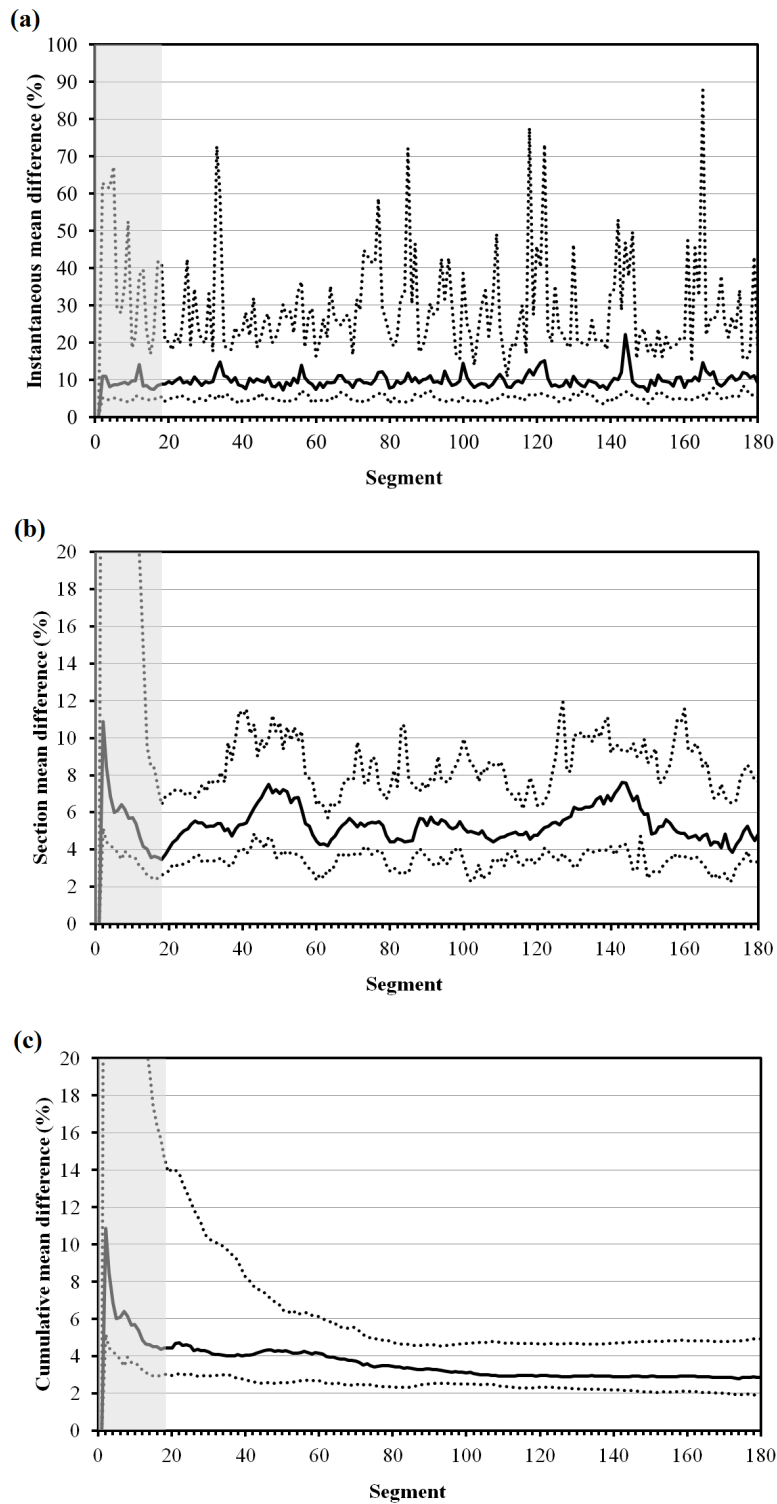


Figure 6. Mean absolute difference between predicted and measured images, as a percentage of the global maximum intensity of the predicted image, for the 46 image acquisitions over the 13 patients as a function of segment number, viewed (a) instantaneously, (b) in sections, and (c) cumulatively. The median, minimum and maximum values are shown in each case. The grey region at the left of each plot shows the segments excluded from the analysis for patients with 180 segments. Note that the scales differ.

Table 1. Distribution of errors for all segments after 10% for all of the patient images.

ANALYSIS	STATISTIC	MEDIAN	MINIMUM	MAXIMUM
		ERROR (%)	ERROR (%)	ERROR (%)
Instantaneous	Central signal	2.1	-87.3	2530.5
Instantaneous	Mean signal	-7.8	-70.2	80.8
Instantaneous	Mean difference global	9.7	3.5	87.9
Instantaneous	Mean difference local	23.4	9.3	187.4
Section	Central signal	0.5	-20.1	27.5
Section	Mean signal	-4.3	-23.4	10.6
Section	Mean difference global	5.3	2.2	11.9
Section	Mean difference local	14.3	6.7	29.7
Cumulative	Central signal	-0.1	-32.8	9.5
Cumulative	Mean signal	-2.4	-19.2	2.9
Cumulative	Mean difference global	3.3	1.9	14.0
Cumulative	Mean difference local	9.6	5.0	23.9

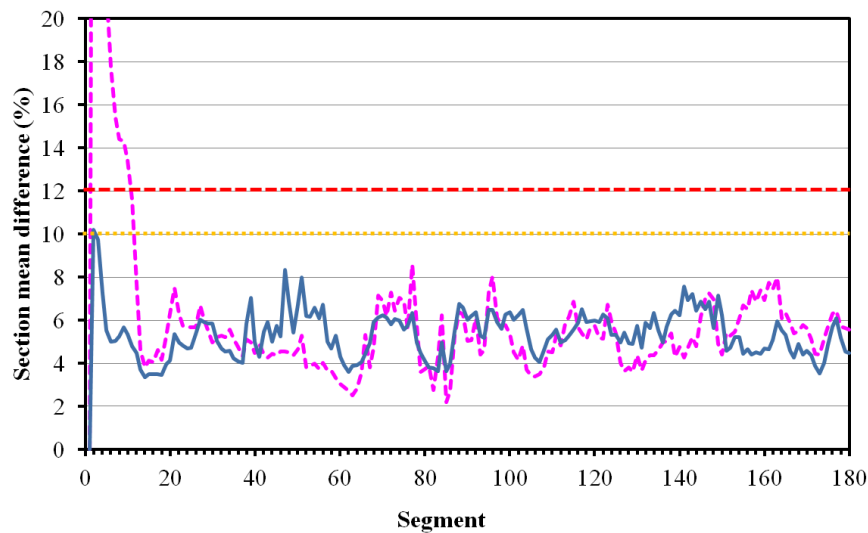


Figure 7. Mean absolute difference between predicted and measured section images as a percentage of the global maximum intensity of the predicted image for one of the patient prescriptions delivered to the patient (blue solid lines) and to a water-equivalent phantom (purple dotted lines), as a function of segment number. The orange dotted line shows the suggested caution tolerance of 10% and the red dashed line shows the suggested error tolerance of 12%.

3.4. Error images

The variations of the maximum value of the section-specific mean absolute difference between predicted and measured images for the various types of deliberately introduced errors are shown in figure 8. The suggested tolerances based upon the results of section 3.2 and 3.3 are also shown. The method is not very sensitive to monitor unit errors, requiring a change of 10% before the method flags up the error unequivocally. The aperture opening error has a larger effect on the mean image difference, with 4 mm of MLC leaf position error in each leaf bank leading to an identifiable increase in difference. An aperture shift of 6 mm likewise leads to an observable increase in mean difference between predicted and measured images. The air gap has relatively little effect on the measured images, but this has relatively little effect on the treatment plan also. The segment at which the error tolerance is first exceeded depends on the magnitude of the error, but in general, if the error is detected at all, it is detected within a few segments of the start of monitoring.

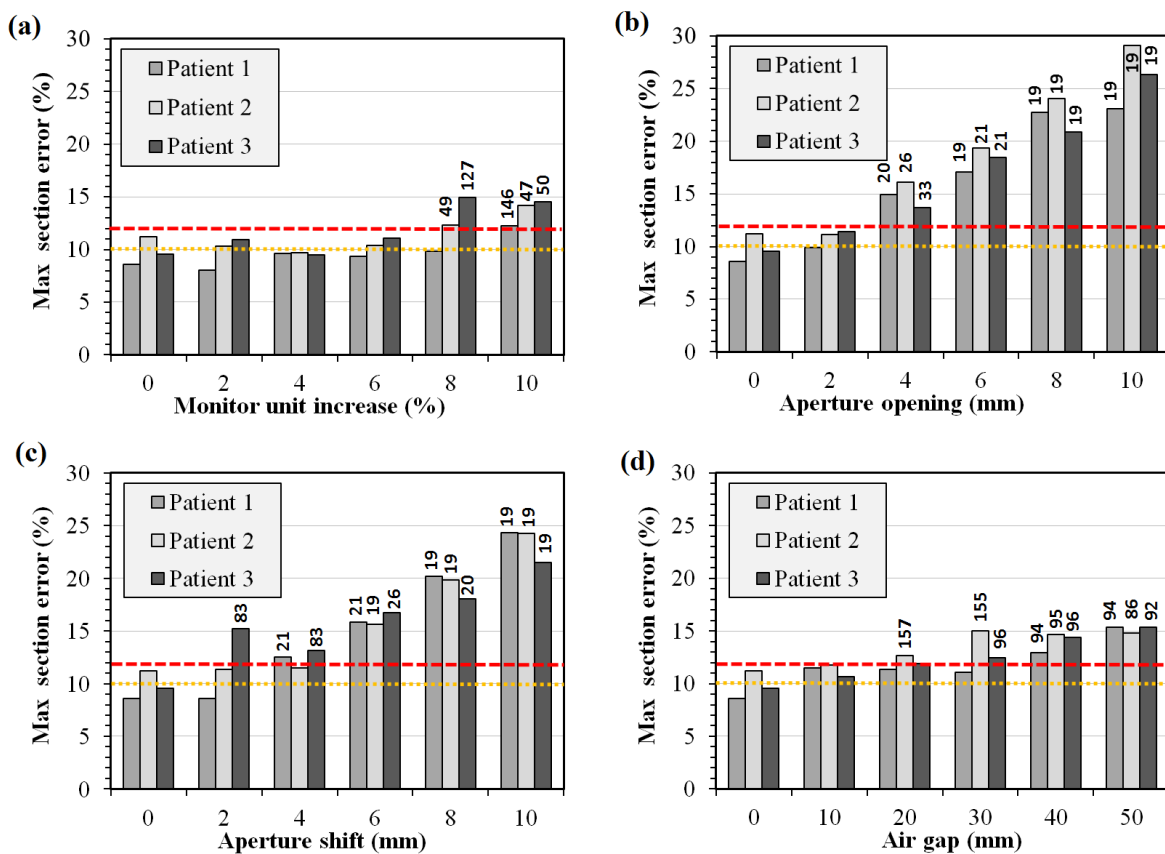


Figure 8. Maximum section-specific mean absolute difference between predicted and measured images, for (a) monitor unit errors, (b) aperture size errors, (c) aperture position errors, and (d) air gap (anatomical) errors. The orange dotted line shows the suggested caution tolerance of 10% and the red dashed line shows the suggested error tolerance of 12%. The numerals indicate the segment (starting at 19) at which the error tolerance is first exceeded (if applicable). The left-most case in each graph is for the delivery without error.

The above errors are present throughout delivery of the treatment plan. However, the catastrophic errors occur midway through delivery. The responses to these errors are shown in figure 9. These major errors are detected by the software immediately.

4. Discussion

The system described in this paper provides information on delivery accuracy during VMAT treatment. In the context of dynamic IMRT fields, a synchronization method is necessary to determine which segment a given acquired image belongs to. However, in the context of VMAT, the gantry angle is stored with the image, so that the segment can be simply determined. The method described above provides image analysis within 100 ms, so as to complete before the next image is acquired. This is achieved using a single-threaded implementation, so as to run on an iViewGT computer

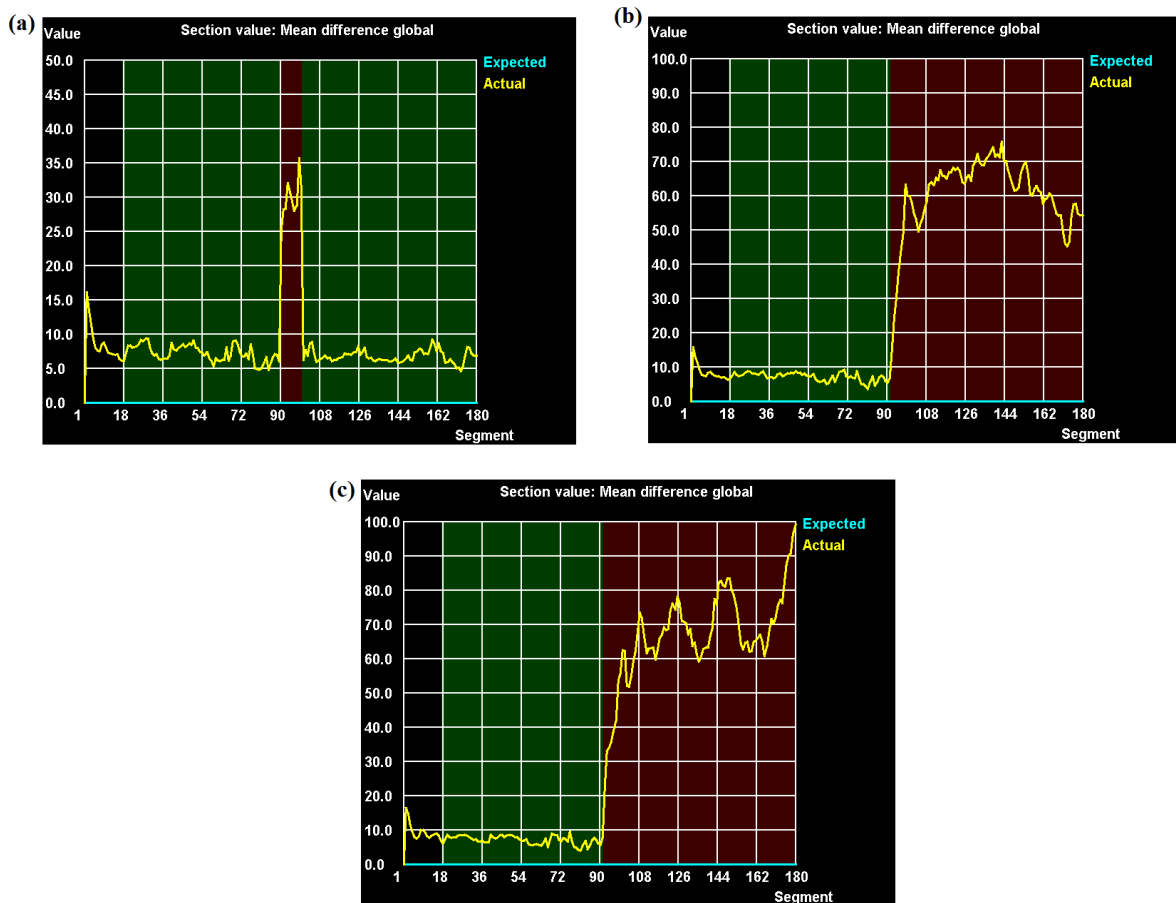


Figure 9. Section-specific mean absolute difference between predicted and measured images, for (a) $10\times$ too many monitor units at control point 90, (b) $2\times$ too many monitor units at all control points after control point 90, and (c) an erroneous $200\text{ mm} \times 200\text{ mm}$ aperture at all control points after control point 90. Threshold 10%, caution level 10%, error level 12%. Note that the vertical scales differ between parts of the figure.

without consuming too many resources. The method is less computationally demanding than the back-projection method of Spreeuw *et al* (2016) and more comprehensive than the point-based method of Fidanzio *et al* (2014).

The performance of the method depends on the accuracy of the underlying image prediction model. The forward-projection method used in this study has been previously compared with a commercial back-projection algorithm so that the performance for analysis of whole fractions is understood (Bedford *et al* 2018). The prostate cases are relatively homogeneous in density but the VMAT aperture shapes are relatively complex for this site, so the overall complexity is moderate. The deconvolution and reconvolution of scatter from the dose in the isocentre plane ensures that the forward prediction method is accurate for inhomogeneous environments such as lung. The planning CT is used for image prediction in this study. Although prediction based on cone-beam CT image might give better agreement with measured images, any anatomical changes between planning and treatment cannot be detected.

The images analysed in this study show a substantial variation in comparison with the predicted images, even for a normal delivery. This results from the variation of accelerator dose rate and MLC leaf positions over the finite period of time taken to acquire a single image frame, as well as image lag or ghosting (McCurdy and Greer 2009) which are not corrected for in this study. In order to minimize the impact of segment-to-segment variation, a given acquired image is divided according to its gantry angle and those of its preceding and succeeding frames into the corresponding range of segments. This approach is also reported by Podesta *et al* (2014) and other studies have used equivalent methods (Woodruff *et al* 2013, Cools *et al* 2017). The gantry speed on the VersaHD accelerator is much more variable than on the Clinac Trilogy accelerator, and it is therefore important to identify any possible errors in gantry angle position. The section images are used to minimize the variability from segment to segment, but the results from each individual segment can always be viewed if required.

Ten segments have been found empirically to be useful for the section images, but other values could be used, such as 2° (Woodruff *et al* 2013). Other studies use frame averaging, which has a similar effect, as well as reducing the number of forward projections required (McCowan and McCurdy 2016). An overall cumulative analysis of the treatment fraction is the extension of this approach to its limit, but its use then precludes a real-time approach and also does not show any errors at specific gantry angles. Using a running cumulative analysis is a possibility, although this is not as sensitive to errors at specific gantry angles as a section approach.

Several methods of image quantification are used in the literature. The gamma index (Low *et al* 1998) is widely used (Mans *et al* 2010, Podesta *et al* 2014, Olaciregui-Ruiz *et al* 2019), together with dosimetric indices in the cases where dose within the patient is being reconstructed by the back-projection method. The chi index (Bakai *et al* 2003) is also used as a faster surrogate for the gamma index (Woodruff *et al* 2015, Fuangrod *et al* 2016). In the present study, a simple percentage

difference is used, partly because of its advantage in terms of speed, but also because it does not tolerate spatial differences in the small VMAT segments which are boosting a particular part of the PTV and which should therefore be accurately positioned. This may result in greater variability in the difference between measured and predicted images.

Understanding the impact of image differences has been facilitated in this work by a retrospective analysis of correct and incorrect deliveries. This is similar to Fuangrod *et al* (2016), who use a process capability index, based on a control limit of several standard deviations of a sample data set to define specific tolerances. Alert criteria adapted specifically for each treatment site (Olaciregui-Ruiz *et al* 2019) or based on specific tests, each focusing on a particular error (Passarge *et al* 2017), could also be used to improve the error sensitivity. Although the true value of the method lies in detecting errors as soon as they occur, the method is also naturally time-resolved, with images analysed according to segment. Schyns *et al* (2016) show that this allows more consistent fail rates for erroneous treatments. In particular, they show that errors which are very apparent at specific parts of a VMAT arc cancel with other errors so that the overall integrated image at the end of treatment is normal. These authors also show that the correlation between gamma fail rates and dose-volume metrics is poor for VMAT. Although cone-beam CT is currently the best method of detecting anatomical changes during treatment, time-resolved EPID dosimetry may also improve the detection of these changes (Persoon *et al* 2016).

The results of deliberately introducing errors into the delivered treatment plans are broadly consistent with other studies. A change in monitor units of 8-10% is needed for the error to be identified clearly, which is slightly less sensitive than when analysing the integrated EPID signal (Bedford *et al* 2018). Fuangrod *et al* (2016) are able to identify a monitor unit error of 5% in the real-time context. Interestingly, the study of Mijnheer *et al* (2018) using back-projection is also not able to detect a 5% change in monitor units in some of the cases examined. As those authors indicate, this error is dependent on the baseline agreement of the predicted and measured images: for example, if the measured image is 2% lower in intensity than the predicted image, a 5% increase in monitor units only raises the signal to 3% too high, which is too small to be detected. In the study of Bojchko and Ford (2015) using the same back-projection method as Mijnheer *et al* (2018), monitor unit errors are relatively easily detected.

The present study is in agreement with the other studies (Bojchko and Ford 2015, Bedford *et al* 2018, Mijnheer *et al* 2018) in that it is most sensitive to MLC leaf positioning errors. This is particularly so for an increase in aperture size, where the increase in output factor contributes to the error as well as the increase in irradiated area of the EPID panel. For this reason, a relatively simple system to detect MLC shapes without comparison of image intensity would be valuable (Fuangrod *et al* 2014). An air gap of the order of 40 mm \times 50 mm is detectable by the method described here, which is comparable with the 10 mm change in overall patient size considered by other authors (Bojchko and Ford 2015, Mijnheer *et al* 2018). These authors show that a change in patient size of

10 mm is generally observable, although for a prostate case, the gamma statistic does not reach the action level. For the major errors considered, representing what might occur should a prescription not transfer correctly, the error is immediately detected.

5. Conclusions

A system has been established, in conjunction with a commercial portal imager, for comparing predicted and measured images at each segment during delivery of a VMAT arc, so as to detect errors before the entire fraction has been delivered. The concept of verifying the arc by section is shown to be useful. Using this concept, the predicted and measured images are compared using the mean absolute difference for a running sum of 10 segments. The method is demonstrated for a cohort of prostate patients, with images acquired during delivery of the plan to the patient and to a phantom. A suitable caution level for the mean absolute image difference over 10 segments is found to be 10% and an appropriate error level is found to be 12%. These levels are conservative, being designed to not cause the termination of any normal treatment. Using these tolerance levels, a range of minor delivery errors can be detected before the end of the fraction, while major errors are detected immediately.

Acknowledgments

The authors acknowledge funding from the National Institute for Health Research (NIHR) Biomedical Research Centre at the Royal Marsden NHS Foundation Trust and the Institute of Cancer Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- Bakai A, Alber M and Nüsslin F 2003 A revision of the γ -evaluation concept for the comparison of dose distributions *Phys. Med. Biol.* **48** 3543-53
- Bedford J L, Hanson I M and Hansen V N 2018 Comparison of forward- and back-projection in vivo EPID dosimetry for VMAT treatment of the prostate *Phys. Med. Biol.* **63** 025008
- Bojchko C and Ford E C 2015 Quantifying the performance of *in vivo* portal dosimetry in detecting four types of treatment parameter variations *Med. Phys.* **42** 6912-18
- British Institute of Radiology, Institute of Physics and Engineering in Medicine, National Patient Safety Agency, Society and College of Radiographers and The Royal College of Radiologists 2008 *Towards Safer Radiotherapy* (London: The Royal College of Radiologists)
- Cools R A M, Dirkx M L P and Heijmen B J M 2017 A novel method for sub-arc VMAT dose delivery verification based on portal dosimetry with an EPID *Med. Phys.* **44** 5556-62
- Fidanzio A, Porcelli A, Azario L, Greco F, Cilla S, Grusio M, Balducci M, Valentini V and Piermattei A 2014 Quasi real time in vivo dosimetry for VMAT *Med. Phys.* **41** 062103
- Fuangrod T, Greer P B, Woodruff H C, Simpson J, Bhatia S, Zwan B, vanBeek T A, McCurdy B M C and Middleton R H 2016 Investigation of a real-time EPID-based patient dose monitoring safety system using site-specific control limits *Radiat. Oncol.* **11** 106
- Fuangrod T, Woodruff H C, Rowshanfarzad P, O'Connor D J, Middleton R H and Greer P B 2014 An independent system for real-time dynamic multileaf collimation trajectory verification using EPID *Phys. Med. Biol.* **59** 61-81
- Fuangrod T, Woodruff H C, van Uytven E, McCurdy B M C, Kuncic Z, O'Connor D J and Greer P B 2013 A system for EPID-based real-time treatment delivery verification during dynamic IMRT treatment *Med. Phys.* **40** 091907
- Hanson I M, Hansen V N, Olaciregui-Ruiz I and van Herk M 2014 Clinical implementation and rapid commissioning of an EPID based *in-vivo* dosimetry system *Phys. Med. Biol.* **59** N171-9
- Low D A, Harms W B, Mutic S and Purdy J A 1998 A technique for the quantitative evaluation of dose distributions *Med. Phys.* **25** 656-61
- Mans A, Remeijer P, Olaciregui-Ruiz I, Wendling M, Sonke J-J, Mijnheer B, van Herk M and Stroom J C 2010 3D Dosimetric verification of volumetric-modulated arc therapy by portal dosimetry *Radiother. Oncol.* **94** 181-7
- McCowan P M and McCurdy B M C 2016 Frame average optimization of cine-mode EPID images used for routine clinical in vivo patient dose verification of VMAT deliveries *Med. Phys.* **43** 254-61
- McCurdy B, Greer P and Bedford J 2018 Electronic portal imaging device dosimetry *Clinical 3D Dosimetry in Modern Radiation Therapy* ed B Mijnheer (Boca Raton, FL: Taylor and Francis) pp 169-98

- McCurdy B M C and Greer P B 2009 Dosimetric properties of an amorphous-silicon EPID used in continuous acquisition mode for application to dynamic and arc IMRT *Med. Phys.* **36** 3028-39
- Mijnheer B, Beddar S, Izewska J and Reft C 2013 In vivo dosimetry in external beam radiotherapy *Med. Phys.* **40** 070903
- Mijnheer B, Jomehzadeh A, González P, Olaciregui-Ruiz I, Rozendaal R, Shokrani P, Spreeuw H, Tielenburg R and Mans A 2018 Error detection during VMAT delivery using EPID-based 3D transit dosimetry *Phys. Med.* **54** 137-45
- Olaciregui-Ruiz I, Rozendaal R, Mijnheer B and Mans A 2019 Site-specific alert criteria to detect patient-related errors with 3D EPID transit dosimetry *Med. Phys.* **46** 45-55
- Passarge M, Fix M K, Manser P, Stampanoni M F M and Siebers J V 2017 A Swiss cheese error detection method for real-time EPID-based quality assurance and error prevention *Med. Phys.* **44** 1212-23
- Persoon L C, Podesta M, Nijsten S M, Troost E G and Verhaegen F 2016 Time-resolved versus integrated transit planar dosimetry for volumetric modulated arc therapy: patient-specific dose differences during treatment, a proof of principle *Technol. Cancer Res. Treat.* **15**(6) NP79-87
- Podesta M, Nijsten S M J J G, Persoon L C G G, Scheib S G, Baltes C and Verhaegen F 2014 Time dependent pre-treatment EPID dosimetry for standard and FFF VMAT *Phys. Med. Biol.* **59** 4749-68
- Schyns L E J R, Persoon L C G G, Podesta M, van Elmpt W J C and Verhaegen F 2016 Time-resolved versus time-integrated portal dosimetry: the role of an object's position with respect to the isocenter in volumetric modulated arc therapy *Phys. Med. Biol.* **61** 3969-84
- Spreeuw H, Rozendaal R, Olaciregui-Ruiz I, González P, Mans A and Mijnheer B 2016 Online 3D EPID-based dose verification: Proof of concept *Med. Phys.* **43** 3969-74
- van Elmpt W, McDermott L, Nijsten S, Wendling M, Lambin P and Mijnheer B 2008 A literature review of electronic portal imaging for radiotherapy dosimetry *Radiother. Oncol.* **88** 289-309
- Wendling M, McDermott L N, Mans A, Olaciregui-Ruiz I, Pecharromán-Gallego R, Sonke J-J, Stroom J, van Herk M and Mijnheer B J 2012 *In aqua vivo* EPID dosimetry *Med. Phys.* **39** 367-77
- Woodruff H C, Fuangrod T, Rowshanfarzad P, McCurdy B M C and Greer P B 2013 Gantry-angle resolved VMAT pretreatment verification using EPID image prediction *Med. Phys.* **40** 081715
- Woodruff H C, Fuangrod T, Van Uytven E, McCurdy B M C, van Beek T, Bhatia S and Greer P B 2015 First experience with real-time EPID-based delivery verification during IMRT and VMAT sessions *Int. J. Radiat. Oncol. Biol. Phys.* **93** 516-22