CZECH TECHNICAL UNIVERSITY IN PRAGUE

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING

DOCTORAL THESIS

Pseudo-3D IMRT Verification with Film and Its Sensitivity to Errors Compared to 2D Methods

Prague 2022 Tereza Hanušová

Bibliografický záznam

Autor	Ing. Tereza Hanušová České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra dozimetrie a aplikace ionizujícího záření
Název práce	Pseudo-3D verifikace IMRT pomocí filmu a její citlivost k chybám ve srovnání s 2D metodami
Studijní program	Doktorský
Studijní obor	Radiologická fyzika
Školitel	Ing. Ivana Horáková, CSc. Státní ústav radiační ochrany, v.v.i.
Školitel specialista	Ing. Irena Koniarová, Ph.D. Fakultní nemocnice Královské Vinohrady
Akademický rok	2021/2022
Počet stran	139
Klíčová slova	IMRT, radiochromický film, gama analýza, radioterapie

Bibliographic entry

Author	Ing. Tereza Hanušová Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Dosimetry and Application of Ionizing Radiation
Title of Thesis	Pseudo-3D IMRT Verification with Film and Its Sensitivity to Errors Compared to 2D Methods
Degree Programme	Doctoral
Field of Study	Radiological Physics
Supervisor	Ing. Ivana Horáková, CSc. National Radiation Protection Institute, v.v.i.
Supervisor Specialist	Ing. Irena Koniarová, Ph.D. Royal Vinohrady Teaching Hospital
Academic Year	2021/2022
Number of Pages	139
Key Words	IMRT, radiochromic film, gamma analysis, radiotherapy

Abstrakt

S příchodem složitějších radioterapeutických technik, jako je radioterapie s modulovanou intenzitou svazku a oblouková terapie s modulovanou intenzitou, vzrostla i potřeba měřit dávkové distribuce v celém 3D prostoru. Na trh se dostalo několik komerčních zařízení, která se pokoušejí měřit a vyhodnocovat 3D dávkové distribuce. Žádné z nich ale nepředstavuje skutečný 3D dozimetr. Všechna zařízení používají softwarové algoritmy, aby 3D dávkovou distribuci vymodelovaly na základě naměřených 2D dat. Uživatel tak může ve 3D kvantitativně srovnat pouze predikovanou a namodelovanou dávkovou mapu. Jediný skutečný 3D dozimetr je polymerní nebo radiochromický gel, ten však stále ještě není vhodný pro rutinní klinické použití.

Cílem této práce bylo vytvořit jednoduchou dozimetrickou metodu, která by umožnila kvantitativní 3D vyhodnocení pouze měřených dat, bez nutnosti použít jakýkoliv rekonstrukční algoritmus. Naše metoda používá vodě-ekvivalentní deskový fantom a radiochromický film, takže je dostupná všem klinikám s využitím jejich stávajícího vybavení. Data jsou zpracovávána v programu MATLAB (The MathWorks, Inc., United States) a kvantifikována jak pomocí 2D, tak pomocí 3D gama analýzy. Stejně tak by však pro tento účel mohl být využit komerční software dostupný v nemocnici. Hlavní výhody naší metody jsou výborné rozlišení v rámci jedné roviny a vodě-ekvivalentnost celého detekčního systému. Hlavní nevýhodou je vyšší časová náročnost každého měření oproti komerčním elektronickým detektorům. Ve srovnání s gelovou dozimetrií je naše metoda přesnější a méně náročná.

Další výhodou této metody je možnost vyhodnotit tutéž naměřenou sadu dat jak ve 2D, tak ve 3D, s použitím stejného softwarového kódu. To neumožňuje žádné komerční zařízení a ve vědecké literatuře také nebylo provedeno žádné porovnání tohoto typu. Porovnání uváděná ve vědecké literatuře jsou ovlivněna inherentními rozdíly mezi 2D a 3D daty a příslušnými softwarovými algoritmy.

Z našich výsledků vyplývá, že 3D hodnocení by mělo být prováděno alespoň při zavádění nových technik nebo nového dozimetrického vybavení do klinické praxe. Dále by mělo být využito pro verifikaci individuálních léčebných plánů, pokud výsledky získané jinými metodami jsou diskutabilní. Testování dávkových distribucí v jediné 2D rovině v 3D prostoru nebo pole po poli nemusí být dostačující kvůli souhře chyb různého typu.

Abstract

With the advent of complex radiation therapy techniques like IMRT (intensity modulated radiation therapy) and VMAT (volumetric modulated arc therapy), the need to measure dose distributions in 3D has increased greatly. Several commercial systems have been released that attempt to measure and evaluate 3D dose distributions. None of them, however, is a real 3D dosimeter. All of them use software algorithms to model 3D dose distributions based on measured 2D data. Thus, the user can quantify discrepancies between the predicted and modeled 3D dose maps only. The only true 3D dosimeter is polymer or radiochromic gel, which is not yet suitable for routine clinical use.

The aim of this work was to develop a simple dosimetric method that would enable quantitative 3D evaluation of measured data only, without the use of any reconstruction algorithms. Our method uses a water equivalent slab phantom and radiochromic film, so it is accessible to all clinics using their existing resources. Data are evaluated in MATLAB (The MathWorks, Inc., United States) and quantified both with 2D and 3D gamma analysis. However, commercial software available at hospitals could also be used for this purpose. The main advantages of our method are excellent spatial resolution within each plane and water equivalence of the detection system. The main disadvantage of the method is some extra time needed for each measurement compared to commercial electronic devices. Compared to gel dosimetry, our method is more precise and less demanding.

Another advantage of the method is the possibility to evaluate the very same measured data set in 2D and in 3D, using the same software code. This cannot be done with commercial equipment and no comparison of this sort has been performed in scientific literature to the author‘s best knowledge. Comparisons from literature are influenced by the inherent differences between the 2D and 3D data and corresponding software algorithms.

Based on our findings, a 3D evaluation is recommended at least when new techniques or new detection equipment are introduced into clinical practice. It is also recommended for individual treatment plan verification where results obtained otherwise are questionable. Testing dose distributions in a single 2D plane in the 3D space or field-by-field might not be sufficient because of the interplay of different types of error.

List of the the the the contents

List of abbreviations

2D	two dimensional
3D	three dimensional
3D CRT	three dimensional conformal radiationrapy
ACPDP	ArcCHECK Planned Dose Perturbation algorithm
ASCII	American standard code for information interchange
CCD	charge-coupled device
CP	control point,,
CT	computed tomography
DD	dose difference
DICOM	digital imaging and communications in medicine
dpi	dots per inch
DTA	distance-with-agreement
DVH	dose-volume histogram
EPID	electronic portal imaging device
FGD	Fricke gel dosimeter
GPU	graphics processing unit
HU	Hounsfield unit
IMRT	intensity modulated radiation therapy
linac	linear accelerator
MLC	multi-leaf collimator
MRI	magnetic resonance imaging imaging
MU	monitor unit
OAR	organ on the risk
OCT	optical computed tomography
PDD	percent depth dose
PGD	polymer gel dosimeter
PMMA	polymethylmethacrylate
PTV	planning target volume
QA	quality assurance
ROI	region of interest
RT	radiation therapy
SAD	source at axis distance
SIB	simultaneous integrated boost
SPD	solid plastic dosimeter
SSD	source from surface distance
STD	standard deviation
TPR	tissue-phantom ratio
TPS	treatment planning system
VMAT	volumetric modulated arc therapy
XCT	X-rays computed tomography

Definition of terms

Gamma score depends	The percentage of points in a 2D or 3D dose distribution where the the gamma index are lower than or equal to 1. In the literature and in clinical practice this parameter is also sometimes referred to as gamma pass rate or gamma passing rate.
Field-by-field verification	Each field of the IMRT (undertensity modulated radiation therapy) plan is measured separately (most often with gantry at 0°) and evaluated with 2D gamma analysis.
Plane-by-plane verification	In this work we refer to plane-by-plane, verification where the whole IMRT plan is measured in a 3D or pseudo-3D manner, but separate 2D planes, in the 3D space (in transverse, sagittal, and coronal directions) are evaluated with 2D gamma analysis only. Or with 3D gamma analysis – specifically with our 3D MATLAB (The MathWorks, Inc., USA) code – but 2D gamma scores are obtained for each plane based on the gamma indices computed with the 3D code.
Pseudo-3D verification	Pseudo-3D refers to the situation when the whole 3D space is partly (not fully) covered with measured points, and the whole plan is delivered at once with original gantry angles,. In this work, pseudo-3D measurements are performed in several different planes of the I'mRT Phantom (Scanditronix Wellhofer North America, USA). When these measurements are evaluated with 3D gamma analysis, a pseudo-3D gamma score is obtained. (Our 3D gamma analysis in MATLAB is not referred to as pseudo-3D because the code is fully 3D.) Individual planes can also additionally be evaluated by 2D gamma analysis.
Pseudo-3D gamma score	The percentage of points in a pseudo-3D dose distribution with gamma index lower than or equal to 1. This is the situation when the whole 3D space is partly (not fully) covered with measured points.
Global (plan) dose verification	The whole treatment plan is delivered (not field-by-field) with original gantry angles.
Local normalization	Specifically,, for our pseudo-3D measurements, local normalization means that, the each plane in the 3D space is normalized to its own dose value. (This is different to the usual definition of local normalization for gamma analysis.)
Global normalization	Specifically for our pseudo-3D measurements, global normalization means that all, points in the 3D space are normalized to the same dose value.

1 Introduction

Intensity modulated techniques such as IMRT (intensity modulated radiation therapy) and VMAT (volumetric modulated arc therapy) have become the standard for modern radiation therapy. Due to their complexity, many clinics perform the ArcCHECK patient-specific plan verification. In some countries, this is even mandatory. Common practice is to deliver planned dose distributions to a commercial QA (quality assurance) device prior to patient treatment and evaluate their agreement with TPS (treatment planning system) predictions using the well,-established gamma analysis method [1]. This can be done either in 2D, evaluating only selected dose planes, or in 3D.

Although commercial devices are capable of performing 3D gamma analysis, the detector arrays (ion chambers, or semiconductor detectors) are always arranged in a 2D manner. Thus, the there are software algorithms, which often require a commissioning procedure, that recalculate the measured dose into 3D. Several drawbacks of this approach have been extensively reported in literature [2-5]. The spatial resolution of such devices is poor and they never perform a true 3D measurement but rather calculation. Thus, the results might depend more on the detector design and calculation algorithm ands models than on the treatment plan and accelerator delivery. Therefore, these methods should preferably rather be called quasi-3D. There is no true 3D dosimeter except for gel, which is too demanding for routine use in the clinic and its accuracy, is still insufficient for clinical purposes [6].

One goal of this study is to find a dosimetric method of plan verification that would address the above mentioned issues. The method should provide measured doses in the whole 3D space with resolution at least comparable to current 2D electronic detectors. That is, comparison of planned TPS doses and measured doses should be based purely on measured results and not on calculation algorithms. Moreover, the method should be accessible to all clinics using existing resources.

Another goal is to give comprehensive explanation of differences between 2D and 3D gamma analysis performed on the same data set and with the same software. To compare, sensitivity of 2D and 3D verification methods to errors, it is important to compare exactly the same data (measured and TPS exported dose planes) with 2D and 3D gamma analysis, which is not possible with electronic detectors without reconstruction algorithms applied. Post-irradiation darkening is also important to use the same code for 2D and 3D gamma analysis and to have control over all parameters, which is not the case with commercial software. An in-house MATLAB (The MathWorks, Inc., USA) script will be developed for this purpose, in our work. This approach will give a clear answer to the question whether current 2D verification methods applied in clinical practice are sufficient to detect errors (whether their sensitivity is sufficient) or whether a 3D method should be used,. And this is the main goal of our study. Although many works have dealt with this issue, e.g. [7-10], none of them used their own 2D and 3D gamma analysis performed on exactly the same *measured* (not reconstructed) data set.

A promising detector with good spatial resolution is radiochromic film. In clinics it is already used for a number of applications,, including Gamma Knife dosimetry [11, 12], robotic radiosurgery [13], brachytherapy [14, 15], in vivo dosimetry [16] or patient-specific plan verification [17, 18], so, no additional resources would be required. Therefore, our method will be based on the widely used EBT3 Gafchromic film (Ashland Inc., USA). Although our evaluation will be done in MATLAB, most types of commercial software available in hospitals are also capable of evaluating film dosimetry.

Therefore, the goals of this work and the activities carried out to achieve these goals are the following:

1) Develop a new pseudo-3D IMRT verification method
   - Establish film dosimetry for IMRT verification in 2D under site-specific conditions (including in-house written scripts, for evaluation)
   - Compare the established film dosimetry method to current clinically used alternatives in 2D
   - Develop a pseudo-3D film dosimetry method for IMRT verification (including evaluation software)
   - Compare the established film dosimetry method to the current clinically used alternative in 3D
2) Compare the new pseudo-3D method to current clinically used IMRT verification tools (different QA systems will be compared)
   - For clinical IMRT and VMAT plans as well as plans with introduced errors
3) Compare 2D and pseudo-3D verification on the same set of data and with the same software algorithm (using only one QA system - the developed method)
   - For clinical IMRT plans as well as plans with introduced errors
   - Evaluate sensitivity[1] to errors and draw recommendations for clinical practice

[1] Sensitivity in this work is not a statistical measure of the performance of a binary classification test because the type of data obtained do not allow to perform such a quantitative evaluation.

2 Theoretical background

2.1 Radiochromic film dosimetry

Properties of different types of Gafchromic detectors (today available from Ashland), particularly of the successive generations of EBT film (EBT, EBT2, EBT3 and EBT-XD) are different. For radiotherapy purposes, the most widely used radiochromic film at the moment is the EBT3 type (used in this work), which suits best the needs of radiotherapy applications and works well in the desired dose range. For higher-dose applications, the EBT-XD film is also used.

EBT3 is a self-developing film suitable for use in high-energy photon beams. The dose range that can be covered with EBT3 is 0.2 Gy – 10 Gy, and making it well suited for radiation therapy applications where doses around the 1-2 Gy are most often measured. Each film sheet comprises two 125 µm thick polyester substrates and a 28 µm thick active layer consisting of, among others, the active component and a marker dye. The active monomer is the lithium salt of pentacosa-10,12-diynoate (LiPAD). Ionizing radiation initiates solid-state polymerization of the monomer, resulting in film darkening that is visible immediately after irradiation and is proportional to dose. However, the the manufacturer recommends to allow stabilization of the response for at least 6 hours before quantification of dose. Different papers, on the other hand, report quite different stabilization time, periods for various reasons [19-22], ranging from a few minutes to 72 hours. It has been shown, to change rapidly in the first 6 hours after irradiation and almost stabilizes at 24 hours. The change in net optical density from 24 hours to 48 hours is about 2 % [23, 24].

Compared to the previous OCTAVIUS EBT2 film, EBT3 film's active layer contains microscopic silica spheres on the surface to eliminate Newton Rings. The side orientation on the scanner should be eliminated in the EBT3 film due to a symmetrical layer configuration [25]. However, film response still depends on its orientation with respect to scanning direction because the active component has a hair-to be structure [26].

The advantages of radiochromic film for radiotherapy applications including IMRT and VMAT plan verification, are, above all the high spatial resolution and near-water equivalence. Film is a natural 2D dosimeter, unlike arrays of point detectors such as ionization chambers or silicon diodes, where the perturbation effects to conditions in water are not negligible. The manufacturer also declares energy independence, (less than 5% in optical density between 100keV and 18 MeV), dose rate independence (less than 5% in net optical density for exposures with 3.4 Gy/min and 0.034 Gy/min) and good stability in room light, although long exposures to visible light and temperatures above 25°C are not recommended. This has been confirmed in scientific papers [27, 22].

Quantification of dose must be done with a flatbed 48 bit (16-bit per channel) RGB color scanner suitable for EBT3 film scanning. The recommended by manufacturer is EPSON Expression 11000XL or 10000XL Photo scanner (Seiko Epson Corporation, Japan). But different scanners work well with film according to literature [25, 28], such as EPSON V700, EPSON V750, or the newly introduced EPSON V800 and EPSON 12000XL [29]. Flatbed scanners are equipped with a light source, which may be a fluorescent lamp or light emission diodes, and a light detector, usually a charge-coupled device (CCD). They can be used in reflection mode, which is less common in film dosimetry, or transmission mode. The pixel values obtained by the scanner are then converted to dose values with a proper calibration function. Alternatively, optical density (OD) can first be calculated from scanner values (SV) with the formula

$$OD = -log\left(\frac{SV}{SV_{max}}\right)$$

and then the dependence of OD on dose can be fitted with a calibration curve. Rational functions are often used to fit the calibration points with a curve because they represent well the physical behaviour when the film is exposed to higher doses (where the response reaches almost a constant value). This type of calibration can be used for example in the manufacturer's commercial software FilmQA Pro (Ashland Inc., USA). On the other hand, polynomial functions would not be valid outside the the calibration range. Polynomial functions also require more dose points. In the literature, however, polynomial functions are more often used. According to [25] this is because they work well both for transmittance and reflectance measurements as well as optical density conversion to dose, unlike rational functions. There is no need to work with high-doses and outside the calibration range in clinical practice. In addition to rational functions and polynomials of different order, other types of functions can be used [29].

Conversion of scanner response to dose can be done using only one of the color channels, [25, 26, 29], usually the red one, or using the multichannel approach [31], where either two [32-34] or all three channels are combined. It is generally assumed that multichannel calibration method [31] gives better results than single-channel (usually using the red channel) [31, 32, 34-36]. The multichannel approach enables to use the yellow marker dye incorporated in the active layer for nonuniformity corrections. It can compensate for the non-uniform response of the scanner as well as finger-prints, dust, and scratches [31, 34, 36, 37]. The multichannel method separates the real dose signal from any non-dose based contributions to the signal. It uses all three color channels and each of the channels has its own calibration curve. It is assumed that the magnitude of dose does not depend on the selection of a particular color channel, so the dose is obtained from each color channel and dose differences between the channels are minimized, thus obtaining the real dose map. However, without the commercial software available, the multichannel method is not very simple to implement and would not be preferred for routine clinical purposes. While the most simple approach might be to take the (weighted) average of dose derived from all channels [34, 38], the state of the art methods use different perturbation models published by Méndez et al. [32, 39] or Pérez Azorín [40]. Scientific literature continues to use the single channel method as well because better results and/or lower uncertainties can be achieved depending, on the particular situation, for example on the scanner type or magnitude of dose [25, 37, 41, 42]. As a reported in the review by Devic et al. [25], the widely implemented Micke [31] method is actual doses just a special case of a more general multichannel method described by Méndez et al. [32] and the multichannel method can produce larger uncertainties than the single channel method. The overall uncertainty of film dosimetry depends on many aspects in the irradiation, calibration and evaluation process, including the evaluation software itself. While for single channel dosimetry, several works have proposed how to calculate the uncertainty budget [43-47], for multichannel approach this is very difficult [29, 48] and papers on this topic are actually lacking [29]. Vera-Sánchez [41] have proposed a Monte Carlo approach for uncertainty analysis that can be applied both for single channel and multichannel dosimetry. Also,, the selection of the multichannel model has affect on the accuracy of results [49]. If a wrong model is chosen, errors of the different color channels can be combined, possibly increasing the overall error [36]. The original Micke [31] model is implemented in the commercial FilmQA Pro software (Ashland) while the Multigaussian method of Méndez et al. [39] is implemented in a cloud computing web application Radiochromic.com [50].

Flatbed scanners used for film dosimetry should be characterized for nonuniformity, noise, short and long term reproducibility and effect of scanner warm-up. Some authors also advise checking geometrical distortion as well [37]. As higher level of noise is associated with better resolution [48, 51], 72 dpi scanning resolution is advised. The level of noise may also impact the results of gamma analysis [52], therefore, application of denoising techniques is advised. This can be done by taking an average response of multiple scans of the same film piece or by applying a smoothing filter [25, 36, 51, 53]. Interestingly, newly introduced LED-based scanners, such as EPSON V800 and EPSON Expression 12,000 XL, are reported to have a more pronounced nonuniform response compared to the fluorescent lamp scanners, such as EPSON V700 and EPSON Expression 11,000 XL [54, 55]. The LED based scanners, on the other hand, tend to have lower random noise [41, 48]. Other important characteristics to consider is the so-called lateral response artifact (LRA). It means that the response of the scanner in the direction perpendicular to the scanning axis parabolically decreases [56, 57] and this is more pronounced for higher doses [42]. While one way to overcome this effect is the use of the multichannel approach, several methods for LRA reduction have been proposed in literature [36, 58-60].

Another issue to mention is the inter-scan variability of film-scanner response which is caused by the instability of the CCD detector and electronics of the scanner as well as film to light distance (curling of the film on the scanner bad) [61]. This can be overcome by using a control film piece, by recalibration methods and by using a glass plate on top of the films placed on the scanner bad (or alternatively, by sticking the film to the scanner bad by tape) [61-63].

To overcome the non-uniformity of film itself, some authors advise to scan each piece before irradiation and to subtract the response of the blank film from the irradiated one [37].

Even though a code of practice from the AAPM (American Association of Physicists in Medicine) has been published in 2020 [28], which updates the previous publication from 1998 [20], it seems reasonable, having in mind all the above mentioned issues, to establish our own methodology for film dosimetry that is relevant with regard to the particular equipment available in our laboratory and to the clinical settings chosen for the experiments. The AAPM report, in our opinion, does not give the user enough guidance and detailed explanations, which are necessary for successful practical implementation of film dosimetry in the clinic. Further in the text, our own metodology for film handling, film calibration and evaluation is developed including the estimate of uncertainties. These are important with regard to gamma analysis criteria, particularly the dose difference criterion, chosen for our further experiments. Our methodology aims to be as simple as possible to be easily implementable in routine clinical practice.

2.2 Current 3D dosimetry systems

Commercial QA devices for 3D dosimetry usually consist of an electronic detector array and evaluation software, sometimes accompanied by a suitable phantom. The following are examples of such devices currently on the market: COMPASS (IBA Dosimetry GmbH, Schwarzenbruck, Germany), Delta^4 (ScandiDos, Uppsala, Sweden), OCTAVIUS system (PTW, Freiburg, Germany), ArcCHECK with 3DVH (Sun Nuclear Corporation, Melbourne, FL, USA), DosimetryCheck (MathResolutions, LLC, Columbia, MD, USA). There are also commercially available 3D gel dosimeters, but these are rarely used in clinical practice.

3D gel dosimeters

The only true 3D dosimeter with very good spatial resolution and water equivalence is polymer gel dosimeter (PGD), Fricke gel dosimeter (FGD), and solid plastic dosimeter (or solid radiochromic polymer, SPD). In PGD, polymerization of monomers occurs after irradiation. This leads to different optical, mechanical, and magnetic properties compared to nonirradiated material [64, 65]. In FGD, new ferrous ions are created after irradiation, which also changes its magnetic properties. The concentration of ions is proportional to absorbed dose [66]. After irradiation of SPD, both polymerization and change of colour takes place [67]. While PGD and FGD can be quantified with magnetic resonance (MRI), SPD cannot. The change in relaxation time is usually quantified for PGD, while for FGD the change in T1 relaxation time is more pronounced. Other options for dosimeter quantification are optical computed tomography (OCT) or X-ray computed tomography (XCT), which is a promising option thanks to the availability of diagnostic CT (computed tomography) scanners in hospitals. An OCT commercially available system is Vista^{TM} by Modus Medical, Inc., USA or OCTOPUS by 3D Dosimetry, Inc., USA. The OCT system uses a thin laser beam and a point detector, which are moved along the transverse and longitudinal direction while the sample is rotated. It can also use a mirror to sweep the ray of light to speed up the process. Other systems use a broad cone beam of laser light together with a CCD camera as the detector. The XCT can be used for PGD because its density is changed after irradiation. The quantification of the density change is based on different photon attenuation in the material [6].

The 3D image must be reconstructed and calibration between the change of dosimeter properties (optical density, concentration of ferrous ions, etc.) and dose must be performed.

Gel development and applications in the medical field have been studied for more than 20 years now [6] but are rarely used in routine clinical practice. The main causes according to [6] are low precision compared to other options (at best 5% uncertainty with 95% confidence), time needed for preparation and evaluation of the dosimeter and its high cost (particularly for routine applications such as verification of IMRT and VMAT plans). Commercially available gel dosimeters are the following: Presage (Heuris Inc., USA), ClearView (Modus Medical, Inc., USA), BANG, and CrystalBall (3D Dosimetry Inc., USA). Commercial analysis software also exists which is not only capable of processing the raw image data obtained with MRI, OCT, or XCT and converting them to dose, but also has other features like comparison of measured dose to calculated dose with gamma analysis or dose volume histograms. Such software comes with the previously mentioned OCTOPUS system by 3D Dosimetry, Inc. Another one is PolyGeVero by GeVero Co., Poland and RTcompare by RTSafe, P.C., USA.

ArcCHECK and 3DVH

ArcCHECK is an acrylic cylindrical phantom, 21 cm in diameter and with an air cavity of 15 cm in diameter in the center to decrease weight (see Figure 2-1). It is designed both for patient-specific plan verification (ideal for VMAT or TomoTherapy, Accuray Inc., USA) or for quality assurance tests. It comprises 1386 diodes with an active volume of 0.019 mm^3 and 1 cm spacing. The detectors are incorporated in the wall of the cylinder. Thus, when irradiating the phantom from any direction, the beam travels through 2 layers of diode detectors, one on the entrance side of the beam and the other on the exit side. The diodes are shifted relative to each other to improve spatial resolution. Each 10 x 10 cm^2 area of ArcCHECK actually contains 221 detectors for monitoring beam's eye view

doses. This can be further improved by a merge feature that comes with the ArcCHECK software, resulting in a coverage of 442 detectors in an area of 10 x 10 cm^2. In contrast to the OCTAVIUS 4D system, ArcCHECK does not comprise any inclinometer and gantry angle is independently calculated from the measured entrance and exit doses. The advantage of the ArcCHECK layout is that both high-dose regions at the entrance side and low-dose regions at the exit side are measured. Errors in beam delivery should manifest in a similar way at both sides, however, the TPS dose calculation in high dose as well as low-dose regions can be checked. The disadvantage of the phantom layout for IMRT or VMAT plan verification is that the basic phantom measures the dose regions around PTV (planning target volume) and usually outside PTV. However, the manufacturer also offers solid inserts where different types of detectors (ionization chambers, film) can be placed to measure doses closer to the isocentre.

Regarding the software by Sun Nuclear Corporation, Melbourne, FL, USA that comes with ArcCHECK, either a 2D evaluation can be performed on the cylinder shell unfolded into a 2D plane (see Figure 2-1) or a 3D evaluations, including 3D gamma analysis, can be performed in the whole 3D space. It is evident that for both 2D and 3D evaluation, a software algorithm is needed to create a 2D dose distribution or a 3D dose distribution from the measured doses on the cylinder shell. The software needs the patient plan and the patient dose from the TPS as input to be able to calculate a 3D volumetric dose distribution based on the measured dose points. Even the 2D unfolded dose map must be corrected, for example for the curved distance between diodes. ArcCHECK also needs some sort of commissioning. One of the most important steps that can bring along systematic errors as high as 8% [68] is the assignment of the phantom density in TPS. The manufacturer recommends to assign a density close to PMMA (polymethyl methacrylate), i.e. 1.19 g/cm^3 and treat the phantom as a homogeneous medium. Absolute dose calibration is also needed based on a reference ion chamber response. Angular dependence of the diodes and field size dependence have been reported [69, 70].

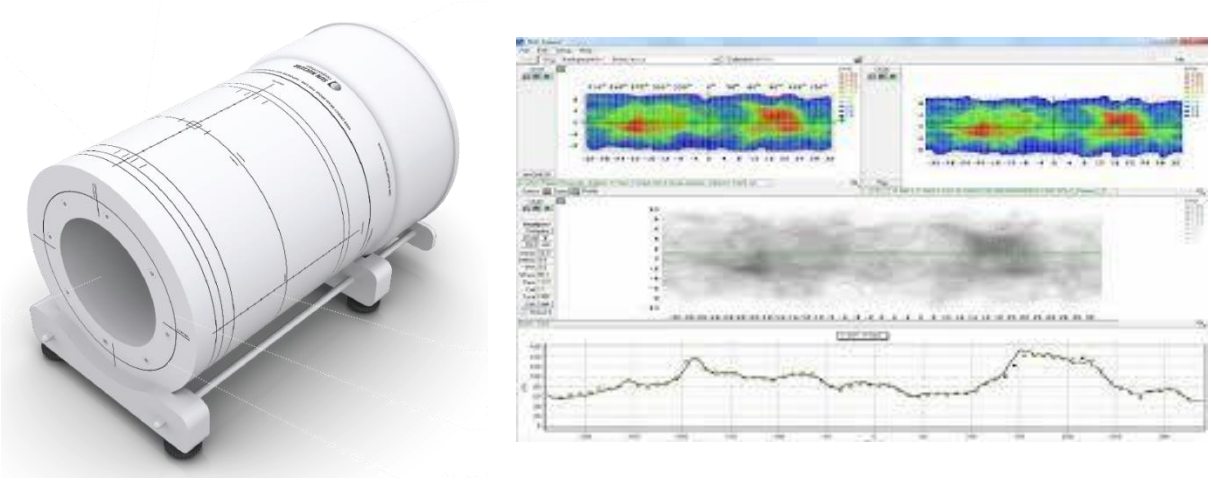


Figure 2-1: The ArCHECK cylindrical phantom and unfolded 2D dose distribution during evaluation [71].

Optionally, 3DVH software (Sun Nuclear Corporation, Melbourne, FL, USA) can be purchased where ArcCHECK measurements are used to calculate dose distribution in a patient geometry. Thus, the user can compare for example planned and delivered dose-volume histograms (DVH). The algorithm responsible for measurement-guided dose reconstruction in a patient CT data set is called ArcCHECK Planned Dose Perturbation (ACPDP). To use it, it is necessary to perform irradiation

with the PMMA plug inserted. [72]. The data acquired by ArcCHECK are inherently time-resolved. ACPDP is a convolution algorithm using the time-resolved measured data and linac-specific and energy-specific beam model acquired on a set of machines of the same type. The algorithm is comprehensively described in [72]. For calculation of dose in patient, it requires the following inputs in DICOM format (digital imaging and communications in medicine): patient plan, structures, dose, axial images, ArcCHECK plan, dose and 4D measurement and the ACPDP model. Results are not independent of the TPS and the same correction factors are used for patient dose calculation as for ArcCHECK phantom dose calculation [73].

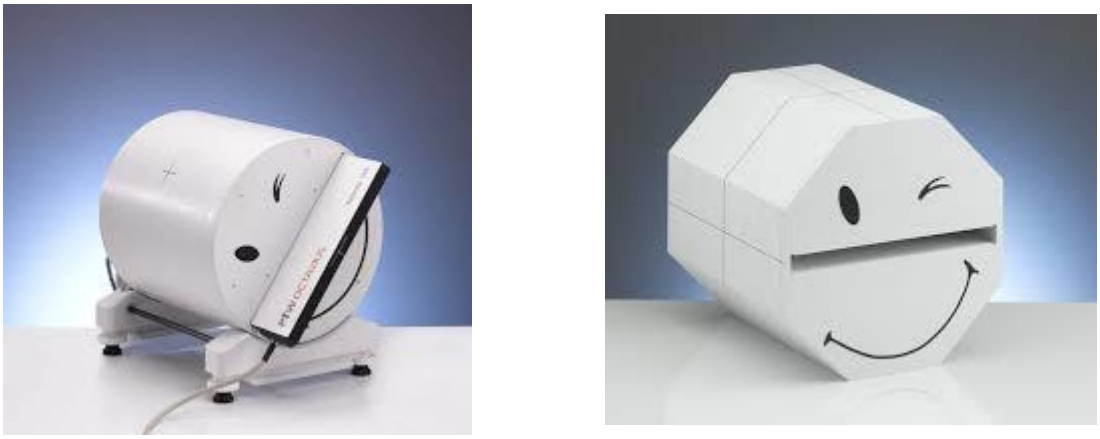
The OCTAVIUS system

OCTAVIUS 4D is a dosimetric system consisting of a cylindrical phantom and a 2D array of ionization chambers (see Figure 2-2 (a)) [74]. Different types of PTW arrays can be inserted – OCTAVIUS 729 with 729 ionization chambers, OCTAVIUS 1500 with 1405 chambers and OCTAVIUS 1000 SRS for stereotactic applications with 977 liquid ionization chambers covering a smaller area than the previous two. The spatial resolution of these arrays is: 10 mm for OCTAVIUS 729, 7.1 mm for OCTAVIUS 1500, and 2.5 mm (5 mm on periphery) for OCTAVIUS 1000 SRS. The OCTAVIUS 4D phantom rotates synchronously with the gantry and uses an inclinometer to read the gantry position. It is designed for VMAT techniques and the measured data are time-resolved and correspond to a specific gantry angle. At measurement, the beam is always perpendicular to the detector plane and there is no need for angular dependence corrections. This was a major drawback of the previous release designed for IMRT – a static OCTAVIUS II phantom with an octogonal shape (see Figure 2-2 (b)).

The algorithm for 3D dose reconstruction in the phantom volume is based on percent depth dose curves, (PDD) obtained for the phantom at commissioning. They should be measured in a water phantom at SSD (source to surface distance) 85 cm and for several field sizes [73]. As a part of commissioning, PTW also recommends to set the phantom relative electron density in the TPS to 1.016 [73]. For 3D dose calculation, the algorithm uses following information: dose points measured with any of the OCTAVIUS detector arrays, field size (also determined from measurement), gantry angle (known thanks to the inclinometer) and a PDD curve converted from water to the phantom material for the corresponding field size. For each measured point the algorithm actually reconstructs the dose along one ray line based on the appropriate PDD values. This is done for all points and all gantry angles resulting in a 3D dose grid. Scatter is also accounted for. For one detector point it is done with the help of the surrounding detectors.Thus, the algorithm is independent of data from the TPS, unlike ArcCHECK and COMPASS.

VeriSoft (software by PTW, Freiburg, Germany) can also optionally reconstruct the 3D dose in the patient. It is done similarly as in the phantom but the algorithm takes into account variable patient SSD and inhomogeneities in patient tissues. Patient CT data must be imported into the software to be able to do this. The algorithm additionally uses or computes the following information: TPR (tissue phantom ratio - calculated from PDD), electron densities — converted from Hounsfield units (HU) — and water depths. It is again possible to obtain dose volume histograms in VeriSoft based on the actually delivered or to overlay the passing and failing points (from gamma analysis) over the patient‘s CT image. Here, unlike ArcCHECK, the patient dose in VeriSoft is calculated independently of the TPS.

Regarding gamma analysis, VeriSoft (in combination with the OCTAVIUS 4D system) offers slice-by-slice analysis where 2D planes are extracted from the calculated dose volume and compared to corresponding 2D planes from the TPS (either in transverse, sagittal or coronal direction). 2D gamma analysis can be performed in each plane as well as 3D gamma analysis in the whole measured and reconstructed dose space.



(a) (b)

Figure 2-2: The OCTAVIUS 4D phantom (a) and the older version OCTAVIUS II (b) [75].

Delta^4

Delta^4 Phantom+ is a dosimetric system consisting of two orthogonal 2D arrays of semiconductor p-Si detectors mounted in a cylindrical phantom. The phantom is 22 cm in diameter and 40 cm in length (see Figure 2-3). The detector planes can either be positioned under 45° to the coronal plane or directly into the coronal and sagittal planes. There are 1069 detectors in total. The detection area of each of the planes is 20 x 20 cm^2. The spacing between detectors is 5 mm in the central 6 x 6 cm^2 area and 10 mm in the outer area. However, it is possible to merge measurements for large fields up to 38 cm in the longitudinal direction. It is also possible to increase the resolution in the central area to 2.5 mm by merging. Unlike with ArcCHECK, isocentric dose distributions are directly measured in 2 orthogonal planes but unlike the OCTAVIUS system, the detectors are not at all times in perpendicular orientation to the radiation beam, and so the device does not take into account all gantry angles [76]. Dose can also be measured with an ion chamber inside the phantom. The standard material of the phantom is PMMA but a Plastic Water DT model is also available for good water equivalence in a wide energy range. The phantom has 22 cm in diameter and 40 cm in length.

Software module Delta^{4DVH} Anatomy (which comprises all functions of modules Delta^{4DVH} and Delta^{4DVH} Professional), also by Sun Nuclear Corporation, Melbourne, FL, USA, allows to evaluate the clinical significance of measured discrepancies by reconstructing the 3D dose distribution in patient's anatomy based on the 2D dose distributions measured with the detector arrays in the Delta^4 Phantom+. The software has an independent calculation algorithm calculating dose in the patient from energy fluence [76]. More specifically, it converts the doses measured by the detectors to doses in the patient's CT in two steps. First, energy fluence is estimated based on the measured doses using a kernel-based pencil beam algorithm. The same pencil beam algorithm is

used to recalculate the energy fluence to dose in the patient's anatomy. To commission this algorithm, 3D dose distributions for several open fields must be exported from the TPS and head scatter factors for the same field sizes must be known [76]. Planned dose distributions from the TPS can be quantitatively compared to the Delta^{4DVH} Anatomy calculated dose distributions in terms of DVHs and gamma analysis performed separately for each patient structure (OARs – organs at risk, PTV and body structure) [76].

Figure 2-3: The Delta^4 Phantom+ [77].

With the Delta^4 dosimetric system, it is also possible to recalculate dose in 3D in the phantom (based on a ray-tracing algorithm [78, 79]. It is possible to perform 2D gamma analysis in the 2 detector planes. The average time reported for patient anatomy dose calculation is 25.5 minutes compared to 5.3 minutes for phantom 3D dose calculation [78].

COMPASS

COMPASS is a verification system capable of dose calculation in patient's CT that is based on either MatriXX measurements or Dolphin measurements (both detectors by IBA Dosimetry GmbH, Schwarzenbruck, Germany).

MatriXX is a 2D array of ionization chambers which is mounted on the gantry in a holder while performing VMAT or IMRT verification. Thus, the beam is always perpendicular to the detector array. It comprises 1020 vented parallel ion chambers and covers the field size of 24.4 x 24.4 cm^2. The spacing between chambers is 7.62 mm centre to centre, the chamber diameter is 4.5 mm. The material used as built-up on top of the chambers is 3 mm ABS Tecaran (density 1.06 g/cm^3). MatriXX can be equipped with an angle sensor.

Dolphin is a transmission detector also mounted on the gantry, directly on the linac head, while performing verification. This detector comprises 1513 air-vented parallel plate ion chambers

of 3.2 mm in diameter and covers the whole treatment field size of 40 x 40 cm^2 with 5 mm resolution in the central 15 x 15 cm^2 area. Dolphin also comprises an angle sensor.

A full 3D dose calculation in patient's CT is performed with a collapsed cone algorithm based on the method by Ahnesjö [80], which ensures relatively high accuracy. The 3D dose calculation can serve as a secondary check of TPS calculation only [81] or it can take into account measured data for dose reconstruction. To calculate dose in 3D on patient's CT, the following DICOM objects must be imported into the software: CT images, RT structure set, RT plan and RT dose. The software uses the original RT (radiation therapy) plan information and the software's internal beam model, which includes inhomogeneity corrections, to calculate dose in patient without the help of measured data. If measured data are to be performed into account in the 3D dose calculation, the algorithm performs perturbative corrections based on the measured signals [81]. . In this case, in the first step, the algorithm predicts a fluence map using the linac model and information from the treatment plan. Even though the MatriXX and Dolphin detectors have a rather coarse resolution, this fluence map has a resolution of typically 2 x 2 mm^2 [2]. In the second step, the detector response is predicted and in the third step it is compared to the measured signal. The difference between these two is used to modify the predicted fluence to obtain a reconstructed fluence. The 3D dose reconstruction in patient is based on the reconstructed fluence and uses the collapsed cone convolution algorithm [2] which is potentially more accurate than computation algorithms in the previously described systems. The resolution of the resulting 3D dose grid is typically 2 x 2 x 2 mm^3, even though the measured 2D dose planes (with the MatriXX or Dolphin) are coarser [2]. This causes acceptable inaccuracies in the reconstructed doses [2]. The COMPASS dose calculations are not completely independent of the treatment planning system. Evaluation of the calculated or reconstructed dose distribution in the patient's CT images can be performed with the help of DVHs, 3D dose difference maps or 3D gamma index maps (with local or global normalization). The system also allows automatic PTV and OAR evaluation.

For commissioning of the system, the following beam data must be measured: dose profiles, percentage depth dose curves and output factors. It also requires absolute dose calibration and geometry of the linac as input [81].

Dosimetry check

Dosimetry Check is a software solution that uses the existing linac hardware or second-party detectors to reconstruct 3D dose in the patient. The dose image can be obtained by an electronic portal imaging device (EPID) (either pre-treatment or during treatment) or by an array of ion chambers or diode detectors. Based on the measured data, the software computes a fluence map in air. The beam images are divided into multiple small beamlets and each beamlet is assigned its weight corresponding to the measured fluence map. The fluence map is calibrated to so-called relative monitor units based on a 10 x 10 cm^2 square field with 100 MU delivered. The algorithm reconstructs dose in the patient based on the fluence map using a pencil beam approach. The newest version of the software 5.4 uses a collapsed cone algorithm and takes advantage of a GPU (graphics processing unit) computation. In addition to existing beam data (percent depth doses, output factors) and CT images of the patient, additional data must be measured at commissioning. These include calibrating the EPID, collecting a series of square field images or collecting square field images behind an increasing thickness, of water, if transit measurements are required. These data

are used to create the measured source model. The reconstructed 3D dose in the patient can be evaluated by 2D dose profiles, isodose overlays, 3D gamma analysis (including a map of gamma indices on the patient's anatomy and 3D gamma histograms), dose volume histograms, and beam statistics. The software is vendor independent and is suitable for Elekta (Stockholm, Sweden), Varian (Varian Medical Systems, Palo Alto, CA, USA), Siemens (Munich, Germany) and also TomoTherapy (Accuray Inc., USA) [10, 82].

2.3 Pseudo-3D radiochromic film dosimetry in the literature

Attemps of using radiographic or radiochromic film for 3D dosimetry have already been made. McCaw et al. [83] proposed a film stack dosimeter which might be used for 3D megavoltage photon beam dosimetry. However, its dimensions are too small to be used for verification of IMRT or VMAT plans in the clinics and it uses many film sheets placed only in one direction, which might not be cost-effective. Others attempted to create a rotational cylindrical phantom with only one sheet of film placed in it, which is less resource-intensive, but in this study it was meant for 2D evaluation only, the user being able to select the plane of interest anywhere between sagittal and coronal by rotating the phantom to respective position prior to measurement [17]. Different groups have loaded film into a cylindrical phantom, either several sheets of film rolled up at different radial distances from the centre and additional axial or coronal planar measurements with film [84] or films rolled up at one radial distance and additional film sheets placed into axial planes in order to measure planar doses [85]. Other groups used one sheet of film coiled into a spiral in a cylindrical phantom [86, 87]. Pseudo-3D dosimetry with radiographic or radiochromic film, with a single film sheet or a film array, has also been used for stereotactic radiosurgery [88], brachytherapy [14, 15, 89], and proton therapy [90-92].

2.4 Correlations between 2D and 3D dosimetry published in the literature

3D evaluation is preferred over 2D evaluation because 2D might be insufficient, as reported in the literature [7, 8, 93-95]. Several recent works have compared 2D and 3D results and/or have tried to find correlations between them, for example [7-9]. However, authors typically compare two different detection systems or softwares, such as 3DVH software for 3D gamma analysis versus Patient software for 2D gamma analysis, both from Sun Nuclear Corporation, Melbourne, FL, USA [7]; the COMPASS system (IBA Dosimetry GmbH, Schwarzenbruck, Germany) for 3D and EBT2 film for 2D evaluation [8] or EPID backprojected doses for 3D gamma and MapCHECK detector (Sun Nuclear Corporation, Melbourne, FL, USA) for 2D gamma evaluation [9]. Authors who used software from the same vendor [7] found a strong correlation between 2D and 3D gamma pass rates while authors using different systems found only weak or no correlation between 2D and 3D gamma evaluation [8] or found significant correlation only in some investigated parameters [9]. As reported by Xing et al. [96], who compared VeriSoft 3D evaluation (PTW, Freiburg, Germany) with 3DVH Sun Nuclear 3D evaluation, although the two systems calculate accurate gamma indices compared to manual calculation, the gamma pass rates are different due to different interpolation of raw data and implementation of algorithms. Users of commercial systems do not have detailed knowledge of the hardware and software they handle, so it is sometimes difficult to find the causes of encountered discrepancies.

3 Materials and methods

This chapter describes the developed pseudo-3D IMRT verification method and also the key steps that was carried out to establish the method. First, film dosimetry was implemented and tested under given experimental conditions. This included testing film directional dependence, one of the most important factors that could influence the results in 3D. Then the film method was tested in 2D (field-by-field) against a clinically used verification tool for IMRT plans (PTW Seven29 ionization chamber array, PTW, Freiburg, Germany). The film method was also tested in 3D against a clinically used verification tool for VMAT plans (PTW OCTAVIUS 1500 ionization chamber array placed into the OCTAVIUS 4D Modular phantom, PTW, Freiburg, Germany). Scripts for film processing, 2D gamma analysis and 3D gamma analysis were written in MATLAB. The 2D gamma analysis code was verified against a commercially available alternative (OmniPro I'mRT, IBA Dosimetry GmbH, Schwarzenbruck, Germany). The 3D gamma analysis code was verified against manual calculation for a smaller 3D dose matrix. Finally, pseudo-3D measurements of clinical and error-induced plans could be performed and evaluated with the MATLAB scripts and various comparisons of pseudo-3D, 2D plane-by-plane and 2D field-by-field approaches could be performed. All the IMRT plan measurements and evaluations that were performed throughout the process are summarized in Table 3-1. They are further described in the following chapters with reference to Table 3-1.

Table 3-1: All IMRT plan measurements and evaluations performed in this work.

Process	IMRT plans	Detector	Phantom	Measure-ment geometry	Gamma score	Evaluation software	Purpose
METHODS							
Establishing film dosimetry ***Part A***	Patient no. 4, 8 and 9	EBT3 film	RW3 slabs, I'mRT Phantom	Pseudo-3D global verification	2D plane-by-plane	MATLAB, FilmQA Pro	Establish our film dosimetry protocol
2D gamma analysis in MATLAB ***Part B***	Clinical plans – Patients no. 1 to 6	EBT3 film	I'mRT Phantom	Pseudo-3D global verification	2D plane-by-plane	OmniPro I'mRT + MATLAB	Compare our MATLAB 2D code against a commercial program
RESULTS							
Benchmar-king film dosimetry ***Part C***	Clinical plan – Patient no. 5 and 9 + Error-induced plan – Patient no. 4	EBT3 film	RW3 slabs	2D field-by-field	2D field-by-field	OmniPro I'mRT + MATLAB	Compare film dosimetry to a clinically used IMRT verification tool
			I'mRT Phantom	Pseudo-3D global verification	Pseudo-3D, 2D plane-by-plane		
		PTW Seven29	RW3 slabs	2D field-by-field	2D field-by-field	VeriSoft v. 3.1 and v. 8.0	
		PTW OCTAVIUS 1500	OCTAVIUS 4D	Pseudo-3D global verification	Pseudo-3D, 2D plane-by-plane		

Table 3-1: All IMRT plan measurements and evaluations performed in this work - continued.

Process	IMRT plans	Detector	Phantom	Measure-ment geometry	Gamma score	Evaluation software	Purpose
Pseudo-3D verification ***Part D***	Clinical plans – Patients no. 1 to 9	EBT3 film	I'mRT Phantom	Pseudo-3D global verification	2D plane-by-plane	MATLAB 2D and 3D code	Compare 2D and pseudo-3D evaluation of the same measured data set
					Pseudo-3D		
Pseudo-3D verification ***Part E***	Clinical plans – Patients no. 1 to 6	EBT3 film	RW3 slabs	2D field-by-field	2D field-by-field	OmniPro I'mRT + MATLAB 2D code	Compare field-by-field to pseudo-3D verification for clinical plans
			I'mRT Phantom	Pseudo-3D global verification	2D plane-by-plane		
		PTW Seven29	RW3 slabs	2D field-by-field	2D field-by-field	VeriSoft v. 3.1	
Pseudo-3D verification ***Part F***	Error induced IMRT plans – Patients no. 4 and 5	EBT3 film	I'mRT Phantom	Pseudo-3D global verification	2D plane-by-plane and pseudo-3D	MATLAB 2D and 3D code	Compare field-by-field to pseudo-3D verification for error-induced plans
Pseudo-3D verification ***Part G***	Error induced VMAT plan – Patient no. 9	EBT3 film	I'mRT Phantom	Pseudo-3D global verification	2D plane-by-plane and pseudo-3D	MATLAB 2D and 3D code	Compare the pseudo-3D film method to a pseudo-3D commercial alternative for an error induced plan
		PTW OCTAVIUS 1500	OCTAVIUS 4D			VeriSoft v. 8.0	

3.1 Establishing methodology for IMRT film dosimetry

Even though many works have been published (see chapter 2.1) on the topic of radiochromic film dosimetry, proper film handling, scanning and calibration, each site must take into account their own working conditions and equipment and should adjust guidelines published in literature in order to achieve good results in terms of accuracy, precision, reproducibility but also time required for the process. Therefore, in our work we established our own methodology for radiochromic film dosimetry using MATLAB. The methodology is easy to implement at our clinical site, suits the purpose of this work, is optimal in terms of time and accuracy and the user has control over the evaluation software. Our methodology is described in the following chapters. These can serve as a practical guideline to clinics who wish to establish their routine protocols for film dosimetry.

If not explicitly specified otherwise, the following gamma analysis parameters are used throughout this chapter: global relative gamma, normalization to the 75th percentile of the dose distribution,

tolerance criteria 3%/3 mm, threshold 10%, search radius 4.5 mm, predicted dose as reference, resolution of matrices, adjusted according to measured distribution.

3.1.1 Film irradiation and scanning

Irradiation of calibration and clinical films was carried out on a Siemens Artiste linear accelerator and a Varian TrueBeam linear accelerator at the Thomayer Hospital in Prague, using the photon energy 6 MV and 18 MV. For each combination of linear accelerator/energy/film batch, a separate calibration curve was obtained and then used for clinical films irradiated with the same combination of linear accelerator/energy and film batch.

EPSON Perfection V700 Photo scanner (Seiko Epson Corporation, Japan) was used in this study together with EBT3 films from two different batches. Film pieces of 6 x 6 cm^2 were used for calibration and film pieces of sufficient size (starting from 8 x 10 cm^2) were used for measurement of clinical plans, so that the central area with clinically relevant doses was captured on film.

Scanning parameters were as follows: Professional mode, Color negative film, Film with film area guide, 48-bit colors, 72 dpi, uncorrected images saved in tiff format, portrait orientation, transmission mode. Films were placed exactly in the centre of the scanner area with the help of markers. The same orientation of all calibration and clinical (IMRT and VMAT) films was maintained at the accelerator and at scanning. The red channel data were used in all cases, for calibration and evaluation (see chapter 3.1.14 for comparison of single channel and triple-channel methods). Proper scanner warm-up was ensured by scanning 5 blank scans prior to film scanning (see chapter 3.1.9). In all cases, film edges possibly damaged by cutting were excluded from the region of interest.

3.1.2 Film calibration

Calibration film pieces of the size 6 x 6 cm^2 were irradiated with the following doses: 0, 0.1, 0.3, 0.5, 0.8, 1, 1.2, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 and 5.5 Gy in an RW3 slab phantom in reference conditions. For both types of linear accelerator the reference conditions were SAD = 100 cm, field size 10 x 10 and depth of 5 cm (the 6 MV beams) or 10 cm (the 18 MV beams). The linac output was checked with an ionization chamber before film calibration in an RW3 phantom. The difference between the slab phantom material and water was not the same for the spectra of Siemens Artiste and Varian TrueBeam. The difference between slab material and water was taken into account at calibration (each combination of linear accelerator/energy and film batch had its proper calibration curve) by experimentally measured correction and then all predictions in the treatment planning systems for our further experiments were done in water.

Calibration films were scanned according to chapter 3.1.1 72 hours after irradiation and processed in MATLAB. The mean scanner value in the red channel for each calibration film was obtained. After excluding the film edges from evaluation, the region of interest from which the mean value was computed was 5.8 x 5.8 cm. The known dose values were plotted against the mean scanner values for all calibration films and fitted with a 4th order polynomial. An example of a calibration curve obtained in MATLAB for the 18 MV beam is shown in Figure 3-1 (a). The accuracy of each calibration curve was evaluated with an independent set of films irradiated (in the same conditions as the calibration films) with known doses. The relative deviations from the known dose values were calculated and this is shown in Figure 3-1 (b) for the same 18 MV curve. Results shown in Figure 3-1 (b) are similar, for example, to a paper by Howard et al. [97]. Doses above 0.5 Gy are within 3% of the expected value and doses above 1 Gy are within 2%. The gamma analysis threshold of 10% used later

in our experiments guarantees that doses below 0.3 – 0.5 Gy are not taken into account in gamma evaluation. Different calibration functions were tested this way (2nd and 3rd order polynomial and different types of rational functions) and the best results were seen with the 4th order polynomial. The determination of dose measured with the independent set of films is, of course, also influenced by the uncertainty of our film dosimetry method, which is quantified in chapter 3.1.15. This type of single channel calibration could be used in MATLAB and OmniPro I'mRT.

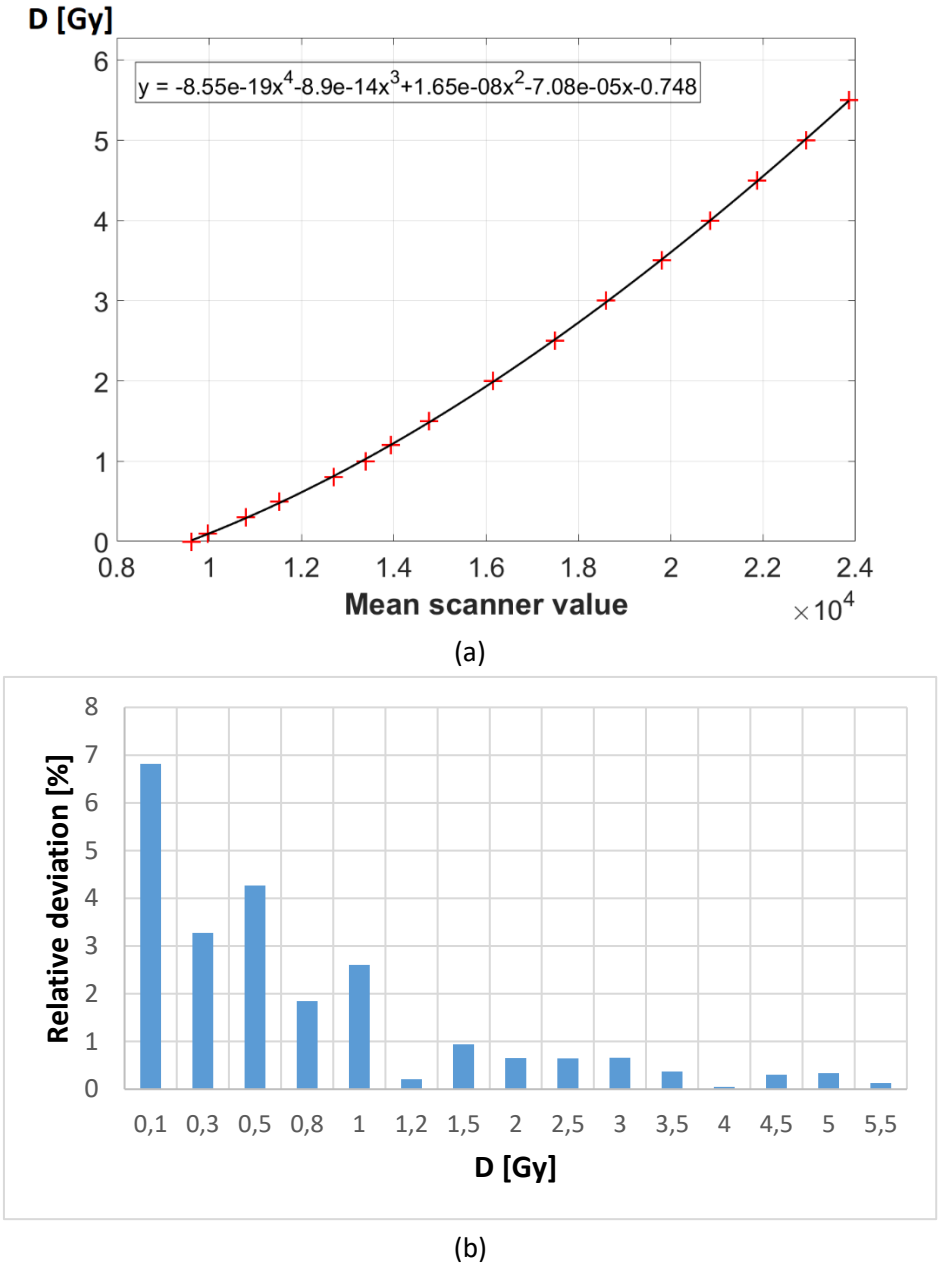


Figure 3-1: An example of a calibration curve obtained in MATLAB for an 18 MV beam (a) and the relative deviation from known dose values after fitting this function to an independent set of irradiated films (b).

A different approach had to be taken in the FilmQA Pro software that was used mainly to compare the single channel and tripple channel methods. To our best knowledge, a 4th order polynomial calibration function cannot be used within the FilmQA Pro software version 2016. Different types of functions available in the FilmQA Pro software were therefore investigated and the accuracy of each calibration function was evaluated by calculating the relative deviations from the known dose values for an independent set of films, as described above. Based on this, a 2nd order polynomial calibration function available within FilmQA Pro was chosen for evaluations that were performed with this software. The accuracy of this type of function was also comparable to the 4th order polynomial used within MATLAB and OmniPro I'mRT.

3.1.3 Energy, dose rate and temperature dependence of EBT3 film

The energy dependence of EBT3 film has been studied in literature and it has been shown that in the megavoltage range, the EBT3 film is energy independent [22, 98, 99]. The same is true for dose rate dependence [22, 99] and temperature dependence [97] (unless extreme temperatures are encountered). In this work, two photon energies were used, namely 6 MV and 18 MV available on the Siemens Artiste (Siemens, Munich, Germany) and TrueBeam (Varian Medical Systems, Palo Alto, CA, USA) linear accelerators in the Thomayer Hospital. The clinically used dose rate (in reference conditions) for each energy and type of accelerator is 500 MU/min for 18 MV and Siemens Artiste, 300 MU/min for 6 MV and Siemens Artiste, 500 MU/min for 18 MV and 6 MV on TrueBeam. For each combination of energy/dose rate/type of accelerator, a separate calibration curve was established and subsequently used for film evaluation. Therefore, irradiation conditions in terms of energy spectrum and dose rate comming out of the linac head were identical for experimental films and corresponding calibration films. Films were handled in room temperature conditions ranging from 20°C to 26°C.

3.1.4 LRA

The lateral response artifact of the scanner was tested by scanning the same piece of film in the center of the scanner and in two lateral positions perpendicular to the central scanning axis. The mean pixel value of each film in each lateral position was compared to the mean pixel value of the same film placed at the central axis. Calibration films were used for this purpose and several dose levels (0, 0.5, 1, 2, 3, 4 Gy). The ROI for the determination of mean pixel value was therefore 5.8 x 5.8 cm^2. Given the dimensions of the EPSON V700 Photo scanner, the two lateral positions were chosen to be 5 cm left and 5 cm right to the central axis of the scanner (position of the center of each film). The resulting values listed in Table 3-2 were taken into account in the uncertainty budget (see chapter 3.1.15). The scanner seems to have a non-uniform lateral response. Signals on the right side are higher than signals on the left side of the scaning area (in the central part of the scanner).

Table 3-2: Difference of mean pixel values taken 5 cm left (PV_{left}) and 5 cm right (PV_{right}) from the central scanning axis relative to the central part (PV_{centre}) for several dose levels.

Dose [Gy]	**0**	**0.5**	**1**	**2**	**3**	**4**
Relative difference $100 * \left(1 - \frac{PV_{left}}{PV_{centre}}\right)$ %	-1,5	0,3	0,3	0,7	0,0	1,0
Relative difference $100 * \left(1 - \frac{PV_{right}}{PV_{centre}}\right)$ %	-2,8	-1,1	-1,7	-0,2	-1,1	-0,9

3.1.5 Resolution, de-noising

The recommended resolution of 72 dpi [28, 29] was used for scanning all films in order to avoid too much noise, which can have impact on comparisons between measured and planned doses. This translates into spatial resolution of approximately 0.35 mm. A median filter of 5 x 5 pixels was applied for all evaluations [28, 29, 36].

3.1.6 Post-irradiation darkening

Based on the recommendations from the literature [28, 29] and also for practical reasons, the period between film irradiation and scanning was chosen to be 72 hours, when the film response should be more or less stable. The same time period was used for all calibrations and experimental measurements, so the post-irradiation darkening should have little impact on results.

3.1.7 Film-to-light distance

The effect of taping the film to the scanner bed was tested. First, each calibration film was put on the scanner without taping it to the scanner bed and scanned. Calibration curve was obtained from these calibration films as mean pixel value versus dose, fitting the points with a 4th degree polynomial (the calibration procedure from chapter 3.1.2 was followed). Then, films with known doses (0, 0.1, 0.3, 0.5, 0.8, 1, 1.2, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 and 5.5 Gy) were scanned in the same way and converted into dose with this calibration curve. The deviation of the obtained dose value from the known dose value (in %) was evaluated. The same calibration and evaluation procedure was done with the same films, this time taped to the scanner bed. This was done for the 18 MV energy. For the untaped films, the mean relative dose deviation from the expected value was found to be 1.54 %, ranging from 0.05 % to 6.82 % (worse results for lower doses). For the taped films, the mean relative deviation was 2.4 %, ranging from 0.01 % to 7.65 %. Similar results were obtained using a glass plate on top of the films. None of these techniques seems to improve results of film dosimetry according to our protocol and with the film batches tested; therefore, all experimental films in this work were scanned without a glass plate or without taping.

3.1.8 Angular dependence

It needs to be pointed out that film calibration geometry was different to the measurement geometry in case of pseudo-3D measurements in the I'mRT Phantom (see chapter 3.3.2). For calibration, films were placed into the coronal plane and irradiated with gantry at 0°. For the pseudo-3D measurements, films were placed into coronal, transverse or sagittal planes and irradiated with different gantry angles. So the angular dependence of film response had to be evaluated. An RW3 slab phantom was set up and one 6 x 6 cm^2 film piece placed into the coronal plane in the reference depth of 10 cm with SAD = 100 cm. It was irradiated with 150 MU with an 18 MV photon beam and a field size of 10 x 10 cm^2. The 18 MV energy was chosen for this experiment because it will be used for clinical plans in our further experiments (as wll as 6 MV). 18 MV beams are now less common in radiotherapy clinics, and while a lot of data on angular independence of EBT3 film have been published for lower energies [26, 28, 100], for 18 MV this information is, to our best knowledge, lacking. This procedure was done for gantry angles 0, 40, 60 and 90°. Film response was recalculated to dose values in the same way as for film calibration for each gantry angle. For this experiment, only a small area in the centre of the film was averaged, comparable to the sensitive volume of the Farmer ionization chamber, 3 mm x 23 mm - was averaged to obtain a mean dose value. A Farmer-type ionization chamber was then placed in exactly the same setup and the point dose was estimated. The dose values for film and ion chamber were compared for each gantry angle.

Film directional dependence is shown in Figure 3-2. It shows dose deviations for three gantry angles from gantry position at 0°. Phantom attenuation was subtracted by comparing the values to point doses measured with an ionization chamber in the same setup. The relative standard uncertainty of relative dose determination with film for a dose of 2 Gy was estimated to 1.85% in chapter 3.1.15 (one sigma; even a higher relative uncertainty is valid for 1.5 Gy). The uncertainty level for 2 Gy is shown in Figure 3-2 with two horizontal lines. It can be seen that film response deviation for different gantry angles falls well within the total uncertainty, which is in agreement with literature [26]), where EBT3 film angluar dependence was studied for 4 MV photons.

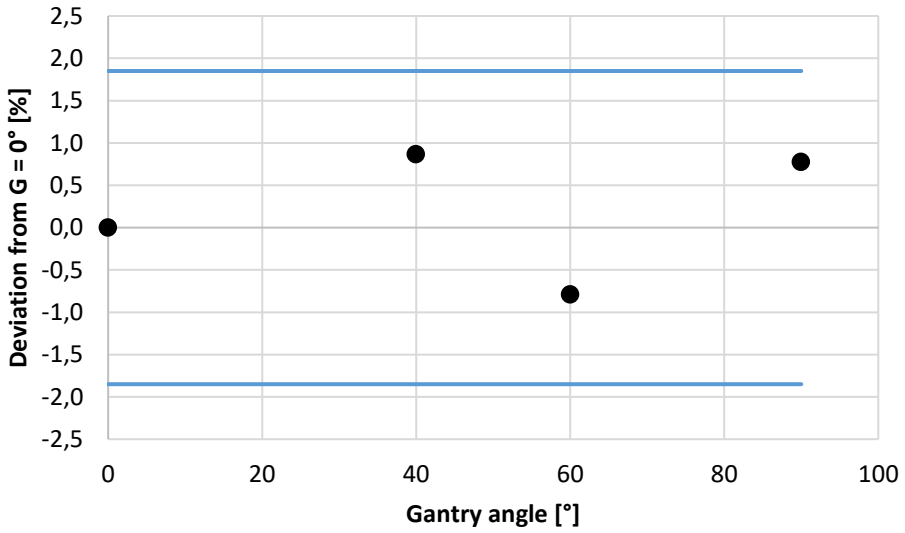


Figure 3-2: EBT3 film angular dependence measured with four different gantry angles. Horizontal lines show relative standard uncertainty of dose estimation with film for 2 Gy.

3.1.9 Scanner warm-up

To account for the effect of scanner warm-up, 16 scans of a blank piece of calibration film (6 x 6 cm^2) were taken subsequently after the scanner was turned on. The mean pixel value in the red channel was evaluated for each scan. While the scanner response went up during the first 5 scans, it remained stable from the 6th scan on. Therefore, in our film protocol, 5 blank scans are always taken after the scanner is turned on and before scanning the experimental films.

3.1.10 Scanner and film uniformity

Scanner uniformity was determined by scanning a small piece of unirradiated EBT3 film (approximately 4 x 4 cm^2, excluding the film edges from the scanning region of interest — ROI) in various places over the scanner area. The mean value of scanner response over this 4 x 4 cm^2 area for each scanned image was denoted M and then uniformity H_s over a given region was calculated as

$$H_S = \sqrt{\frac{(M_i - \bar{M})}{n-1}} \tag{3.1}$$

where n is the number of scanned film images within that given region.

The central region of the scanner (16 x 16 cm^2) was found to be uniform with a relative standard deviation of 1.23%. When the whole scanner area was assessed, the relative standard deviation increased to a value of 3.07%. Therefore, the central area must always be used for scanning. Scanner non-uniformity was accounted for in the uncertainty budget (see chapter 3.1.15).

Film uniformity was assessed with the calibration film pieces scanned in the center of the scaner. Therefore, film uniformity was assessed for the calibration dose levels (0, 0.1, 0.3, 0.5, 0.8, 1, 1.2, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 and 5.5 Gy) and the film size was 6 x 6 cm^2 for each piece (film edges were excluded for evaluation). This comprised any possible scanner non-uniformity within the central 6 x 6 cm^2 region and also the linac field non-uniformity within this region for a 10 x 10 cm^2 radiation field. The relative standard deviation (in scanner values, not in dose values) was calculated for the pixels of each calibration film. The relative standard deviation for the calibration films ranged from 0.06% to 0.49% with a mean value of 0.13%. The highest value 0.49% was seen for the nonirradiated 0 Gy film (which did not comprise the nonuniformity of the linac radiation field). Higher values of relative standard deviation were generally seen for lower doses and lower values for higher doses. Film non-uniformity was also accounted for in the uncertainty budget (see chapter 3.1.15).

3.1.11 Scanner reproducibility

Short term reproducibility of the scanning process was assessed by scanning the same piece of film (doses of 0.3, 1, 2, 5 and 8 Gy) five times in sequence and determined as the relative standard deviation of the mean value. The maximum value of short term reproducibility found for the above mentioned doses was 0.24%. Long term reproducibility was assessed by scanning the same piece of film (doses of 0.3, 1, 2, 5 and 8 Gy) on several subsequent days 1 month after irradiation. Long term stability of the scanner was found to be less than 0.23% for all doses between 0.5 and 8 Gy. However, for 0.3 Gy it was as bad as 20%. The gamma threshold of 10% used for our further experiments ensures that doses below 0.3 – 0.5 Gy are not taken into account in gamma evaluation. Scanning reproducibility was also reflected in the total uncertainty of dose estimation with film in chapter 3.1.15.

3.1.12 Inter-scan variability

To investigate the influence of the interscan variability in our film dosimetry protocol on gamma results, the same irradiated piece of film was scanned 10 times on different days within a one month period. This test was performed 4 years after irradiation of the film piece, when the changes of the film response should have been reasonably stable [28]. The isocentric coronal plane of an IMRT prostate plan measured in a pseudo-3D manner (see chapter 3.3.2 for details) from Patient no. 4 was used for this purpose (18 MV, Siemens Artiste step-and-shoot IMRT). 2D gamma analysis was performed with our 2D Matlab gamma code (see chapter 3.3.3 for details) taking the first scan of the 10 subsequent scans as the reference dose distribution and each of the 9 remaining scans as the evaluated dose distributions. The gamma scores and gamma statistics obtained with 3% and 3 mm criteria and 2% and 2 mm criteria (and the rest of the gamma parameters kept as described at the begining of chapter 3.1) can be found in Table 3-3. Because gamma scores are in all cases 100% and mean and maximum gamma index values are very low, it was concluded that the interscan variability does not play an important role in our film protocol and will not be explicitly accounted for in the uncertainty budget. The uncertainty budget already comprises short-term and long-term scanner reproducibility and inter-film variability (see chapter 3.1.15).

Table 3-3: Gamma scores obtained for 10 repeated scans of the same film piece. The first scan was taken as the reference.

Scan no.	3% and 3mm criteria				2% and 2 mm criteria			
	Minimum gamma index	Maximum gamma index	Mean gamma index	Gamma score [%]	Minimum gamma index	Maximum gamma index	Mean gamma index	Gamma score [%]
1	0.00	0.28	0.08	100	0.00	0.43	0.13	100
2	0.00	0.27	0.08	100	0.00	0.41	0.12	100
3	0.00	0.29	0.07	100	0.00	0.44	0.11	100
4	0.00	0.23	0.08	100	0.00	0.35	0.12	100
5	0.00	0.27	0.08	100	0.00	0.41	0.12	100
6	0.00	0.35	0.08	100	0.00	0.52	0.12	100
7	0.00	0.27	0.08	100	0.00	0.40	0.11	100
8	0.00	0.29	0.07	100	0.00	0.44	0.11	100
9	0.00	0.32	0.07	100	0.00	0.48	0.11	100

3.1.13 Subtracting blank film prior to irradiation

In order to see whether improvement in gamma scores could be achieved with single channel dosimetry by scanning each piece of film before irradiation and subtracting the blank film from the irradiated one, this was done for two clinical plans: IMRT sliding window technique, Varian TrueBeam, a 6 MV glioblastoma plan (Patient no. 8) and an 18 MV prostate plan (Patient no. 7). All pieces of film were first scanned in the same way as described in chapter 3.1.1. Then, films were loaded into the I'mRT Phantom (Scanditronix Wellhofer North America, USA) for pseudo-3D measurements as described in chapter 3.3.2 and irradiated with the 6 MV and the 18 MV clinical plan on a Varian TrueBeam accelerator. Our 2D Matlab code (described in chapter 3.3.3) was used to perform 2D gamma analysis of 15 planes for each plan (5 coronal, 5 sagittal and 5 transverse planes) following the methodology described in chapter 3.3 and using single channel dosimetry with the red channel, 3 % and 3 mm criteria. The first set of gamma scores was obtained with the irradiated films as usual (without subtracting the blank films) by comparing them to the predicted TPS values, Gamma Score 1. The second set of gamma scores was obtained after the blank piece of film was subtracted from the corresponding irradiated piece of film in terms of scanner values, also comparing them to the predicted TPS values – Gamma Score 2. All parameters were left the same for both analyses. This required proper alignment of the film images (using markers on film) and adjustment of the calibration curve (the mean value of an unexposed calibration film was subtracted from the calibration points). The difference between Gamma Score 1 and Gamma Score 2 was evaluated for 15 planes of the 6 MV plan and 15 planes of the 18 MV planes, 30 planes in total. Also the mean gamma index in these 30 planes was compared.

The difference between the two sets of gamma scores, taken as Gamma Sore 2 minus Gamma Score 1 (therefore positive values are in favor of the subtraction method), ranged from -7,9 % to + 13,9 % for the 6 MV plan, with a mean value of + 0,5 %. Improvement was seen for 6 planes out of 15 and lower gamma scores were seen for 9 planes out of 15 with the subtraction method. For the 18 MV plan, the differences in gamma score ranged from -5,6 % to 1,3 % with a mean value of -0,6 %. Improvement was seen for 7 planes out of 15 and lower scores were obtained for 8 planes out of 15 with the subtraction method. The difference in the mean gamma index was on the order of +/-0,01

for nearly all the 30 planes, except for 3 planes in the 6 MV plan where the difference was on the order of +/-0,1. Based on these results, it was concluded that subtracting the blank film from the irradiated one is not necessary in our case and this step was not included in our methodology. It needs to be pointed out that these conclusions are valid under our specific experimental conditions.

3.1.14 Comparison of single channel and multichannel methods

The effect of using either single channel calibration or tripple channel calibration on the resulting gamma scores for our experimental setup and particular methodology was investigated in the FilmQA Pro software version 2016 (Ashland Advanced Materials, New Jersey, USA). The same set of calibration films (doses of 0, 0.1, 0.3, 0.5, 0.8, 1, 1.2, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 and 5.5 Gy) was used for both calibration methods. A calibration function called Dose quadratic versus color in the FilmQA Pro software, with the formula $D(x) = A + B \cdot x + C \cdot x^2$, where *D* stands for dose, *x* stands for pixel value and *A, B* and *C* are parameters of fit, was used because a 4th order polynomial (which was used in the rest of our experiments, see chapter 3.1.2) is not available in the software.

First, the Dose mapping method within the software was set to Dose map using tripple channel uniformity optimization, where all three channels are used for conversion of scanned film pixel values to dose values. Gamma analysis was then performed for three patient plans measured in pseudo-3D manner (see chapter 3.3) and 2D gamma scores were calculated for 15 planes (5 transverse, 5 sagittal and 5 coronal) for each patient, 45 planes in total. The patients were Patient no. 4 with introduced error according to chapter 3.3.5.2 (prostate with lymph nodes, 18X, Siemens Artiste, XiO, step-and-shoot IMRT), Patient no. 9 (prostate only, 18X, Varian TrueBeam, Eclipse, VMAT) and Patient no. 8 (glioblastoma, 6X, Varian TrueBeam, Eclipse, sliding window IMRT) according to chapter 3.3.5.1. The plan with the introduced error was included in order to see the behaviour of the two methods for lower gamma scores. The parameters of gamma analysis were set as close as possible to the OmniPro I'mRT and Matlab parameters (see chapters 3.2.6 and 3.3): normalization to the maximum dose of the predicted dose distribution, global gamma, relative gamma, Gamma in normalized form (using the equation of Low et al. [1] – one of the options available in FilmQA Pro), search radius of 4.5 mm, threshold 10% of maximum predicted dose. Prior to evaluation, a median filter of 5 x 5 pixels was used to smooth the measured dose distributions. Films were matched to the predicted dose planes with the help of fiducials on film and with the help of dose profiles where necessary, but usually matching the image centers of the predicted and measured dose distributions was enough. Unlike in OmniPro I'mRT, FilmQA Pro software adjusts the resolution of the measured dose distribution to the coarser resolution of the predicted matrix by averaging neighboring pixels. This cannot be modified by the user. Thus, the resolution of film (72 dpi corresponding to approximately 0.35 mm) was adjusted to the 1 mm dose grid (in case of XiO) or the 2.5 mm dose grid (in case of Eclipse). Both the criteria 3 %/3 mm and 2 %/2 mm were investigated.

Next, the Dose mapping method was set to Dose map from single channel calibration, where the software uses the red channel (not adjustable by the user) and the process of gamma analysis was repeated with exactly the same conditions.

Results are shown in Table 3-4.

Table 3-4: Comparison of the single channel and the tripple channel approach in Film QA Pro for three patients. Patient no. 4 is a plan with introduced error.

Gamma scores – 3 %/3 mm															
Single channel calibration (red channel)															
Plane no.	Transverse planes					Sagittal planes					Coronal planes				
Patient no. ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	87.69	96.45	93.58	88.24	91.62	87.52	92.56	71.93	95.2	89.24	93.54	94.61	83.59	82.5	81.89
9	54.16	99.80	99.64	99.24	97.99	99.12	100.0	99.77	98.74	99.55	98.91	99.90	99.90	100.0	99.42
8	93.67	99.41	95.75	98.31	92.60	99.41	99.96	99.84	99.63	99.08	89.28	98.84	99.78	99.84	99.39
Tripple channel calibration															
Plane no.	Transverse planes					Sagittal planes					Coronal planes				
Patient no. ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	96.61	91.81	97.62	93.75	99.26	92.77	96.28	99.01	98.10	92.21	96.97	95.33	91.83	80.97	84.76
9	53.62	99.81	99.75	99.27	97.60	98.86	99.96	98.65	98.33	97.66	99.27	100	99.96	99.98	99.24
8	95.47	99.76	97.52	96.50	91.98	98.04	99.56	99.17	97.32	95.79	91.06	97.41	99.76	99.57	96.78
Gamma scores – 2 %/2 mm															
Single channel calibration (red channel)															
Plane no.	Transverse planes					Sagittal planes					Coronal planes				
Patient no. ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	67.74	74.89	74.25	74.24	75.25	70.43	70.44	51.53	74.26	71.86	80.72	70.81	61.25	65.25	58.32
9	43.70	96.28	96.85	88.91	87.52	94.9	99.09	94.65	94.49	95.57	84.09	99.06	95.05	93.92	90.97
8	79.23	94.21	86.28	90.12	80.79	93.69	97.81	95.08	95.61	92.08	89.93	95.83	93.79	96.11	97.16
Tripple channel calibration															
Plane no.	Transverse planes					Sagittal planes					Coronal planes				
Patient no. ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	82.40	71.16	86.70	76.35	94.44	75.91	87.83	91.95	84.41	69.27	81.91	76.51	79.23	62.49	68.42
9	42.63	90.43	92.15	93.53	89.19	89.6	96.62	96.8	88.33	89.89	77.08	98.16	97.02	95.69	91.63
8	85.11	92.87	87.49	87.56	81.38	90.14	95.81	91.26	90.24	83.53	87.14	93.55	96.1	95.39	92.73

Table 3-4 indicates that single channel dosimetry is comparable to tripple-channel dosimetry under our experimental conditions and with our methodology for the clinical plans of Patient no. 8 and 9. No advantage can be seen for either of the approaches in terms of gamma score improvement. It might seem that for error-induced plan of Patient no. 4 (less agreement between the measured error-induced dose distribution and the TPS error-free distribution), the triple channel method gives slighty higher gamma scores. A statistical test is needed to make a decision whether gamma scores in general (for all three patients from Table 3-4) are depenent on the method (single-channel or tripple-channel) or not.

A statistical chi-square test of independence was performed with the data in Table 3-4. This test is used to determine whether two categorical variables are related or not. Our first categorical variable was a set of four gamma score bins: gamma score < 90%, gamma score between 90% and 94%, gamma score between 94% and 97%, and gamma score > 97%. In other words, we had four categories of gamma scores. The second categorical variable was the method used to determine the gamma scores, either the single channel or the triple-channel method. A contingency table with frequencies for each combination of categories was created and is shown in Table 3-5. The chi square test of independence was performed in MATLAB by writing our own script (there is no built-in

function within the Statistical Toolbox). Our null hypothesis was that gamma scores and the method used to determine them (single or tripple channel) are independent at a 5% significance level. The so-called expected values were computed as the product of Row totals and Column totals divided by the Grand total (see Table 3-5) for each cell in Table 3-5. The assumption of the test is that each expected value is larger than 5, which was fulfilled with our data.

Table 3-5: Contingency table with frequencies for the chi square test of independence comparing the single channel and the tripple channel approach in Film QA Pro.

Gamma score category (GS)	Single channel method	Tripple channel method	Row totals
GS < 90%	33	28	61
90% <GS < 94%	12	18	30
94% < GS < 97%	16	15	31
GS > 97%	29	29	58
Column totals	90	90	**Grand total:** 180

We calculated the test statistics using the formula

$$\chi^2 = \sum_{i,j=1}^{n} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{3.2}$$

Where *O* is the observed count, *E* is the expected count, *i* and *j* are row and column indices and *n* is the number of rows or columns in Table 3-5. The p-value for 3 degrees of freedom was estimated with the MATLAB function *chi2cdf*. The degrees of freedom *df* are based on the number of rows *r* and columns *c* in the contingency table:

$$df = (r-1)\cdot(c-1) \tag{3.3}$$

The computed p-value for our data was 0.6499 and the null hypothesis that gamma scores are not dependent on the estimation method was not rejected. In order to see the behaviour for lower gamma scores, that were typically obtained with the error-induced plan of Patient no. 4, different types of bins were used for the first categorical variable, starting e.g. with gamma score <80% or gamma score <75%. It turned out that the type of binning does not affect the decision on the null hypothesis. Although there were some differences in the p-value (the p-value ranged from 0.2496 to 0.6921), the null hypothesis was not rejected for any of the settings.

In the context of other works published on this topic, many papers have concluded that triple channel dosimetry performs better, e.g. [35, 39, 53]. Other papers have shown that accuracy and precision of single versus triple channel methods may vary depending on the film and scanner models and on the dose range of interest [37, 41]. Mathot et al. [101] have compared single channel and multichannel methods within the FilmQA Pro software and concluded that the multichannel approach gives higher gamma pass rates for local gamma (3%/3 mm). However, they also found [101] (for two different scanners) that intensive use of the scanner (which is the case with our EPSON V700 Photo scanner) makes multichannel dosimetry less accurate because of tilting in the blue profile in the lamp direction. Howard et al. [97] reported comparable results for single channel dosimetry in ImageJ and multichannel dosimetry in FilmQA Pro. Palmer et al. [42] have even reported better passing rates for single channel dosimetry (98.0% compared to 71.0% global gamma score,

3%/2 mm), which might be explained by a lack of separation between individual channels for higher doses [29] because they worked with stereotactic plans. Pócza et al. [49] state that the choice of the right multichannel model is crucial, otherwise errors of the different color channels (that are correlated,) may be combined in a way that leads to an increase in the overall error [36]. Dufek et al. [37] found that for EPSON 11,000 XL, the multichannel method gives higher gamma passing rates within FilmQA Pro while for EPSON V750 scanner, the red channel method (after subtracting blank film from the irradiated film) gives higher gamma passing rates. González-López et al. [48] have shown that the level of noise may influence single channel and multichannel methods in a different way. While the fixed pattern noise is less pronounced with the multichannel dosimetry, the random noise is less pronounced with the single channel method. Shameem et al. [55] have shown that while the V800 EPSON scanner is noisier in the blue and green channels than the V700 EPSON scanner, in the red channel it is the other way round. The characteristics of the particular scanner apparently play an important role when comparing single channel and multichannel dosimetry methods.

We conclude that it is reasonable to use the single channel method for our further experiments.

3.1.15 Uncertainty budget

The following methodology of uncertainty estimation is valid for our single channel calibration method performed in MATLAB. The assessment of uncertainty for tripple channel methods is less straightforward and a comprehensive uncertainty estimation may even be impossible [41, 102]. The correlations of responses in individual color channels should be considered, as well as specialities in software/algorithm and multichannel model performance, which are not known to the user of commercial software [29]. Almady et al. [103], for example, compared three different software tools (FilmQA Pro, Radiochromic.com and VeriSoft) and found up to 2.4% large differences in global gamma scores, without reporting achieved precision. Vera-Sánchez et al. [41] proposed a Monte Carlo method of uncertainty estimation for radiochromic film dosimetry (both single channel and tripple channel), being the only paper dealing with multichannel dosimetry uncertainties to date. They reported total uncertainty of 2-3% (except for very low doses below 1 Gy) for the triple channel method as well as for the single-channel method in the red and the green channels. The blue channel had larger uncertainties. While for EPSON V800 scanner, the multichannel methods were found to have lower uncertainties, for EPSON 10,000 XL, lower uncertainties were seen with single channel dosimetry and doses above 4 Gy. For these reasons, single channel and triple channel dosimetry was compared within the same software only (FilmQA Pro, see chapter 3.1.14 above) to minimize possible differences between the methods and uncertainties for the triple channel method are not reported.

The main sources of uncertainty for radiochromic film dosimetry, also according to previous publications [43-47], are the following: scanner and film uniformity (of the whole film sheet that is to be cut into pieces and of a small piece of film locally), reproducibility of scanning, inter-film reproducibility, effect of delay between irradiation and scanning, room light effects and film orientation. The latter two were left out from the uncertainty budget because in the film dosimetry procedure used here they can be neglected. The above affect the scanner values obtained during film scanning. Another source of uncertainty that has to be treated separately is the uncertainty of fit when fitting the calibration data with a curve. The magnitude of uncertainties depends on the protocol, the dose range, the color channel, the film model and the scanner model [29]. If uncertainty of absolute dose estimation is of interest, the uncertainty of dose measured by an

ionization chamber must be included. The expected overall uncertainty for single channel film dosimetry is around 2% [29].

The total uncertainty associated with dose estimation from radiochromic film dosimetry in this work was computed as the quadratic sum of experimental dose uncertainty and fitting uncertainty, similarly to [45]. Additionally, for absolute dose estimate, the ionization chamber dose uncertainty was added. However, in the work [45], they estimated dose calibration curve as a relation between dose and net optical density. Here it was estimated as the relation between dose and scanner value and a fourth order polynomial was considered as the proper calibration function.

Table 3-6: The standard uncertainties in scanner values (SV) or dose values (Gy) and the relative uncertainties for absolute dosimetry estimated for our EBT3 film method.

Source of uncertainty	**Standard uncertainty**	**Relative standard uncertainty [%]**
Scanner reproducibility – short term	7.68 SV	0.04
Scanner reproducibility – long term	2.36 SV	0.01
Inter-film reproducibility	7.68 SV	0.04
Scanner uniformity	83.33 SV	0.60
Whole film uniformity	13.44 SV	0.07
Piece of film uniformity	18.03 SV	0.13
Postexposure delay	2.36 SV	0.01
Combined uncertainty σ_{SV}	87.06 SV	0.64
Experimental dose uncertainty σ_{Dexp}	0.03 Gy	1.44
Uncertainty of fit σ_{Dfit}	0.02 Gy	1.17
Ionization chamber σ_{ion}	0.03 Gy	1.33
Total uncertainty (k=1)	**0.05 Gy**	**2.28**
Expanded total uncertainty (k=2)	0.09 Gy	4.56

Thus, the total uncertainty of dose estimation is given by

$$\sigma_D = \sqrt{\sigma_{Dexp}^2 + \sigma_{Dfit}^2 + \sigma_{ion}^2} \tag{3.4}$$

Where σ_{Dexp} and σ_{Dfit} are the experimental and fitting uncertainties and σ_{ion} is the uncertainty of absolute dose measurement with an ionization chamber at the Thomayer Hospital. The experimental uncertainty is given by

$$\sigma_{Dexp} = (a \cdot 4SV^3 + b \cdot 3SV^2 + c \cdot 2SV + d) \cdot \sigma_{SV} \tag{3.5}$$

Where *a, b, c* and *d* are coefficients of the calibration polynomial converting scanner values (*SV*) to dose values and *SV* is the scanner value for a piece of film irradiated with 2 Gy (here *SV* stands for

physical quantity but in Table 3-6 it is also used as physical unit). Uncertainty in this case is a function of dose and the values given in Table 3-6 are valid for 2 Gy which is the most representative dose for our experiments. The uncertainty of scanner value σ_{SV} takes into account the main sources of experimental uncertainty described above.

The uncertainty of fit σ_{Dfit} was estimated in MATLAB with the help of the optional output parameters of the *polyfit* function. This uncertainty given in Table 3-6 is again valid for the dose of 2 Gy.

It was assumed that the sources of uncertainty are not correlated.

Table 3-6 shows the uncertainty budget for absolute dose determination with our EBT3 film method. The one sigma total uncertainty (2.28%) falls below the dose difference criterion chosen for gamma analysis in our experiments (3%). The expanded total uncertainty (k = 2) does not. However, these uncertainty values are valid for a dose of 2 Gy, while for higher doses the relative uncertainty slightly decreases and for lower doses it rapidly increases, as can be seen in Figure 3-3. On the other hand, the absolute standard uncertainty in Gy actually increases with dose (see Figure 3-4). In IMRT clinical plan verification, the dose on film was approximately in the range 0-5 Gy. The gamma analysis threshold for film dosimetry was 10% of the maximal dose, which is approximately 0.5 Gy – doses below this value are not taken into account in gamma index calculation. Moreover, the uncertainty budget did not take into account the covariance terms. It was assumed that the sources of uncertainty are not correlated which is not necessarily true [46]. This would further decrease the uncertainty – Bouchard et al. [46] estimated that for a third order fit function and depending on several parameters, especially dose, neglecting all covariance terms leads to an overestimation of uncertainty by the factor 1.8-5. Therefore, a 3% dose difference criterion seems optimal for our gamma analysis and it is easily comparable to the clinical results in the Thomayer Hospital in Prague, where the same criterion is used. It needs to be pointed out that the use of relative gamma (as done in our work) is not so dependent on absolute dose calibration. Doses are normalized to a high dose value representing the maximum dose (or similar) in each dose distribution. Therefore, large relative uncertainties for low doses do not matter. Also, for relative gamma analysis the uncertainty σ_{ion} should be left out from equation 3.3 and the total uncertainty for a dose of 2 Gy then would be 1.85% for k = 1 and 3.70% for k = 2. Uncertainties for other dose levels would change similarly (but in case of an ionization chamber, the relative uncertainty does not depend on the level of dose).

The uncertainty of dose determination estimated for our film method is in agreement with numbers published in literature [25, 43, 47], even though each group uses different equipment and methodology.

The sources of uncertainty listed in Table 3-6 are believed to be the ones that should be taken into account in our particular experimental conditions. Different types of uncertainties are taken into account in different scientific papers [43, 46, 47]. Sometimes also the irradiation process associated uncertainties are considered, such as linac output reproducibility, stochastic nature of dose deposition, or dose variation within region of interest. Linac output for Siemens Artiste was stable, both the short-term and the long-term stability (considering the experimental period for which one film batch and one calibration curve was used) was within 0.5%. For Varian TrueBeam it was within 1%, but this was taken into account during the film calibration procedure. Stochastic nature of dose deposition was overcome by calibrating with relatively large pieces of film. Dose variation due to

non-ideal field flatness was neglected because for a 10 x 10 cm^2 field on the Siemens Artiste and Varian TrueBeam machines, field flattness was around 1 % for both energies (defined in the flattened region of dose profiles as 100 - D_{min}/D_{max}*100) and the calibration film pieces were of the size 5.8 cm x 5.8 cm.

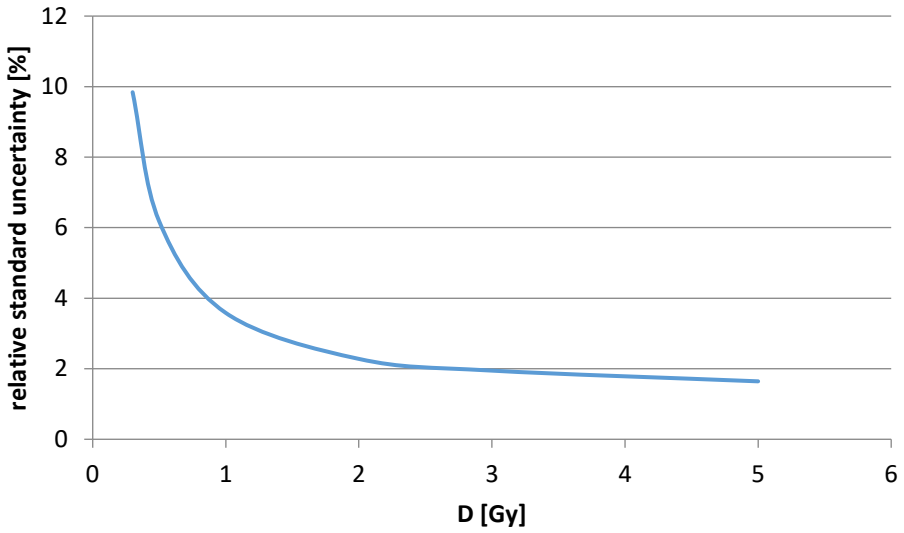


Figure 3-3: Relative standard uncertainty of absolute dose determination with film as a function of dose.

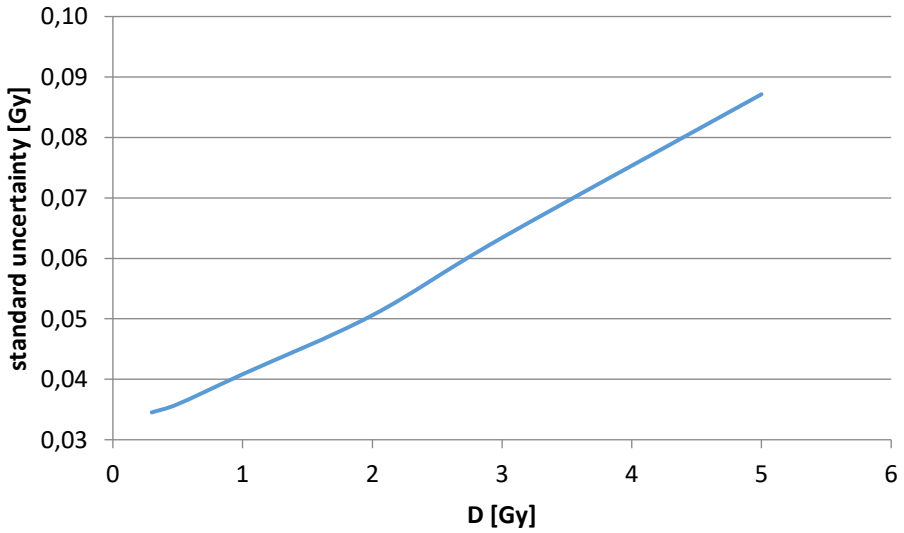


Figure 3-4: Standard uncertainty of absolute dose determinaton with film as a function of dose.

In order to see how the uncertainty in relative dose estimation propagates to gamma scores obtained with our film dosimetry method, one patient was selected (Patient no. 7, 18 MV prostate, IMRT, Varian TrueBeam – see details in chapter 3.3.5.1) and gamma analysis was performed 50 times with our 2D MATLAB code (see chapter 3.3.3 for details). There were no changes in the script

between the runs except that random Gaussian noise with a mean of 0 and a standard deviation estimated in this chapter for relative dose measurements (0.037 Gy) was added to the dose data in each film. This was done for 5 transverse, 5 sagittal and 5 coronal planes measured with film in a pseudo-3D manner (see chapter 3.3.1 and 3.3.2 for the measurement setup). In total, we got 50 values of gamma score for each of the 15 planes, affected by relative dose uncertainty. The maximum difference in gamma score (G_{max} minus G_{min}; G_{max} is the maximum gamma score value from the 50 subsequent runs and G_{min} is the minimum value) for each of the 15 planes are shown in Figure 3-5. The maximum difference was 4.4% in plane no. 5, followed by a value of 2.8% in plane no. 1. But these planes lied outside the high-dose area, beyond the edge of the prostate PTV, and the gamma score values for these two planes were 70% and 78%, respectively. In the rest of the planes, the difference in gamma scores were below 2%, in most cases below 1%. This is a very rough estimation of the possible propagation of uncertainties. The resulting gamma scores depend in a complex way on the input dose distributions, so for plans of a very different complexity, for example, the estimation could be different.

Having understood the performance of the films and scanner, it was decided to use doses of typically 2-3 Gy to irradiate films with clinical IMRT fields or plans, where possible, and to use a 10% threshold in gamma analysis.

Another source of uncertainty that can affect the results of gamma analysis is uncertainty in position of the pieces of film irradiated with IMRT fields/plans within a phantom and uncertainty in position of the phantom itself. The overall uncertainty in position should be compared to the selected criterion for gamma analysis – 3 mm DTA (distance-to-agreement).

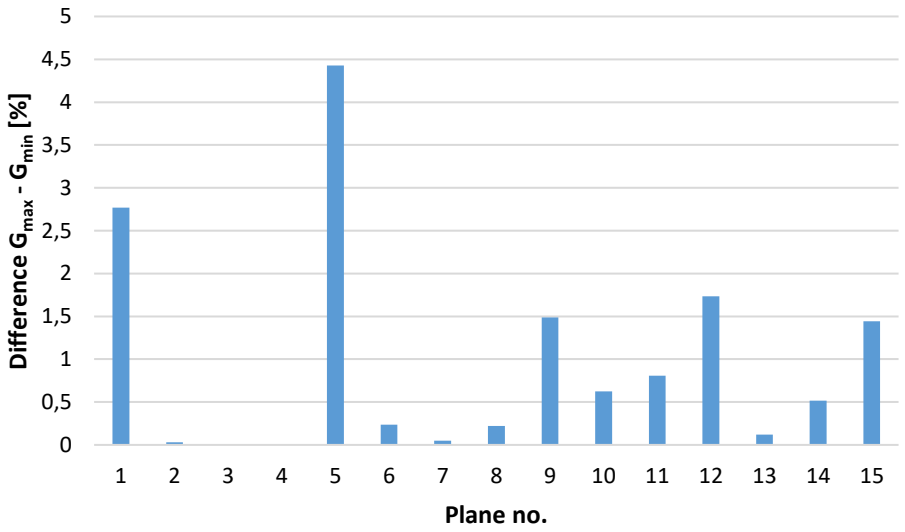


Figure 3-5: Difference between the maximum gamma score (G_{max}) and the minimum gamma score (G_{min}) in 50 runs of the MATLAB 2D code for gamma analysis for Patient no. 7. Results are shown for 5 transverse (1-5), 5 sagittal (6-10) and 5 coronal planes (11-15) measured with EBT3 film in a pseudo-3D manner.

Phantoms used in this study were the RW3 slab phantom and the I'mRT Phantom (Scanditronix Wellhofer North America, USA). Details are given in chapters 3.2 and 3.3. For the RW3 phantom, irradiation was performed with gantry at 0° only, so positioning of the phantom could have been

affected by uncertainties in SSD. The I'mRT Phantom was positioned into the isocentre with the help of lasers. Film pieces were positioned within the RW3 slab phantom for field-by-field dosimetry using markers on film and light field. For pseudo-3D dosimetry, film pieces were positioned within the I'mRT Phantom using markers on film and markers on the phantom slabs (drawn on the surface). When filling the I'mRT Phantom with slabs, the thickness of five pieces of film loaded at a time caused that one of the 2 mm thick slabs had to be replaced with a 1 mm thick slab. This resulted in 1 mm additional position uncertainty in one of the directions. (See schematic drawing in Figure 3-7 in chapter 3.3.2.)

The overall estimated position uncertainty for 2D field-by-field dosimetry is 1.4 mm. The maximum position uncertainty for 3D dosimetry in the I'mRT Phantom is estimated to 2 mm. Therefore, the 3 mm gamma analysis criterion seems reasonable. Note, however, that dose distributions were matched with the help of dose profiles during gamma analysis in those situations where an obvious shift in position occurred during measurement (except for intentionally introduced position errors).

3.1.16 Summary

Our own film methodology was established for EBT3 radiochromic films and EPSON Perfection V700 Photo scanner. In our film protocol, we use 16 dose points (0 – 5.5 Gy) for calibration and a 4th order polynomial calibration function is obtained in MATLAB (dose versus scanner value) for each combination of linear accelerator/energy/dose rate/film batch. We use the red channel only, films are scanned one by one in the center of the scanner in transmission mode at 72 dpi. Prior to film scanning, 5 blank scans are obtained in order to warm up the scanner. All films are scanned 72 hours after exposure. Before evaluation, a 5 x 5 pixels median filter is applied. Short-term and long-term reproducibility of the scanner and scanner uniformity are taken into account in the uncertainty budget, as well as whole-film and piece-of-film uniformity, inter-film reproducibility and post exposure delay. The following aspects are not taken into account in our film protocol because it has turned out that they have little effect: lateral response artifact, film-to-light distance, angular dependence, subtraction of a blank film prior to irradiation and using multiple channels for evaluation.

The overall uncertainty of absolute dose determination with our EBT3 film method was estimated to 2.28% (k = 1) and 4.56% (k = 2). These numbers are valid for a dose of 2 Gy on film and are comparable to other published methodologies [25, 47, 104]. Covariance terms were not included in the uncertainty estimation, but this might greatly improve these numbers. For relative applications, the uncertainty of film dosimetry according to our protocol is 1.85% (k = 1) and 3.70% (k = 2). Our film methodology is optimized for the given experimental conditions (EBT3 film, EPSON Perfection V700 Photo scanner).

3.2 Benchmarking film dosimetry against current practice

Several studies have been carried out [105-108] and most of them conclude that electronic detectors (such as ionization chamber arrays or diode arrays) perform better than film (EBT2 or EBT3), due to uncertainties inherent in film itself, like heterogeneity and noise. The aim of this part of our study is to compare radiochromic film dosimetry as implemented in this work to the ionization chamber array PTW Seven29 placed in an RW3 slab phantom (field-by-field verification) and to the ionization chamber array PTW OCTAVIUS Detector 1500 placed into the OCTAVIUS 4D Modular phantom (pseudo-3D verification), in terms of their performance for IMRT and VMAT plan verification. The aim

is to prove that film performs well enough to be used for our further experiments. The tested detector and phantom configurations are given in Table 3-7.

Table 3-7: Phantom and detector setup for benchmarking film dosimetry and tested patients (details about patient plans are given in Table 3-9 andTable 3-11).

Technique	Detector	Phantom	Gantry position	Detector position	Tested patients	What was tested
Field-by-field	PTW Seven29	RW3 slabs	0°	isocentre + different depths	Patient no. 5 (clinical IMRT, 9 fields, Siemens)	All fields 45 planes - film 25 planes - PTW Seven29
	EBT3 film			isocentre	Patient no. 4 (error-induced IMRT, 7 fileds, Siemens)	3 fields with 1% MU error
						3 fields with 1 mm leaf error
Pseudo-3D	PTW OCTAVIUS Detector 1500	OCTAVIUS 4D	VMAT dual arc 179° - 181°	rotating in isocentre	Patient no. 9 (clinical VMAT, 2 arcs, TrueBeam),	The whole plan
	EBT3 film	I'mRT Phantom		5 transverse, 5 sagittal and 5 coronal planes		

Measurements were carried out on Siemens Artiste (the IMRT field-by-field plan) and TrueBeam (the VMAT plan) in the Thomayer Hospital in Prague, using the photon energy 18 MV. This energy was clinically used for the majority of IMRT patients at the site – for pelvis cases. Due to the amount of measured data and the time available at the linac, all measurements could not be performed on the same day. However, relevant quality assurance tests were taken regularly to minimize uncertainty caused by varying machine performance. Particularly, the dose output, MLC (multi-leaf collimator) positioning and gantry angle were tested and fell within 1%, 1 mm and 0.5° tolerance limits.

3.2.1 IMRT and VMAT plans

3.2.1.1 Clinical plans

The evaluated detectors for field-by-field IMRT verification (EBT3 film and PTW Seven29 array) were irradiated with an IMRT clinical step and shoot plan for prostate and lymph nodes prepared for a Siemens Artiste photon energy 18 MV. Treatment planning was performed in XiO 4.80. The plan contained 9 fields that were distributed evenly in the whole 360°. The gantry angles used for fields no. 1, 2, 3, 4, 5, 6, 7, 8 and 9 were 0, 40, 80, 120, 160, 200, 240, 280 and 320°, respectively. General

High Smoothing was applied during optimization and other optimization parameters were left default. The prescribed dose was 44 Gy in 22 fractions. Altogether, 45 planes were tested with film and 25 planes (4 fields at 5 depths and 5 fields only at isocentre) were tested with the PTW Seven29 array. (This plan is used further in the work and is denoted as Patient no.5. – see Table 3-9.)

The evaluated detectors for VMAT pseudo-3D verification (EBT3 film and PTW OCTAVIUS Detector 1500) were irradiated with a VMAT clinical plan for prostate prepared for a Varian TrueBeam photon energy 18 MV. Treatment planning was performed in Eclipse 16.1 and the dose distribution was calculated with the AAA (Analytical Anisotropic Algorithm) algorithm v. 16.1. The plan contained 2 full arcs (179° - 181°) with a rotated collimator (30° and 330°). The prescirbed dose was 70 Gy in 28 fractions. (This plan is used further in the work and is denoted as Patient no.9 – see Table 3-9.)

3.2.1.2 Error-induced plan

Another clinical pelvic plan was chosen where errors were introduced in order to test the ability of different detection systems to reveal the errors. (One error-induced plan was used for benchmarking purposes while more error-induced plans were used in our further experiments with the benchmarked method.) The original clinical plan is also used further in this work and is denoted as Patient no. 4 (see Table 3-9). Patient no. 4 was a simultaneous integrated boost (SIB) planned in XiO 4.80 for prostate, seminal vesicles and lymph nodes, to be treated on Siemens Artiste with the energy 18 MV and with 82 Gy (for the high-dose prostate region) in 41 fractions and had only 7 treatment fields with gantry angles 45, 80, 160, 315, 280, 200 and 0°. No smoothing was applied during optimization in this case. The type of introduced errors was chosen so that the errors would already have some non-negligible clinical impact on DVH parameters and at the same time they could remain hidden by normal QA procedures.

Errors were introduced into fields no. 1, 2 and 4. The first type of error was the manual change of several segment weights within the treatment field. For each field, number of MUs was changed in two largest segments,, the resulting MUs for these segments were 1% lower in fields no. 1 and 4 and 1% higher in field no. 2. It must be noted that in XiO this kind of change affects all field segments. Their MU weights are changed so that the total dose weight of the field in Gy in an arbitrary point chosen by the algorithm remains constant. However, total number of MUs also changes. For field no. 1 the total beam MUs were 85 originally and 89.3 after modification, for field no. 2 it was 130.5 before and 122.2 after and for field no. 4 it was 89.3 before and 94.4 after change. This information is also summarized in Table 3-11.

The second type of error was a manual change of leaf positions. For fields no. 1, 2 and 4, leaf positions were changed for all segments and all leaves in the 10 cm central area (-5 cm and +5 cm around the isocentre, MLC leaves are 0.5 cm wide at isocentre). In field no. 1, each leaf bank was shifted 1 mm so that the apertures became 2 mm narrower. In field no. 2 the shift was 2 mm and in field no. 4 it was 3 mm in the same way. It must be noted that after this kind of change, XiO recalculates weights of all segments in the field so that the total beam weight in Gy in an arbitrary point chosen by the algorithm remains constant. This results in MU change in all segments and in the treatment field in total, affecting the relative dose distribution within the field. Thus, total beam MUs changed as follows: for field no. 1 - 85 MU before and 99.5 MU after shift, for field no. 2 - 130.5 MU before and 117.4 MU after shift and for field no. 4 -89.3 MU before and 109.1 MU after shift. Unlike in 3D CRT (3D conformal radiotherapy) fields, MLC leaf positions cannot be modified manually

for IMRT fields/segments/control points in Mosaiq version 2.50.0507 (Impac Medical Systems, Inc.). Simulating this type of error separately (due to erroneus linac leaf travel or calibration) would require modification of DICOM plan in a third-party software and reconfiguration of Mosaiq import options, which was not desirable. This information is also summarized in Table 3-11.

The choice of these errors to be tested is supported, for example by a multicentre study [109] where similar errors in Elekta step-and-shoot IMRT plans were detected with the verification system used by the authors (IQM large-area ionization chmaber mounted on the linac head). They introduced 1 mm leaf bank shift error and 1-3 MU error per beam, these errors already had a non-negligible impact on DVH parameters, similarly to our observation with Siemens Artiste.

3.2.2 PTW Seven29 measurements in an RW3 slab phantom

The procedure for IMRT plan verification consisted of measuring the isocentric dose maps of individual fields with gantry at 0° in an RW3 slab phantom (30 x 30 x 30 cm^3) at reference depth (10 cm for 18 MV photons) using the PTW Seven29 ionization chamber array. Additionally, for some fields the PTW detector was placed to different depths in the RW3 phantom, corresponding to the depths measured with film (see explanation for film in section 3.2.4). Predicted dose maps were exported from the treatment planning system XiO with the Modulation QA tool. They were calculated for a virtual water phantom of 30 x 30 x 30 cm^3 in size. Proper calibration of the detector array was ensured, taking into account the difference between water and RW3 material. A reference depth of 10 cm in the solid phantom for 18 MV photon beams and a reference field of 10 x 10 cm^2 were used to irradiate the central chamber with 2 Gy, the rest of the chambers were adjusted in the PTW MultiCal software (PTW, Freiburg, Germany) relatively to the central chamber with the help of a reference calibration matrix covering the whole area of the detector. The uncertainty of absolute dose measurement (k = 2) with the PTW Seven29 array at the Thomayer Hospital is 3.5% (based on the assessment of uncertainty for the cross-calibration procedure).

3.2.3 PTW OCTAVIUS 1500 measurements in the OCTAVIUS 4D phantom

To measure a VMAT plan in a pseudo-3D manner, the OCTAVIUS Detector 1500 was placed into the rotating cylindrical OCTAVIUS 4D Modular phantom by PTW. This phantom includes an inclinometer attached to the gantry and the phantom rotates simultaneously with the gantry, so that the beam is always perpendicular to the detector array. During commissioning of the OCTAVIUS 4D system at the Thomayer Hospital, a set of PDDs was obtained in a water phantom for different field sizes and each energy, down to a 2 x 2 cm field size, and an SSD corresponding to the geometry of the OCTAVIUS 4D phantom (SSD = 85 cm). A proper material density and HU value was set to the virtual image of the OCTAVIUS 4D phantom in the TPS. Because of simultaneous use of the AAA and Acuros XB algorithms, the OCTAVIUS 4D phantom in the TPS was assigned 2 HU. For our experiemnts, only the AAA algorithm was used. A calibration plan with reference conditions (10 x 10 cm field size, SSD = 85 cm, gantry in zero position) was prepared in the TPS Eclipse for each energy. Before the irradiation session, the PTW OCTAVIUS 1500 ionization chamber array placed into the OCTAVIUS 4D phantom was calibrated, taking into account the daily linac output, temperature and pressure. VeriSoft v. 8.0 reconstructs 3D dose distribution in the phantom based on the measured 2D dose images and the set of PDDs from commissioning. The accuracy of the reconstructed dose distribution according to the manufacturer (personal communication) is 2% - 6% depending on plan complexity and resolution of the used ionization chamber array.

3.2.4 Field-by-field measurements with film

Films were cut into pieces of 8.4 x 6.8 cm^2. The sheets were placed in several coronal planes into an RW3 slab phantom (SAD — source to axis distance — setup with SAD = 100 cm, isocentre at 10 cm depth). The measured coronal planes were at the depths of 7, 8, 10, 12 and 13 cm. The initial idea of placing films to different depths and irradiating them with gantry at 0° was to identify whether all film pieces can give comparable results for a given field. Agreement of the measured and calculated dose matrix can differ at different depths, depending for example on the TPS dose calculation algorithm and its performance in the patient and phantom materials. Each tested field was irradiated 10 times in order to obtain doses of 2-3 Gy on film (maximum dose usually around 4.5 Gy). This is because of the poor performance of film for low doses. Again, predicted dose maps were exported from XiO using the Modulation QA tool and a virtual water phantom of the size 30 x 30 x 30 cm^3. Difference between water and RW3 material was taken into account at film calibration.

3.2.5 Pseudo-3D measurements with film for benchmarking

To measure the VMAT prostate plan of Patient no. 9, films were cut into pieces of 8.4 x 10.1 cm^2. They were placed into 5 tranverse, 5 sagittal and 5 coronal planes of the I'mRT phantom (each direction at a time) and irradiated with the original VMAT plan. Thus, a measured pseudo-3D distribution was obtained and could be analzyed with 3D gamma analysis. The pseudo-3D method is further described in chapter 3.3.

3.2.6 2D gamma analysis in OmniPro I´mRT and VeriSoft v. 3.1

Gamma analysis was performed to quantify the agreement between predicted (from TPS) and measured dose maps. Film measurements were evaluated in OmniPro I'mRT software version 1.7. All measured IMRT fields were fitted with the calibration curve in MATLAB using the polyval function. The gamma criteria were 3% dose difference and 3 mm distance to agreement, search distance 4.5 mm, avoiding signals with lower doses than 10% of the maximal predicted dose. Lower threshold than 10% for films would not be practical due to poor film performance for low doses. Relative global gamma was performed (i.e. each of the compared distributions was normalized to a chosen value in that distribution) and both the predicted and the measured dose distributions were normalized to the average value of the chosen region of interest. ROI was chosen so that film edges and film markers would be excluded from evaluation. Predicted and measured dose distributions were matched with the help of film markers and with the help of X and Y dose profiles. Dose grid of the predicted matrices was converted to the same pixel distance as for the scanned (measured) matrices (0.352778 mm) using linear interpolation. Measured dose matrices were smoothed with a median filter of the size 5 x 5 pixels. (Gamma analysis was also performed with our 2D MATLAB code described in chapter 3.3.3, with the same parameters as in OmniPro I'mRT, except for normalization. Due to the non-reproducibility of the manually chosen ROI, matrices were normalized to their 75^{th} percentile in MATLAB. This was only done to see if there is any significant difference between the two softwares.)

PTW Seven29 measurements were evaluated in PTW VeriSoft version 3.1. Gamma criteria were also 3% dose difference and 3 mm distance to agreement, avoiding signals lower than 5% of the maximum predicted dose. Higher threshold than 5% of maximum dose is not possible in the VeriSoft version used. The whole detector area was evaluated and relative global gamma evaluation was performed. Search distance is not adjustable. Normalization of the matrices was left default, to the maximum dose point in the distribution. However, normalization does not affect

the result of gamma calculation in this version of VeriSoft, it only affects visual apperance of the matrices.

The clinically used tolerance limit of a 90% gamma score (percentage of points in the evaluated dose matrix with the gamma index lower than 1) was kept here for guidance. The aim was to compare the verification method as a whole – as it is or would be (in case of film) used clinically (particularly under the conditions at the Thomayer Hospital in Prague). The aim was not to analyze the differences between the detectors' and software's performance (with many user adjustable options), this would require a different study design. This is also the reason why results for film and MATLAB are not shown. The goal here is to show that the clinical method (2D field-by-field gamma analysis with PTW Seven29 and VeriSoft v. 3.1) and the film method being developed (2D field-by-field gamma analysis with EBT3 film) give similar results within their uncertainties or that the new method performs better.

3.2.7 3D gamma analysis in VeriSoft v. 8.0

The newer version of VeriSoft v. 8.0 that was used together with the OCTAVIUS 1500 ionization chamber array and the OCTAVIUS 4D phantom for 3D gamma analysis has more user-adjustable options. The main parameters for gamma analysis were used as described in chapter 3.2.6 for VeriSoft v. 3.1. (3%, 3mm, global relative gamma, normalization to the maximum dose in the predicted distribution). These are used clinically. However, the threshold dose could be changed to 10%. The gamma method implemented according to Depuydt et al. [110] was selected. Predicted dose distribution was exported from TPS Eclipse in DICOM format as RT Dose and RT Plan. Both files are needed to place the dose into the coordinate system of the phantom correctly in VeriSoft v. 8.0. Dose distributions – the predicted and the reconstructed one – can then be compared in VeriSoft v. 8.0. Either 2D plane-by-plane evaluation is possible (with 2D gamma index for each plane) or 3D gamma analysis can be performed. The 3D gamma analysis shows an overall 3D gamma score but also passing percentages for different isodose levels (i.e. how many points at a certain dose level pass the gamma test). The clinical tolerance for the 3D gamma score, as well as for the individual gamma scores at each dose level and for the 2D plane-by-plane gamma scores is set to 95%. The goal here, again, is to show that the clinical method (pseudo-3D verification with OCTAVIUS 4D equipped with OCTAVIUS 1500 detector evaluated in VeriSoft v. 8.0) and the film method being developed (pseudo-3D verification with EBT3 film placed in different planes of the I'mRT Phantom evaluated in MATLAB) give similar results within their uncertainties, or that the new method performs better.

3.3 Pseudo-3D verification with film

3.3.1 The pseudo-3D method

Our pseudo-3D film dosimetry method for IMRT plan verification uses EBT3 film sheets placed in several transverse, sagittal and coronal planes in the water equivalent I'mRT Phantom, using only its central cubic part. Each of the 3 directions must be measured separately in the phantom, due to its design (see Figure 3-6). Thus, the whole treatment plan must be delivered 3 times to obtain measured dose values in each of the 3 directions. The plan can be delivered with its original gantry angles. The measured pseudo-3D dose distribution (covering the whole 3D space) is built-up in MATLAB using the 3 separate measurements. The predicted pseudo-3D dose distribution is built-up in MATLAB using 2D planes exported from the TPS (the 2D planes correspond to the film sheet positions in the 3D space at measurement). Gamma analysis is performed in MATLAB using either 2D evaluation of each plane in the 3D space (2D plane-by-plane gamma analysis) or using

3D evaluation. The 2D evaluation with our 2D MATLAB code needs a 2D measured dose matrix and a 2D predicted dose matrix as input (several 2D planes can be evaluated at once). It gives a map of gamma indices for each plane, taking into account only dose values in the given plane (predicted and measured) and a 2D gamma score is computed for each plane. The 3D evaluation with our 3D MATLAB code needs the built-up pseudo-3D measured dose matrix and the built-up pseudo-3D predicted dose matrix as input. It gives a map of gamma indices in all measured points in the 3D space, taking into account all measured/predicted dose values from the pseudo-3D dose matrices previously built-up in MATLAB. 2D gamma scores can then be computed for each plane in the 3D space (plane-by-plane evaluation) but this time based on the pseudo-3D map of gamma indices (which is numerically different to the previous situation, even in the same points in the 3D space, also the 2D gamma score is computed with different input – see chapters 3.3.2, 3.3.3 and 3.3.4 for explanation). A pseudo-3D gamma score — a single number for the whole treatment plan — can also be obtained. It is computed as the percentage of points with gama index lower than or equal to one in the pseudo-3D matrix of gamma indices. The gamma pass/fail decision criterion for our pseudo-3D method used in this work is a 90% gamma score (2D or pseudo-3D) for Siemens Artiste and 95% for Varian TrueBeam. These values were used clinically at the Thomayer Hospital in Prague, so this criterion is used for comparison purposes. If the method is used clinically for a period of time and a larger amount of data is obtained, a different and perhaps more suitable criterion can be established for the particular hospital.

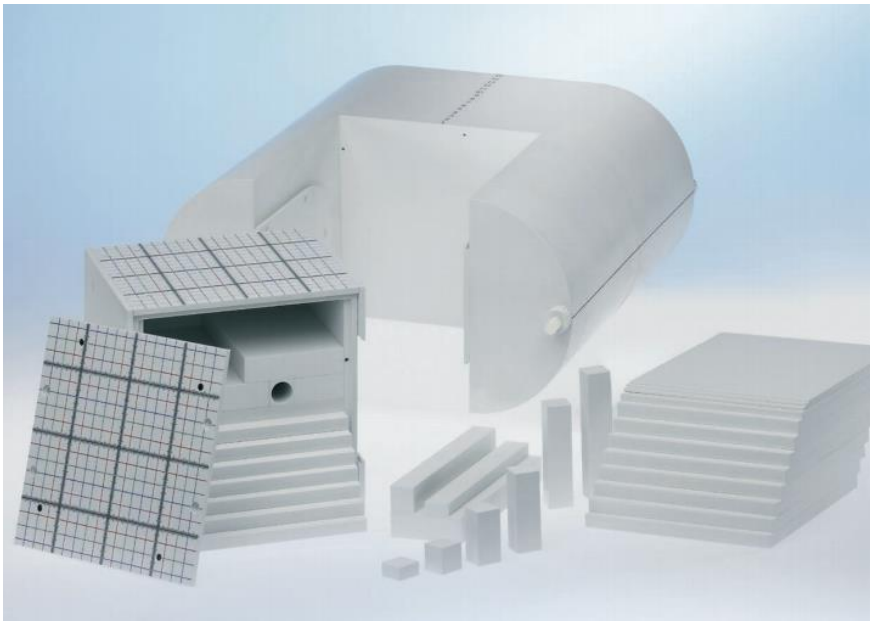


Figure 3-6: The I'mRT Phantom (Scanditronix Wellhofer North America, USA) [111].

The phantom and film setup for measurements, as well as the export of 2D planes from the TPS, are described in detail in chapter 3.3.2. Our 2D MATLAB code including gamma analysis parameters is described in chapter 3.3.3. Our 3D MATLAB code including gamma analysis parameters is described in chapter 3.3.4. IMRT plans that were measured with our pseudo-3D method are described in chapter 3.3.5.

3.3.2 Pseudo-3D measurement with film

Films were cut into pieces of 8.4 x 10.0 cm^2 and placed into an IMRT cube phantom into 5 transverse, 5 coronal and 5 sagittal planes close to the isocentre. One film sheet was always placed directly to the isocentric plane, the rest was plus and minus 2 cm and plus and minus 4 cm from the isocentric plane in the adjacent transverse, coronal and sagittal planes. This is schematically drawn in Figure 3-7. Films were stuck onto the phantom slabs with the help of markers drawn on the film sheets and markers drawn on the individual slabs. Then the phantom was loaded with all slabs and films and positioned into the isocentre of the linac with the help of lasers. For each direction (transverse setup, coronal setup and sagittal setup), the phantom was loaded again with a new set of films and the whole treatment plan was delivered at once with fields in their original gantry angles. Thus, a dose of approximately 2-3 Gy was deposited on film in most (but not all) cases. The IMRT slab phantom is filled with slabs of 1 cm thickness but 0.5, 0.2 and 0.1 cm slabs are also available. Because the film thickness is not negligible, having 5 sheets of film inside the phantom meant that one of the 1 cm slabs had to be replaced by the thinner ones. Therefore, there was a little air gap left (approx. 1 mm) and the films in coronal, sagittal and transverse planes might have been slightly shifted relative to each other in the 3D space.

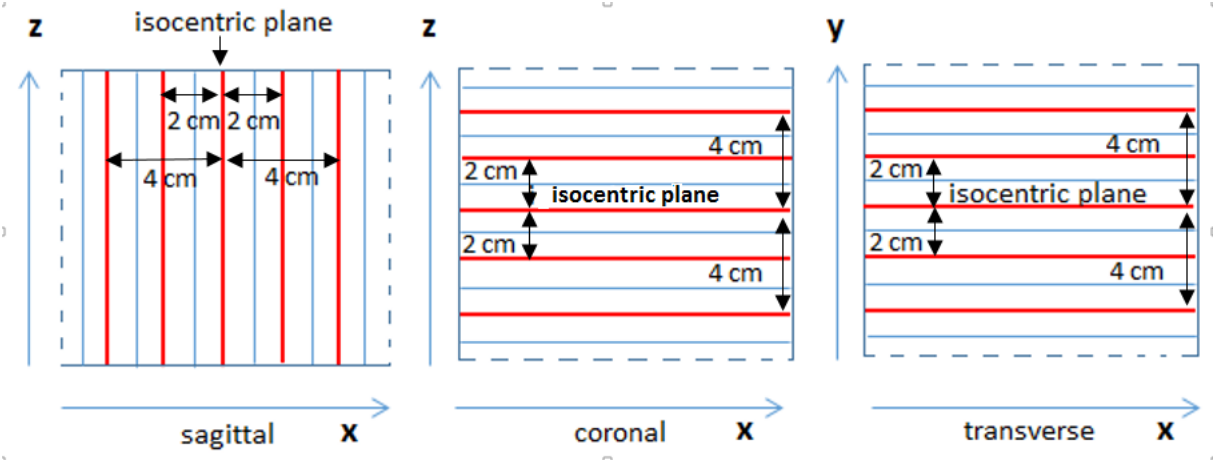


Figure 3-7: I'mRT Phantom loaded with film for 3D measurements. X is the latero-lateral direction on the patient couch, Y is the cranio-caudal direction, Z is the anterior-posterior direction. Red planes indicate film positions. (There were more than 2 slabs next to the last film position – the outer dimensions of the cube are 18 x 18 x 18 cm^3, each slab is 1 cm thick. One of the 1 cm slabs was replaced with 5 mm + 2 x 2 mm slabs because of the film sheets' total thickness.)

Predicted dose maps in the measured planes for Siemens Artiste were exported from XiO using the Dose Plane Export feature where the user can choose the direction of 2D map export. A virtual water phantom of the size corresponding to the I'mRT Phantom was used to keep the film calibration and the difference between water and RW3 material valid. Another option is to export 3D dose map directly from XiO. However, our measured dose matrix was sparse (2 cm resolution in adjacent planes) and the goal was to compare only measured dose points with corresponding predicted doses. So the rest of the dose data would have been unused. Also, at the time of our experiment, the licence for DICOM RT Dose export was not available. For Varian TrueBeam, predicted dose was exported from Eclipse in the measured planes in 2D in DICOM format. The selected export options were absolute dose, planar plan dose, 20 x 20 cm size, resolution of 512 x 512 pixels. A virtual water phantom of the size corresponding to the I'mRT Phantom was used, similarly to XiO.

3.3.3 2D gamma analysis in MATLAB

In order to evaluate the irradiated films with gamma analysis, the scanned films were fitted with an appropriate calibration curve (for each film batch, linac and energy) in MATLAB using the function *polyval* and measured dose values were obtained. A code for 2D and 3D gamma analysis was written in MATLAB for the purpose of this study.

$$r(\vec{r_e},\vec{r_r}) = |\vec{r_e} - \vec{r_r}|$$

$$\delta(\vec{r_e},\vec{r_r}) = D_e(\vec{r_e}) - D_r(\vec{r_r})$$

$$\Gamma(\vec{r_e},\vec{r_r}) = \sqrt{\frac{r^2(\vec{r_e},\vec{r_r})}{\Delta d^2} + \frac{\delta^2(\vec{r_e},\vec{r_r})}{\Delta D^2}} = 1 \tag{3.6}$$

$$\gamma(\vec{r_r}) = \min\{\Gamma(\vec{r_e},\vec{r_r})\}\forall\{\vec{r_e}\}$$

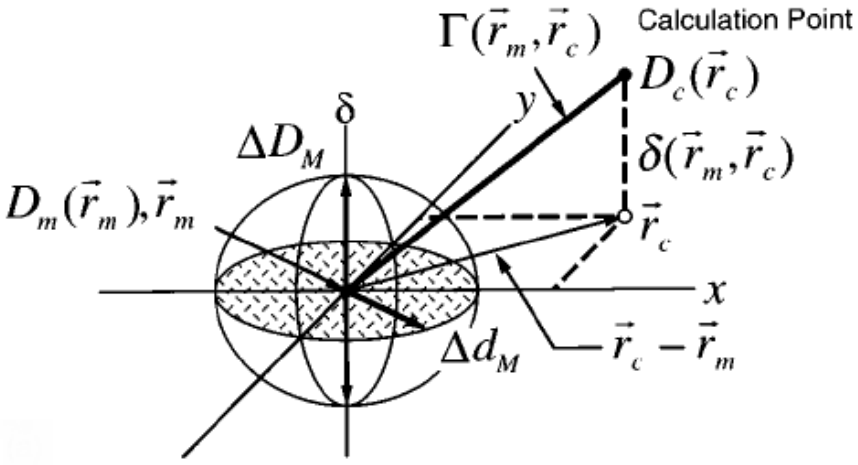


Figure 3-8: The gamma analysis concept by Low et al. [1].

2D gamma analysis according to [1] uses the equation of a 3D ellipsoid where two axes represent spatial distance and the third axis represents dose – see Figure 3-8 and equation 3.6. Thus, both dose difference and spatial difference in the compared dose distributions can be taken into account. One of the dose distributions is a reference one – in our case the planned dose distribution exported from the TPS. Each point of this distribution is assigned a gamma index $\gamma(\vec{r_r})$ by searching the evaluated dose distribution — in our case the measured one — for similar doses and finding the minimum values. Gamma index is smaller than 1 if the point falls within the volume of the ellipsoid with boundaries defined by the dose difference ΔD and distance-to-agreement (DTA) Δd criterion. Either the whole matrix of doses can be searched or the search distance may be restricted by a user given parameter to an area around the given point. The gamma score is a percentage of points in the gamma index map that passes the criterion $\gamma(\vec{r_r}) \leq 1$.

Gamma analysis as described by [1] was implemented in our code. Discretization and false negative gamma index values discussed by [110] were not considered for the following reasons: Spatial resolutin within the film plane was 0.35 mm. As noted in (Low and Dempsey), the error of gamma caused by pixelization is reduced to 0.2 if resampling to a resolution of 1 x 1 mm^2 takes place. In the third dimension, where film planes were 2 cm apart, the solution of [110] would affect minimum

of points. Another important reason was the comparison of results to OmniPro I'mRT version 1.7. There, the conventional gamma evaluation is also implemented according to Low et al. Another option in the software is a so-called digital gamma, which takes into account discrete distributions. However, this option was not used in our evaluation. Lastly, we were interested in the actual values of the gamma indices and the solution of Depuydt et al. [110] gives only a pass or fail information. It should also be noted that the choice of the reference matrix plays an important role. If the measured dose distribution was taken as reference, this would in some cases significantly affect the gamma index map and the gamma score.

The input to our MATLAB 2D code, as usual, is a predicted 2D dose distribution exported from the TPS and a 2D dose distribution measured with film (the scanned RGB film image of which the red channel was used). Then a series of operations is performed (adjustment of matrix orientation – flipping, eventually rotation, eventually space shifts according to dose profiles, dose images and markers on film, choice of ROI, size adjustment of the predicted matrix, interpolation of the predicted matrix, smoothing of the measured matrix, normalization of matrices). Then the gamma function is called with selected dose difference and DTA criteria and threshold. The output of the code is a 2D map of gamma indices, 2D gamma score, gamma histograms (and any other desired output can be added). The same process can be used for 2D plane-by-plane evaluation (where 2D gamma scores are computed for various planes in the 3D space), as well as for field-by-field evaluation of a treatment plan.

In order to check proper funcionality of our 2D code, a manual calculation of gamma indices and gamma score was performed for artificial small dose matrices (an artificial predicted matrix and a corresponding artificial measured matrix). The same calculation was performed with our 2D MATLAB code and it gave the same results.

Selected film data was also evaluated in OmniPro I'mRT software version 1.7 (only in 2D) to be able to compare our MATLAB code with existing commercial software. OmniPro I'mRT was chosen because most parameters in the evaluation can be comprehensively adjusted by the user, unlike with FilmQA Pro or VeriSoft. Thus, most evaluation parameters could be the same for MATLAB and OmniPro, except for adjustments that are done graphically with a hand tool in OmniPro I'mRT and therefore are not perfectly reproducible. In chapter 3.1 it turned out that multichannel analysis was not needed in our case.

All measured dose planes in the I'mRT Phantom in 3D space (5 coronal, 5 sagittal and 5 transverse planes for each patient) for Siemens Artiste were evaluated with the above described 2D gamma analysis in MATLAB as well as in OmniPro I'mRT and results were compared quantitatively. This process refers to Part B of Table 3-1. This data set was chosen for evaluation of our 2D MATLAB code against a commercial programme because it contained the most data.

In both softwares, the gamma criteria were 3% dose difference and 3 mm distance to agreement, search distance 4.5 mm, avoiding signals with lower doses than 10% of the maximal predicted dose. Lower threshold than 10% for films would not be practical due to poor film performance for low doses. Relative global gamma was performed (global gamma by definition means that normalization is performed globally, it is not locally variable). ROI was chosen so that film edges and film markers would be excluded from evaluation. It needs to be pointed out that

this is not perfectly reproducible in MATLAB, because in OmniPro it is done interactively with a hand tool. In both softwares, predicted and measured dose distributions were matched with the help of film markers and with the help of X and Y dose profiles. Dose grid of the predicted matrices was converted to the same pixel distance as for the scanned (measured) matrices (0.352778 mm) using linear interpolation. Measured dose matrices were smoothed with a median filter of the size 5 x 5 pixels. In OmniPro I'mRT, dose distributions were normalized to a selected point in a high-dose homogeneous region (both the predicted and the measured dose maps were normalized to the same point) inside the region of interest (ROI). This was not reproducible in the subsequent evaluation in MATLAB and the normalization method was chosen to be the 75th percentile of the ROI. This is the most suitable type of normalization for film dosimetry because the dose matrices remain noisy even after smoothing. We intended to use this type of normalization for comparisons between different film dosimetry methods later in our work (field-by-field, 2D plane-by-plane and pseudo-3D). The differences in gamma scores between the two softwares thus reflect also the effect of different normalization and how it affects the gamma score. This is something we wanted to check in this part of our study.

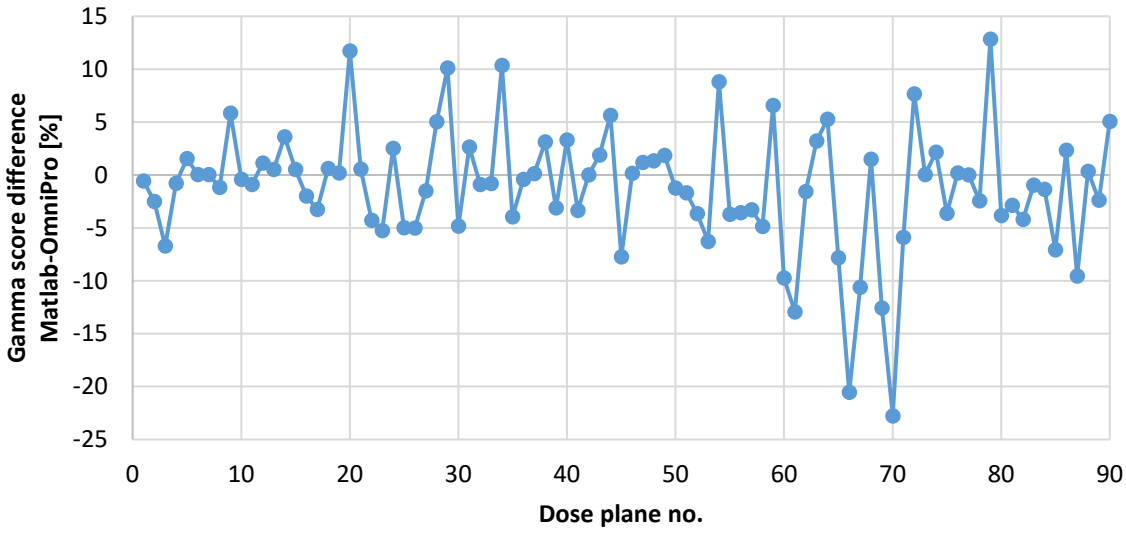


Figure 3-9: 2D gamma score difference between MATLAB and OmniPro I'mRT for all 2D planes (transverse, sagittal, coronal) for all patients in the I'mRT Phantom.

The magnitude of gamma score differences for individual planes is plotted in Figure 3-9. These can be explained by the nonreproducible parameters, namely normalization and ROI. Bigger differences were seen for homogeneous dose distributions because there is a more pronounced difference between the normalization strategies. A point dose normalization to a dose point in a high dose homogeneous region is not so dependent on the point choice in a homogeneous distribution. On the other hand, the 75th percentile might be quite different for a homogeneous dose distribution compared to a distribution with very low as well as very high doses. Relative gamma analysis is generally dependent on normalization and Table 3-8 reveals how big the differences are. Furthermore, as noted by Low and Dempsey [112], if noise is present in the evaluated dose

distribution, the gamma results are underestimated compared to the no-noise condition. In our case, measured film doses were smoothed with a median filter, however, some noise still remained in the evaluated distributions. Thus, the algorithm had more options to find the minimal combination of dose difference and spatial distance between the reference and evaluated distributions for a given point. This is valid for both OmniPro and MATLAB evaluation, however, normalization is also sensitive to this behaviour. It should be pointed out that normalization in relative gamma affects the values in equation 3.5. The largest encountered difference was 23% in gamma score but for most cases it was below 5%. Taking into account all parameters that can influence the resulting gamma score even within the same software (some of them mentioned above), the agreement between the OmniPro software and MATLAB code is very good. The gamma score values from which the differences in Figure 3-9 are computed are given in Table 3-8.

Table 3-8: 2D gamma score values for 90 planes in the 3D space as computed with OmniPro I'mRT and with our 2D MATLAB code.

OmniPro I'mRT															
Plane no.	Transverse planes					Sagittal planes					Coronal planes				
Patient no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	99.4	98.5	99.6	98.1	98.1	97.0	91.3	83.8	99.1	96.7	95.2	94.3	96.9	96.4	90.6
2	70.6	84.0	91.9	98.1	92.3	97.3	93.0	98.7	99.0	92.0	85.2	81.7	68.7	90.1	73.1
3	94.6	86.0	99.2	92.4	94.2	93.1	93.7	93.0	80.8	82.1	93.8	88.5	83.4	100.0	98.8
4	78.7	70.3	85.7	98.6	96.2	90.7	86.1	94.1	98.9	91.3	94.4	85.5	99.9	93.1	98.6
5	98.0	88.8	88.4	99.4	100.0	97.9	99.9	98.4	99.9	99.9	100.0	97.8	100.0	81.0	97.9
6	98.8	83.9	87.7	99.9	99.8	99.0	93.3	99.9	95.9	85.8	93.6	99.0	98.1	96.7	90.6
MATLAB 2D code															
Plane no.	Transverse planes					Sagittal planes					Coronal planes				
Patient no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	98.8	96.0	92.9	97.3	99.6	96.7	99.1	82.6	97.1	96.6	89.7	97.5	97.4	97.9	95.8
2	90.3	94.9	92.5	84.2	82.4	92.6	94.8	93.4	95.5	92.3	68.1	88.6	73.8	91.8	80.4
3	96.8	91.5	98.4	96.4	90.6	81.7	80.9	96.2	90.5	96.4	95.4	88.5	85.3	94.1	86.1
4	96.3	99.8	87.0	72.1	77.5	89.7	95.3	87.8	95.0	87.0	95.0	89.8	95.1	92.1	84.7
5	87.0	97.8	91.6	94.1	90.2	79.3	89.3	99.9	87.3	75.1	92.0	88.7	100.0	100.0	96.4
6	100.0	99.9	85.3	96.7	95.0	83.0	91.7	99.0	91.9	91.9	92.9	87.2	98.5	96.6	98.6

Also the statistical chi square test of independence was performed to see whether the difference between the OmniPro gamma scores and the MATLAB gamma scores from Table 3-8 is statistically significant or not. The chi square test of independence was used because it is suitable for this type of data, as already discussed in chapter 3.1.14. This test is used to determine whether two categorical variables are related or not. Our first categorical variable was a set of five gamma score bins: gamma score ≤ 80%, gamma score between 80% and 90%, gamma score between 90% and 94%, gamma score between 94% and 97% and gamma score > 97%. In other words, we had five categories of gamma scores. The second categorical variable was the software used to determine the gamma scores – either OmniPro I'mRT or MATLAB. A contingency table with frequencies for each combination of categories was created. The chi square test of independence was performed in MATLAB by writing our own script (there is no built-in function within the Statistical Toolbox). Our null hypothesis was that gamma scores and the software used to determine them (OmniPro or MATLAB) are independent at a 5% significance level. The so-called expected values were computed.

The assumption of the test is that each expected value is larger than 5, which was fulfilled with our data. **The p-value for the chi square test of independence was 0.0616 and therefore we cannot reject the null hypothesis. The differences in gamma scores between OmniPro I'mRT and MATLAB in** Table 3-8 **are not statistically significant.**

3.3.4 3D gamma analysis in MATLAB

To extend the idea of 2D gamma analysis into 3D, the equation of a 4D hyperellipsoid must be used, where three axes represent spatial distance and the fourth axis represents dose. Equation 3.5 remains the same, however, $\overrightarrow{r_r}$ and $\overrightarrow{r_e}$ are three dimensional vectors. The same gamma parameters were kept for our 3D gamma analysis code in MATLAB.

One approach could be to evaluate all points in the whole 3D space, perform interpolation between the measured film sheets and export the TPS predicted dose as a 3D cube. This is the approach of most commercial QA devices, but in our study we aimed at avoiding any computational algorithms and wanted to perform the evaluation with measured dose points only. Therefore, the evaluated 3D matrix was build from the measured film sheets in 3 directions and the reference 3D matrix was build of 2D planes exported from the TPS, corresponding to the measured ones. The 3D gamma code was written so that any number of points in the 3D space (up to all points in the 3D space) can be evaluated in the future. In our study, film planes were 2 cm apart. So the 3D gamma calculation here took into account, for each point in the reference distribution, those points in the evaluated distribution that were available within the search distance (chosen to be 4.5 mm). The evaluated distribution consisted of only those points in the 3D space where doses were actually measured with film. So the resulting gamma indices were also available at those points only. The overall 3D gamma score also took into account only those points where doses were measured.

It should be noted that if the search distance remains constant for both concepts, in our case 4.5 mm, in 2D it is a square neighbourhood 4.5 x 4.5 mm^2 and in 3D it is a cube neighbourhood 4.5 x 4.5 x 4.5 mm^3. Therefore, in a given dose plane, the set of gamma indices is the same for 2D and 3D gamma analysis, but in 3D there are additional points $\overrightarrow{r_e}$ available in other dose planes. Because the minimum value of gamma is searched, 3D gamma analysis should always give more passing points than 2D gamma analysis. In this feasibility study, films were placed 2 cm apart in adjacent planes, so the 3D matrix was rather sparse. However, the search distance was 4.5 mm in all directions, so there were at least some additional points $\overrightarrow{r_e}$ available in 3D gamma analysis for almost half of the points $\overrightarrow{r_r}$. The code was written as fully 3D, so if more film sheets closer to each other are placed in the IMRT cube phantom in the future, this code can be used for a truly 3D gamma analysis.

The input to the code in this case were the 2D measured dose planes in the 3D space and the corresponding 2D dose planes exported from the TPS. The same operations as described in 3.3.3 were performed with the matrices and a predicted and measured 3D dose cube was formed. Because the films were 2 cm apart in the phantom and the inplane resolution within the dose plane was approx. 0.35 mm, the 3D dose cubes actually contained a lot of points where dose values were assigned to zero (and thus fell below the 10% threshold for gamma analysis). These points were not reflected in the evaluation. Two types of normalization were used, which in this work are reffered to as local normalization and global normalization. Local normalization in this study means that each dose plane is normalized to its 75^{th} percentile. Global normalization means that all dose points in the 3D space are normalized to the 75^{th} percentile of the whole 3D matrix.

Some additional issues in 3D had to be addressed. The scanned images of films had to be rotated and flipped to form a 3D matrix with the same orientation of individual film sheets as during measurement. The orientation also had to be put in agreement with the predicted dose planes exported from the TPS. In the IMRT cube phantom, all films could not be irradiated at once but coronal, sagittal and transverse direction had to be measured separately (each plan was delivered 3 times to obtain the resulting 3D dose matrix). There might have been slight fluctuations in the linac output and leaf positioning and also the reproducibility of phantom positioning and film positioning inside the phantom had to be considered as described in chapter 3.3.2. As a result, the cross sections of the 3 different directions might have contained slightly different doses. Each cross section is represented by one row or one column of values in the 3D matrix in MATLAB.

To check these assumptions, all profiles overlapping in the isocentre (transverse versus sagittal profile, sagittal versus coronal profile and coronal versus transverse profile; in the isocentre for both directions and with a defined offset from the isocentre in one of the directions) were compared for all patients. This is schematically drawn in Figure 3-10.

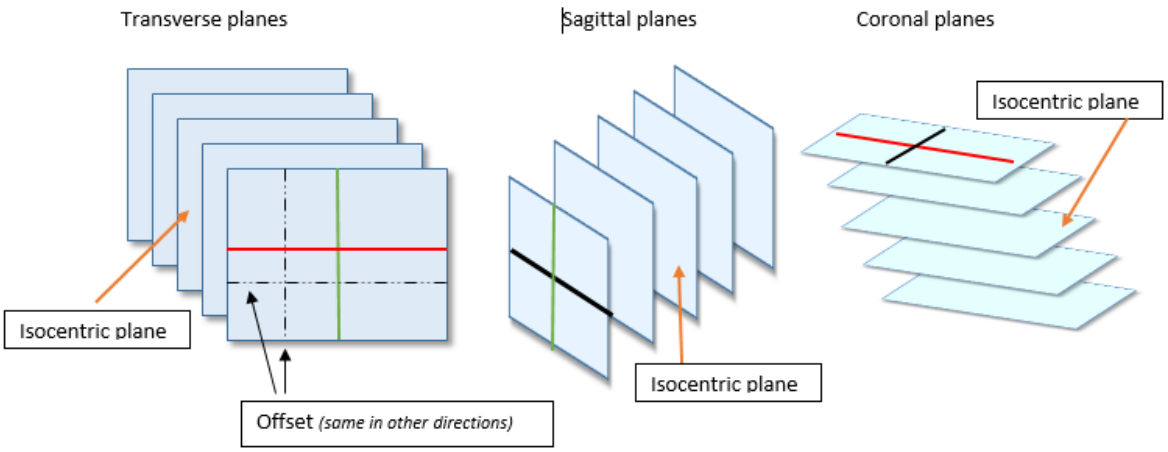


Figure 3-10: Comparison of overlapping profiles in different directions (for illustration). Lines of same colors correspond to each other.

Dose profiles usually agreed in shape, therefore there was almost no spatial shift in films. If so, this could be adjusted by the user in MATLAB, similarly to the feature in OmniPro. However, for some cases, there was a large systematic dose shift. The disagreement in the worst case was 28% deviation in dose – this was observed when different film batch (and hence a different calibration curve) was used for transverse, sagittal and coronal planes (for Patients no. 1 and 6). For the ideal case when all planes were irradiated and evaluated in one session (Patients no. 4 and 5), the agreement was as good as 1-2% dose difference. Different batches were not used on purpose but because the previous batch was expended. This should be avoided, as it magnifies uncertainties, even though a calibration curve was obtained for each film batch in our study. However, the worst result (28%) must have been caused by an unknown error (error in geometry, damaged piece of film…). Other causes of discrepancy could be: fluctuations in linac output (within 0.5% for Siemens Artiste and the given time period and within 1.0% for Varian TrueBeam – however, the Varian data were not used in this chapter) and leaf positioning (within 1 mm for Siemens Artiste and within 0.5 mm for Varian TrueBeam), reproducibility of phantom and film positioning, different storing conditions between

film irradiation and scanning (ambient temperature), little air gap left in the IMRT cube phantom between slabs. (Uncertainties were discussed in chapter 3.1.15.) An example of overlapping profiles is shown in Figure 3-11.

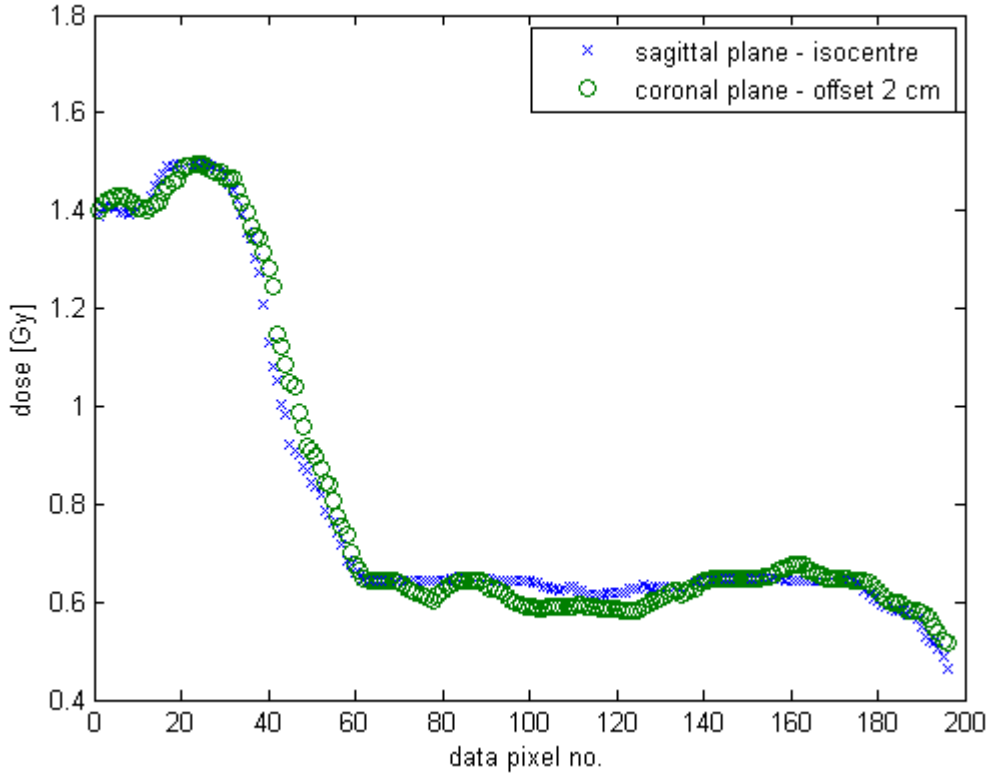


Figure 3-11: Example of overlapping profiles in the sagittal direction (plane in isocentre) and coronal direction (plane 2 cm from isocentre) for Patient no. 1.

The overall average disagreement for all compared profiles was -0.648% and -0.014 Gy (median 0.146% and 0.009 Gy, 25th percentile -6.349% and -0.107 Gy, 75th percentile 2.050% and 0.037 Gy). This means that large disagreement can be attributed to uncertainties in the process which can be avoided. These numbers were also influenced by the noise that was left on film even after smoothing with a filter of 5 x 5 pixels.

Surprisingly, some profiles did not agree even in the dose distributions exported from the TPS XiO (they were exported as individual 2D planes, not as a whole 3D dose distribution). An example is given in Figure 3-12 for Patient no. 1. One of the reasons might be the geometrical uncertainty in reconstructed CT images in sagittal and coronal planes which is quite pronounced in TPS XiO 4.80. However, in order to explain this behaviour, detailed knowledge of the software algorithms would be required.

Consequently, for the purpose of this study, average dose values of both (or all 3) planes were used for gamma calculation at the cross sections. Relative gamma evaluation and local normalization (as defined here) were used to overcome the problem of slightly different doses on films for patients that were not measured in one session.

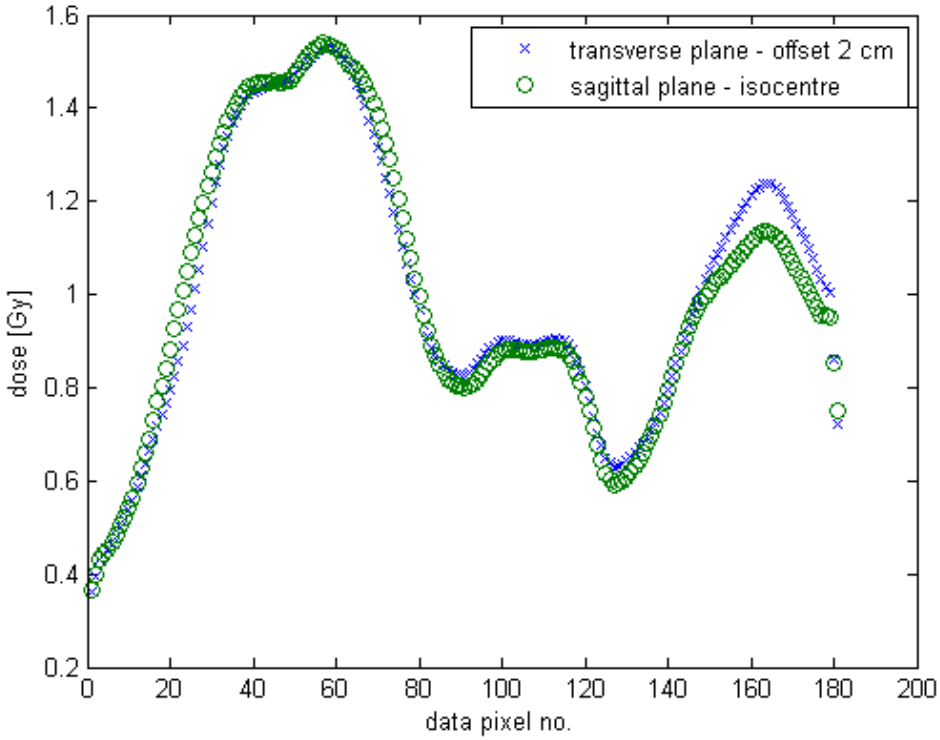


Figure 3-12: Dose disagreement in overlapping profiles from 2 different dose planes exported from the TPS for Patient no. 1 — transverse plane 2 cm from isocentre compared to sagittal plane in the isocentre.

3.3.5 IMRT and VMAT plans

3.3.5.1 Clinical plans

All measurements were carried out on the Siemens Artiste and Varian TrueBeam linear accelerators in the Thomayer Hospital in Prague. Six IMRT clinical step-and-shoot treatment plans, two IMRT sliding window plans and one VMAT plan were used in this part of the study. These included five pelvic plans for prostate and lymph nodes using 18 MV photon beams (Siemens Artiste), one head-and-neck case using 6 MV photons (Siemens Artiste), one glioblastoma case using 6 MV beams (Varian TrueBeam) and two prostate cases using 18 MV beams (prostate only, Varian TrueBeam). Head-and-neck patients were treated rarely with Siemens Artiste at this site, this is why only one case was included. Radical treatment of glioblastoma cases only started when Varian TrueBeam was commissioned. VMAT technique has been commissioned very recently (summer 2022). Treatment planning was performed in XiO version 4.80 in case of the Siemens Artiste patients and Eclipse version 16.1 in case of the TrueBeam patients. The IMRT plans contained 7-9 fields that were distributed in the whole 360°. Two of these plans had not previously met the clinical tolerance criteria for gamma analysis and Siemens Artiste irradiation (gamma score of 90%, i.e. gamma index smaller than 1 for at least 90% of points) and were not used for treatment. The VMAT plan contained 2 full arcs with collimator angles 30° and 330°. Table 3-9 and Table 3-10 summarize the significant information. Some of the patients were already used in the previous parts of our work.

Table 3-9: Clinical IMRT plans used in this study. SIB – simultaneous integrated boost.

Siemens Artiste patients						
Patient no.	**Patient 1**	**Patient 2**	**Patient 3**	**Patient 4**	**Patient 5**	**Patient 6**
Diagnosis	Head-and-neck	Prostate and lymph nodes	Prostate and lymph nodes	Prostate and lymph nodes	Prostate and lymph nodes	Prostate and lymph nodes
Dose prescription	25 x 2 Gy	22 x 2 Gy	22 x 2 Gy	41 x 2 Gy (SIB)	22 x 2 Gy	22 x 2 Gy
Energy	6 MV	18 MV	18 MV	18 MV	18 MV	18 MV
Technique	IMRT step-and-shoot 7 fields	IMRT step-and-shoot 7 fields	IMRT step-and-shoot 7 fields	IMRT step-and-shoot 7 fields	IMRT step-and-shoot 9 fields	IMRT step-and-shoot 9 fields
Passed clinical limits?	Yes	No	No	Yes	Yes	Yes

TrueBeam patients			
Patient no.	**Patient 7**	**Patient 8**	**Patient 9**
Diagnosis	Prostate	Glioblastoma	Prostate
Dose prescription	28 x 2.5 Gy	15 x 2.7 Gy	28 x 2.5 Gy
Energy	18 MV	6 MV	18 MV
Technique	IMRT sliding window 7 fields	IMRT sliding window 5 fileds	VMAT 2 arcs
Passed clinical limits?	Yes	Yes	Yes

Table 3-10: Gamma score values for plans that failed clinical verification.

Patient no.	**Patient 2**	**Patient 3**	
Field no.	Field 6	Field 1	Field 7
Gamma score — clinical verification	87.77%	88.9%	88.2%

3.3.5.2 Error-induced plans

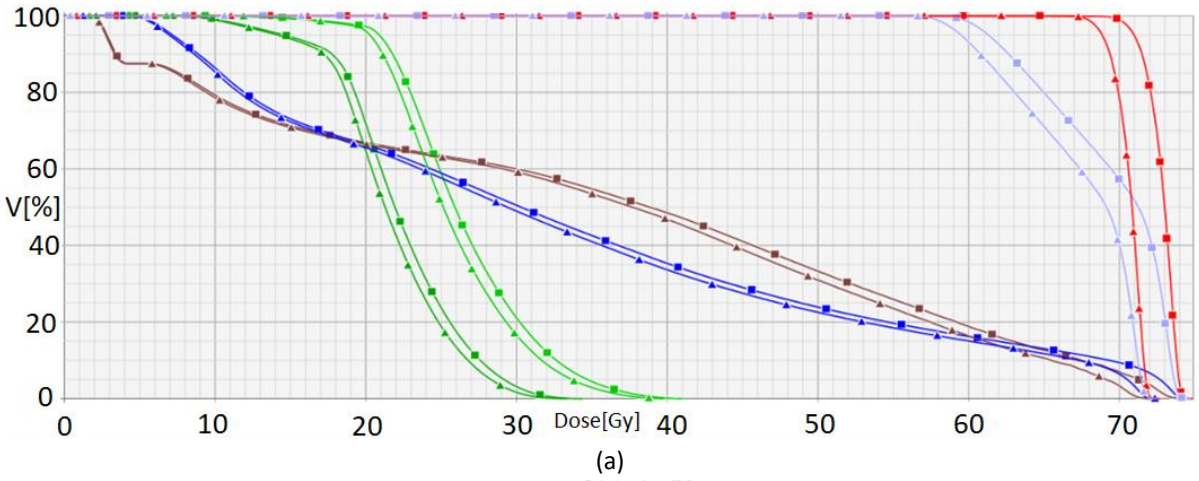
Error-induced plans were also created in order to test the ability of the pseudo-3D method to find these errors. Plan of Patient no. 4 was modified as described in chapter 3.2.1.2, i.e. 1% MU error and 1-3 mm leaf position error were introduced into 3 treatment fields. However, because it is not directly possible to model separately an MU change and MLC leaf position change in XiO, the second error-induced plan was created by modifying MLC leaf positions only. Plan of Patient no. 5 was modified so that a systematic leaf position error would be introduced into treatment fields no. 2, 4 and 5. All leaves in all segments (both leaf banks) were shifted 1 mm to the right. Again, this resulted in MU per segment recalculation for all segments and in beam total MU change, while the weight of the field in Gy in an arbitrary point chosen by the algorithm remained constant. However, because this time it was a systematic shift, the total number of MUs per field for Patient no. 5 remained almost the same, except for field no. 4. Both the error-induced plans with their parameters are summarized in Table 3-11. A 1 mm error in a single leaf position (resulting in a possible 2 mm error in

the field size modelled by MLC) is the tolerance limit for Siemens Artiste quality control tests and this systematic shift of all leaves might already cause some clinically significant discrepancies, as discussed in chapter 4.2.3.2.

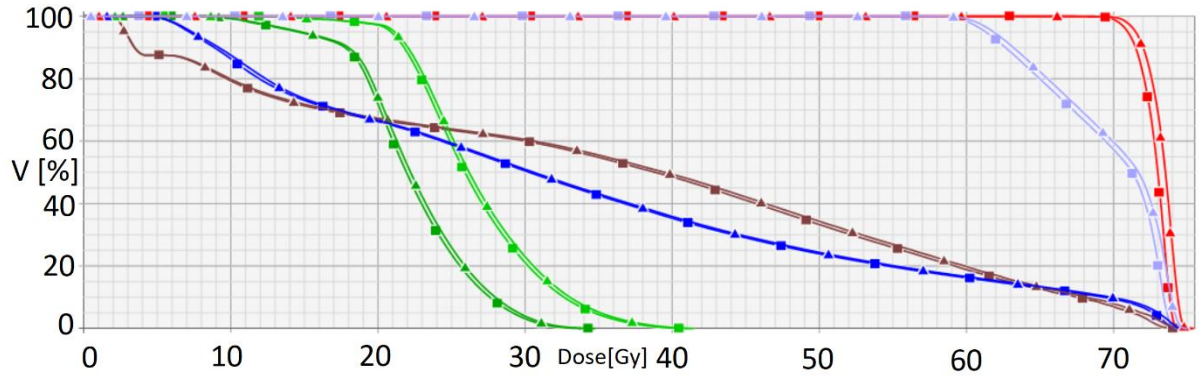
Table 3-11: Error-induced IMRT plans.

	MU error			MLC leaf position error		
Patient no. 4	Field 1	Field 2	Field 4	Field 1	Field 2	Field 4
	-1%	+1%	-1%	1 mm each leaf bank	2 mm each leaf bank	3 mm each leaf bank
				10 cm central area, both leaf banks, fields more narrow		
Original field MUs	85.0	130.5	89.3	85.0	130.5	89.3
Modified field MUs	89.3	122.2	94.4	99.5	117.4	109.1
Patient no. 5	Field 2	Field 4	Field 5	Field 2	Field 4	Field 5
	none	none	none	1 mm each leaf bank	1 mm each leaf bank	1 mm each leaf bank
				all leaves in all segments, both leaf banks – shift to the right (+X)		
Original field MUs	46.6	68.4	38.9	46.6	68.4	38.9
Modified field MUs	-	-	-	46.6	68.4	38.8

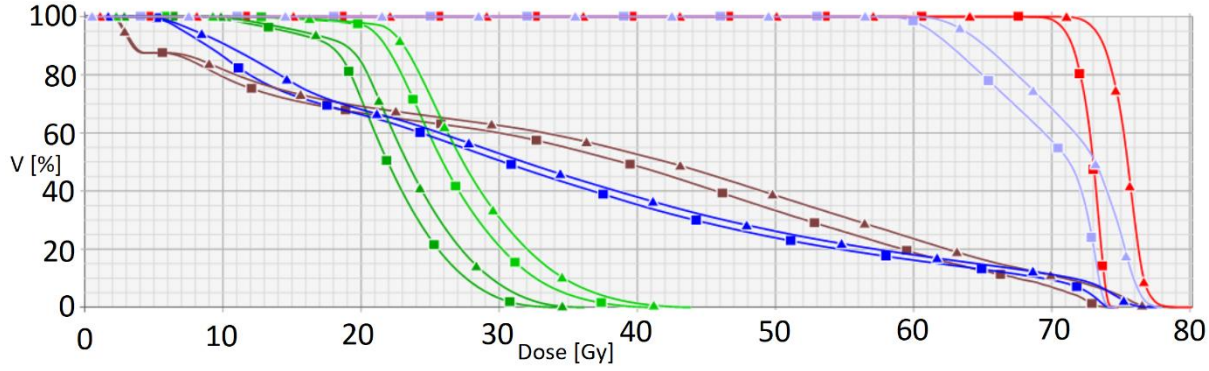
Additionally, the VMAT plan of Patient no. 9 was modified in 3 different ways. The first type of introduced error was a systematic leaf bank shift in each control point. The X1 MLC leaf bank was shifted in one of the two treatment arcs so that the beam opening was 3 mm larger in each control point. In TPS Eclipse, the dose was recalculated but the number of MUs in each arc remained the same as in the original plan. This type of error resulted in increased DVH parameters, such as 6.6% increase in PTV maximum dose, 3.4% increase in PTV mean dose and 6.7% increase in rectum mean dose. The difference in DVH is illustrated in Figure 3-13 (c). The second type of error was the same as the first one but the shift in the MLC leaf bank position was only 1 mm. This type of error should not theoretically be detectable with 3%/3 mm criteria according to [113] and even though some change can be seen in the resulting DVH (Figure 3-13 (b)), it is not clinically so important. The increase in PTV maximum dose was only 0.8%, in PTV mean dose 0.7% and in rectum mean dose 1.5%. (Note: In TPS Eclipse, the MLC shift is not associated with an MU change, unlike in XiO. Moreover, the same errors in VMAT plans are less pronounced than in IMRT plans [114-116]. The third type of error was an MU change of 3% in each arc. The total number of MUs in each arc was manually decreased by 3% and the dose was recalculated. The DVH was affected and approximately 3% systematic drop in dose (resulting in e.g. 3% lower PTV mean dose and PTV maximum dose, as well as 3% lower rectum mean dose) was seen due to this error, so it was considered clinically relevant. Figure 3-13 shows the difference in DVH for each of the modelled errors. The type and magnitude of introduced errors were chosen based on previous published studies (113, 117, 118). A gantry angle error could also be modelled for a VMAT plan but it turned out, according to Mijnheer et al. [117], that even a gantry angle error of 5° does not have a clinical impact. Moreover, it was not detectable by their verification technique (EPID 3D transit dosimetry).



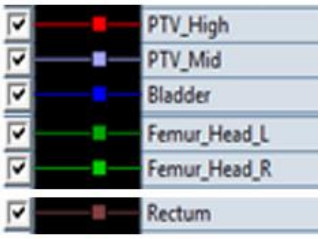
(a)



(b)



(c)



(d)

Figure 3-13: Dose-volume histograms for error-induced VMAT plans for Patient no. 9 compared to the original clinical VMAT plan (a) for the 3% MU error in each arc, (b) for the 1 mm leaf bank sift in one arc, (c) for the 3 mm leaf bank shift in one arc. (d) The legend shows the considered organs at risk and tumor volumes. Squares indicate the original VMAT plan and triangles indicate the error-induced plan.

Yet another type of error could be a change in the collimator angle. According to Mijnheer et al. [117], only a relatively large collimator error was detectable with their EPID 3D transit dosimetry technique (they introduced a change in the collimator angle from 340° to 20°). These types of errors were therefore not considered in this study. Table 3-12Table 3-12 summarizes the introduced errors for the VMAT plan.

Table 3-12: Error-induced VMAT plan

	MU error	**MLC leaf position error**	
Patient no. 9	3% decrease in both arcs	1 mm X1 leaf bank shift (all leaves) In each control point In 1 arc *Segments wider*	3 mm X1 leaf bank shift (all leaves) In each control point In 1 arc *Segments wider*

4 Results and discussion

4.1 Benchmarking film dosimetry against current practice

The following results 4.1.1, 4.1.2 and 4.1.3 refer to Part C of Table 3-1. The purpose of this part of the study is to compare film dosimetry and clinically used alternative for field-by-field and pseudo-3D IMRT and VMAT verification.

4.1.1 Clinical plans

4.1.1.1 Field-by-field method

Table 4-1 shows gamma scores obtained for the tested detectors and phantoms. It only shows results for the combination of EBT3 film and OmniPro I'mRT software because this would probably be the clinically used option. Gamma scores for EBT3 film calculated in MATLAB were very similar. Differences between our MATLAB 2D code and OmniPro I'mRT were quantified and discussed in chapter 3.3.3 (because there were most data available for evaluation).

Table 4-1: 2D field-by-field gamma score values (comparison to TPS) for EBT3 film and PTW Seven29 array obtained in OmniPro I'mRT and VeriSoft v. 3.1 for Patient no. 5.

Gamma score [%] (OmniPro I'mRT – EBT3 film; VeriSoft v. 3.1 – PTW Seven29)										
Detector and phantom	**Depth in phantom**	**Field no.**								
		1	2	3	4	5	6	7	8	9
EBT3 film in RW3 slabs	7 cm	99.3	93.5	99.4	99.3	99.1	100.0	99.6	100.0	99.8
	8 cm	98.8	96.7	98.0	93.6	98.6	99.9	98.5	99.0	99.7
	isocentre	99.7	96.0	94.8	97.4	96.3	96.0	95.7	99.7	97.7
	12 cm	99.3	99.4	92.7	100.0	95.0	96.4	91.7	89.5	99.3
	13 cm	97.8	92.3	99.6	97.2	97.4	96.9	93.4	99.7	98.2
PTW seven 29 in RW3 slabs	7 cm	94.5	96.6	96.6	97.4	-	-	-	-	-
	8 cm	95.3	96.4	96.7	96.7	-	-	-	-	-
	isocentre	92.2	91.7	95.2	92.3	92.3	91.6	92.7	95.7	93.0
	12 cm	95.7	95.1	97.4	95.5	-	-	-	-	-
	13 cm	92.4	92.3	95.8	94.5	-	-	-	-	-

The subsequent figures show the OmniPro I'mRT values.Figure 4-1 shows gamma pass rates for EBT3 film (in combination with OmniPro I'mRT) and PTW Seven29 array (with VeriSoft v. 3.1) for all 9 fields of the IMRT plan of Patient no. 5 measured at isocentre. EBT3 film seems to give higher gamma scores (the highest percentage of points with $\gamma < 1$), even though this is not so evident for some fields. The dependence of gamma score values on the method with which they were obtained can be tested with the statistical chi square test of independence that is suitable for this type of data (two categorical variables) and was already used and explained in chapter 3.1.14. We will only consider those fields and depths where data are available for both methods (EBT3 film and PTW Seven29 array and corresponding software). The first categorical variable will be the method used (EBT3 film or PTW Seven29 array). The second categorical variable will be gamma score bins. We will consider the following categories, in order to have at least 5 expected values in each category, as required by the test: Gamma score < 93%, gamma score between 93% and 95%, gamma score between 95% and 97% and gamma score > 97%. The contingency table is displayed in Table 4-2. Our null hypothesis is that gamma scores and the method with which they were obtained are independent at a 5%

significance level**. With the data in Table 4-2, the p-value is 0.0031 and we reject the null hypothesis. Therefore, we might think that the EBT3 method gives higher gamma scores than the PTW Seven29 array method.**

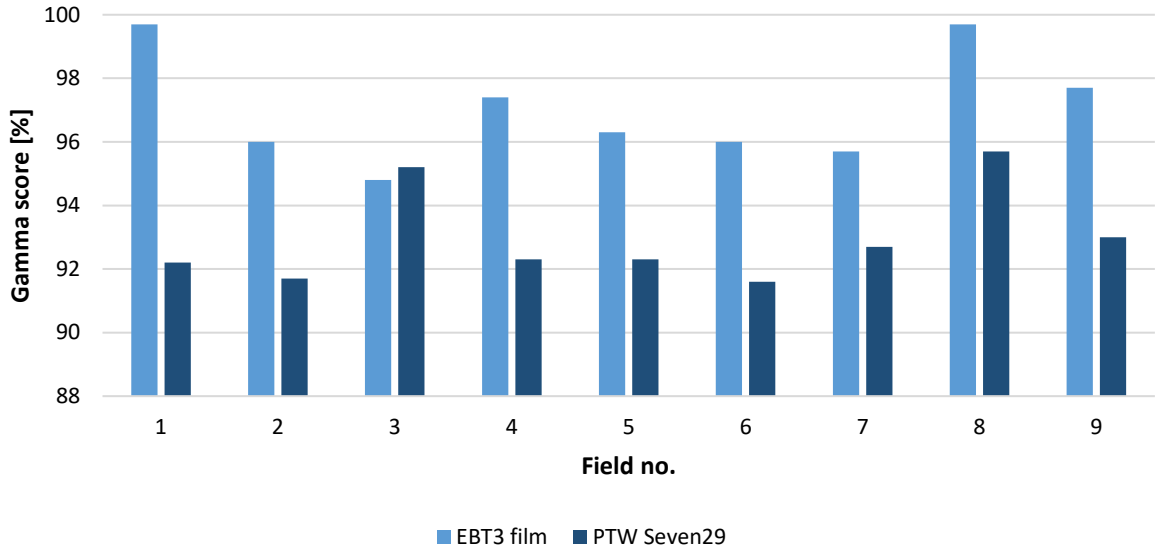


Figure 4-1: Gamma scores for EBT3 film and PTW Seven29 shown for 9 individual fields of the IMRT plan measured in an RW3 slab phantom at isocentre.

For comparison with literature, there is a paper by Amin et al. [116], who used a combination of the PTW Seven29 array and the OCTAVIUS 4D phantom for verification of IMRT and VMAT cases planned in the Eclipse TPS for the Unique accelerator. They report 3D global gamma scores for various treatment sites with 3%/3 mm criteria and 10% threshold calculated in VeriSoft. Their passing rates are roughly similar to ours. Even though their values represent 3D gamma scores and in Table 4.1 we show field-by-field values, we give the information here only to indicate that the lower gamma score values might be attributed to the characteristics of the PTW Seven29 array itself. When the PTW Seven29 array was replaced with the PTW OCTAVIUS 1500 detector (better spatial resolution) at the Thomayer Hospital, it turned out during commissioning that gamma scores rised approximately 3% for the very same measured patient cases. Looking at Figure 4.1, if values were increased this way, they would be closer to the EBT3 film resluts for most of the fields. This could not be experimentally proven for our data because shortly after the new PTW array was purchased, the Siemens Artiste linear accelerator was decommissioned.

Table 4-2: Contingency table with frequencies for the chi square test of independence comparing the EBT3 field-by-field method and the PTW Seven29 field-by-field method.

Gamma score category (GS)	EBT3 method	PTW Seven29	Row totals
GS < 93%	4	9	13
93% <GS < 95%	4	9	13
95% < GS < 97%	7	7	14
GS > 97%	10	0	10
Column totals	25	25	**Grand total: 50**

Furthermore, Table 4.1 shows the range of gamma scores for each type of detector obtained for different depths. The PTW Seven29 detector was placed to 5 different depths in the RW3 slab phantom (7, 8, 10, 12 and 13 cm) in 4 fields and only to the isocentre (10 cm depth) for the other 5 fields. EBT3 film was placed to 5 different depths for all 9 fields. The minimal gamma score for EBT3 film is lower than for the PTW Seven29 in the RW3 slab phantom, even though EBT3 film generally seems to give higher passing rates. It seems that the PTW Seven29 array gives more consistent values at different depths than film. Figure 4.2 shows an example of gamma scores obtained at different depths for one of the fields where the behaviour was the most pronounced. The question arises which of the methods is closer to reality and whether the EBT3 method is trustworthy.

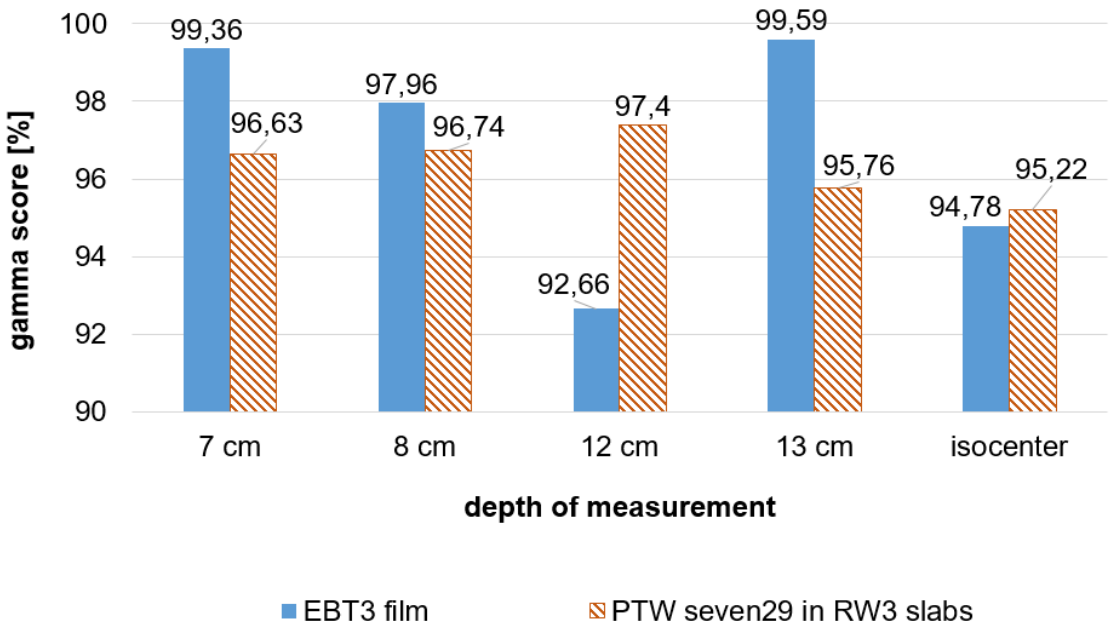


Figure 4-2: Example of measurements at different depths – field no. 3. Gamma scores at different depths for field no. 3 measured with EBT3 film and the PTW Seven29 detector placed in the RW3 phantom.

Regarding reproducibility of the PTW detector measurement in clinical conditions, the same IMRT step-and-shoot field measured several times in sequence with the PTW Seven29 detector can lead to variable gamma scores, the variability being as high as 2% in gamma score. Here, the variability among different depths is larger, the maximum difference being 5.1% in gamma score for Field no. 4. (This was calculated as the maximum minus minimum gamma score for Field no. 4 and all depths.) Therefore, variability in different depths is most likely caused by other reasons than reproducibility of measurement or linac delivery (the PTW detector was placed at one depth at a time while with film dosimetry all depths could be measured at once).

The maximum difference in gamma scores for the EBT3 film method is 10.5% for Field no. 8. However, there is no data for the PTW array for this field. In Field no. 1 to Field no. 4, where data are available for both methods in different depths, the largest difference in gamma scores for the EBT3 method is 7.1% for Field no. 2. For most fields, the maximum difference in gamma scores at different depths is larger for film than for the PTW array. One of the reasons for larger differences for film might be the fact that film is supposed to be more sensitive to leaf positioning errors than the PTW

array. The manufacturer of Siemens Artiste guarantees 1 mm reproducibility for each leaf (and during periodic QA checks this limit was always met). Nevertheless, this means a possible error of up to 2 mm in the position of each two opposing leafs in the step-and-shoot plan. The film dosimetry method is supposed to be more sensitive to this type of error thanks to its good spatial resolution. The 1 cm spacing between chambers of the PTW detector and their square shape is less sensitive to these fluctuations in leaf positions for highly modulated fields. The film data for all depths were acquired in one single irradiation, however, the magnitude of leaf positioning error can still propagate differently to gamma scores computed at different depths. In chapter 3.1.15 it was shown that a difference of up to 4.5% in gamma score is caused by the uncertainty of measurement with film. For the PTW array, this is around 2% (based on repeated measurements and gamma calculation – however, this comprises the uncertainty of the linac delivery, which is actually being checked with the PTW array). In the view of these numbers, larger differences in gamma scores at different depths for film are expected.

It must be pointed out that the softwares used for gamma analysis are not directly comparable, which might be another reason for the different numbers. Higher threshold than 5% of maximum dose is not possible for gamma analysis in the VeriSoft version used. But lower threshold than 10% for films would not be practical due to poor film performance for low doses. Also, the implementation of algorithms in both softwares might play a role. For example, the search distance is not adjustable in VeriSoft v. 3.1 and is not known to the user. Also, interpolation between measurement and prediction dose points is a part of the algorithm and cannot be influenced by the user. Different normalization of the matrices was applied in OmniPro I'mRT and VeriSoft v. 3.1. Point dose normalization in film dosimetry is very sensitive to the point selection due to film noise (even though the evaluated matrices were smoothed prior to evaluation). Also the evaluated regions of interest differed. The aim of the study was to compare the different verification methods as a whole, because it would not be practical to perform gamma analysis for film and ion chamber matrix data in exactly the same way. The goal of this part of the study (benchmarking) was to show if film dosimetry can perform equally well or even better than electronic devices for IMRT field-by-field verification and if it can be used for our further experiments. A different type of study would have to be carried out to quantitatively evaluate particular differences among the detectors or softwares, similar to those published in [105-108, 119].

Based on the results of the chi-square test of independence we assume that the film method can be used for field-by-field experiments further in our study but it must be kept in mind that there is larger variability in gamma scores due to uncertainties in film dosimetry than for the electronic devices.

4.1.1.2 Pseudo-3D method

Results from the OCTAVIUS 4D measurements (with OCTAVIUS 1500 detector) of the VMAT plan and 3D gamma calculation in VeriSoft v. 8.0 are shown in Table 4-3. The overall 3D gamma score calculated by VeriSoft v. 8.0 is shown as well as 3D gamma scores for various dose levels available in the software. Results of our pseudo-3D film dosimetry method (evaluated with our 2D and 3D MATLAB codes) are shown in Table 4-4. 2D plane-by-plane gama scores calculated either with our 2D MATLAB code or with our 3D MATLAB code are given. 2D plane-by-plane gamma scores in the same transverse, sagittal and coronal planes calculated with VeriSoft v. 8.0 and based on the OCTAVIUS 4D measurements are shown for comparison. The overall 3D gamma score calculated with our 3D

MATLAB code is also given in Table 4-4, for global as well as local normalization. Global normalization means that all dose points in the 3D space are normalized to the 75th percentile of the whole 3D matrix. Local normalization in this study means that each dose plane is normalized to its 75th percentile.

Table 4-3: Results from the OCTAVIUS pseudo-3D method for a clinical VMAT plan (Patient no. 9). The overall 3D gamma score is shown as well as gamma scores at various isodose levels available in VeriSoft v. 8.0 (comparison to TPS).

VeriSoft v. 8.0 (OCTAVIUS 4D)									
3D gamma score [%]	99.5								
Dose level [% of normalization value]	10	30	50	70	80	85	90	95	100
Gamma score [%]	99.4	99.5	99.8	99.6	99.5	99.4	99.3	99.1	100.0

Table 4-4: 2D plane-by-plane gamma scores from the pseudo-3D film method calculated with our 2D MATLAB code and 3D MATLAB code and the overall pseudo-3D gamma score obtained from MATLAB with local and global normalization. 2D plane-by-plane gamma scores calculated in VeriSoft v. 8.0 from the OCTAVIUS 4D measurements are shown in the same planes for comparison. All comparisons are done against TPS.

2D gamma score	**MATLAB 2D code Gamma score [%]**					**MATLAB 3D code, local normalization Gamma score [%]**				
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	99.0	99.4	98.0	99.2	27.8	-	99.4	99.2	99.7	-
sagittal	99.2	99.4	96.7	99.0	95.7	-	99.5	98.2	99.1	-
coronal	98.4	100.0	98.5	98.0	94.7	-	100.0	98.8	98.0	-
VeriSoft v. 8.0 (from OCTAVIUS 4D) Gamma score [%]						**Pseudo-3D gamma score (MATLAB 3D code) [%]**				
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	Local normalization:				
transverse	100.0	100.0	99.0	100.0	100.0	99.1				
sagittal	99.9	99.9	99.9	99.9	100.0	Global normalization:				
coronal	100.0	100.0	99.8	100.0	100.0	99.1				

In case of this VMAT prostate plan, all gamma scores (2D and pseudo-3D) are close to 100% for both investigated methods. The only exception is the transverse plane 4 cm away from the isocentre towards the patient's head where an extremely low number is encountered with the film method. This is because this plane lied in a very low dose region outside the PTV where the dose was below 20% of the prescribed dose, i.e. below 0.5 Gy. While the ionization chamber array method gives a 100% agreement for this plane, the film method is not suitable for such low doses, as already discussed e.g. in chapter 3.1.15. Also in the sagittal and coronal planes +4 cm away from the isocentre (beyond the edge of the PTV) somewhat lower gamma scores were obtained with the film method. Low dose regions might thus be considered a limitation of the film method. Even though there is not enough data and therefore we cannot decide based on a statistical test the agreement of the two methods for a simple small prostate VMAT plan seems very good, given the uncertainties of the methods. (Uncertainties of the film method were discussed in chapter 3.1.15. For the OCTAVIUS 4D measurement and reconstruction of 3D dose inside the phantom and calculation of 3D gamma analysis, the uncertainty is not known. However, it is certainly larger than 2%).

4.1.2 Error-induced plan

EBT3 film dosimetry (field-by-field verification in an RW3 slab phantom) was compared to the clinical verification method (field-by-field verification with PTW array Seven29 in an RW3 slab phantom) for a clinical plan of Patient no. 4 where MLC positioning errors and dose errors were introduced as described in chapter 3.2.1.2. The compared measurement setups were described in chapter 3.2. The goal of this part of the study was to tell whether our film dosimetry method would reveal the introduced errors similarly to the clinically used method.

Table 4-5: Gamma scores obtained with EBT3 film and OmniPro I'mRT. Difference is calculated as gamma score from original TPS x detector measured dose comparison minus gamma score from original TPS x modified TPS dose comparison.

Field no.	Gamma score [%]						
	Original clinical plan	1% MU change			MLC position change		
	Original TPS dose x Film dose	Original TPS dose x Film dose	Original TPS dose x Modified TPS dose	Difference	Original TPS dose x Film dose	Original TPS dose x Modified TPS dose	Difference
1	99.7	82.8	100.0	-17.2	73.2 (1 mm error)	81.8	-8.5
2	98.0	60.1	56.4	3.7	49.9 (2 mm error)	51.7	-1.8
4	98.7	81.3	94.9	-13.6	79.4 (3 mm error)	73.3	6.1

Measured dose distributions for 3 error-induced fields were compared to original (error-free) TPS dose distributions. Gamma scores for the two methods are given in Table 4-5 and Table 4-6. Gamma scores for each field are also given for original TPS dose distribution compared to modified TPS dose distribution because these gamma scores might be expected with an ideal delivery of the error-induced dose distribution (if no linac or measurement discrepancies were present). Moreover, each detector was evaluated in its corresponding software, i.e. OmniPro I'mRT for films and VeriSoft v. 3.1 for the PTW array, and the evaluated ROIs differed for film and PTW array. The aim of this study was to compare the verification methods as a whole in the way they would be used clinically and to point out what the gamma score values, widely used as the decision pass/fail criterion, tells the user. The aim was not to evaluate the detector and software performance separately. For the same reason, results obtained with our 2D MATLAB code are not shown (they are similar to OmniPro I'mRT gamma scores and OmniPro I'mRT is more likely to be used clinically with film). Thus, gamma scores in Table 4-5 and Table 4-6 are not directly comparable. Besides the differences between the two softwares and ROIs that were already discussed, it needs to be pointed out that VeriSoft v. 3.1 gives gamma index values in chamber positions only (and the resolution is thus very poor compared to film).

The values that could be directly compared for the different detectors and softwares were gamma score differences between 1) TPS dose distribution comparison (original TPS dose x modified TPS dose) and 2) measured dose distribution comparison (original TPS dose x detector dose), calculated as 2) minus 1). These differences are shown in Figure 4-3 (MU change) and Figure 4-4 (MLC position change) as well as in Table 4-5 and Table 4-6.

From Table 4-5 and Table 4-6 and Figure 4-3 and Figure 4-4 it is evident that for 1% MU change, the radiochromic film method (including OmniPro I'mRT software) shows relatively large differences in gamma scores between original TPS x modified TPS dose and original TPS x film dose, while the PTW array method (with VeriSoft v. 3.1) gives smaller differences for all fields. For MLC position change, film and PTW array method give similar results in terms of gamma score differences (last column in Table 4-5 and Table 4-6) for fields no. 1 and 2. However, for field no. 4 results are very different. The film method even gives a positive difference, meaning that gamma score from original TPS dose x modified TPS dose comparison was lower than from original TPS dose x measured dose comparison. In other words, it appears that certain errors visible in TPS error-induced dose distribution were not revealed by measurement. This is because for OmniPro, the combination of DD (dose difference) and DTA might be different (and therefore the minimum gamma index lower for TPS x measured dose comparison), because for TPS x TPS dose comparison, matrices of 1 mm resolution were compared while for TPS x measured dose comparison, matrices of 0.35 mm were compared and more dose points were available for the search algorithm. Gamma score, being a percentage of points with gamma index lower than 1, could also be different depending on the number of passing points relative to the total number of points, which was different in each case.

Table 4-6: Gamma scores obtained with PTW Seven29 and VeriSoft v. 3.1. Difference is calculated as gamma score from original TPS x detector measured dose comparison minus gamma score from original TPS x modified TPS dose comparison.

Field no.	Gamma score [%]						
	Original clinical plan	1% MU change			MLC position change		
	Original TPS dose x Seven29 dose	Original TPS dose x Seven29 dose	Original TPS dose x Modified TPS dose	Difference	Original TPS dose x Seven29 dose	Original TPS dose x Modified TPS dose	Difference
1	93.5	87.4	92.6	-5.2	44.2 (1 mm error)	49.5	-5.3
2	94.6	82.4	81.0	1.3	55.2 (2 mm error)	56.4	-1.2
4	96.7	88.4	92.8	-4.4	34.2 (3 mm error)	46.4	-12.2

In VeriSoft v. 3.1, both comparisons (TPS x TPS dose and TPS x measured dose) use the same number of points because the software first recalculates dose into the PTW array chamber positions only (thus always using 729 points). Even here, for 1% MU change and field no. 2, the gamma score difference in Table 4-6 is a positive value. This is a feature of gamma analysis, the combination of DD and DTA might be different and therefore the minimum gamma index for a given point lower/higher when slightly different dose distributions (in this case TPS error-induced dose and measured error-induced dose) are compared. This is valid for both softwares, of course.

The PTW array method gave remarkably lower gamma scores for MLC leaf position change for fields no. 1 and 4 than the film method (both TPS x TPS dose comparison and TPS x measured dose comparison) indicating that the detector array resolution of 1 cm (with 0.5 cm interpolation) is less suitable for detecting this type of error.

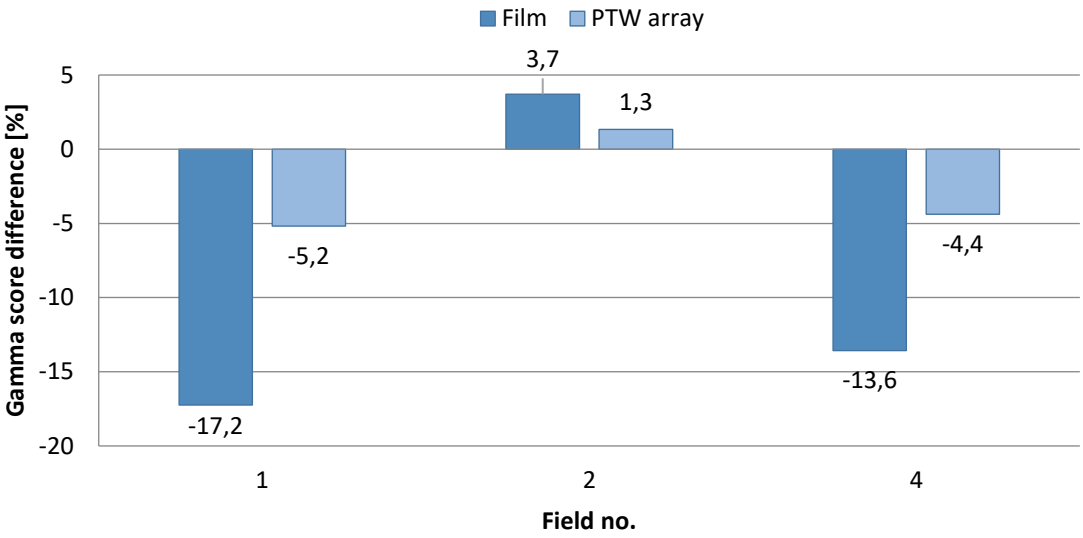


Figure 4-3: Gamma score differences for 1% MU change in 3 error-induced fields calculated as gamma score from original TPS x detector measured dose comparison minus gamma score from original TPS x modified TPS dose comparison.

It needs to be pointed out again that both the 1% MU change in some of the field segments and the MLC leaf position change in some or all of the field segments result in dose distribution recalculation in Xio. The MUs per segment change for all segments as well as the total field MUs, even when nothing but a leaf position is changed manually. The value that remains constant is the total field weight in Gy in an arbitrary point chosen by the optimization algorithm in XiO. Therefore, MLC leaf position change results also in a dose change in the TPS dose distribution, affecting all segments in the field, not only the modified ones.

Both methods would probably reveal all types of introduced errors because for all 3 fields they showed lower gamma scores than for the original plan verification. However, for Field no. 2 the drop in gamma scores due to the introduced erros was not above the uncertainty level for any of the methods. Both methods behave differently for the above mentioned reasons. There is not enough data in this part of the study (i.e. comparison of two methods for field-by-field verification for the error-induced plan) to perform statistical tests.

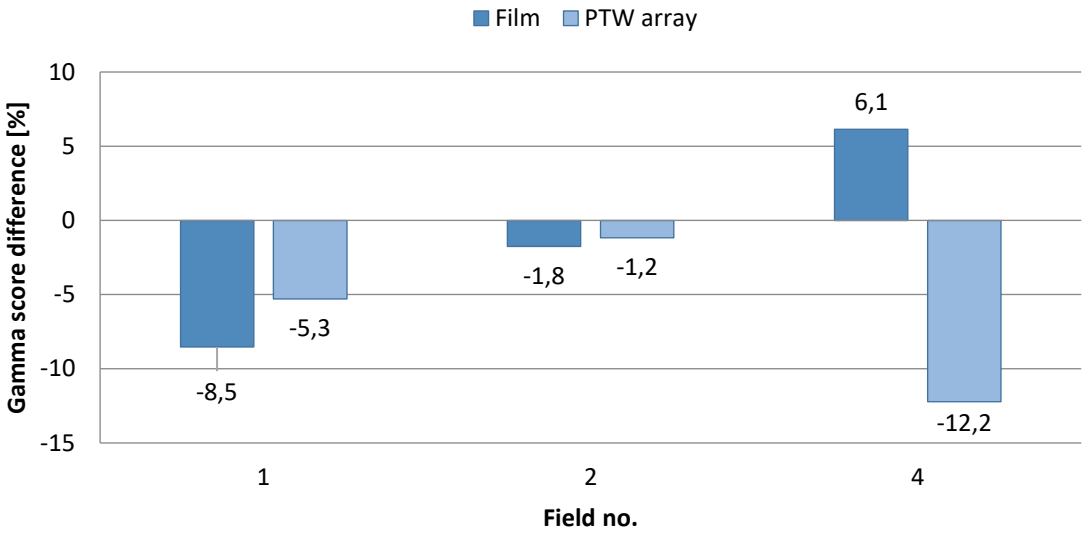


Figure 4-4: Gamma score differences for MLC positioning errors in 3 error-induced fields calculated as gamma score from original TPS x detector measured dose comparison minus gamma score from original TPS x modified TPS dose comparison.

4.1.3 Summary

For the clinical IMRT plan and field-by-field verification, higher gamma scores were obtained with EBT3 film than with the PTW Seven29 placed into an RW3 slab phantom. For the VMAT clinical plan and pseudo-3D verification, the pseudo-3D film method and the OCTAVIUS 4D method gave very similar results. For the error-induced plan, both the EBT3 film method and the clinically used method (PTW Seven29 placed into an RW3 slab phantom) revealed the introduced errors for field-by-field verification. However, for one of the fields and both methods, the drop in gamma score value was not above the uncertainty level. But this behaviour was the same for both methods. Results could be influenced by the use of different software and approaches to gamma analysis. Our results indicate that EBT3 film can be an excellent tool when proper film handling and an adequate methodology are applied. This finding is contrary to some previous studies [106, 108]. Film and portal dosimetry is recommended by [121, 122] for commissioning of new techniques thanks to their better spatial resolution. Only a few studies can actually be found testing the PTW Seven29 detector against EBT3 film. More works focus on the PTW Seven29 detector comparison with older types of radiographic and radiochromic film (EBT, EBT2, EDR2) [119, 123-127]. The reason might be that EBT3 film was commercially released later than the PTW Seven29 detector but the PTW Seven29 detector is still commercially available (now under the name OCTAVIUS Detector 729) and widely used. Therefore, the contribution of this work might be significant in the field. Because results of clinical plan verification depend on the type of detection system used, good understading of the detector's and software's performance is needed to safely and reliably check individual clinical plans.

4.2 Pseudo-3D verification with film

The results in the following chapters refer to parts D, E, F and G of Table 3-1. Various comparisons of pseudo-3D and field-by-field verification are done. If not specified otherwise, the 3%/3 mm criteria for gamma analysis are reported and other parameters of gamma analysis are according to chapters 3.2.6, 3.2.7, 3.3.3 and 3.3.4. The question is whether the pseudo-3D method can give the user more information (or more relevant information) than the 2D method. It is obvious that with the pseudo-

3D method, more data points (more measured dose points and more gamma indices) are available. However, in clinical practice, a simple criterion is needed that would tell the user (very often not a medical physicist who performs the verification) whether the IMRT plan passes or not. In the case of patient-specific plan verification, it is important to detect relevant errors for every single plan and at the same time have as little false positive results (= plan failing tolerance even though delivered dose distribution in patient is acceptable) as possible.

The answer to this question is sought with the following results. Specifically, the following questions will be addressed:

1) *Is it enough to measure dose distribution in one plane in a field-by-field manner?*
- This is current practice in many clinical centres.
2) *Is it enough to measure dose distribution in one plane when a whole treatment plan is delivered?*
- If so, this would save time and material (in case of film dosimetry).
- The assumption could be that errors coming from slightly different leaf positions, gantry sag or dose error specific for step-and-shoot techniques (beam switching on and off after a small number of monitor units) might already be revealed in a single dose plane.
3) *Or is it more appropriate to check several planes in a 3D space with 2D gamma analysis?*
4) *Or is it more appropriate to calculate a pseudo-3D map of gamma indices and a pseudo-3D gamma score?*

4.2.1 Pseudo-3D verification results

This part refers to Part D of Table 3-1. The purpose here is to compare 2D and pseudo-3D evaluation of the same measured data set (EBT3 film measurements in the I'mRT Phantom and MATLAB evaluation). The MATLAB computed gamma scores are given in Table 4-7. It must be pointed out that the 2D gamma scores in Table 4-7 computed with the 2D MATLAB code were obtained with the same size of matrices as in the case of the MATLAB 3D code. These were square matrices to form a 3D cube. Otherwise, the numbers would be affected by the shape and size of the matrices as well as the choice of ROI and normalization, so they would not be directly comparable. The numbers are therefore different from Table 3-8 in chapter 3.3.3 for MATLAB.

One of the sagittal planes for Patient no. 3 and 3D code and one of the sagittal planes for Patient no. 4 and 2D code in Table 4-7 have extremely low gamma scores. This is due to normalization for gamma analysis. If a different normalization was applied, the gamma score could rise close to 90%. But the same normalization was kept for the same data throughout the study for comparison purposes. In most cases, however, unexpectedly low gamma scores are encountered in planes that are 4 cm distant from the isocentre (e.g. transverse planes of Patient no. 7 and Patient no. 9). This is because these planes in the 3D space lay beyond the edge of the PTV in a region of very low doses (the mean dose being below 0.5 Gy). It has been shown in this work that measuring such low doses can be problematic with film dosimetry and in chapter 4.1 it has been shown that it is a drawback of the film dosimetry method compared to other pseudo-3D methods that are based on ionization chamber arrays.

The subsequent figures illustrate the conclusions drawn from this part of the study.

The overall pseudo-3D gamma score for the 9 investigated patients, computed as the percentage of points with gamma index lower than 1 in the whole 3D space (including only points where data were measured for the percentage calculation), is given in Table 4-8 and Table 4-9. Global normalization does not seem like the method of choice with pseudo-3D film dosimetry proposed in our work. This is mainly due to uncertainties associated with 3 separate irradiations of film sheets in the I'mRT Phantom to cover the whole 3D space. Local normalization in this study means that each dose plane is normalized to its 75th percentile. Global normalization means that all dose points in the 3D space are normalized to the 75th percentile of the whole 3D matrix.

Figure 4-5 shows a gamma map image of one plane generated with the 3D code, global normalization compared to 3D code, local normalization for a patient measured in one session and for a patient measured in several sessions. Also, local and global normalization gives the user different information. Using global normalization, the user can see areas of agreement and failure in the whole 3D space and compare different parts of the 3D space. Using local normalization, different results might be obtained in high dose and low dose regions (in planning target volume — PTV and organs at risk — OAR) because the dose difference criterion in absolute values becomes different.

The resolution chosen for film scanning in our work was 0.35 mm which is far better than the resolution of electronic detectors. However, this choice brought along several drawbacks. First of all, the gamma analysis computation time increases with increasing number of points in the matrix. Our 3D code ran on a computer with only a dual-core Intel® Core™ i3-3120M processor, 2.50 GHz and 4 GB RAM and the calculation took several hours without parallelization (which would not increase the speed in this case, quite the contrary, as there was a big parallel overhead using only two workers). Making profit of the MATLAB Parallel Computing Toolbox and using a multi-core processor or a GPU would increase the speed so that results would be available in acceptable time. However, if the resolution of the matrices was coarser, it would increase the speed dramatically. Secondly, with such a high resolution, the maximum variable size allowed in MATLAB 2017a was reached on our system.

Table 4-7: 2D gamma scores computed for all planes in the 3D space for all 9 investigated patients (clinical plans). The left part of the table is computed with our 2D MATLAB code, the right part is computed with our 3D MATLAB code. Data for some planes are not available on the right because they fall (spatially) outside the 3D cube.

2D gamma score	**MATLAB 2D code**					**MATLAB 3D code, local normalization**				
Patient no. 1										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	100.0	94.2	89.7	97.2	99.8	-	96.6	93.8	97.7	-
sagittal	99.8	99.3	92.8	97.0	100.0	-	99.3	96.4	98.0	-
coronal	91.7	99.3	98.3	99.9	97.6	-	99.6	99.2	99.9	-
Patient no. 2										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	97.9	90.0	89.9	84.7	87.2	-	90.9	90.4	86.5	-
sagittal	95.8	95.5	64.7	99.6	92.6	-	95.9	67.1	99.5	-
coronal	64.7	93.8	82.9	92.6	74.7	-	93.4	83.4	93.4	-

Table 4-7: 2D gamma scores computed for all planes in the 3D space for all 9 investigated patients (clinical plans). The left part of the table is computed with our 2D MATLAB code, the right part is computed with our 3D MATLAB code. Data for some planes are not available on the right because they fall (spatially) outside the 3D cube - continued.

2D gamma score	**MATLAB 2D code**					**MATLAB 3D code, local normalization**				
Patient no. 3										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	96.8	96.6	97.1	96.3	96.7	-	96.1	97.1	96.2	-
sagittal	92.5	81.6	94.9	88.0	98.3	-	28.6	95.9	90.4	-
coronal	98.1	100.0	90.4	97.6	84.5	-	100	91.8	97.9	-
Patient no. 4										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	95.5	100.0	89.9	78.5	85.6	-	99.9	90.4	79.2	-
sagittal	17.8	97.1	94.0	97.3	93.2	-	97.0	94.4	97.4	-
coronal	92.9	94.3	97.7	98.4	83.3	-	95.2	97.6	98.4	-
Patient no. 5										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	99.3	100.0	99.8	99.0	97.5	-	99.7	99.8	98.8	-
sagittal	78.2	88.4	100.0	89.3	65.2	-	88.6	100.0	90.3	-
coronal	98.5	71.7	99.9	98.3	98.8	-	72.3	99.9	98.3	-
Patient no. 6										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	99.9	99.8	83.1	100.0	84.6	-	100.0	83.5	100.0	-
sagittal	92.1	99.1	97.9	94.4	98.7	-	99.2	98.0	94.8	-
coronal	95.5	99.4	99.7	100.0	99.9	-	99.4	99.7	100.0	-
Patient no. 7										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	51.6	93.7	99.9	96.1	28.4	-	94.6	99.9	96.5	-
sagittal	89.7	98.6	97.2	78.4	83.0	-	98.6	98.1	83.0	-
coronal	85.6	76.0	97.5	89.5	47.5	-	82.0	98.3	90.5	-
Patient no. 8										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	85.0	98.9	85.9	98.9	88.0	-	99.3	90.5	98.9	-
sagittal	84.4	98.8	98.8	96.8	96.4	-	99.1	98.9	97.5	-
coronal	93.3	99.9	99.6	96.2	99.3	-	99.9	99.6	97.6	-
Patient no. 9										
Planes:	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm	-4 cm	- 2 cm	iso	+ 2 cm	+ 4 cm
transverse	99.0	99.4	98.0	99.2	27.8	-	99.4	99.2	99.7	-
sagittal	99.2	99.4	96.7	99.0	95.7	-	99.5	98.2	99.1	-
coronal	98.4	100.0	98.5	98.0	94.7	-	100.0	98.8	98.0	-

*Table 4-8: The overall pseudo-3D gamma score for 9 patients with **local normalization** as defined in this study.*

Patient no.	1	2	3	4	5	6	7	8	9
pseudo-3D gamma score (local normalization) **[%]**	97.8	83.8	77.6	94.4	94.2	97.1	93.5	97.9	99.1

*Table 4-9: The overall pseudo-3D gamma score for 9 patients with **global normalization** as defined in this study.*

Patient no.	1	2	3	4	5	6	7	8	9
pseudo-3D gamma score (global normalization) **[%]**	66.3	72.8	42.2	92.1	65.0	86.1	93.5	98.1	99.1

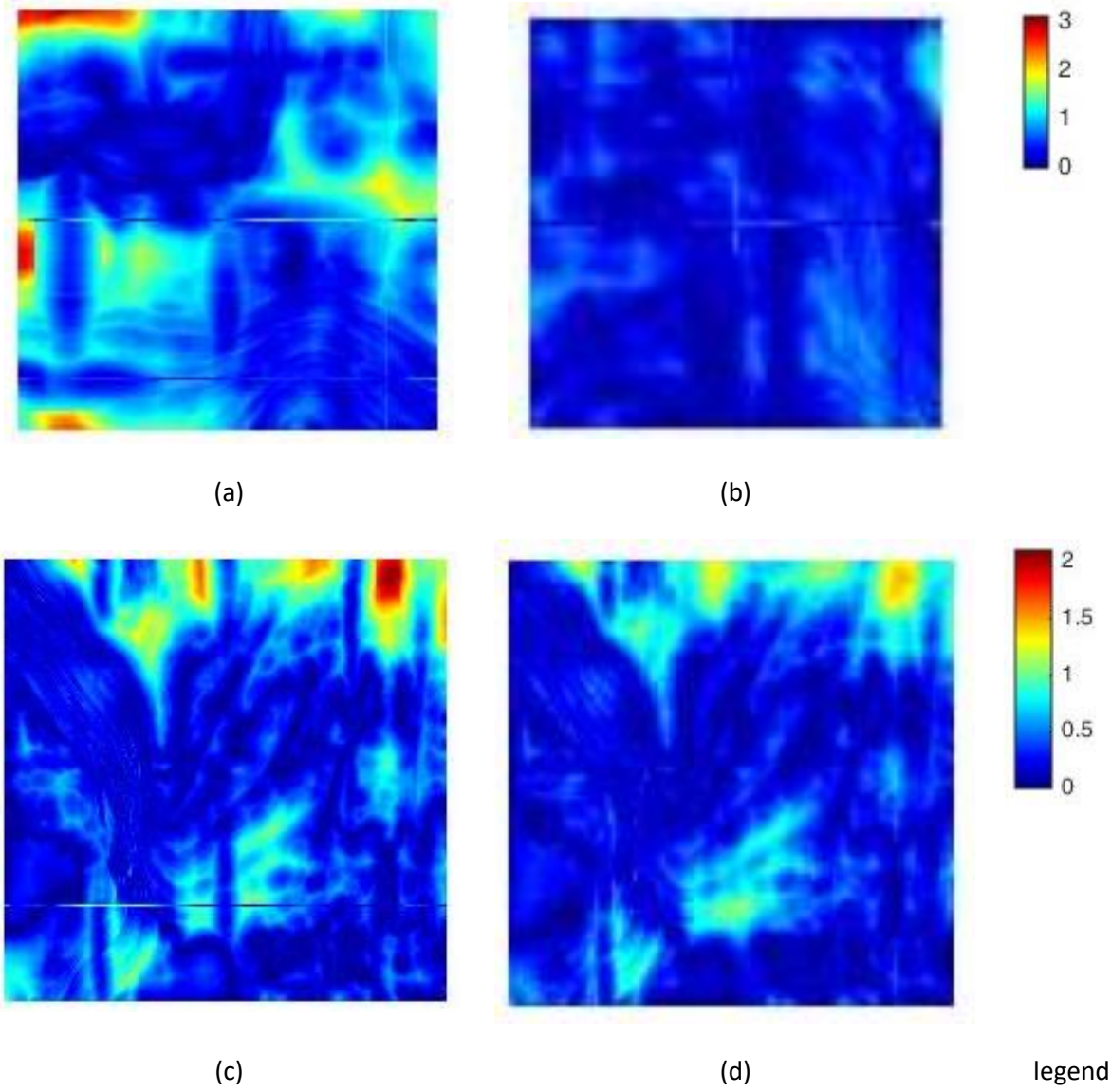


Figure 4-5: Map of gamma indices calculated with our 3D code for (a) isocentric sagittal plane for Patient no. 1 – measured in two sessions with different film batch, global normalization (b) the same plane using local normalization; and (c) a sagittal plane 2 cm from the isocentre for Patient no. 4 – measured in one session with the same film batch, global normalization (d) the same plane using local normalization.

Figure 4-6 shows examples of gamma index maps in individual planes as calculated by our 2D gamma code, 3D gamma code with local normalization and 3D gamma code with global normalization. Note that the planes correspond to sagittal and transverse planes actually measured with film, no interpolation was performed, so where measured data were missing in the 3D matrix, the gamma index was equal to zero. The difference between 2D gamma and 3D gamma with local normalization

can only be seen in areas of cross section with another plane and within the defined search distance (4.5 mm) around them. Gamma index in these areas should normally always be lower in the 3D approach than in the 2D approach because there are more available dose points likely to produce a lower gamma index than the existing values within the 2D plane. However, in some cases the gamma index was lower in the 2D approach. This is because of using average doses at the cross sections in the 3D approach, averaging the transverse, sagittal and coronal planes, both in the measured and TPS exported reference dose planes (where differences could be seen as well). Thus, for a given dose point, the algorithm in 3D may compute different dose difference values and thus also the combination of dose difference and DTA (the minimal value) may become higher than in the 2D case. It should be noted that this could also affect the dose points outside the intersecting columns or rows but where the intersection lay within the search distance and the ideal corresponding dose point was found at the intersection. The 3D gamma calculation with global normalization can produce a similar or a quite different gamma map than with local normalization, as shown in Figure 4-6. This is due to the chosen dose difference limit (in our case 3%) which normalizes the dose difference values in equation 3.5 and depends on normalization of the matrices. Therefore, further comparisons between 2D and pseudo-3D results in this text are performed with 3D, local normalization only, if not specified otherwise.

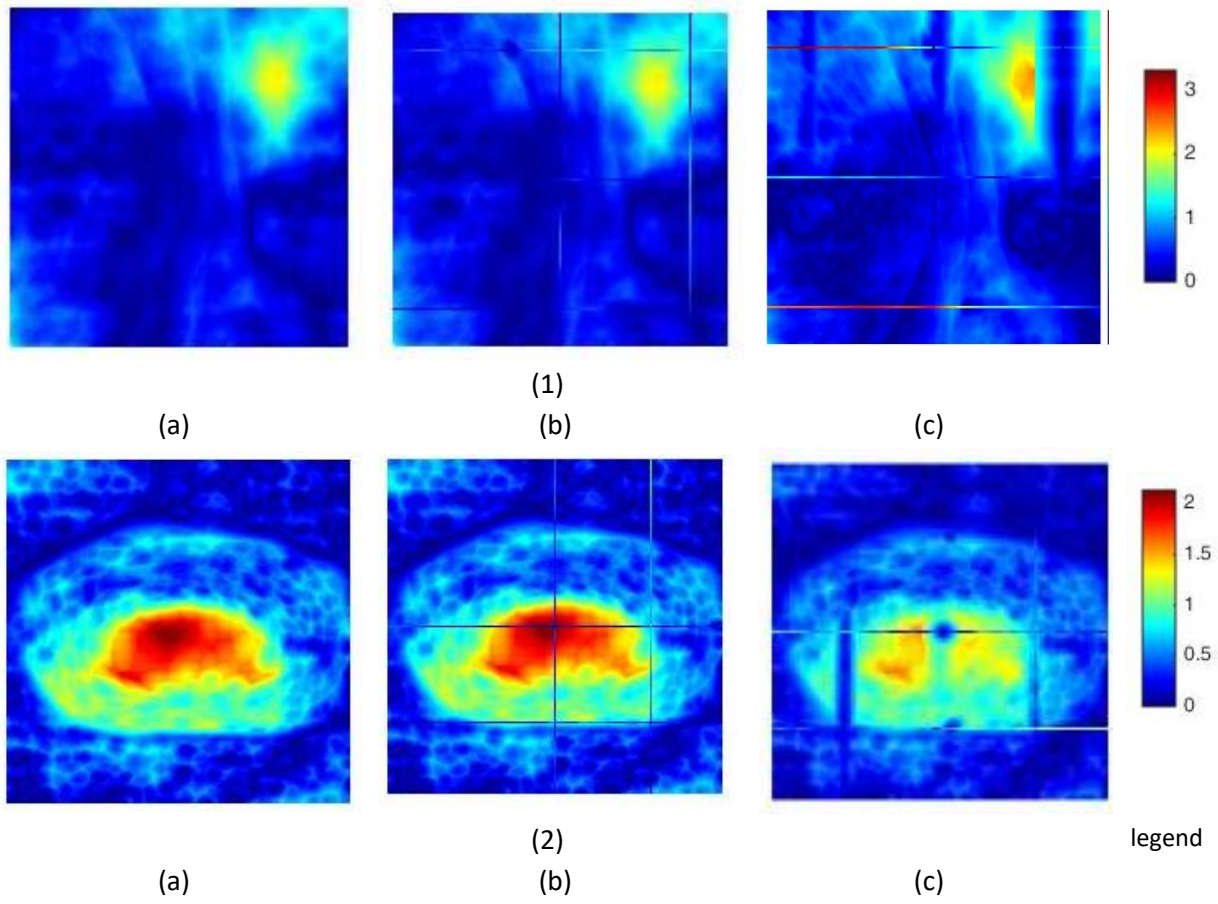


Figure 4-6: Example of gamma distributions for (1) Patient no. 5 and a sagittal plane 2 cm from isocentre (a) as calculated by our 2D gamma code, (b) as calculated by our 3D gamma code with local normalization, (c) as calculated by our 3D gamma code with global normalization; (2) Patient no. 6 and a transverse plane 2 cm from isocentre (a) as calculated by our 2D gamma code, (b) as calculated by our 3D gamma code with local normalization, (c) as calculated by our 3D gamma code with global normalization. Stripes clearly visible in b) and c) are located at places of cross section with another plane.

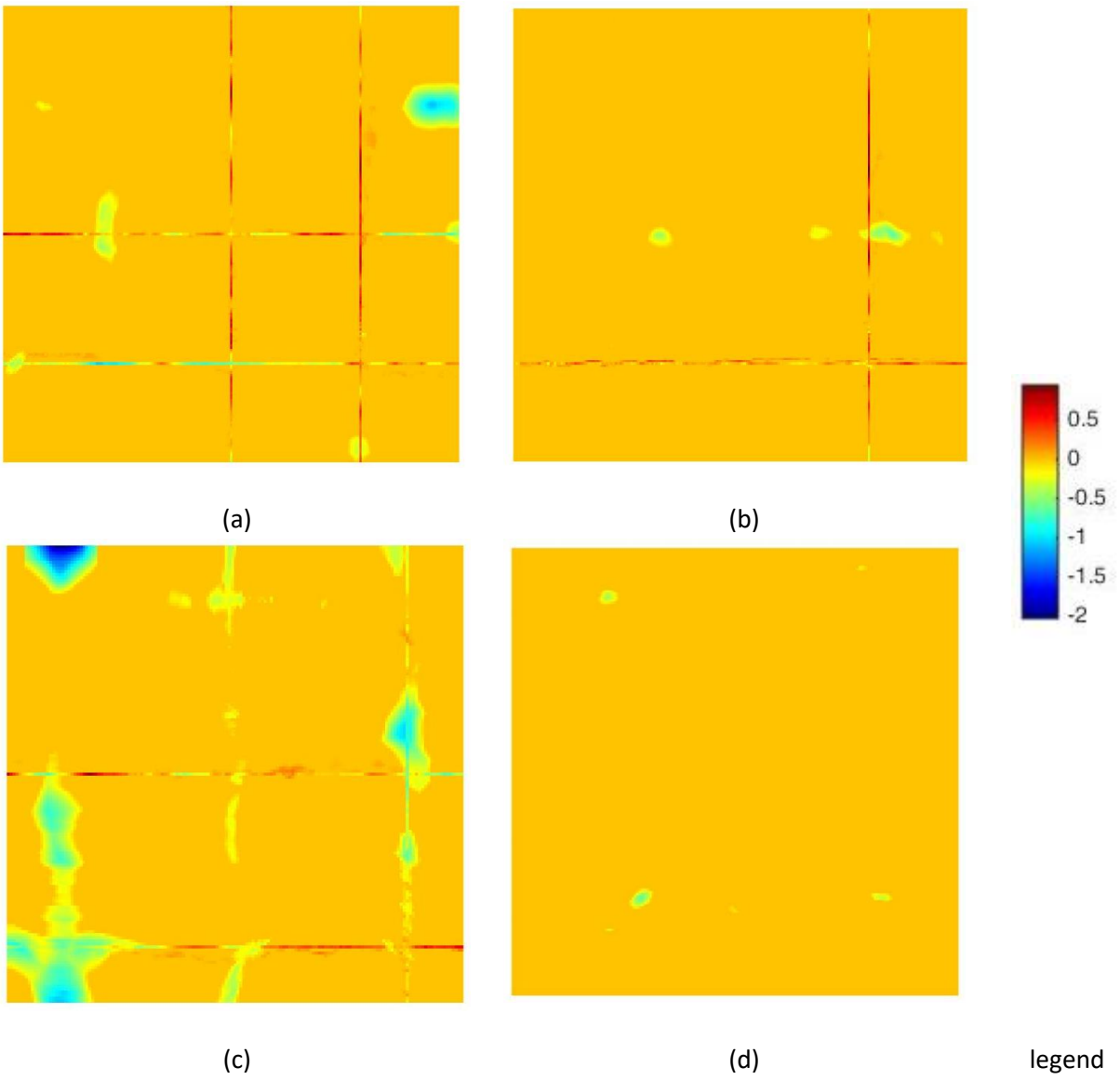
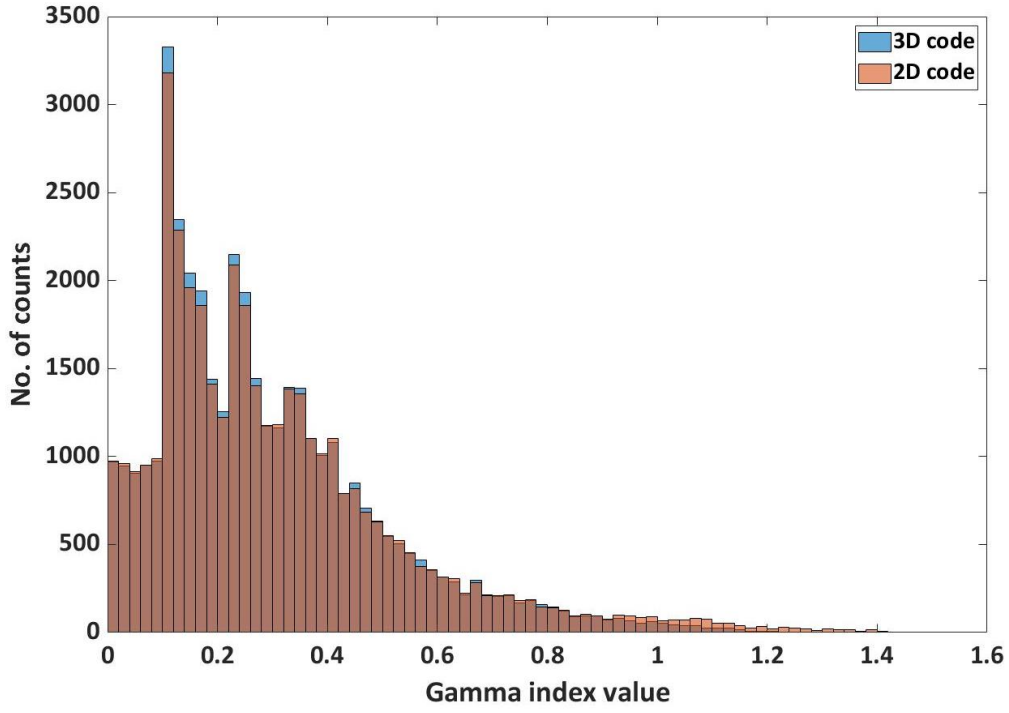


*Figure 4-7: Maps of gamma index differences between the 3D and 2D gamma code applied to the same data. (a) Patient no. 4, coronal plane 2 cm from isocentre, 3D-2D gamma score difference **0.9%**, (b) Patient no. 4, coronal plane at isocentre, 3D-2D gamma score difference **0.1%**, (c) Patient no. 1, sagittal plane at isocentre, 3D-2D gamma score difference **3.7%**, (d) Patient no. 4, coronal plane 2 cm from isocentre, 3D-2D gamma score difference **0%.***

Figure 4-7 shows examples of differences in gamma index maps for selected planes computed for the same data with the 3D code and with the 2D code. This figure illustrates that even if differences in gamma index are relatively small, they can have a significant impact on the resulting gamma score (percentage of points with gamma index lower than or equal to 1) or vice versa. In Figure 4-7 (b) and Figure 4-7 (d) , there is only a very small area of gamma index differences and the difference in gamma score is also small, almost equal to 0% in both cases. Figure 4-7 (a) illustrates the most typical map of differences, where the difference in gamma score is close to 1% between the 3D and the 2D code computed values. Figure 4-7 (c) shows a case with a large area of gamma index difference between the 3D and the 2D code and the gamma score difference is 3.7%. One thing to note is that in any of the cases, except for the 0% difference, gamma score could fall below the tolerance limit in

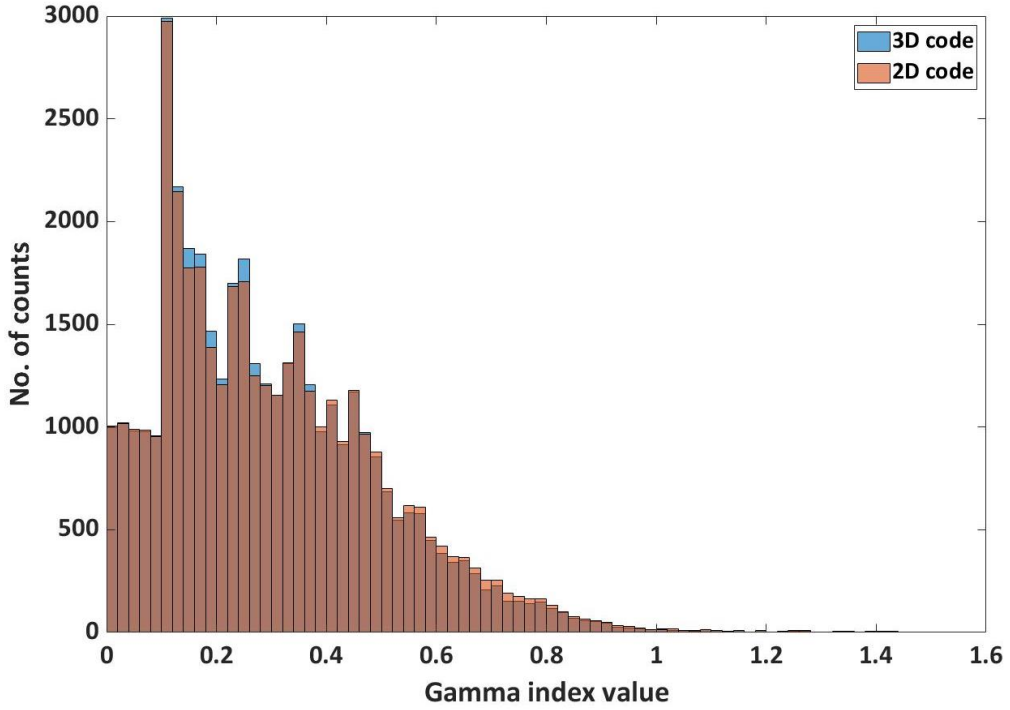
the 2D approach, while in 3D it could pass these tolerance criteria, depending on the chosen value of the tolerance limit. Another thing to note is that in Figure 4-7 (c), there are relatively big areas of gamma index difference greater than 1 (while a gamma index of 1 is the threshold between pass and fail for that point). Still, the overall difference in gamma score is only 3.7%. It should be understood that despite of the magnitude of gamma index differences between the pseudo-3D and 2D approach, the values can still fall below 1 (or above 1) for both calculations and the gamma score can remain unchanged. Or the other way round, even for very small differences between gamma indices in corresponding points, the gamma scores can be different because the values exceed the value of 1 in one of the cases. The gamma score only on the number of points that pass or fail, not on the magnitude of differences.

Gamma score values tell the user what fraction of points in the dose distribution pass the gamma test. They do not tell how much the failing points differ from expected values and whether the dose in these points is higher or lower than expected. Therefore, other quantitative measures can be helpful to evaluate when performing gamma analysis, such as gamma index statistics and gamma histograms. These were also evaluated for all planes and all patients, both with the 2D gamma code and with the 3D gamma code. Figure 4-8 shows the difference in gamma histograms between the 2D code and the 3D code, as an example, for Patient no. 9 and isocentric transverse plane and Patient no. 8 and isocentric coronal plane. Gamma histograms for all clinical plans looked similar, regardless of the technique, linear accelerator or treatment site.

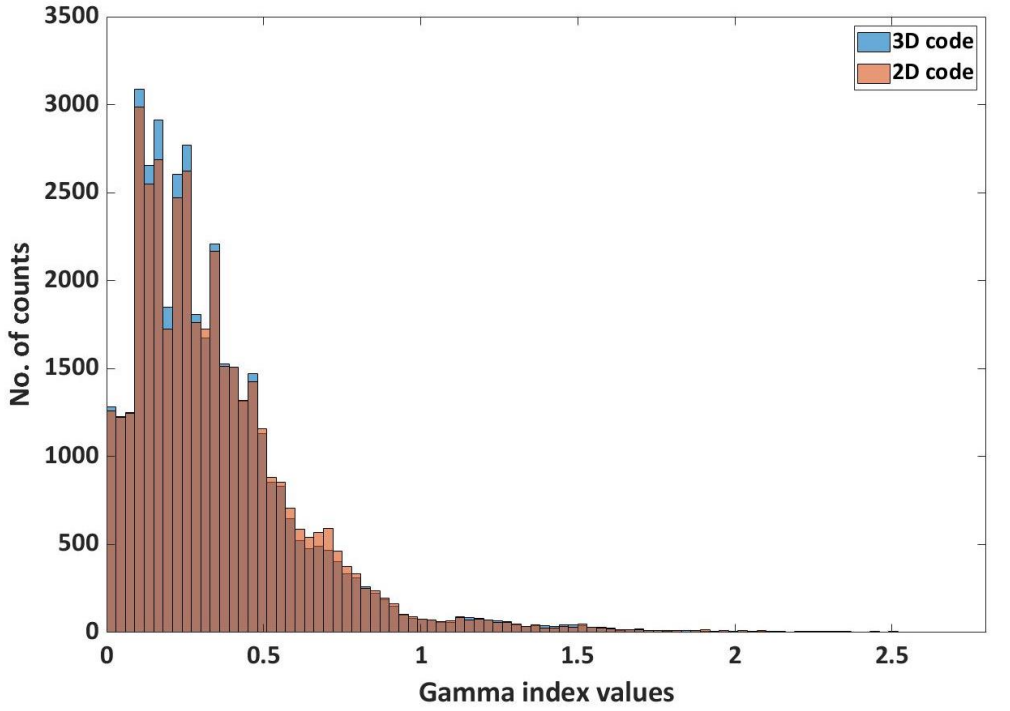


(a)

Figure 4-8: Gamma histograms showing the difference between the 2D code-based and 3D code-based computation of gamma indices. (a) Patient no. 9 and isocentric transverse plane, (b) Patient no. 8 and isocentric coronal plane, (c) Patient no. 9 with introduced error – 3 mm leaf bank shift – and isocentric sagittal plane. Blue data represent 3D code-calculated values and orange data represent 2D code-calculated values – to be continued on next page.



(b)



(c)

Figure 4-8: Gamma histograms showing the difference between the 2D code-based and 3D code-based computation of gamma indices. (a) Patient no. 9 and isocentric tranverse plane, (b) Patient no. 8 and isocentric coronal plane, (c) Patient no. 9 with introduced error – 3 mm leaf bank shift – and isocentric sagittal plane. Blue data represent 3D code-calculated values and orange data represent 2D code-calculated values - continued.

Patient no. 9 with introduced error (3 mm leaf bank shift in one arc of the VMAT plan) and sagittal isocentric plane is also shown in Figure 4-8 to illustrate the different distribution of gamma indices for an error induced plan (results for this error-induced VMAT plan are discussed further in chapter 4.2.4). Figure 4-9 shows the differences in gamma index maps calculated with the 3D and with the 2D code (computed as 3D minus 2D) for one patient and 9 planes. It is only shown for one patient for illustration but the behaviour was the same for all patients. Each point in the figure shows an average difference of gamma indices in the given plane (no. 1-3 transverse planes, no. 4-6 sagittal planes and no. 7-9 coronal planes around the isocentre). Maximum and minimum values in each plane are indicated. However, most values are close to zero (the first and the third percentile is close to zero). Again, some values could fall below zero, meaning that gamma index computed with the 2D code was lower than with the 3D code, due to averaging of doses in intersecting columns or rows.

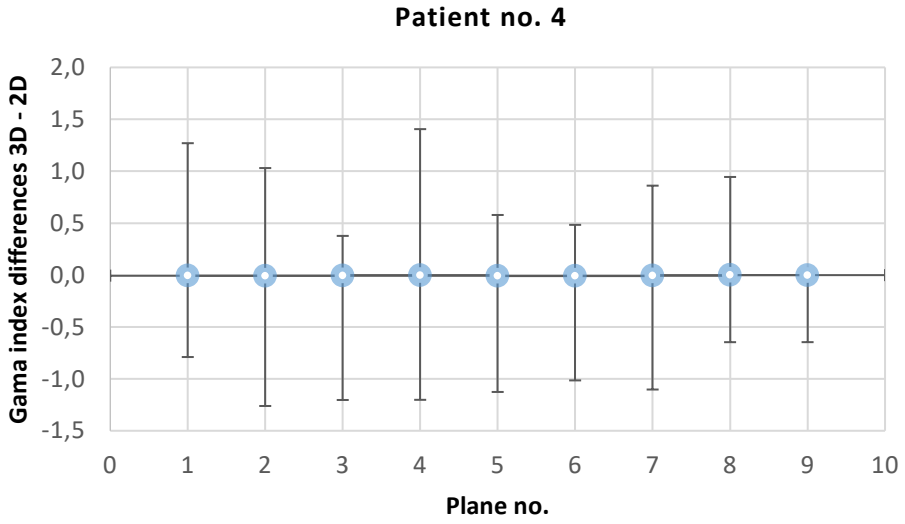


Figure 4-9: Differences in gamma indices in 2D planes computed with the 3D and 2D code for Patient no. 4 (calculated as 3D-based minus 2D-based map of gamma indices for each plane). Shown for 3 transverse (no. 1-3), 3 sagittal (no. 4-6) and 3 coronal (no. 7-9) planes around the isocentre. The figure shows average difference for each plane (close to zero) and maximum and minimum values.

As already mentioned, in some cases, the 3D gamma code could produce higher values of gamma indices than the 2D code due to averaging doses at cross sections in the 3D dose matrix. Therefore, for a few cases the gamma score for a given plane as computed with the 2D code was slightly better than the gamma score based on the 3D code, local normalization. However, generally, 3D gamma analysis on the same data set should always give higher passing rates than 2D gamma analysis, simply due to the fact that the algorithm has higher chance of finding a better combination of dose difference and spatial distance in 3D. Even though our 3D matrix was rather sparse (adjacent planes 2 cm apart and search distance 4.5 mm) and only less than half of the reference points could search their corresponding doses in adjacent planes, in some cases there was a relatively big difference between 2D code-based and 3D code-based gamma scores, already. This would be even more pronounced if a better 3D resolution was chosen, e.g. 1 cm between adjacent planes. Results of gamma score differences are shown in Figure 4-10 for all planes, computed as 3D code-based minus 2D code-based gamma score for a given plane. (For clarity, 3D code-based gamma score means that

gamma indices were computed with the 3D code and the 2D gamma score for a given plane was calculated as percentage of points with these gamma indices lower than 1.)

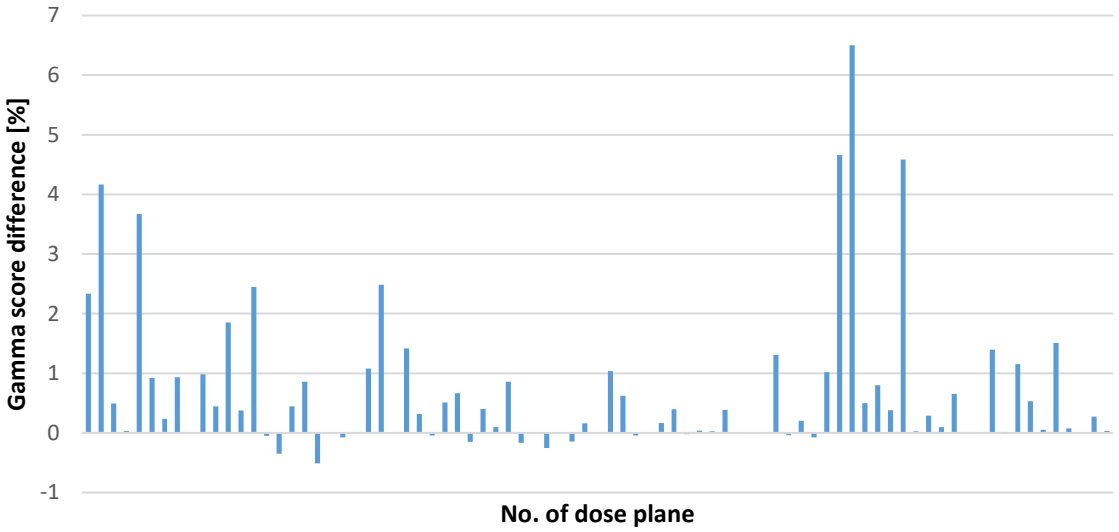


Figure 4-10: Differences in gamma scores for all investigated planes as obtained with the 2D gamma code and with the 3D gamma code, local normalization (gamma score computed for the given plane, difference computed as 3D code-based gamma score minus 2D code-based gamma score in that plane).

A statistical test was also performed to see whether the difference between 3D-code based and 2D-code based gamma scores in 2D planes is statistically significant. The chi square test of independence was used because it is suitable for this type of data, as already discussed in chapter 3.1.14. This test is used to detemine whether two categorical variables are related or not. Our first categorical variable was a set of four gamma score bins: gamma score ≤ 90%, gamma score between 90% and 94%, gamma score between 94% and 97% and gamma score > 97%. In other words, we had four categories of gamma scores. The second categorical variable was the method used to determine the gamma scores – either the 3D MATLAB code or the 2D MATLAB code. Only those planes that have a gamma score computed with both the 2D code and the 3D code were included in the test, i.e. 81 planes in total. A contingency table with frequencies for each combination of categories was created and is shown in Table 4-10. The chi square test of independence was performed in MATLAB by writing our own script (there is no built-in function within the Statistical Toolbox). Our null hypothesis was that gamma scores and the method used to determine them (3D code or 2D code) are independent at a 5% significance level. The so-called expected values were computed. The assumption of the test is that each expected value is larger than 5, which was fulfilled with our data.

Table 4-10: Contingency table with frequencies for the chi square test of independence comparing the 2D MATLAB code and the 3D MATLAB code.

Gamma score category (GS)	2D MATLAB code	3D MATLAB code	Row totals
GS < 90%	18	10	28
90% <GS ≤ 94%	6	10	16
94% < GS ≤ 97%	12	12	24
GS > 97%	45	49	94
Column totals	81	81	**Grand total:** 162

The p-value for the chi square test of independence is 0.3265 and therefore we cannot reject the null hypothesis that the gamma scores and the code that is used to determine them are independent at a 5% significance level. Even though from Figure 4-10 and Table 4-7 it might seem that the differences between the 2D code- and the 3D code-computed gamma scores are big enough, they are not statistically significant. This might be due to the sparse matrix used (films 2 cm apart and search distance 4.5 mm, so that differences in gamma index could be seen at plane intersections and their neighbourhood within 4.5 mm). Clinically, we are interested in particular results for particular patients, so Figure 4-10 probably shows more valuable results than the statistical tests.

To see whether the differences between the 3D code-computed and the 2D-code computed gamma scores in 2D planes are higher for some patients or some of the directions (transverse, sagittal and coronal), gamma score differences are also plotted in Figure 4-11. There, planes no. 1-3 correspond to transverse direction, planes no. 4-6 correspond to sagittal direction and planes no. 7-9 correspond to coronal direction. Note that even though 5 planes were measured for each patient and each direction, due to the rectangular shape of the films, only 3 planes around the isocentre intersect the other directions in the 3D space.

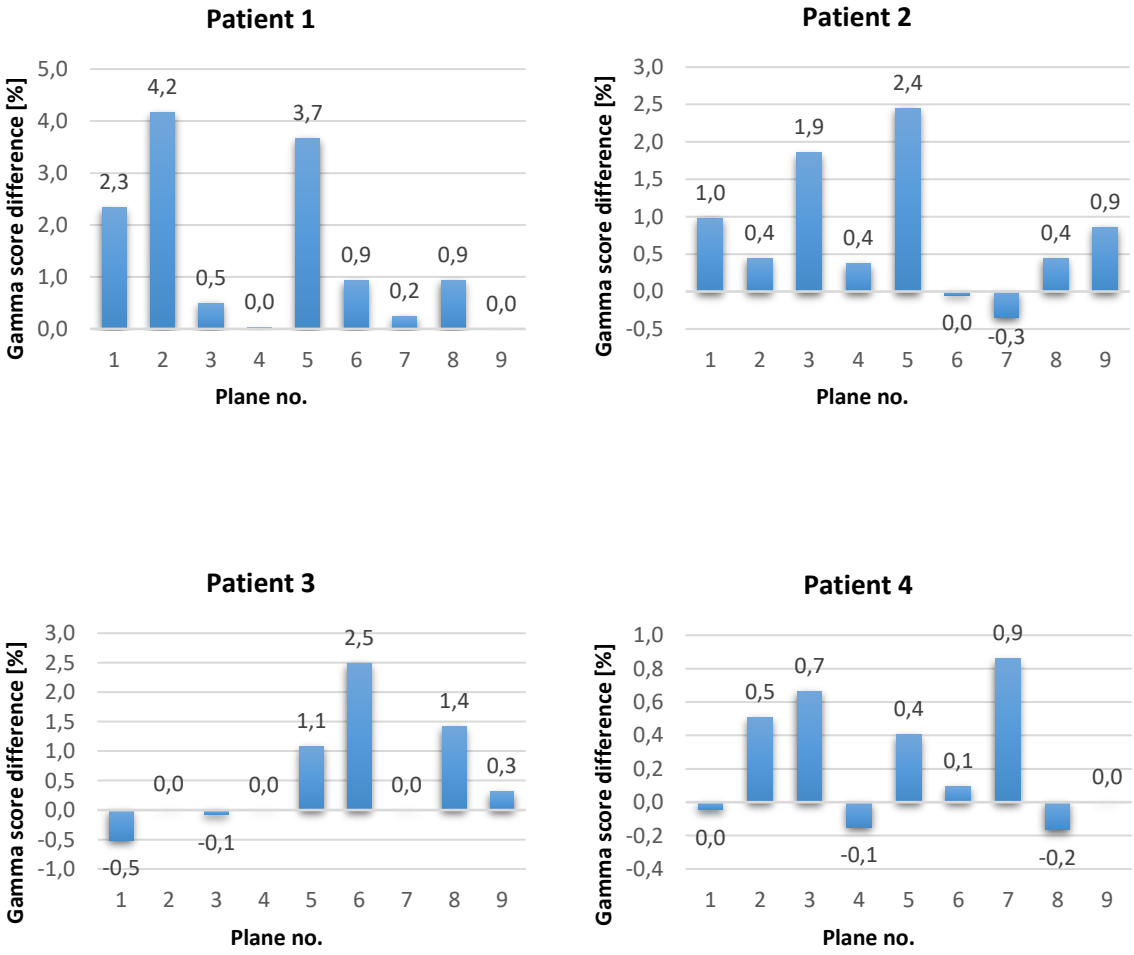


Figure 4-11: Differences in gamma scores for individual planes and 7 investigated patients, computed as 3D code-based gamma score minus 2D code-based gamma score. No. 1-3 transverse planes, no. 4-6 sagittal planes, no. 7-9 coronal planes – to be continued on next page.

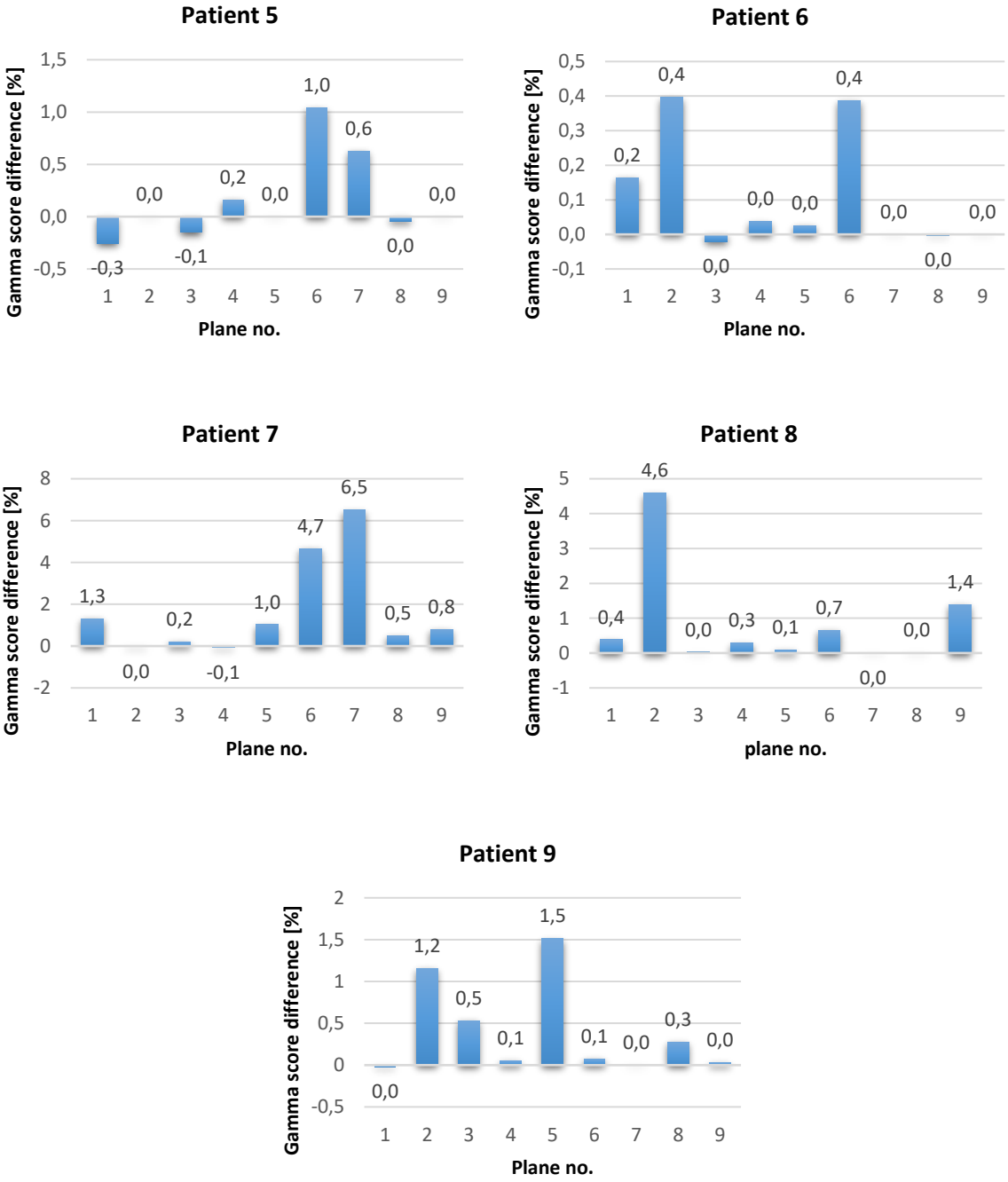


Figure 4-11: Differences in gamma scores for individual planes and 7 investigated patients, computed as 3D code-based gamma score minus 2D code-based gamma score. No. 1-3 transverse planes, no. 4-6 sagittal planes, no. 7-9 coronal planes - continued.

Qualitatively, it seems that the largest differences in 2D code-based and 3D code-based gamma scores are seen for Patient no. 1 (the Siemens Artiste head-and-neck 6 MV step-and-shoot IMRT plan) and also for Patient no. 7 (the TrueBeam prostate 18 MV sliding window IMRT plan). It seems, therefore, that the amount of difference is not associated with a particular technique, treatment site or energy. However, there is not enough data to confirm this by a statistical test. The same is true for different directions (transverse, sagittal and coronal). Qualitatively, it seems that the magnitude of the 2D code-based and 3D code-based gamma score difference is not affected by a particular

direction, but there is not enough data to perform a statistical test (the expected values from the chi square test of independence would not be above 5 in all categories).

The maximum gamma score difference (for 9 patients, 81 evaluated planes – 9 planes for each patient) between gamma scores computed with our 3D and 2D gamma code was 6.5% and minimum gamma score difference was -0.51% (the difference calculated as 3D minus 2D gamma score, most of the gamma scores were higher for the 3D code).

4.2.2 Comparison of field-by-field and pseudo-3D verification for clinical plans

This part of the study refers to part E of Table 3-1 and shows 2D gamma scores only. They are either computed for the field-by-field approach or for the global plan dose verification in the I'mRT Phantom, but the 15 planes (5 transverse, 5 sagittal and 5 coronal) are evaluated in 2D only (plane-by-plane). Therefore, this part of the study is only based on the Siemens Artiste data, because the Varian TrueBeam patients were not measured field-by-field with film dosimetry (for the VMAT plan this would not even be possible). Evaluation of the same data with the MATLAB 3D code was already done in chapter 4.2.1. Here, gamma analysis is performed either in VeriSoft 3.1 (for the PTW array) or in OmniPro I'mRT and MATLAB using our 2D code (for film). Table 4-11 and Table 4-12 show the field-by-field gamma scores relevant in this part of the study. MATLAB and OmniPro I'mRT values of 2D plane-by-plane gamma scores for all planes in the 3D space were already shown in chapter 3.3.3 in Table 3-8. The difference between the 2D plane-by-plane gamma scores given in Table 4-7 in chapter 4.2.1 and in Table 3-8 in chapter 3.3.3 is actually the size of the evaluated ROI. While in Table 4-7 in chapter 4.2.1 it is a square ROI taken from the built-up 3D matrix, in Table 3-8 in chapter 3.3.3 it is a rectangular ROI corresponding to the whole dose distribution measured with each sheet of film in the 3D space. Here, numbers from chapter 3.3.3 will be used. The subsequent figures aim to illustrate the interpretation of results and because OmniPro I'mRT and our MATLAB 2D code give similar gamma scores, the figures only show results for OmniPro I'mRT. The statistical tests, however, are based on the MATLAB 2D code values. The purpose here is to compare field-by-field and pseudo-3D verification for clinical plans.

Interestingly, the pseudo-3D gamma scores in Table 4-8 (given already in chapter 4.2.1) correspond to the results previously obtained from clinical plan verification (2D dosimetry with the PTW Seven29 and VeriSoft version 3.1, in the isocentric plane) that can be seen in Table 4-12 here. If the same limit of a 90% gamma score is kept as the decision criterion, the same patients (Patient no. 2 and no. 3) would fail the verification and their IMRT plans would not be used clinically. As listed in Table 4-12, the clinical 2D gamma scores for some of the fields were around 88% for Patients no. 2 and 3. (Patient no. 4 also had one gamma score lower than 90% at clinical verification, but the failing points were outside the field boundaries and the field itself had a low weight so this plan was applied clinically.) If 2D evaluation of different dose planes in the 3D space is also considered (gamma scores from Table 3-8 in chapter 3.3.3 calculated in MATLAB and OmniPro I'mRT), the following conclusions can be drawn: All patients have some failing planes in their clinical plan (with gamma score lower than 90%) as calculated with both softwares. The lowest numbers can be seen for Patients no. 2, 3 and 4. This appears to be in accordance with the clinical results (field-by-field verification with PTW Seven29 and VeriSoft v. 3.1). Of course, more clinical plans would have to be tested and a proper tolerance limit would have to be estimated for pseudo-3D verification with radiochromic film. But at first glance, the pseudo-3D film dosimetry method would play a similar role in the decision-making

process as the current clinically used 2D method. The following results try to look at the numbers statistically (where possible – sufficient amount of data).

Table 4-11: Field-by-field 2D gamma scores measured with EBT3 film in the isocentric plane in the RW3 phantom and evaluated in OmniPro I'mRT and MATLAB.

OmniPro I'mRT									
Field no. / **Patient no.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	97.9	99.5	98.1	100.0	98.5	98.5	100.0	-	-
2	93.7	99.2	97.0	99.9	99.7	99.3	97.2	-	-
3	83.3	99.9	98.7	98.3	99.9	87.5	97.5	-	-
4	99.7	98.0	98.3	98.7	99.6	99.7	92.8	-	-
5	99.7	96.0	94.8	97.4	96.3	96.0	95.7	99.7	97.7
6	98.3	94.5	98.7	94.2	98.9	93.3	97.4	99.9	98.7
MATLAB 2D code									
Field no. / **Patient no.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	99.3	99.9	100.0	99.9	99.0	99.1	99.3	-	-
2	96.4	93.6	98.1	99.9	99.3	96.5	98.7	-	-
3	95.3	98.4	99.2	94.1	100.0	92.0	95.7	-	-
4	99.7	99.2	95.5	90.86	98.4	89.3	97.7	-	-
5	97.8	92.89	100.0	100.0	100.0	99.1	99.8	100.0	99.9
6	100.0	100.0	99.2	97.6	100.0	99.0	98.8	100.0	98.3

Table 4-12: Field-by-field 2D gamma scores measured with the PTW Seven29 array in the isocentric plane in the RW3 phantom and evaluated in VeriSoft v. 3.1 (comparison to TPS).

VeriSoft v. 3.1									
Field no. / **Patient no.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	93.4	96.3	100	97.2	98.1	98.8	95.4		
2	91.1	99.6	95.7	97.3	97.3	88.1	94.1		
3	88.9	93.0	95.0	90.6	94.0	94.7	88.2		
4	93.5	94.6	92.7	96.7	96.2	91.8	87.0		
5	92.2	91.7	95.2	92.3	92.3	91.6	92.7	95.7	93.0
6	93.5	91.2	96.0	95.3	90.9	92.0	92.3	95.7	95.0

Verification results both with film and PTW Seven29 are summarized in Figure 4-12. The first two boxes of the box plot show field-by-field results obtained with PTW Seven29 and EBT3 film. The distribution shown for each patient is drawn from 7 individual fields (Patients no. 1 to 4) or 9 individual fields (Patient no. 5 and 6). The other three boxes show results for global plan dose measurement with film in coronal, sagittal and transverse planes. Here, the box is drawn from 5 measured planes for each patient. Figure 4-12 is only meant to guide the eye and visualize graphically the distribution of gamma scores. It is not meant to provide quantitative information on median gamma score values, for example. Even though averaging gamma scores is common practice in published literature, including highly cited papers [36, 37, 68, 101, 109, 116, 118, 120 etc.], it is not

mathematically correct. Gamma scores obtained with PTW Seven29 seem to be lower than with EBT3 film for field-by-field dosimetry but in most cases they are above the clinical tolerance limit of 90%.

Significance of differences for PTW Seven29 and EBT3 film field-by-field was tested with the chi square test of independence, similarly to chapter 4.1.1, but this time for a different set of data. Here the test was performed on larger samples, where all six patients and all their 7 or 9 fields per plan were evaluated. Thus, each sample contained 46 values. Of course, the test statistic can be sensitive to the size of the sample. The first categorical variable contained 4 gamma score bins: gamma score <94%, gamma score between 94% and 96%, gamma score between 96% and 98% and gamma score > 98%. These boundaries were considered so that the condition of the test, that the so-called expected values must be at least 5, would be fulfilled with our data. The second categorical variable was the method with which they were obtained (PTW VeriSoft v. 3.1 + Seven29 and MATLAB 2D code + EBT3 film). The null hypothesis was that gamma score values are not related to the method with which they were measured and calculated. The contingency table is shown in Table 4-13. This table suggests that the frequency of low gamma scores is very high for the PTW Seven29 array while for the EBT3 film, the frequency is the biggest for high gamma scores. **The resulting p-value from the test was very small, p = 2.42 x 10e-08. Therefore, the null hypothesis was rejected at a 5% significance level. We might suppose that the PTW array Seven29 method for field-by-field verification together with VeriSoft 3.1 gives lower gamma scores than the EBT3 film method combined with 2D MATLAB code and that the difference is statistically significant.** The corresponding box plot - again only to guide the eye - is in Figure 4-13.

Table 4-13: Contingency table with frequencies for the chi square test of independence comparing field-by-field verification with EBT3 film and with PTW Seven29.

Gamma score category (GS)	EBT3 film + 2D MATLAB code	PTW Seven29 + VeriSoft v. 3.1	Row totals
GS < 94%	5	24	29
94% <GS ≤ 96%	4	12	16
96% < GS ≤ 98%	5	6	11
GS > 98%	32	4	36
Column totals	46	46	**Grand total:** 92

It must be pointed out again that the detectors used for field-by-field measurements (PTW Seven29 and EBT3 film) are not directly comparable because of the different software used for evaluation. As already mentioned earlier, higher threshold than 5% of maximum dose is not possible in the VeriSoft version used. But lower threshold than 10% for films would not be practical due to poor film performance for low doses. The implementation of algorithms in both softwares might play a role. For example, the search distance is not adjustable in VeriSoft v. 3.1 and is not known to the user. Also, interpolation between measurement and prediction dose points is a part of the algorithm and cannot be influenced by the user. Different normalization of the matrices was applied in OmniPro I'mRT, MATLAB and VeriSoft v. 3.1. Point dose normalization in film dosimetry is very sensitive to the point selection due to film noise (even though the evaluated matrices were smoothed prior to evaluation). So it would be impractical to choose the maximum dose point for normalization. In VeriSoft v. 3.1, there is no other option. Also, the field area

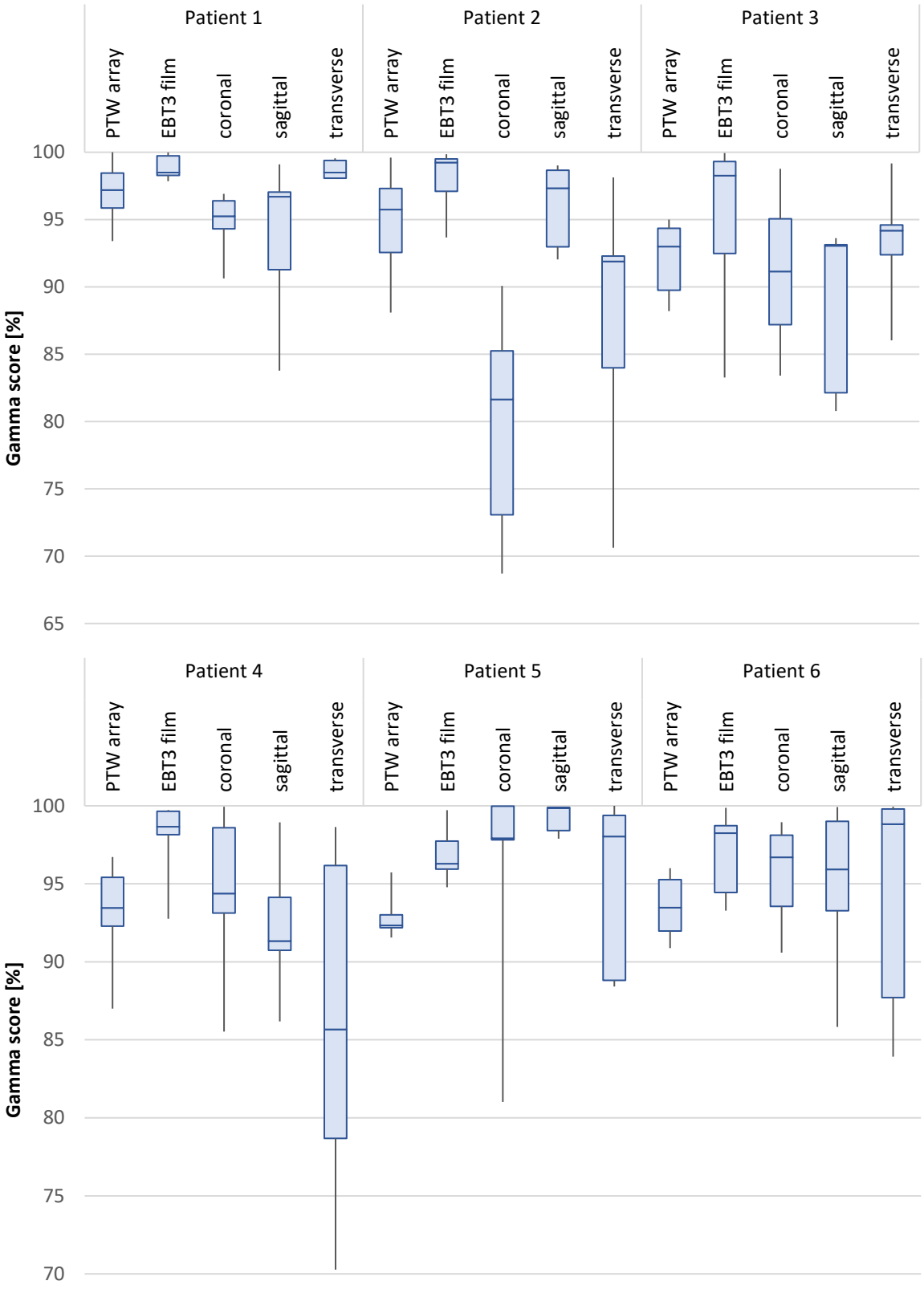


Figure 4-12: Box plot showing verification results for individual Siemens Artiste patients for the five tested setups. PTW array *= PTW Seven29 field-by-field measurements;* EBT3 film *= EBT3 film field-by-field measurements;* coronal *= EBT3 film global dose measurements in coronal planes;* sagittal *= EBT3 film global dose measurements in sagittal planes;* transverse *= EBT3 film global dose measurements in transverse planes. The boxes show the median value, the first and the third quartile and the lines show the minimum and the maximum value of gamma scores. This is only to guide the eye, the figure does not aim to provide quantitative information.*

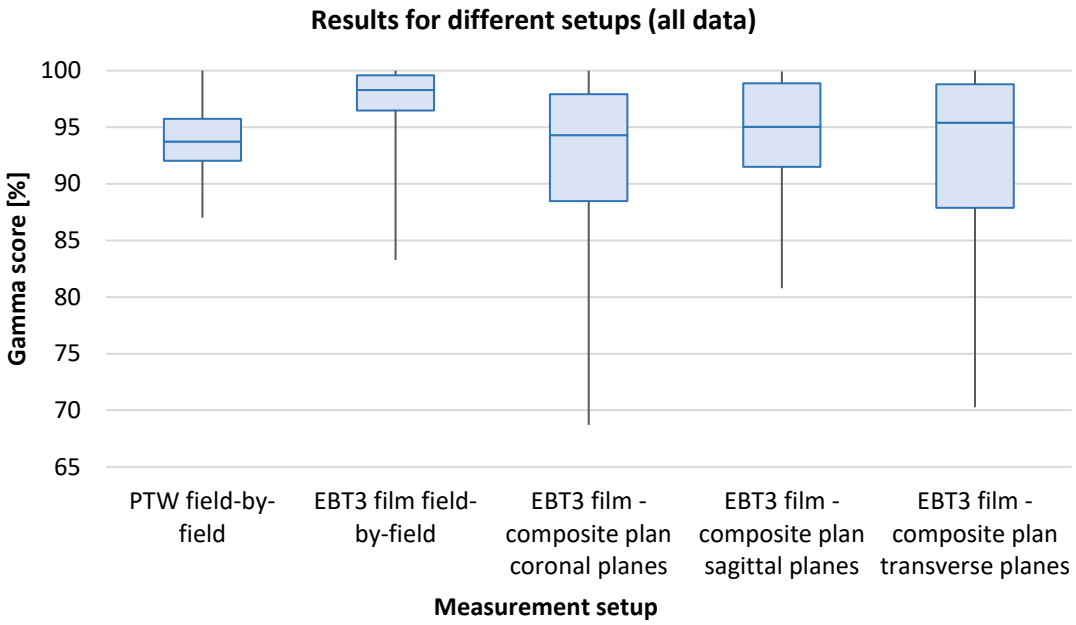


Figure 4-13: Box plot showing the distribution of gamma scores based on all measured data. The first two boxes show all 46 individual fields (for all tested Siemens Artiste patients). The other three boxes show all 30 global (composite) plan measurements in each direction – 5 planes for all 6 Siemens Artiste patients in each box. The boxes show the median value, the first and the third quartile, and the lines show the minimum and the maximum value of gamma scores. This is only to guide the eye, the figure does not aim to provide quantitative information.

evaluated with film and the PTW detector was different due to the different software features. Several studies have been published that address the issue of evaluation software performance [69, 123, 128, 129]. Our comparison rather shows the difference between a clinical verification scenario (PTW array field-by-field measurements as performed at the Thomayer Hospital) compared to what can be achieved with film using a suitable, but different, methodology. However, pseudo 3D measurements and field-by-field measurements with film are directly comparable, as the same methodology was used and film angular dependence was found to be negligible.

Now the question is whether pseudo-3D verification gives different results to field-by-field measurements. These data are directly comparable because they were obtained and evaluated with the same detector and software. It seems that the gamma scores are different, guessing visually from Figure 4-13, where the distribution of 46 field-by-field measurements (all fields for all patients) and 30 global dose measurements (5 planes for each patient measured in coronal, sagittal or transverse direction) for the Siemens Artiste patients is shown for visual guidance of the eye. This would suggest that in the whole 3D space there are errors that cannot be detected with the field-by-field approach. They can occur because smaller errors from different fields add to each other (or they compensate each other) when the whole plan dose distribution is measured. This issue has been investigated before [130, 131] and then discussed further [132-134] with unclear conclusions. EBT3 field-by-field measurements and EBT3 global plan dose measurements can directly be compared to answer the question.

For this purpose, all 46 individual fields measured with film were compared to all 30 coronal (or sagittal or transverse) planes measured with film (5 planes in each direction for each of the 6

patients) for the Siemens Artiste patients by the chi square test of independence. The first categorical variable contained the gamma score categories: gamma score ≤ 94%, gamma score between 94% and 96%, gamma score between 96% and 98% and gamma score > 98%. The second variable was the method considered – EBT3 field-by-field approach or EBT3 pseudo-3D approach. The null hypothesis was that the values of gamma score are not related to the type of verification (field-by-field or pseudo-3D). The so-called expected values met the assumption of the test with these data. The contingency table is shown in Table 4-14. It seems from this table that for the pseudo-3D approach, there is a high frequency of gamma scores below 94% while for the field-by-field approach, there is a high frequency of gamma scores above 98%. **The resulting p-value was again very small (p = p = 2.05 x 10e-09) and we rejected the null hypothesis at a 5% significance level. The gamma scores thus seem to be related to the type of verification method and the field-by-field approach seems to give significantly higher gamma scores than the pseudo-3D approach.**

Table 4-14: Contingency table with frequencies for the chi square test of independence comparing the field-by-field and pseudo-3D approach with EBT3 film.

Gamma score category (GS)	Field-by-field	Pseudo-3D	Row totals
GS < 94%	5	48	53
94% <GS ≤ 96%	4	12	16
96% < GS ≤ 98%	5	16	21
GS > 98%	32	13	45
Column totals	46	89	**Grand total:** 135

The corresponding box plot (only for visualization of the qualitative gamma score distribution) is in Figure 4-13. A different gamma score limit might be appropriate for global plan film measurements in clinical practice than for field-by-field measurements. The use of a single number — the gamma score limit — for pass/fail decisions must be justified by previous investigations into the causes and magnitude of discrepancies observed in different planes of the global dose distribution and in individual fields. The causes of discrepancy and their difference between field-by-field and global plan scenario might lie for example in the design of the particular TPS calculation algorithms. Before an optimization based on a sufficient amount of clinical cases is done, however, the same tolerance level should be considered for field-by-field and global plan dose measurements.

By looking at Figure 4-13 and the distribution of values, the general behaviour of the verification methods (PTW Seven29 with VeriSoft v. 3.1 and EBT3 film with OmniPro I'mRT) can be discussed at least qualitatively. It is not intended here to really quantify quartiles or medians because, as mentioned earlier, in case of gamma scores this would not be mathematically correct. This is just a visual inspection of the behaviour of values. The gamma scores that can be obtained with PTW Seven29 and VeriSoft v. 3.1 are not very variable. However, this is not necessarily true for individual patients (Figure 4-12). (This is, of course, influenced by the small number of values included for 1 patient.) For EBT3 film detector, Figure 4-13 shows that lower gamma scores are obtained more often. It seems that gamma scores that can be obtained with the film method are more variable and the minimal values are generally lower than with the PTW array. This is even more significant for individual patients in Figure 4-12, even though in a few cases the behaviour is the opposite. As already explained in chapter 4.1 (but with different data), the uncertainty in relative dose measurement with film can translate to variable gamma scores and the largest difference seen in

chapter 3.1.15 was 4.5% in gamma score. While for the PTW Seven29 array the variability in gamma score is roughly around 2%. (The term uncertainty of gamma score is not used on purpose.) This behaviour should be taken into account when establishing tolerance limits for film dosimetry in the clinic.

In literature, authors have also addressed the question of different results for different plane directions (coronal, transverse, sagittal) [135]. Their study suggests that results of verification are strongly dependent on the chosen direction and that coronal plane (which is mostly chosen clinically) is insensitive to delivery error. They found no correlation between errors in different verification planes (using film). Here, this issue can be described qualitatively by Figure 4-13, which shows the distribution of results for all patients, and Figure 4-12, where the different directions are compared for each patient individually. In Figure 4-13 it seems that all three directions give similar gamma scores while gamma scores in sagittal planes are less variable than in the other two. However, for individual patients the situation can be different, as shown in Figure 4-12. This figure shows – again only qualitatively to guide the eye - the distribution of gamma scores obtained in 5 parallel planes around the isocentre separated by 2 cm of slabs in the phantom for each patient and each direction (coronal, transverse, sagittal). It can be seen that in most cases gamma scores are very variable among these 5 parallel planes. This could mean that measuring just one plane (e.g. at isocentre) in only one direction would be insufficient to perform verification comprehensively. Dose deviations coming from the treatment planning system might be different for coronal, sagittal and transverse planes because of the planning CT images. While in transverse direction CT slices contain precise information about patient's tissue, coronal and sagittal planes are calculated with a certain geometrical error. Then the precision of dose calculation in discrete coronal and sagittal planes will depend on the particular TPS algorithm design. As has been shown in Figure 3-12 in chapter 3.3.4, in XiO 4.80, differences in dose profiles can be seen at the intersection of two different directions in the TPS. Delivery error, on the other hand, can be affected by performance of the multileaf collimator (MLC), which travels in the patient's latero-lateral direction, never in the craniocaudal direction (if the collimator is kept at 0°). Thus, coronal, transverse and sagittal planes might be affected differently by this type of error. On the other hand, there is a known uncertainty in relative film dosimetry with our method that translates into variability in gamma scores and it has been shown that gamma scores can differ up to 4.5% just due to the uncertainty in measured dose. (The term uncertainty of gamma score is not used on purpose.) Generally, some planes in the 3D space can fall into a region of lower doses and it has been shown in this work that our film dosimetry method does not perform well for such low doses. This is a drawback of the pseudo-3D film dosimetry method as compared to systems based on ionization chamber arrays, as already discussed in chapter 4.1. However, this was not the case for the Siemens Artiste patients that included 5 prostate and lymph node cases and one head and neck case.

The chi square test of independence was performed to see if gamma score values depend on the plane direction. One of the categorical variables were four gamma score bins: gamma score ≤ 90%, gamma score between 90% and 94%, gamma score between 94% and 97% and gamma score > 97% and the other was the plane direction (transverse, sagittal or coronal). All 90 planes measured for the Siemens Artiste patients were included in the evaluation and the so-called expected values were higher than 5. The contingency table is shown in Table 4-15. This table also suggests that the frequency of gamma score values lower than 90% is higher than for the other gamma score bins, at least for sagittal and coronal planes. **The null hypothesis was that gamma score values are**

independent of the direction in which they were measured (transverse, sagittal or coronal). With a p-value of 0.6962 we cannot reject the null hypothesis at a 5% significance level. Our finding is therefore contrary to study [135] because it seems that verification results with film dosimetry are not affected by the chosen plane of measurement.

Table 4-15: Contingency table with frequencies for the chi square test of independence comparing different plane directions measured with EBT3 film in psuedo-3D manner.

Gamma score category (GS)	Transverse planes	Sagittal planes	Coronal planes	Row totals
GS < 90%	7	11	12	30
90% <GS ≤ 94%	7	7	4	18
94% < GS ≤ 97%	8	8	7	23
GS > 97%	8	4	7	19
Column totals	30	30	30	**Grand total:** 90

Here, data from Varian TrueBeam (Patients no. 7 and 8 – IMRT and Patient no. 9 – VMAT) could also be included in the chi square test of independence to test the behaviour of different plane directions, even though it is possible that Varian TrueBeam with TPS Eclipse might behave differently to Siemens Artiste and XiO in this aspect. Anyway, if Varian data is also included in the test, with a p-value of 0.8796 the null hypothesis still cannot be rejected at a 5% significance level. There is not enough data to test this separately for Varian TrueBeam.

4.2.3 Comparison of field-by-field and pseudo-3D verification for error-induced IMRT plans

The following results show two clinical Siemens Artiste plans where leaf position errors of 1 mm (which is within tolerance limits in quality assurance procedures) or 2 and 3 mm were introduced in three fields (for Patient no. 4 and Patient no. 5, see chapter 3.3.5.2). This refers to part F of Table 3-1 and the purpose is to compare field-by-field to pseudo-3D verification for error-induced plans.

The next step could be to analyse whether such an error is clinically significant. Although this study does not focus on this specific question and details are not given in this text, conclusions from a different study based on patient data in the Thomayer Hospital in Prague (and carried out by the author) suggest that these errors are already significant for Siemens Artiste in IMRT mode an TPS XiO. This is an important assumption for interpretation of results in the following text.

Results are shown in terms of gamma score, because this is the most commonly used clinical decision criterion. Additionally, gamma histograms are shown and gamma statistics, such as the mean and the maximum gamma index, are given. Due to low amont of data, statistical tests cannot be performed in this part of the study.

4.2.3.1 Results for Patient no. 4

Table 4-16 and Figure 4-14, which represent values of 2D plane-by-plane gamma scores for the 15 investigated planes in the 3D space, based on our 2D MATLAB code and evaluating the same square ROI as in Table 4-17, show that the introduced errors of 1, 2 and 3 mm in leaf positions, even though only in 3 out of 7 treatment fields, generally resulted in a lower gamma score. (Note that in this case the whole treatment plan was delivered at once.) This is not true in all investigated planes,

but for the isocentric planes, which are usually chosen for verification, this is true in all 3 directions. Table 4-16 also shows that in the original clinical plan only 5 planes failed the clinical tolerance of a 90% gamma score, while in the error-induced plan most of the planes failed this criterion. However, if only one dose plane in the 3D space was chosen for verification, it might not reveal the introduced error (planes no. 2, 5, 7, 9 and 15) because their 2D gamma score would meet clinical criteria. In other words, the positioning errors in all MLC leaves introduced into 3 teatment fields, which are believed to be already clinically significant, might remain unnoticed if only one plane in the 3D space is measured during plan verification.

Table 4-16: 2D plane-by-plane gamma scores calculated with our **2D code in MATLAB** *for 15 planes (1-5 transverse, 6-10 sagittal, 11-15 coronal) in the I'mRT Phantom for Patient no. 4. Measured doses on film are compared to TPS predicted values of the error-free plan. Values failing the clinical tolerance of 90% are marked in grey. 2D gamma score difference for a given plane is also shown (calculated as error-induced minus original).*

Plane no.	2D gamma score [%]		
	Original clinical plan	**Error-induced plan**	**Difference**
1	95.6	81.1	-14.4
2	100.0	95.9	-4.1
3	89.8	86.8	-3.1
4	78.5	84.3	5.8
5	85.6	92.6	7.0
6	17.8	20.2	2.4
7	97.1	91.3	-5.8
8	94.0	81.3	-12.7
9	97.3	92.0	-5.3
10	93.2	80.9	-12.3
11	92.9	68.6	-24.3
12	94.3	90.0	-4.3
13	97.7	13.7	-84.0
14	98.5	85.3	-13.2
15	83.3	94.6	11.3
	10 planes passing 90% tolerance	6 planes passing 90% tolerance	**Average difference:** -10.5%

If gamma score falls below the tolerance level during clinical verification, as in the case of plane no. 13, the physicicst should perform further investigations into the cause of this behaviour – whether this is only caused by properties of the verification method (including pre-set evaluation parameters) or by a true delivery or calculation error.

The same conclusions can be drawn from Table 4-17 and Figure 4-15 which show 2D gamma scores in 9 investigated planes in the 3D space calculated with our 3D code. (The difference is that for 2D gamma score calculation with the 3D code, gamma indices from the 3D calculation are considered. They might differ from 2D calculation in areas of intersection with other planes in the 3D space.) Here, only 9 planes can be calculated because the others do not intersect each other in the 3D cubic matrix.

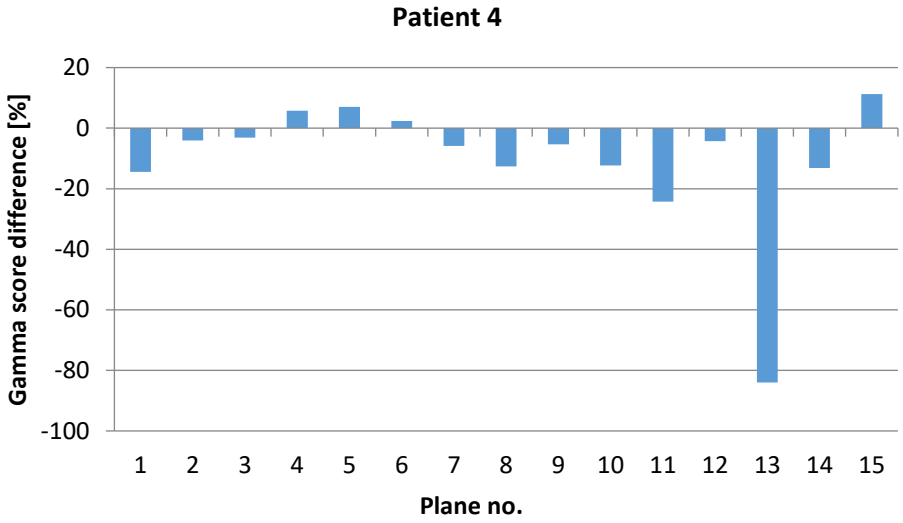


Figure 4-14: 2D gamma score difference for 15 planes (1-5 transverse, 6-10 sagittal, 11-15 coronal) in the I'mRT Phantom for Patient no. 4 (calculated as error-induced plan minus original plan). 2D gamma scores were calculated with our **2D code in MATLAB**.

Table 4-17: 2D plane-by-plane gamma scores calculated with our **3D code in MATLAB** *for 9 planes (2-4 transverse, 7-9 sagittal, 12-14 coronal) in the I'mRT Phantom for Patient no. 4. Measured doses on film are compared to TPS predicted values of the error-free plan. Values failing the clinical tolerance of 90% are marked in grey. 2D gamma score difference for a given plane is also shown (calculated as error-induced minus original).*

Plane no.	2D gamma score [%] (from pseudo-3D calculation, local normalization)		
	Original clinical plan	**Error-induced plan**	**Difference**
2	100.0	87.6	-12.4
3	90.4	96.2	5.8
4	79.2	85.9	6.7
7	97.0	81.9	-15.1
8	94.4	92.4	-2.0
9	97.4	92.5	-4.9
12	95.2	19.0	-76.2
13	97.6	90.3	-7.3
14	98.5	86.9	-11.5
	8 planes passing 90% tolerance	4 planes passing 90% tolerance	**Average difference:** -13.0%

Gamma index statistics – the minimum, the maximum and the mean gamma index – are given in Table 4-18 for the error-induced as well as for the original clinical plan. This refers to gamma indices calculated with our 2D MATLAB code in all 15 measured planes. The differences for each plane calculated as original plan value minus error-induced plan value are also shown for each plane. The information is visually supported by Figure 4-16 where the difference in mean gamma index between

the original and the error-induced plan is plotted. In most (but not in all) planes, the mean gamma index is lower in the error-induced plan (in 10 planes out of 15).

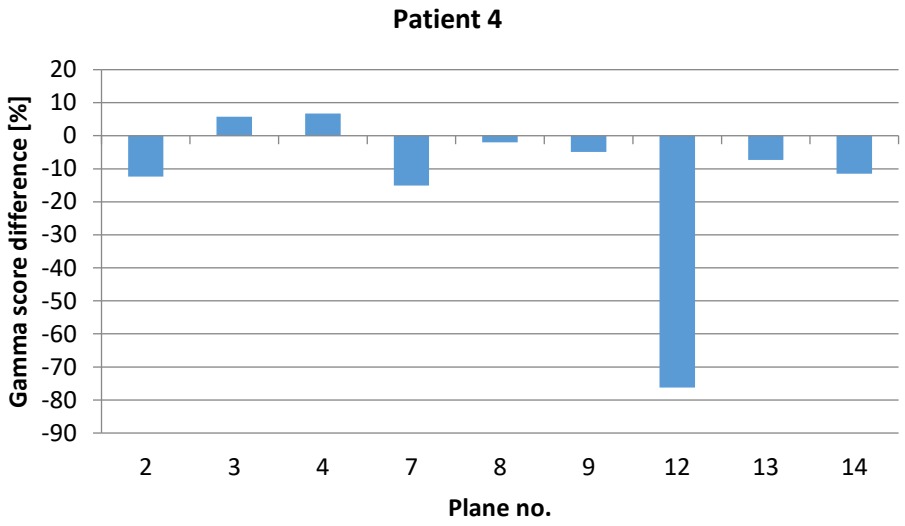


Figure 4-15: 2D gamma score difference for 9 planes (2-4 transverse, 7-9 sagittal, 12-14 coronal) in the I'mRT Phantom for Patient no. 4 (calculated as error-induced plan minus original plan). 2D gamma scores were calculated with our **3D code in MATLAB**.

Table 4-18: Minimum, maximum and mean gamma indices from 2D MATLAB code for the error-induced plan of Patient no. 4 and values from the original clinical plan for comparison. Differences were calculated as original minus error-induced plan.

Plane no.	Original clinical plan			Error-induced plan			Difference		
	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}
1	0.00	2.33	0.37	0.00	2.23	0.60	0.00	0.10	-0.23
2	0.00	1.05	0.31	0.00	1.36	0.43	0.00	-0.31	-0.12
3	0.00	1.83	0.54	0.00	1.90	0.50	0.00	-0.07	0.03
4	0.00	1.82	0.66	0.00	2.24	0.58	0.00	-0.42	0.08
5	0.00	2.05	0.45	0.00	1.72	0.45	0.00	0.33	0.00
6	0.00	6.95	2.26	0.00	3.73	1.78	0.00	3.22	0.48
7	0.00	1.44	0.41	0.00	1.70	0.47	0.00	-0.26	-0.06
8	0.00	1.46	0.47	0.00	1.79	0.61	0.00	-0.33	-0.13
9	0.00	1.48	0.35	0.00	2.09	0.46	0.00	-0.61	-0.12
10	0.00	2.64	0.51	0.00	2.71	0.61	0.00	-0.07	-0.10
11	0.00	1.57	0.42	0.00	5.15	0.80	0.00	-3.58	-0.38
12	0.00	1.86	0.46	0.00	4.24	0.53	0.00	-2.38	-0.07
13	0.00	1.48	0.40	0.00	5.53	2.82	0.00	-4.05	-2.42
14	0.00	1.69	0.30	0.00	1.83	0.54	0.00	-0.14	-0.24
15	0.00	2.11	0.55	0.00	3.96	0.46	0.00	-1.85	0.09

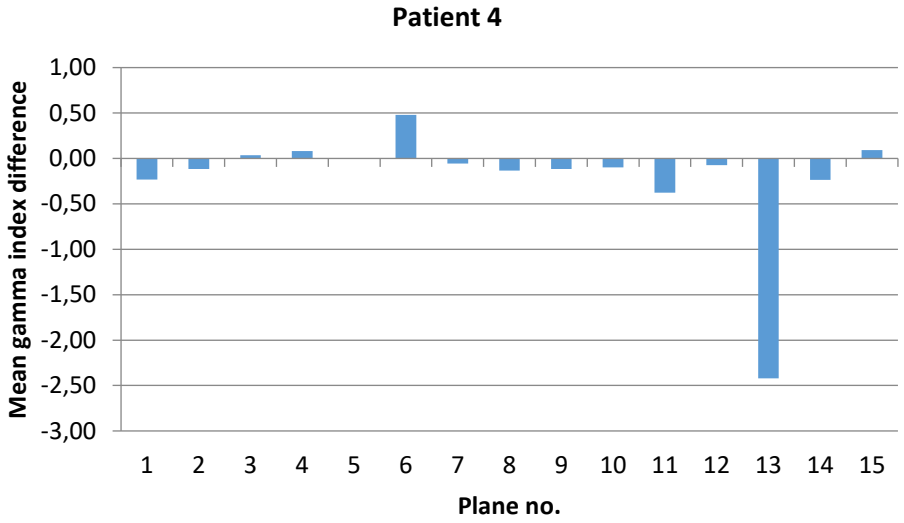
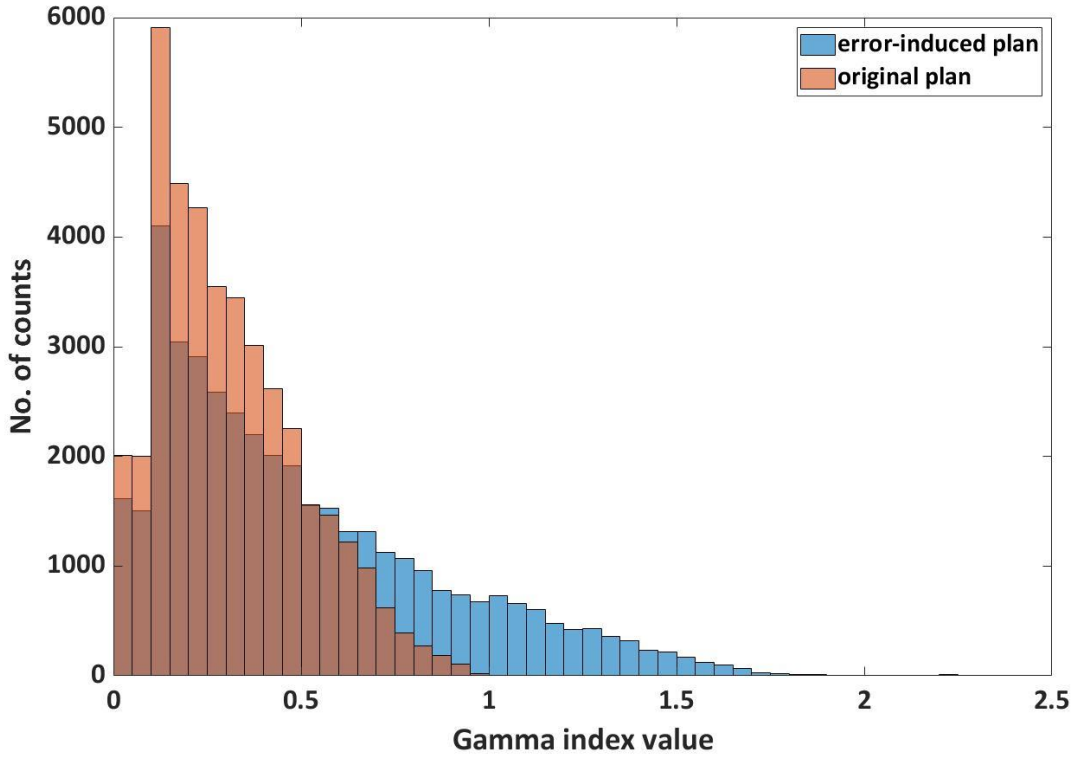
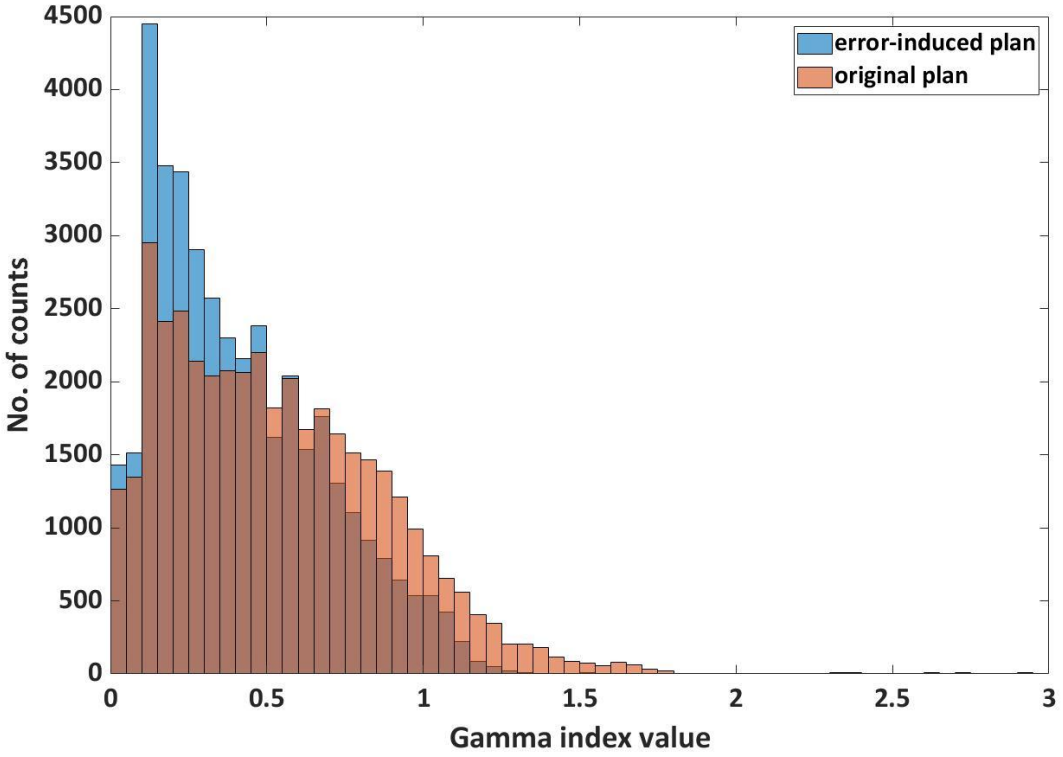


*Figure 4-16: Difference in mean gamma index between the original clinical plan and the error-induced plan for Patient no. 4 (original minus error-induced value in each plane) as calcualted with our **2D MATLAB code**.*



(a)

Figure 4-17: Gamma histograms for plane no. 2 (a) and 3 (b) for the error-induced plan (in blue) and the original clinical plan (in orange) for Patient no. 4 calculated with our 3D MATLAB code – to be continued on next page.



(b)

Figure 4-17: Gamma histograms for plane no. 2 (a) and 3 (b) for the error-induced plan (in blue) and the original clinical plan (in orange) for Patient no. 4 calculated with our 3D MATLAB code - continued.

Table 4-19: The overall pseudo-3D gamma score calculated with our 3D MATLAB code for Patient no. 4. Measured doses on film are compared to TPS predicted values of the error-free plan.

Pseudo-3D gamma score [%]	Local normalization	Global normalization
Original clinical plan	94.4	92.1
Error-induced plan	81.4	87.2

Table 4-20: 2D field-by-field gamma scores for the 3 error-induced fields of Patient no. 4 compared to the same original fields, as calculated by our 2D code in MATLAB and measured field-by-field with film in an RW3 slab phantom and gantry at 0°. Measured doses on film are compared to TPS predicted values of the error-free plan.

Field no.	2D gamma score [%]		
	Original clincial plan	Error-induced plan	Positioning error
1	99.7	83.8	1 mm
2	99.2	57.5	2 mm
4	90.9	83.1	3 mm

Additionally, for illustration of a possible difference between the original and the error-induced plan, Figure 4-17 (a) shows gamma histograms for plane no. 2 where the error-induced plan had a lower gamma score. The values come from our MATLAB 3D code (it corresponds to gamma scores in Table 4-17 and not to the values in Table 4-16 where the MATLAB 2D code was used). For comparison, Figure 4-17 (b) also shows gamma histograms for plane no. 3 where a positive difference in gamma

score was seen, meaning that the original plan had a lower gamma score than the error-induced plan.

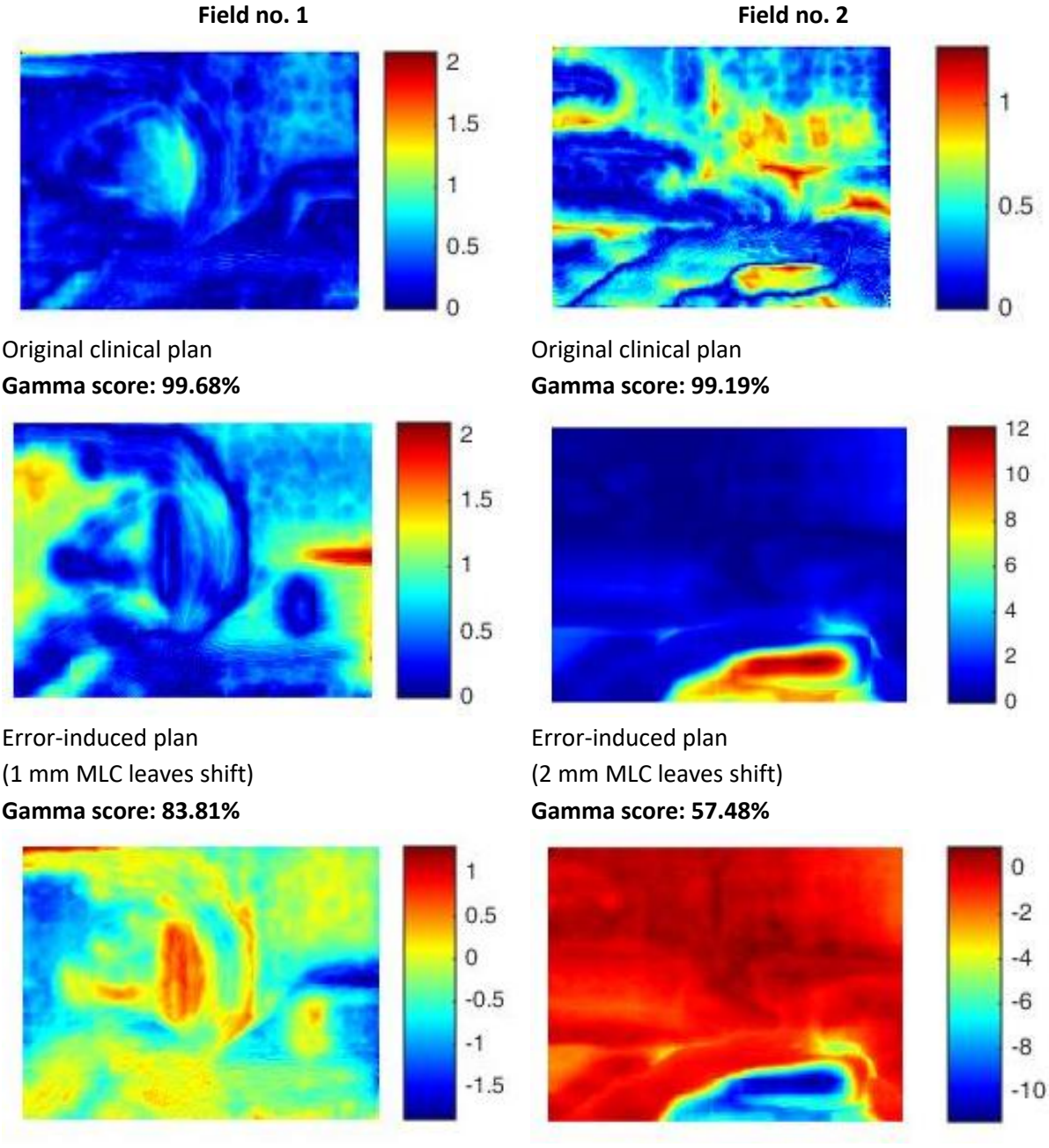


Figure 4-18: Maps of gamma indices (measured field-by-field with film and calculated with our 2D MATLAB code) for fields no. 1 and 2 for Patient no. 4, both for original clinical fields and error-induced fields. Figure also shows maps of gamma index differences calculated as original minus error-induced field.

Table 4-19 clearly shows that for Patient no. 4 the overall pseudo-3D gamma score (that reflects gamma points in the 3D space, but only at those points where dose was actually measured with film) would reveal the introduced error. For the original clinical plan it passes the auxilliary tolerance of 90% while for the error-induced plan it does not, both with local normalization and global

normalization. (Local normalization in this study means that each 2D plane is normalized to its 75th percentile while global normalization means that all planes are normalized to the 75th percentile of the whole 3D space).

It is important to compare these findings with the clinically used field-by-field verification. That is, field-by-field measurements in an RW3 slab phantom in reference depth and gantry always at 0°. Patient no. 4 was already tested in chapter 4.2.1. Here, 2D gamma scores were recalculated with our 2D gamma code in MATLAB and a summary of results is given in Table 4-20 because the only values that are directly comparable with the above mentioned data are gamma scores obtained with film and MATLAB. As already mentioned in previous chapters, some parameters for gamma analysis are not reproducible if a different detector or software is used. Unlike in the 3D space, where the whole treatment plan was delivered and the dose distribution was affected both by the erroneous fields and the original fields, here we can only see the erroneous ones. So the difference between gamma scores of the error-free and error-induced plan should be more pronounced. All three fields of Patient no. 4 revealed the error because their gamma score was lower than the clinical tolerance of 90%.

The introduced error for this patient was thus revealed with both field-by-field verification and the pseudo-3D gamma score as well as with most of the investigated planes in the 3D space when 2D gamma scores were calculated for them, either with our 2D or 3D code in MATLAB. However, generally speaking, in the 3D space the errors are partly compensated by the error-free fields. If the field-by-field verification does not take into account the weight of the individual fields and perhaps also the particular clinical situation, a false positive result might be obtained with field-by-field approach (because a small error could be compensated in the global 3D dose distribuiton).

Gamma index maps for illustration purposes (only for Patient no. 4) are shown in Figure 4-18.

4.2.3.2 Results for Patient no. 5

For Patient no. 5, results are similar to Patient no. 4. For this patient, 1 mm systematic positioning error was introduced into MLC shapes of all segments in 3 treatment fields (out of 9). 2D plane-by-plane gamma scores for 15 planes in 3D space, calculated with our 2D MATLAB code and with the same square ROI as in Table 4-22, were generally lower for the error-induced plan and less planes passed the tolerance of 90% (see Table 4-21 and Figure 4-19). When 2D gamma scores were calculated with our 3D MATLAB code, as in Table 4-22 and Figure 4-20, again, more planes passed the 90% level in the original clinical plan and the gamma scores were generally higher for this plan. The overall pseudo-3D gamma score shown in Table 4-24 would also reveal the induced error if local normalization was applied. Global normalization for this patient should be avoided due to very homogeneous dose distributions, as discussed in the previous chapters (75th percentile of a homogeneous dose distribution is different from the same percentile of a modulated distribution and global normalization might be inappropriate). (Local normalization in this study means that each 2D plane is normalized to its 75th percentile while global normalization means that all planes are normalized to the 75th percentile of the whole 3D space. It should be pointed out that 1 mm leaf positioning error is considered in tolerance during quality assurance procedures.

More interestingly, only one out of 3 error-induced fields would reveal the error on gamma score if field-by-field verification was performed, see Table 4-25. Considering that this patient had 9 treament fields, this means that according to the clinical verification protocol at the Thomayer

Hospital in Prague, this plan would meet the tolerance criteria (only one field failed the gamma score limit of 90% and had a gamma score close to 90%). Supposing that 1 mm leaf positioning error on all leaves might have a clinically relevant impact on dose distribution (as suggested by a different study mentioned earlier), field-by-field verification in this case would bring a false negative result (this plan would be delivered to the patient). While the overall pseudo-3D gamma score with local normalization and 2D gamma scores for the 15 investigated planes in the 3D space would probabaly prevent this plan from being irradiated (depending on specific tolerance levels that have not yet been established for pseudo-3D verification).

*Table 4-21: 2D plane-by-plane gamma scores calculated with our **2D code in MATLAB** for 15 planes (1-5 transverse, 6-10 sagittal, 11-15 coronal) in the I'mRT Phantom for Patient no. 5. Measured doses on film are compared to TPS predicted values of the error-free plan. Values failing the clinical tolerance of 90% are marked in grey. 2D gamma score difference for a given plane is also shown (calculated as error-induced minus original).*

Plane no.	2D gamma score [%]		
	Original clinical plan	**Error-induced plan**	**Difference**
1	99.3	95.9	-3.4
2	100.0	91.7	-8.3
3	99.8	85.8	-14.0
4	99.0	89.7	-9.3
5	97.5	80.7	-16.7
6	78.2	83.0	4.8
7	88.4	93.6	5.2
8	100.0	98.6	-1.4
9	89.3	74.7	-14.5
10	65.2	71.5	6.2
11	98.5	93.6	-4.9
12	71.7	82.0	10.4
13	99.9	84.9	-15.1
14	98.3	70.2	-28.0
15	98.8	90.8	-8.0
	10 planes passing 90% tolerance	6 planes passing 90% tolerance	**Average difference:** -6.5%

Gamma index statistics – the minimum, the maximum and the mean gamma index – are given in Table 4-23 for the error-induced as well as for the original clinical plan. This refers to gamma indices calculated with our 2D MATLAB code in all 15 measured planes. The differences for each plane calculated as original plan value minus error-induced plan value are also shown for each plane. The information is visually supported by Figure 4-21 where the difference in mean gamma index between the original and the error-induced plan is plotted. In most (but not in all) planes, the mean gamma index is lower in the error-induced plan (in 10 planes out of 15).

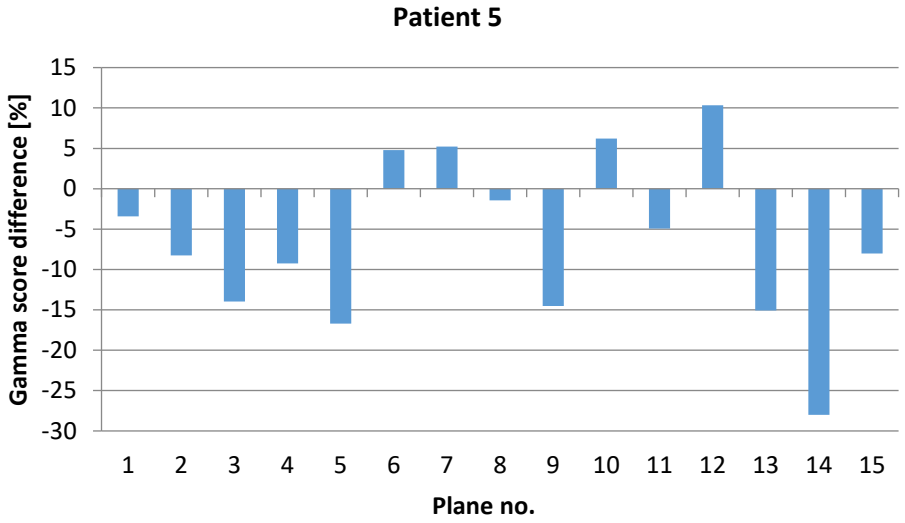


*Figure 4-19: 2D gamma score difference for 15 planes (1-5 transverse, 6-10 sagittal, 11-15 coronal) in the I'mRT Phantom for Patient no. 5 (calculated as error-induced plan minus original plan). 2D gamma scores were calculated with our **2D code in MATLAB**.*

*Table 4-22: 2D plane-by-plane gamma scores calculated with our **3D code in MATLAB** for 9 planes (2-4 transverse, 7-9 sagittal, 12-14 coronal) in the I'mRT Phantom for Patient no. 5. Measured doses on film are compared to TPS predicted values of the error-free plan. Values failing the clinical tolerance of 90% are marked in grey. 2D gamma score difference for a given plane is also shown (calculated as error-induced minus original).*

Plane no.	2D gamma score [%] (from pseudo-3D calculation, local normalization)		
	Original clinical plan	**Error-induced plan**	**Difference**
2	99.7	85.3	-14.4
3	99.8	94.2	-5.6
4	98.8	89.6	-9.3
7	88.6	98.4	9.8
8	100.0	93.3	-6.7
9	90.3	74.8	-15.5
12	72.3	86.2	14.0
13	99.9	82.0	-18.0
14	98.3	71.4	-26.9
	7 planes passing 90% tolerance	3 planes passing 90% tolerance	**Average difference:** -8.07%

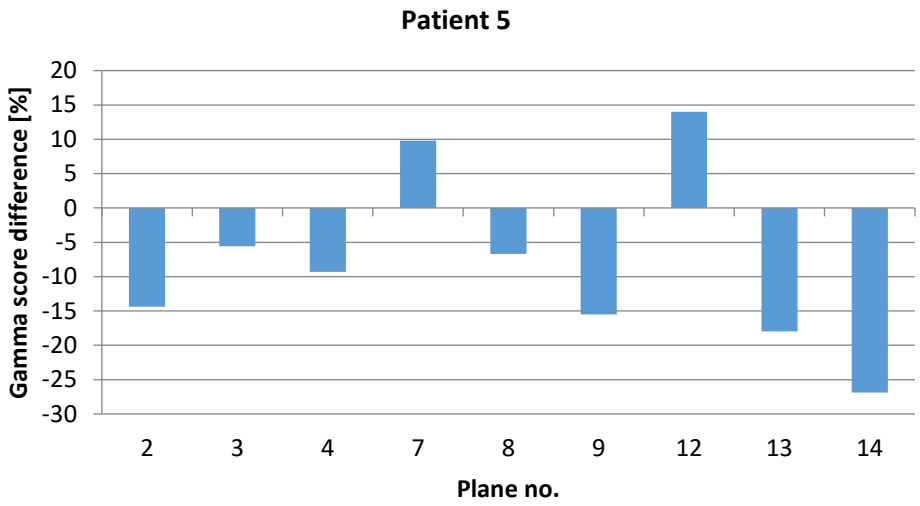


*Figure 4-20: 2D gamma score difference for 9 planes (2-4 transverse, 7-9 sagittal, 12-14 coronal) in the I'mRT Phantom for Patient no. 5 (calculated as error-induced plan minus original plan). 2D gamma scores were calculated with our **3D code in MATLAB**.*

Table 4-23: Minimum, maximum and mean gamma indices from 2D MATLAB code for the error-induced plan of Patient no. 5 and values from the original clinical plan for comparison. Differences were calculated as original minus error-induced plan.

Plane no.	Original clinical plan			Error-induced plan			Difference		
	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}
1	0.00	1.46	0.37	0.00	1.68	0.39	0.00	-0.22	-0.02
2	0.00	0.65	0.17	0.00	2.33	0.54	0.00	-1.68	-0.37
3	0.00	1.08	0.33	0.00	2.04	0.49	0.00	-0.96	-0.16
4	0.00	1.11	0.34	0.00	1.40	0.55	0.00	-0.28	-0.21
5	0.00	1.35	0.43	0.00	3.49	0.63	0.00	-2.14	-0.20
6	0.00	2.24	0.64	0.00	2.58	0.57	0.00	-0.34	0.07
7	0.00	2.14	0.48	0.00	2.35	0.40	0.00	-0.20	0.08
8	0.00	0.79	0.21	0.00	1.77	0.37	0.00	-0.98	-0.15
9	0.00	1.91	0.49	0.00	3.37	0.67	0.00	-1.45	-0.18
10	0.00	2.37	0.82	0.00	3.31	0.76	0.00	-0.94	0.07
11	0.00	1.21	0.39	0.00	1.31	0.50	0.00	-0.09	-0.10
12	0.00	1.93	0.74	0.00	2.18	0.49	0.00	-0.25	0.24
13	0.00	1.03	0.43	0.00	1.97	0.58	0.00	-0.94	-0.15
14	0.00	1.71	0.33	0.00	3.05	0.82	0.00	-1.34	-0.48
15	0.00	1.61	0.35	0.00	2.12	0.47	0.00	-0.52	-0.12

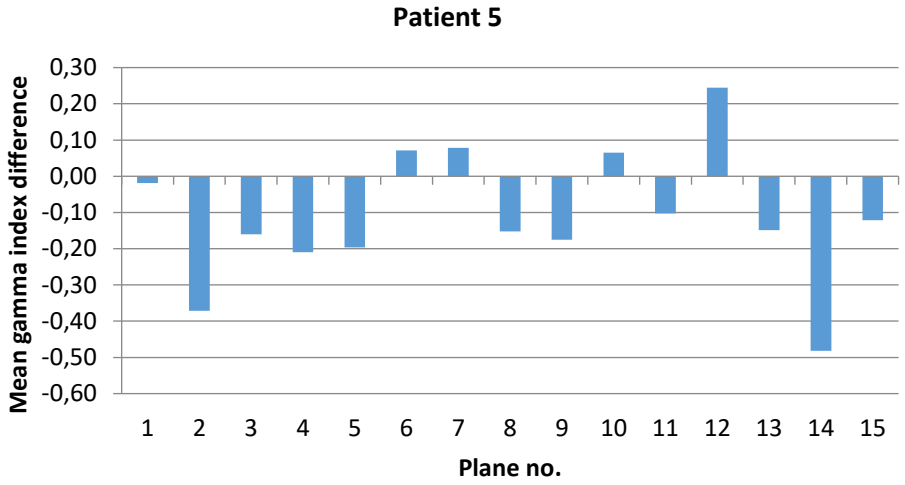
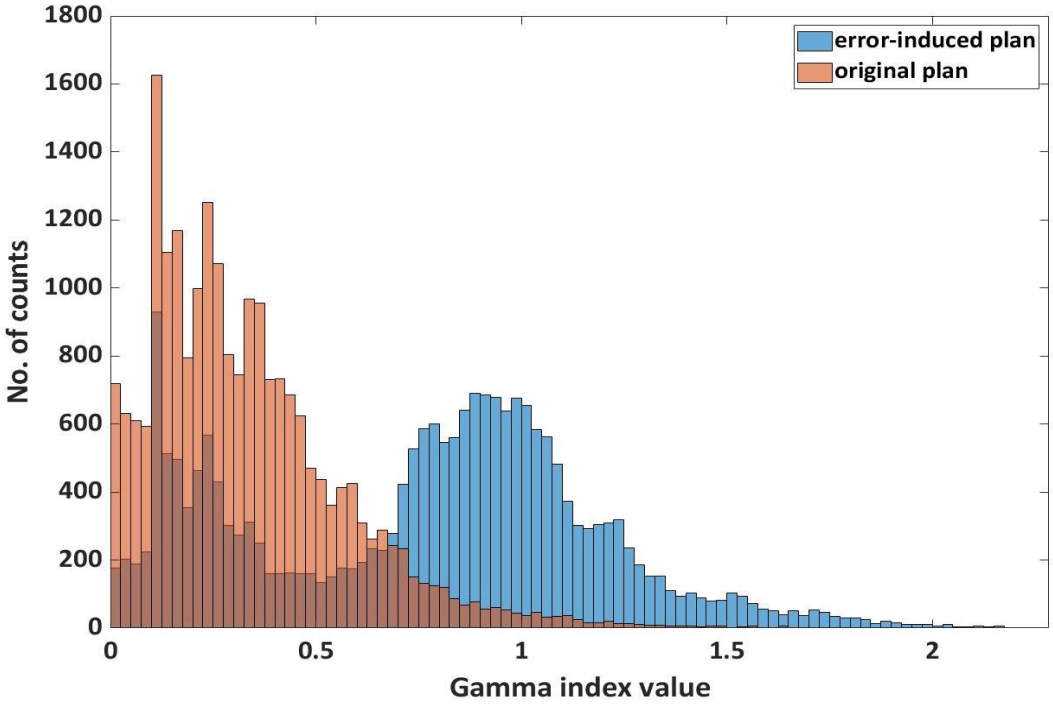


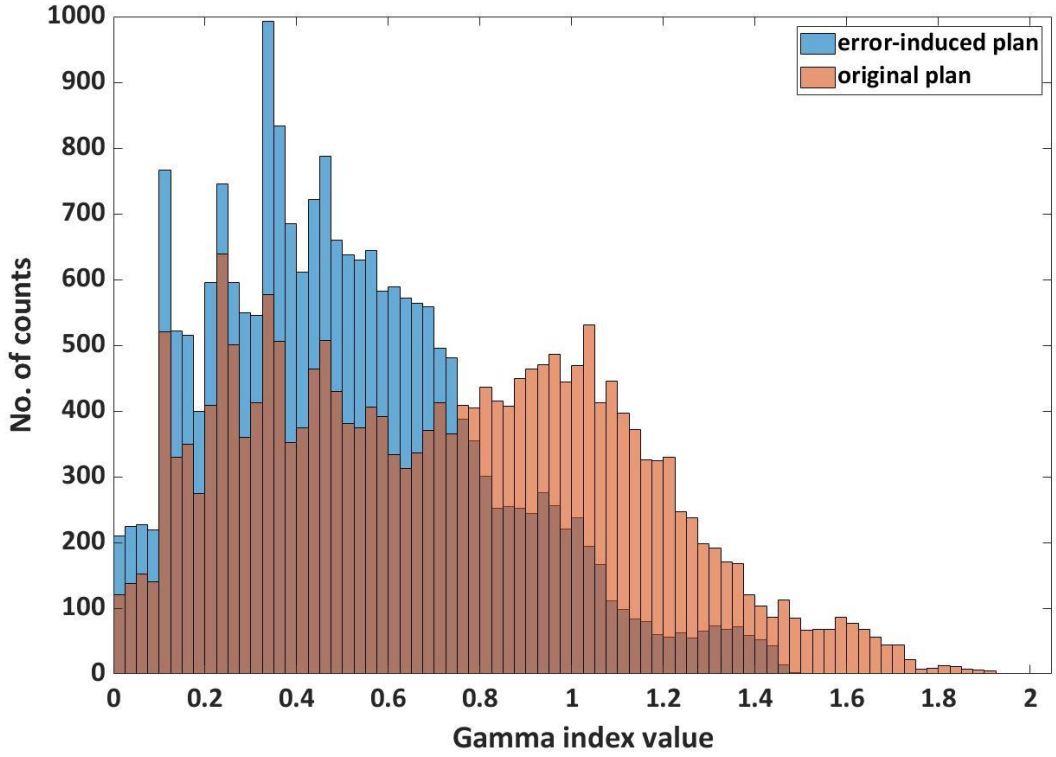
Figure 4-21: Difference in mean gamma index between the original clinical plan and the error-induced plan for Patient no. 5 (original minus error-induced value in each plane) as calcualted with our 2D MATLAB code.

Additionally, for illustration of a possible difference between the original and the error-induced plan, Figure 4-22 (a) shows gamma histograms for plane no. 14 where the biggest difference in gamma score was seen between the original and the error induced plan and the error-induced plan had a lower gamma score. The values come from our MATLAB 3D code and correspond to gamma socres in Table 4-22 (not to values in Table 4-21 computed with the 2D MATLAB code). For comparison, Figure 4-22 (b) also shows gamma histogram for plane no. 12 where a positive difference in gamma score was seen, meaning that the original plan had a lower gamma score than the error-induced plan.



(a)

Figure 4-22: Gamma histograms for plane no. 14 (a) and 12 (b) for the error-induced plan (in blue) and the original clinical plan (in orange) for Patient no. 5 calculated with our 3D MATLAB code – to be continued on next page.



(b)

Figure 4-22: Gamma histograms for plane no. 14 (a) and 12 (b) for the error-induced plan (in blue) and the original clinical plan (in orange) for Patient no. 5 calculated with our 3D MATLAB code - continued.

Table 4-24: The overall pseudo-3D gamma score calculated with our 3D MATLAB code for Patient no. 5. Measured doses on film are compared to TPS predicted values of the error-free plan.

Pseudo-3D gamma score [%]	Local normalization	Global normalization
Original clinical plan	94.2	65.0
Error-induced plan	86.1	80.3

Table 4-25: 2D field-by-field gamma scores for the 3 error-induced fields for Patient no. 5 compared to the same original fields, as calculated by our 2D code in MATLAB and measured field-by-field with film and gantry at 0°. Measured doses on film are compared to TPS predicted values of the error-free plan.

	2D gamma score [%]	
Field no.	Original clincial plan	Error-induced plan
2	92.9	92.8
4	100.0	82.0
5	100.0	92.0

4.2.4 Comparison of the pseudo-3D approach with film and OCTAVIUS 4D for an error-induced VMAT plan

This part of the text refers to Part G of Table 3-1 and aims to compare the pseudo-3D film method to a clinically used, commercially available pseudo-3D alternative – the OCTAVIUS 4D rotating cylindrical phantom equippped with the PTW OCTAVIUS 1500 array of ionization chambers evaluated in VeriSoft version 8.0. Something similar was done in chapter 4.1.1.2 for a clinical plan where both systems seemed to show very good agreement with TPS predictions, except for the film method and planes with very low doses, beyond the edge of the PTV. Now the two methods are compared for VMAT plans with introduced errors. The original VMAT plan for Patient no. 9 (18 MV beam, prostate) was modified as described in chapter 3.3.5.2. Thus, three error-induced plans were obtained – 1 mm X1 leaf bank shift in each control point in one arc, 3 mm X1 leaf bank shift in each control point in one arc and 3% MU error in both arcs.

Results are shown in terms of gamma score, gamma statistics and gamma histograms. There is not enough data in this part of the study to perform statistical tests. Results are shown for 3%/3 mm criteria where not specified otherwise.

Table 4-26 shows gamma scores for all three types of introduced error as well as for the original clinical VMAT plan for Patient no. 9 (already shown in chapter 4.1.1.2), both for the pseudo-3D film method and the pseudo-3D OCTAVIUS 4D method. All the measured dose distributions (the error-induced ones as well as the original clinical one) were compared to the original clinical plan dose prediction form the TPS using gamma analysis. The differences in gamma score in Table 4-26 were calculated as the error-induced minus the original value for each tested plane. First five planes (no. 1 – 5) correspond to transverse planes, where planes no. 1 and 5 are 4 cm away from the isocentre, planes no. 2 and 4 are 2 cm far from the isocentre and plane no. 3 is the isocentric plane. Similarly, planes no. 6 – 10 correspond to sagittal planes and planes no. 7 – 15 correspond to coronal planes. Gamma scores in Table 4-26 were calculated with our 2D MATLAB code (the plane-by-plane approach) and the same planes were evaluated in VeriSoft version 8.0 for the OCTAVIUS system, also as plane-by-plane in 2D.

It is obvious that the 1 mm error was not detected with the OCTAVIUS system and the 3%/3 mm cirteria. For the 3 mm leaf bank shift, which seems to be the most pronounced error (already from the DVH values in chapter 3.3.5.2), there was a drop in gamma score in all 15 investigated planes for the OCTAVIUS system, although only 5 planes dropped below the clinically used 95% tolerance value. Interestingly, in the planes that lay beyond the edge of the PTV, i.e. planes no. 1, 5, 10 and 11, there was no change in gamma score and it remained 100% as in the original clinical plan. For the 3% MU error, a drop in gamma score could be seen in almost all the 15 planes with the OCTAVIUS system but none of the planes dropped below the clinical tolerance of 95%. The gamma score in planes no. 1 and 5 (in the low dose region) keeped to the 100% value and in plane no. 11 (also in the low dose region) it was 99.9% compared to 100% for the error-free plan.

With the film dosimetry method, low to extremely low gamma scores were seen in planes no. 1 and 5 (which were in the low dose area, the mean dose being below 0.5 Gy) for all types of investigated errors. Plane no. 5 had an extremely low gamma score even in the original clinical plan. Low gamma scores were also seen for planes no. 10 and 11, which were also in the low dose area beyond the edge of the PTV. The OCTAVIUS system saw a perfect agreement (100%) in all these cases except for

plane no. 10 and the plan with 3% MU error. Thus, both systems behave differently in low dose areas. The film dosimetry method tends to predict very poor agreement with TPS while the OCTAVIUS 4D method tends to predict perfect agreement. Note that there was a 10% threshold for both methods but the normalization for film dosimetry was to the 75th percentile while for OCTAVIUS system it was to the maximum dose in the predicted matrix.

With the film dosimetry method, even the 1 mm leaf bank shift maked some difference in gamma scores, while with the OCTAVIUS system this type of error remained undetected. Only 8 planes out of 15 met the clinical tolerance of 95% with the film dosimetry method. There was a drop in gamma score in almost all planes. For some planes the drop was larger than the 4.5% value which represents a possible difference in gamma score due to uncertainty in film dosimetry. Plane no. 5 showed a large positive difference meaning that the 1 mm error-induced plan had a higher gamma score than the original one. But this was the plane outisde of the high dose area where the film dosimetry method cannot be relied on. Other planes with positive differences were below the 4.5% value. Interestingly, other studies that had performed a similar experiment and had tried to detect a 1 mm error with another verification system had not succeeded (e.g. [113]).

The 3 mm leaf bank shift caused a drop in gamma score in almost all investigated planes with the film dosimetry method. Plane no. 5 showed a large positive difference, for which the explanation has already been given (low dose beyond the edge of the PTV). The positive difference in plane no. 3 was below the 4.5% value representing the possible variability in gamma score with film dosimetry. In many of the planes, the decrease in gamma score was larger than 4.5%. Only 6 planes passed the cinical tolerance of 95% in the plan with 3 mm leaf errors.

For the plan with the 3% MU error, the film dosimetry method showed slightly higher gamma scores for some of the planes, but these were well below the 4.5% variability level. Except for plane no. 5, as already explained (beyond the edge of the PTV). In some of the planes, the drop in gamma score was larger than 4.5%. Only 9 planes out of 15 passed the clinical 95% tolerance level.

The data also show the need for measuring more planes in the 3D space. The 3 mm error would probably be detected with both methods, but several sagittal and coronal planes would have to be investigated. The 3% MU error would probably be detected with the film method if several sagittal and coronal planes were investigated. With the OCTAVIUS system, this type of error would be detected with the 3%/3 mm criteria only if a higher tolerance level than 95% was used. All the isocentric planes (transverse, sagittal and coronal) had a gamma score value higher than 95% for all types of error with the film dosimetry method. Only one isocentric plane (sagittal plane in the plan with 3 mm error) was below 95% with the OCTAVIUS system. Altogether, only 5 planes dropped below 95% with the OCTAVIUS system and these were sagittal and coronal planes in the plan with 3 mm error.

Table 4-27 shows the overall pseudo-3D gamma score values obtained with the film pseudo-3D dosimetry method (local normalization as defined in this work; global normalization in this particular case gave the same value for the error-free plan) and with the OCTAVIUS 4D pseudo-3D method for the original clinical plan and for all plans with introduced errors. The film dosimetry method gave lower 3D gamma scores for all plans with introduced errors compared to the original clinical plan. Only for the plan with 3 mm leaf position errors this was below 95%. The OCTAVIUS system gave a higher 3D gamma score value for the plan with 1 mm leaf position error compared to the original

Table 4-26: 2D plane-by-plane gamma scores from the pseudo-3D film method calculated by our 2D MATLAB code for the original clinical VMAT plan of Patient no. 9 and three VMAT plans with introduced errors. 2D plane-by-plane gamma scores from VeriSoft v. 8.0 based on OCTAVIUS 4D measurements are shown for the same planes for comparison. Differences in gamma scores are calculated as error-induced minus original value in each plane. All comparisons are done against TPS.

Plane no.	2D gamma score [%]													
	EBT3 film pseudo-3D method							OCTAVIUS 4D pseudo-3D method						
	Original clinical plan	1 mm leaf bank shift	Difference	3 mm leaf bank shift	Difference	3% MU error	Difference	Original clinical plan	1 mm leaf bank shift	Difference	3 mm leaf bank shift	Difference	3% MU error	Difference
1	99.0	34.8	-64.2	43.8	-55.2	44.5	-54.5	100.0	100	0	100.0	0	100.0	0
2	99.4	98.9	-0.5	95.5	-3.9	99.9	0.5	100.0	100	0	97.2	-2.8	99.4	-0.6
3	98.0	98.8	0.8	100.0	2.0	99.7	1.7	99.0	100	1	95.4	-3.6	98.5	-0.5
4	99.2	98.1	-1.1	96.8	-2.4	99.6	0.4	100.0	100	0	98.2	-1.8	98.8	-1.2
5	27.8	53.6	25.8	42.3	14.5	74.9	47.1	100.0	100	0	100.0	0	100.0	0
6	99.2	97.4	-1.8	92.9	-6.3	92.7	-6.5	99.9	100	0	99.7	-0.2	99.6	-0.3
7	99.4	93.2	-6.2	94.1	-5.3	97.1	-2.3	99.9	99.9	0	91.0	-8.9	97.1	-2.8
8	96.7	97.8	1.1	96.8	0.1	98.4	1.7	99.9	99.9	0	90.7	-9.2	97.4	-2.5
9	99.0	97.2	-1.8	31.4	-67.6	99.3	0.3	99.9	100	0.1	89.9	-10	96.8	-3.1
10	95.7	85.1	-10.6	22.9	-72.8	83.1	-12.6	100.0	99.9	-0.1	100.0	0	97.7	-2.3
11	98.4	74.5	-23.9	74.3	-24.1	79.9	-18.5	100.0	100	0	100.0	0	99.9	-0.1
12	100.0	96.1	-3.9	96.2	-3.8	99.7	-0.3	100.0	99.9	-0.1	95.6	-4.4	95.7	-4.3
13	98.5	93.9	-4.6	97.4	-1.1	99.5	1.0	99.8	99.8	0	87.0	-12.8	97.3	-2.5
14	98.0	99.0	1.0	75.1	-22.9	90.6	-7.4	100.0	100	0	94.2	-5.8	99.0	-1
15	94.7	83.8	-10.9	87.2	-7.5	95.6	0.9	100.0	100	0	96.6	-3.4	100.0	0
	13 planes passing 95% tolerance	8 planes passing 95% tolerance	Average difference: -6.7%	6 planes passing 95% tolerance	Average difference: -17.1%	9 planes passing 95% tolerance	Average difference: -3.2%	15 planes passing 95% tolerance	15 planes passing 95% tolerance	Average difference: +0.1%	10 planes passing 95% tolerance	Average difference: -4.2%	15 planes passing 95% tolerance	Average difference: -1.4%

clinical plan. For the other two error-induced plans, the estimated 3D gamma score was lower than for the original clinical plan but none of them was below 95%. It seems that the film dosimetry method might be more sensitive to small errors in leaf positions, as expected, thanks to the better spatial resolution of the method. Of course, more VMAT plans would have to be investigated to confirm this assumption. The low-dose areas (problematic for the film method) are not actually taken into account when calculating the 3D gamma score because these planes lie outside of the 3D cube of used data. Because the 3 mm leaf position error and the 3% MU error are believed to be clinically significant, as shown in chapter 3.3.5.2, it seems that a higher tolerance value must be used for the

evaluation of 3D gamma scores in case of VMAT plans, both for the pseudo-3D film method and for the pseudo-3D OCTAVIUS 4D method, with 3%/ 3 mm criteria.

Table 4-27: The overall pseudo-3D gamma score calculated with our 3D MATLAB code for Patient no. 9. Measured doses with film or with the OCTAVIUS 4D system for the three error-induced plans are compared to TPS predicted values of the error-free plan.

Pseudo-3D gamma score [%]	EBT3 film + MATLAB 3D code, local normalization	OCTAVIUS 1500 detector + OCTAVIUS 4D + VeriSoft v. 8.0
Original VMAT plan	99.1	99.5
1 mm leaf bank shift	97.6	100.0
3 mm leaf bank shift	89.7	96.9
3% MU error	98.3	99.2

Figure 4-23 to Figure 4-26 show the same numbers as Table 4-26 and they are meant to guide the eye. Figure 4-23 shows the gamma score differences between the original clinical plan and the plans with introduced errors for the film dosimetry method (2D plane-by-plane approach and 2D MATLAB code). Figure 4-24 shows the same for the OCTAVIUS 4D method and plane-by-plane evaluation in VeriSoft v. 8.0. Figure 4-25 shows the plan with 3 mm leaf position errors and comparison between the film method and the OCTAVIUS 4D method (differences in gamma score between the error-free and the error-induced plan). Figure 4-26 shows the same for the plan with 3% MU error. The plan with 1 mm leaf position errors is not shown because there is practically no difference between the error-free and error-induced plan for the OCTAVIUS system.

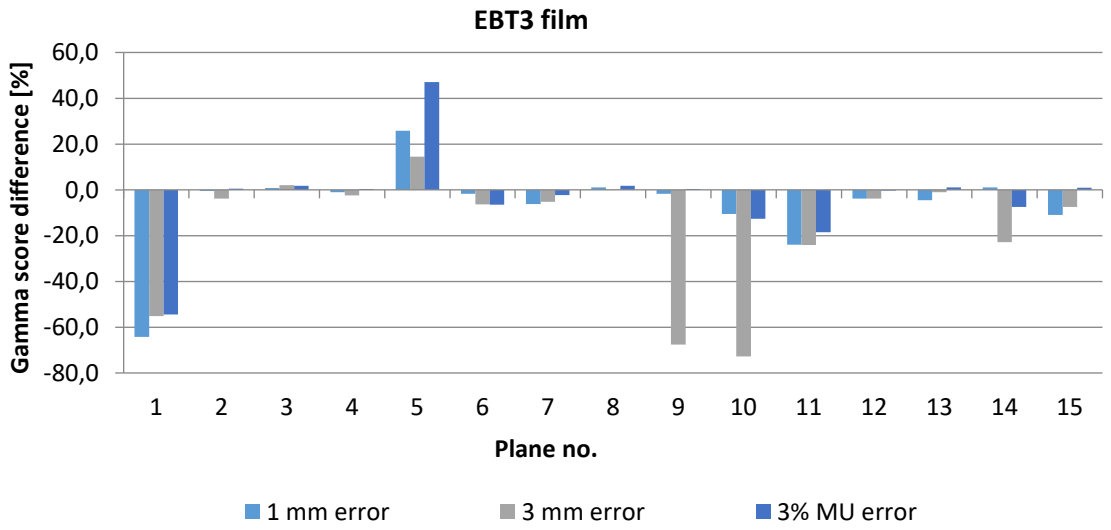


Figure 4-23: Gamma score difference in each plane between the error-induced VMAT plans and the original VMAT plan of Patient no. 9, calculated as error-induced minus original value, estimated with our 2D MATLAB code based on the pseudo-3D film method.

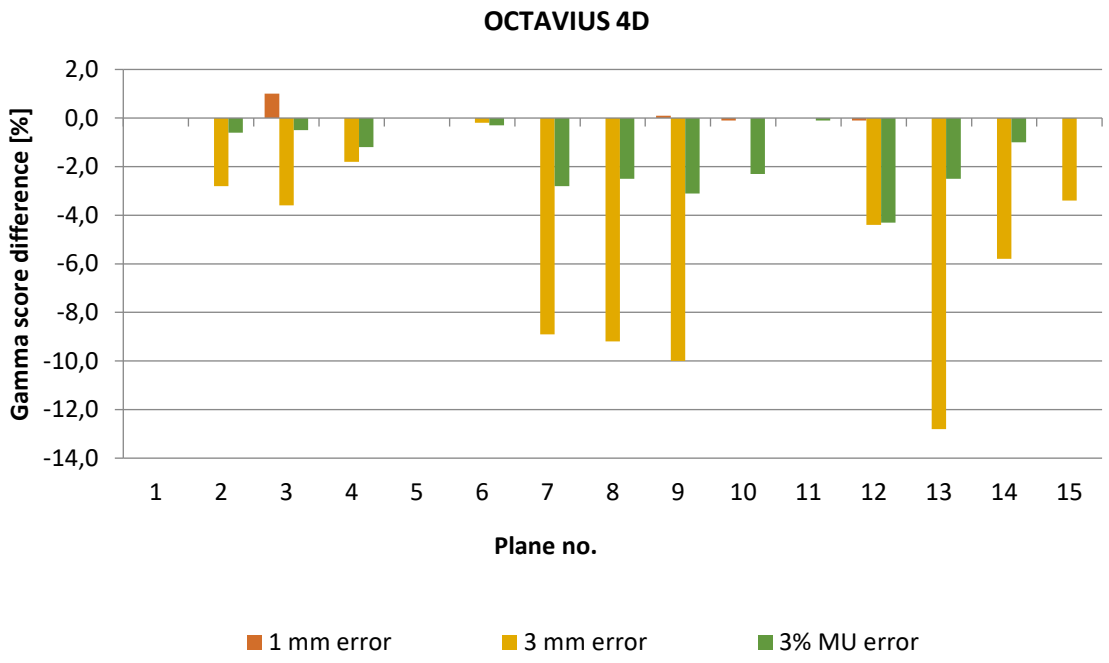


Figure 4-24: Gamma score difference in each plane between the error-induced VMAT plans and the original VMAT plan of Patient no. 9, calculated as error-induced minus original value, estimated with VeriSoft v. 8.0 based on the pseudo-3D OCTAVIUS 4D method.

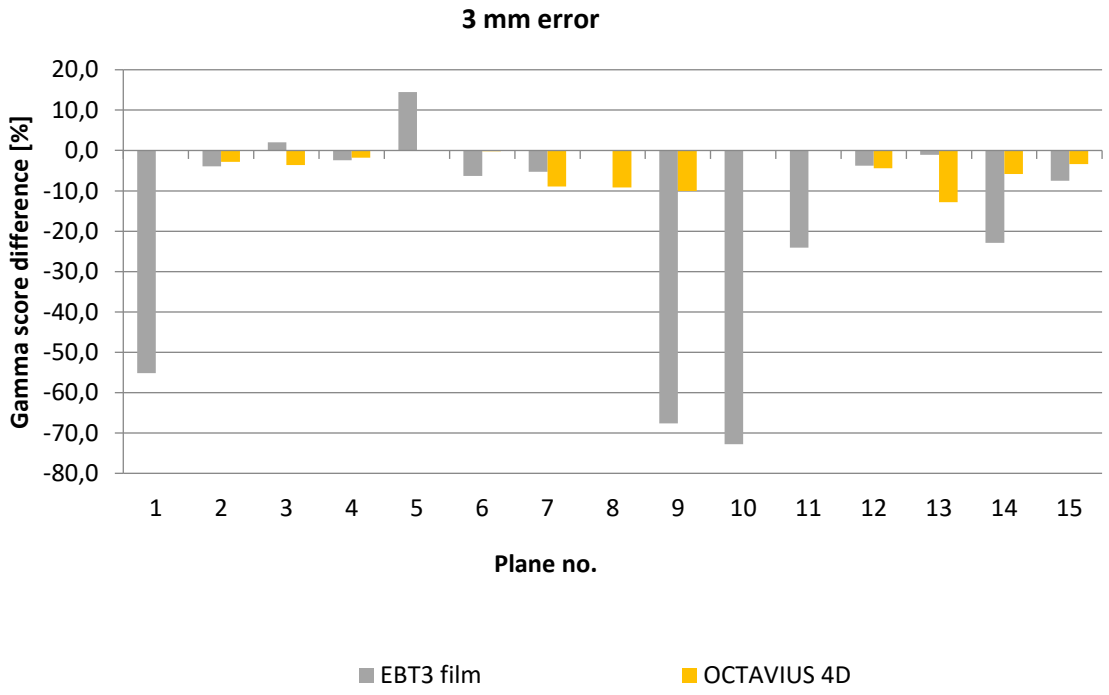


Figure 4-25: Gamma score difference for the error-induced VMAT plan with 3 mm leaf bank shift compared to the original VMAT plan (calculated as error-induced minus original value in each plane) comparing the pseudo-3D film method and the OCTAVIUS 4D pseudo-3D method.

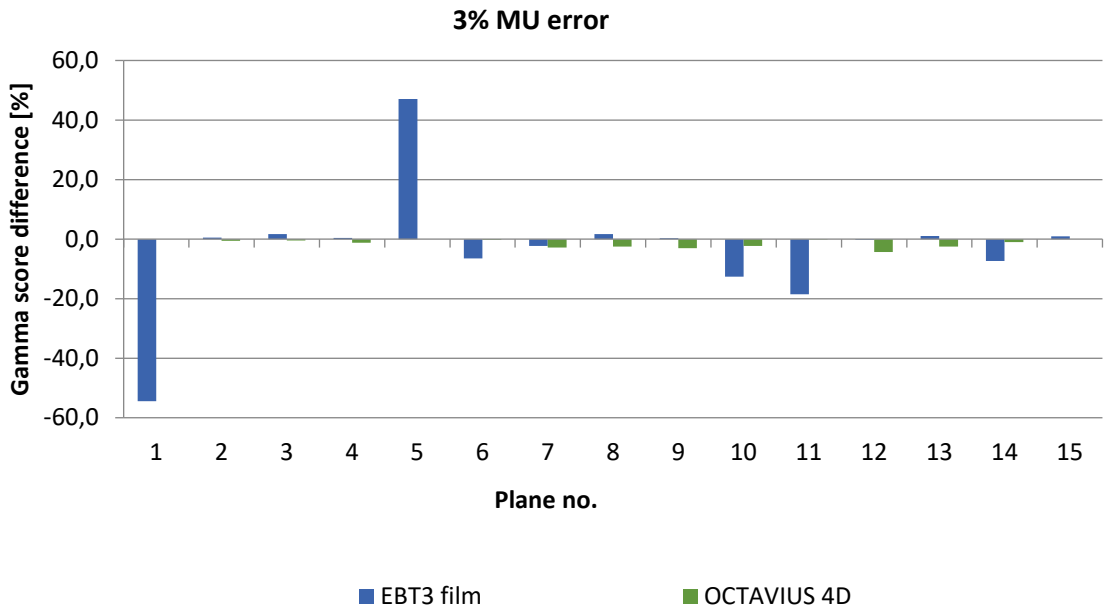


Figure 4-26: Gamma score difference for the error-induced VMAT plan with 3% MU error compared to the original VMAT plan (calculated as error-induced minus original value in each plane) comparing the pseudo-3D film method and the OCTAVIUS 4D pseudo-3D method.

Table 4-28: 2D plane-by-plane gamma scores from the pseudo-3D film method calculated by our 3D MATLAB code for the original clinical VMAT plan of Patient no. 9 and three VMAT plans with introduced errors. Differences in gamma scores are calculated as error-induced minus original value in each plane.

Plane no.	2D gamma score [%] (from pseudo-3D calculation, local normalization)						
	Original clinical plan	1 mm leaf bank shift	Difference	3 mm leaf bank shift	Difference	3% MU error	Difference
2	99.4	99.1	-0.30	96.0	-3.4	99.8	0.4
3	99.2	99.3	0.10	99.8	0.6	99.7	0.5
4	99.7	98.4	-1.30	97.5	-2.2	99.6	-0.1
7	99.5	94.8	-4.70	95.5	-4.0	97.3	-2.2
8	98.2	97.7	-0.50	97.0	-1.2	98.6	0.4
9	99.1	97.3	-1.80	53.5	-45.6	99.2	0.1
12	100.0	96.6	-3.40	96.9	-3.1	99.6	-0.4
13	98.8	96.3	-2.50	98.2	-0.6	99.6	0.8
14	98.0	98.9	0.90	76.7	-21.3	91.7	-6.3
	9 planes passing 95% tolerance	8 planes passing 95% tolerance	Average difference: -1.5%	7 planes passing 95% tolerance	Average difference: -9.0%	8 planes passing 95% tolerance	Average difference: -0.8%

Table 4-28 shows similar information as Table 4-26 for the pseudo-3D film dosimetry method, but this time the numbers were calculated with the 3D MATLAB code, local normalization. (Local normalization in this study means that each dose plane is normalized to its 75th percentile. Global normalization means that all dose points in the 3D space are normalized to the 75th percentile of the whole 3D matrix.) This table shows only those planes that lay in the 3D dose cube which was used to calculate the pseudo-3D gamma score. Planes that lay on the edge of the PTV or beyond (planes 4 cm away from the isocentre) were not taken into account here. Figure 4-27 shows the differences in 2D gamma scores from Table 4-28 (from the 3D MATLAB code, calculated as error-induced value minus original value) graphically to guide the eye.

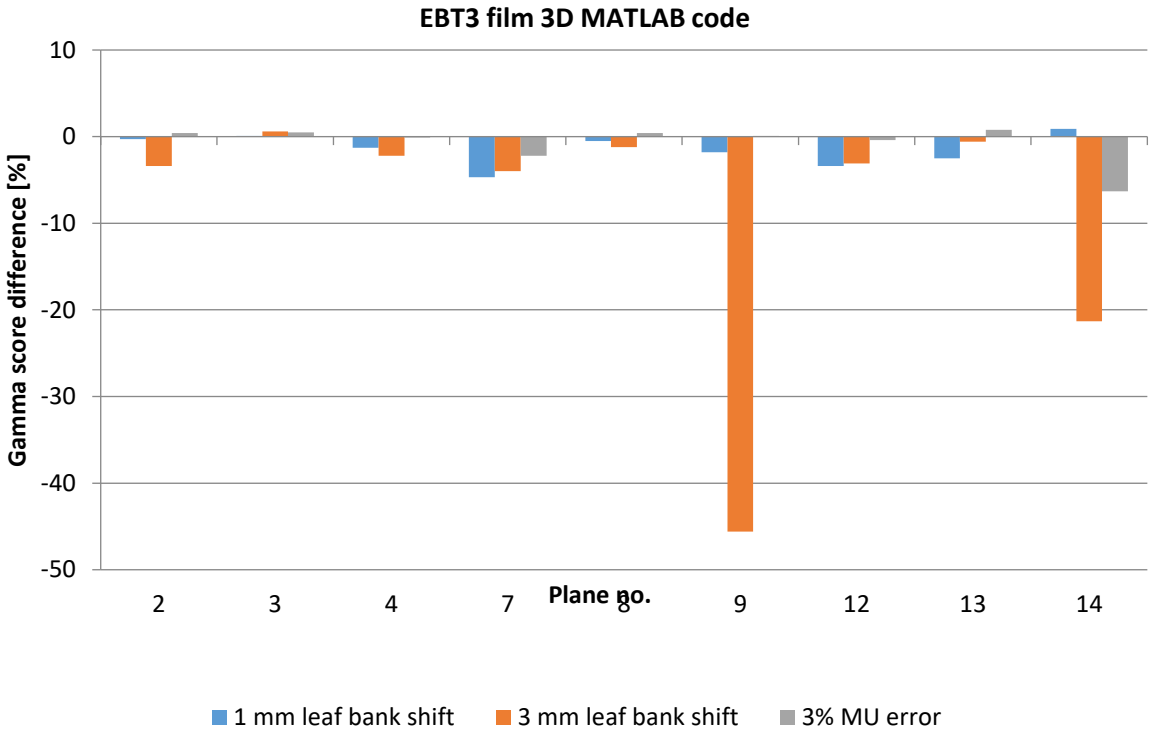


Figure 4-27: Gamma score difference in each plane between the error-induced VMAT plans and the original VMAT plan of Patient no. 9, calculated as error-induced minus original value, estimated with our 3D MATLAB code based on the pseudo-3D film method.

Table 4-29 to Table 4-31 show gamma statistics for the error-induced plans compared to the error-free original clinical plan. The differences in minimum gamma index, maximum gamma index and mean gamma index were calculated as error-free plan value minus error-induced plan value, the differences should normally be below zero. Positive values have already been explained in the text (film dosimetry as implemeted in this work cannot be relied on for very low doses; there is some known variability due to uncertainties in film dosimetry). Figure 4-28 shows the values of mean gamma index differences form Table 4-29 to Table 4-31 graphically to guide the eye.

Table 4-29: Minimum, maximum and mean gamma indices from 2D MATLAB code for the error-induced plan of Patient no. 9 (1 mm MLC leaf bank shift) and values from the original clinical plan for comparison. Differences were calculated as original minus error-induced plan.

Plane no.	Original clinical plan			Error-induced plan			Difference		
	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}
1	0.00	1.64	0.25	0.00	6.24	1.81	0.00	-4.60	-1.56
2	0.00	1.54	0.32	0.00	1.69	0.28	0.00	-0.14	0.04
3	0.00	1.48	0.31	0.00	1.30	0.35	0.00	0.18	-0.04
4	0.00	1.23	0.32	0.00	1.81	0.30	0.00	-0.59	0.03
5	0.00	7.43	2.55	0.00	4.94	1.25	0.00	2.50	1.30
6	0.00	1.59	0.33	0.00	2.40	0.34	0.00	-0.80	-0.01
7	0.00	1.29	0.29	0.00	2.71	0.48	0.00	-1.42	-0.19
8	0.00	1.50	0.39	0.00	3.42	0.37	0.00	-1.91	0.02
9	0.00	1.19	0.41	0.00	2.81	0.38	0.00	-1.62	0.03
10	0.00	2.25	0.37	0.00	2.90	0.58	0.00	-0.66	-0.22
11	0.00	2.49	0.34	0.00	3.94	0.73	0.00	-1.45	-0.39
12	0.00	0.91	0.25	0.00	3.06	0.38	0.00	-2.15	-0.13
13	0.00	1.28	0.43	0.00	2.24	0.43	0.00	-0.96	0.00
14	0.00	3.44	0.40	0.00	2.02	0.35	0.00	1.42	0.05
15	0.00	1.70	0.40	0.00	2.10	0.55	0.00	-0.40	-0.15

Table 4-30: Minimum, maximum and mean gamma indices from 2D MATLAB code for the error-induced plan of Patient no. 9 (3 mm MLC leaf bank shift) and values from the original clinical plan for comparison. Differences were calculated as original minus error-induced plan.

Plane no.	Original clinical plan			Error-induced plan			Difference		
	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}
1	0.00	1.64	0.25	0.00	4.84	1.39	0.00	-3.20	-1.14
2	0.00	1.54	0.32	0.00	2.20	0.37	0.00	-0.65	-0.06
3	0.00	1.48	0.31	0.00	1.11	0.27	0.00	0.37	0.04
4	0.00	1.23	0.32	0.00	2.31	0.35	0.00	-1.08	-0.03
5	0.00	7.43	2.55	0.00	3.69	1.29	0.00	3.74	1.27
6	0.00	1.59	0.33	0.00	3.94	0.47	0.00	-2.35	-0.15
7	0.00	1.29	0.29	0.00	3.00	0.36	0.00	-1.70	-0.07
8	0.00	1.50	0.39	0.00	3.38	0.36	0.00	-1.87	0.04
9	0.00	1.19	0.41	0.00	9.88	2.66	0.00	-8.69	-2.26
10	0.00	2.25	0.37	0.00	13.96	3.17	0.00	-11.72	-2.80
11	0.00	2.49	0.34	0.00	4.66	0.72	0.00	-2.18	-0.38
12	0.00	0.91	0.25	0.00	4.28	0.35	0.00	-3.37	-0.10
13	0.00	1.28	0.43	0.00	4.18	0.39	0.00	-2.91	0.03
14	0.00	3.44	0.40	0.00	6.41	0.72	0.00	-2.97	-0.32
15	0.00	1.70	0.40	0.00	4.37	0.53	0.00	-2.67	-0.13

Table 4-31: Minimum, maximum and mean gamma indices from 2D MATLAB code for the error-induced plan of Patient no. 9 (3% MU error) and values from the original clinical plan for comparison. Differences were calculated as original minus error-induced plan.

Plane no.	Original clinical plan			Error-induced plan			Difference		
	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}	γ_{min}	γ_{max}	γ_{mean}
1	0.00	1.64	0.25	0.00	5.40	1.32	0.00	-3.76	-1.07
2	0.00	1.54	0.32	0.00	1.48	0.23	0.00	0.06	0.09
3	0.00	1.48	0.31	0.00	1.49	0.25	0.00	-0.01	0.06
4	0.00	1.23	0.32	0.00	1.63	0.27	0.00	-0.40	0.05
5	0.00	7.43	2.55	0.00	3.51	0.76	0.00	3.92	1.79
6	0.00	1.59	0.33	0.00	2.30	0.42	0.00	-0.71	-0.09
7	0.00	1.29	0.29	0.00	3.24	0.36	0.00	-1.95	-0.07
8	0.00	1.50	0.39	0.00	1.67	0.39	0.00	-0.17	0.00
9	0.00	1.19	0.41	0.00	2.44	0.35	0.00	-1.25	0.06
10	0.00	2.25	0.37	0.00	3.39	0.59	0.00	-1.14	-0.22
11	0.00	2.49	0.34	0.00	3.76	0.67	0.00	-1.27	-0.33
12	0.00	0.91	0.25	0.00	2.32	0.20	0.00	-1.41	0.05
13	0.00	1.28	0.43	0.00	2.82	0.29	0.00	-1.54	0.14
14	0.00	3.44	0.40	0.00	2.95	0.44	0.00	0.49	-0.04
15	0.00	1.70	0.40	0.00	3.50	0.40	0.00	-1.80	0.00

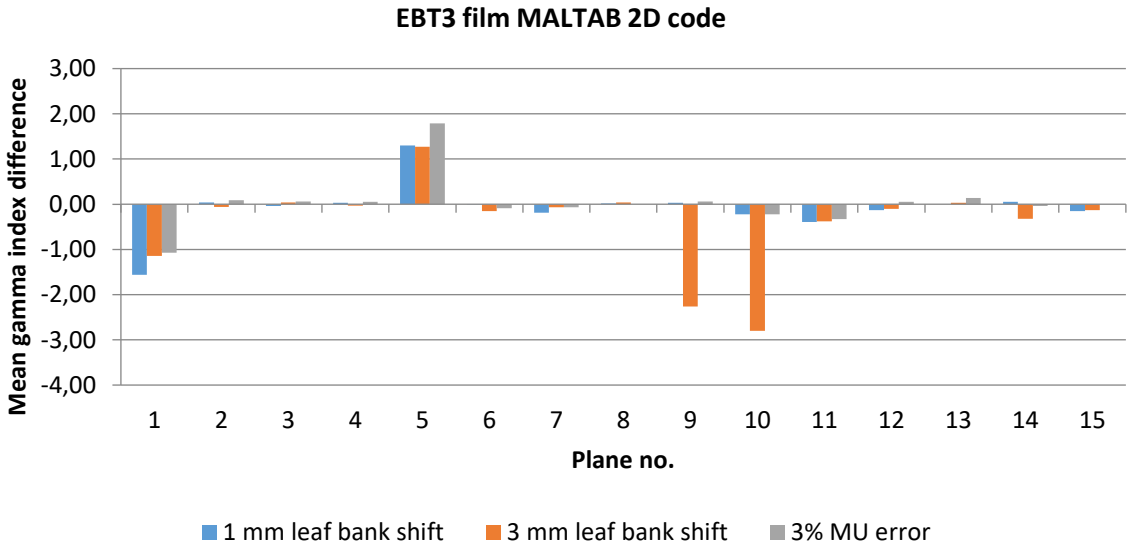


Figure 4-28: Difference in mean gamma index between the original clinical VMAT plan and the error-induced VMAT plans for Patient no. 9 (original minus error-induced value in each plane) as calcualted with our 2D MATLAB code.

Figure 4-29 to Figure 4-31 show the distribution of gamma indices in selected planes for the three types of introduced error in comparison to the gamma index distribution in the same plane in the original clinical plan for illustration (as calculated with our 3D MATLAB code).

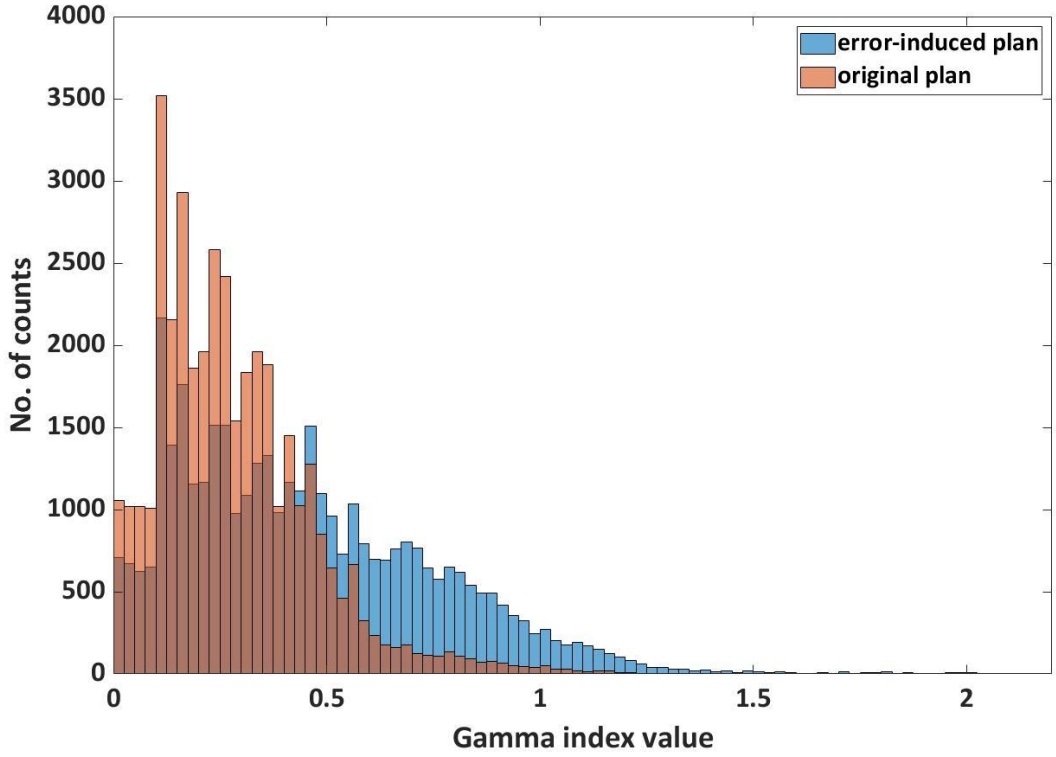


Figure 4-29: Gamma histograms for plane no. 7 for the error-induced VMAT plan with 1 mm leaf bank shift (in blue) and for the original clinical VMAT plan of Patient no. 9 (in orange) as calculated with our 3D MATLAB code.

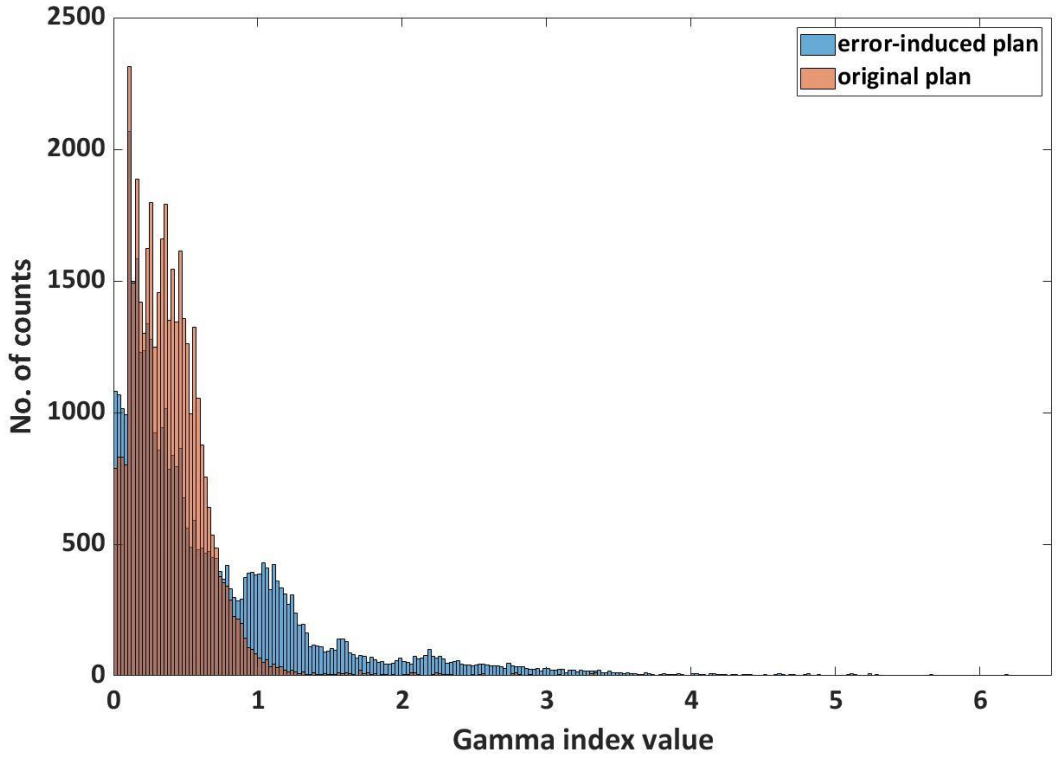


Figure 4-30: Gamma histograms for plane no. 14 for the error-induced VMAT plan with 3 mm leaf bank shift (in blue) and for the original clinical VMAT plan of Patient no. 9 (in orange) as calculated with our 3D MATLAB code.

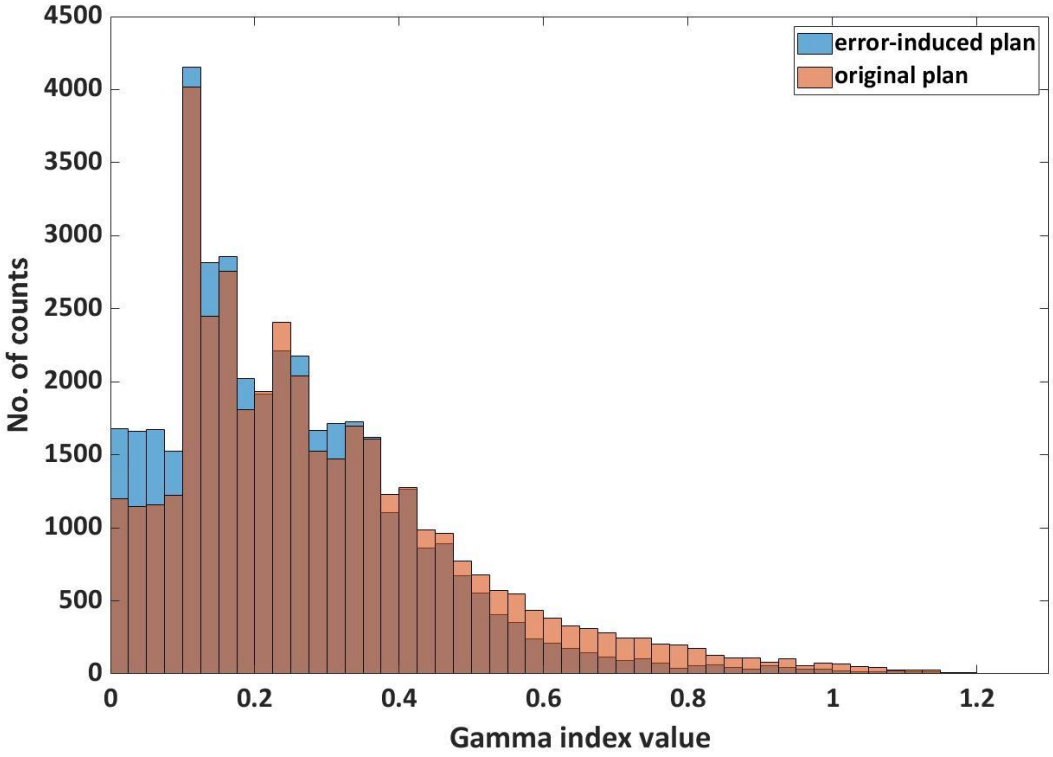


Figure 4-31: Gamma histograms for plane no. 3 for the error-induced VMAT plan with 3% MU error (in blue) and for the original clinical VMAT plan of Patient no. 9 (in orange) as calculated with our 3D MATLAB code.

Additionally, Table 4-32 shows results for 2D plane-by-plane evaluation of the error-induced planes, as well as for the original clinical plan, using 2%/2 mm criteria and the OCTAVIUS 4D pseudo-3D method. Other parameters of gamma analysis were kept the same (threshold 10%, ROI was the whole active area of the detector, normalization to the maximum dose in predicted distribution). For the pseudo-3D film method, 2%/2 mm criteria would not give reliable results because the uncertainty of relative dosimetry with film as established in this work is actually larger than the 2% value (k = 2). Therefore, results are not shown. There is no clinically established tolerance value for gamma score using 2%/2 mm criteria, but with the OCTAVIUS system, all planes were above 95% for the original clinical plan. The 1 mm leaf bank shift showed almost no difference in gamma scores (note that the 2 mm criterion is still larger than the introduced error), so this type of error would remain undetected even with the 2%/2 mm criteria. The 3 mm leaf bank shift caused a large drop in gamma scores compared to the original error-free plan in most planes. The 3% MU error also caused a large drop in gamma scores in many of the tested planes. In this case, more planes would have to be evaluated to reveal the error. These two types of error would thus be revealed with the OCTAVIUS system and 2D plane-by-plane evaluation using 2%/2 mm criteria. Planes no. 1 and 5 that lay in the low dose area beyond the edge of the PTV (transverse planes 4 cm away from the isocentre) had a 100% gamma scores in all cases except for the 3 mm leaf error and plane no. 1.

Table 4-33 shows the psuedo-3D gamma score obtained with the OCTAVIUS system and 2%/2 mm criteria. Again, the 1 mm leaf bank error would not be revealed with the pseudo-3D gamma score, while the other two types of error would be revealed in this case.

Table 4-32: 2D plane-by-plane gamma scores obtained with the OCTAVIUS 4D pseudo-3D method with 2%/2mm criteria for the original clinical VMAT plan of Patient no. 9 and for three error-induced plans. Differences in gamma score are calculated as error-induced minus original value in each plane.

	2D gamma score [%] *2%/2 mm criteria*						
	OCTAVIUS 4D pseudo-3D method						
Plane no.	**Original clinical plan**	**1 mm leaf bank shift**	**Difference**	**3 mm leaf bank shift**	**Difference**	**3% MU error**	**Difference**
1	100	100	0	75.5	-24.5	100	0
2	100	99.8	-0.2	87.1	-12.9	97.2	-2.8
3	99.6	99.6	0	84.9	-14.7	91.9	-7.7
4	99.7	99.6	-0.1	85.5	-14.2	90.3	-9.4
5	100	100	0	100	0	100	0
6	99.3	99.6	0.3	86.3	-13	92.4	-6.9
7	98.3	98.5	0.2	77.2	-21.1	88.0	-10.3
8	98.2	98.0	-0.2	72.3	-25.9	89.0	-9.2
9	98.0	97.3	-0.7	74.5	-23.5	86.1	-11.9
10	99.7	98.2	-1.5	93.8	-5.9	83.8	-15.9
11	100	100	0	95.4	-4.6	92.2	-7.8
12	99.8	99.4	-0.4	75.5	-24.3	87.1	-12.7
13	96.5	96.7	0.2	70.6	-25.9	85.1	-11.4
14	99.7	98.9	-0.8	71.1	-28.6	95.0	-4.7
15	100	99.9	-0.1	76.0	-24	97.8	-2.2
	15 planes passing 95% tolerance	15 planes passing 95% tolerance	Average difference: -0.2%	2 planes passing 95% tolerance	Average difference: -17.5%	5 planes passing 95% tolerance	Average difference: -7.5%

Table 4-33: The overall pseudo-3D gamma score obtained with the OCTAVIUS system for the original clinical VMAT plan of Patient no. 9 and for three error-induced plans with 2%/2 mm criteria.

Pseudo-3D gamma score [%] *2%/2mm criteria*	**OCTAVIUS 1500 detector + OCTAVIUS 4D**
Original VMAT plan	99.6
1 mm leaf bank shift	99.5
3 mm leaf bank shift	85.5
3% MU error	94.6

4.2.5 Clinical example of IMRT commissioning with film

The film dosimetry data from this study for Patients no. 7 and 8 (the IMRT sliding window 18 MV prostate plan and the IMRT sliding window 6 MV glioblastoma plan) were also used for clinical commissioning of the IMRT technique after installation of two Varian TrueBeam accelerators in the Thomayer Hospital in Prague. Compared to the previous linac (Siemens Artiste), the IMRT technique was different (sliding window compared to step-and-shoot), the treatment planning system was different (Eclipse compared to XiO, AAA algorithm compared to Superposition Convolution algorithm), the record-and-verify system was different (ARIA compared to Mosaiq), the detectors and software for patient-specific plan verification were different (PTW OCTAVIUS 1500 detector compared to PTW Seven29, newer version of VeriSoft and also Varian Portal Dosimetry). The 6 MV beam model was based on Golden Data provided by Varian while the 18 MV beam model was based on measured data in the Thomayer Hospital because Varian does not provide data for this energy. During commissioning, it turned out that the 6 MV beam model agreed almost perfectly with all measured data but it was not the case with the 18 MV beam model, even though everything was within tolerances. Moreover, while the OCTAVIUS 1500 detector with VeriSoft v. 8.0 showed almost perfect agreement for field-by-field verification for all tested patients and plans (100% gamma score with 3%/3 mm criteria and gamma score above 95% for 2%/2 mm criteria), the Portal Dosimetry tool from Varian very often showed gamma score values lower than 95% or even 90% with the 3%/3 mm criteria, worse results were seen for the 18 MV beam.

Table 4-35 shows results of field-by-field verification with the Portal Dosimetry tool and 3%/3 mm criteria as well as 2%/2 mm criteria for Patient no. 7 (prostate treated with IMRT and 18 MV beam). Table 4-34 shows field-by-field results obtained for the same plan with the OCTAVIUS system, also applying 3%/3 mm and 2%/2mm criteria. The criteria of gamma analysis for Portal Dosimetry and for the OCTAVIUS 1500 detector with VeriSoft v. 8.0 were the same where possible (normalization to the maximum dose in predicted matrix, 10% threshold, evaluated ROI was the whole active area of the detector – 27 x 27 cm for OCTAVIUS 1500 detector and 40 x 40 cm for EPID). The difference between the OCTAVIUS-based verification and Portal Dosimetry-based verification is, among others, in the algorithm that is used to recalculate the treatment plans in order to prepare a verification plan. In the case of the OCTAVIUS 1500 detector, the plan is recalculated into the phantom geometry (in this case an RW3 phantom and planar measurement with the detector in isocentre in reference depth) with the same dose algorithm that is used to calculate the dose in patient (in this case the AAA algorithm). In the case of Portal Dosimetry, the TPS uses the PDIP (Portal Dose Image Prediction) algorithm that is configured independently of the AAA algorithm (and uses different input data, even though it is partly based on the AAA beam model). Because there was no previously used verification tool that could be relied on and all the verification equipment was commissioned together with the new linacs, there was a need for an independent, previously established method. The pseudo-3D film dosimetry method established in this work was used for this purpose.

Table 4-34: Field-by-field verification results for Patient no. 7 obtained with the OCTAVIUS 1500 detector in an RW3 phantom evaluated in VeriSoft v. 8.0.

Field no.		1	2	3	4	5	6	7
Gamma score [%]	3%/3 mm	100	100	100	100	100	100	100
	2%/2 mm	100	100	96.2	98.7	95.5	97.1	100

Table 4-35: Field-by-field verification results obtained with the Portal Dosimetry tool and the old PDIP model for Patient no. 7.

Field no.		1	2	3	4	5	6	7
Gamma score [%]	3%/3 mm	95.9	97.8	94.4	95.7	96.3	92.7	98.2
	2%/2 mm	87.3	95.0	80.8	84.5	83.9	79.9	95.1
Maximum gamma index (with 3%/3 mm criteria)		2.73	3.50	2.69	2.35	2.46	3.73	2.88
Mean gamma index (with 3%/3 mm criteria)		0.34	0.22	0.42	0.39	0.39	0.44	0.22

Results from the pseudo-3D film dosimetry method shown in Table 4-7 in chapter 4.2.1 for Patients 7 and 8 were, at a first glance, similar to those previously seen with the film dosimetry method for Siemens Artiste. This was confirmed by the statistical chi square test of independence. Two categories of linacs were considered – Siemens Artiste (both the 6 MV and 18 MV beam) and Varian TrueBeam (both the 6 MV and 18 MV beam but only IMRT technique). The other categorical variable were the usual gamma score bins: gamma score < 90%, gamma score between 90% and 94%, gamma score between 94% and 97% and gamma score > 97%. The null hypothesis was that gamma scores obtained with the pseudo-3D dosimetry method for the IMRT technique are independent of the type of accelerator. The contingency table is given in Table 4-36 and the p-value was 0.0821, so the null hypothesis could not be rejected at a 5% significance level. Even though there is little data for the Varian TrueBeam category and gamma scores between 90% and 97%, all the so-called expected values were larger than 5. (Note that for both patients, the lower gamma scores in planes 4 cm distant from the isocentre could be explained by the fact that these planes lay beyond the edge of the PTV in area of very small dose, as already discussed.) It is, of course, important to investigate each treatment case individually as well. This test gives just a general idea that there is probably no reason to think that one linac (in our case the new TrueBeam) performs worse than the other.

Table 4-36: Contingency table with frequencies for the chi square test of independence comparing Siemens Artiste and Varian TrueBeam IMRT patients.

Gamma score category (GS)	**Siemens Artiste**	**Varian TrueBeam**	**Row totals**
GS < 90%	30	13	43
90% <GS ≤ 94%	18	2	20
94% < GS ≤ 97%	23	4	27
GS > 97%	19	11	30
Column totals	90	30	**Grand total:** 120

This result, supported by the excellent results of field-by-field verification with OCTAVUS 1500 and VeriSoft v. 8.0, suggested that the beam model was probably correct and maybe the Portal Dosimetry tool might not have predicted correct gamma scores. There might have been some issues with the EPID aSi-1200 detector or the PDIP model might not have been configured optimally. It was expected, also based on previous experience with Varian Portal Dosimetry, that the Portal Dosimetry tool should give higher gamma scores than what can be seen in Table 4-35 for Patient no. 7 as an example.

The PDIP model was reconfigured with different data provided by Varian. The gamma scores improved substantially. Table 4-37 shows field-by-field gamma scores for Patient no. 7 after improvement of the PDIP model. Table 4-38 shows filed-by-field gamma scores for Patient no. 8 (the glioblastoma case treated with IMRT and 6 MV beam) obtained with the Portal Dosimetry tool after improvement of the PDIP model. (Before the improvement of the model, this patient was only measured with the pseudo-3D film dosimetry method and with the OCTAVIUS 1500 detector in a field-by-field manner. Evaluation of OCTAVIUS measurements in VeriSoft v. 8.0 gave 100% gamma scores for all fields with 3%/3 mm criteria. There is no data available for this patient and the old PDIP model).

Table 4-37: Field-by-field verification results obtained with the Portal Dosimetry tool and the new PDIP model for Patient no. 7.

Field no.		1	2	3	4	5	6	7
Gamma score [%]	3%/3 mm	98.5	98.6	98.4	98.6	99.4	97.5	99.0
	2%/2 mm	94.1	97.1	92.9	94.3	95.3	91.2	97.3
Maximum gamma index (with 3%/3 mm criteria)		2.16	2.97	2.41	1.97	1.87	3.24	2.31
Mean gamma index (with 3%/3 mm criteria)		0.25	0.19	0.29	0.27	0.26	0.30	0.19

Table 4-38: Field-by-field verification results obtained with the Portal Dosimetry tool and the new PDIP model for Patient no. 8.

Field no.		1	2	3	4	5
Gamma score [%]	3%/3 mm	99.8	99.9	99.8	99.9	99.7
	2%/2 mm	95.8	98.8	99.2	99.0	98.6
Maximum gamma index (with 3%/3 mm criteria)		2.04	2.15	1.76	1.06	1.51
Mean gamma index (with 3%/3 mm criteria)		0.24	0.14	0.18	0.18	0.13

4.2.6 Summary

The clinically used field-by-field IMRT dosimetry with PTW Seven29 as implemented at the Thomayer Hospital for Siemens Artiste gives generally lower gamma scores than EBT3 film field-by-field method. At the same time, gamma scores obtained with the PTW Seven29 array are less variable than with film. In literature, works can be found that prove the opposite (gamma scores with film are lower than with electronic devices) [106, 119]. Their results are perhaps obtained with sub-optimal methodology for film dosimetry.

An important finding is that field-by-field verification gives higher gamma scores than if 2D planes in different directions in the 3D space are evaluated when the global plan dose distribution is measured in a pseudo-3D manner (Table 4-14). Measurements of the global plan dose distribution in only one plane might also hide potential errors in accelerator performance and/or treatment planning software (see e.g.Table 4-26) – this is in accordance with literature, e.g. [135].

MLC positioning errors introduced into 3 fields of two clinical IMRT treatment plans (generated in XiO for Siemens Artiste) were not revealed with field-by-field EBT3 film verification and 3%/3 mm criteria for one of the two plans. However, pseudo-3D gamma score calculated with our MATLAB code and local normalization (as defined above) would probably reveal the error. This finding is based on insufficient amount of data, on the other hand, in clinical applications we need to decide adequately for each single patient plan. For clinical plans without introduced erorrs, pseudo-3D gamma scores with local normalization (as defined in this work) seem to give similar pass/fail information as the clinically used field-by-field 2D gamma analysis method with the PTW array Seven29.

For a VMAT plan, the pseudo-3D film dosimetry method was able to detect 1 mm and 3 mm MLC leaf positioning errors (X1 leaf bank shift in each control point in one of two arcs) as well as 3% MU error (in each arc) with 3%/3 mm criteria and 95% tolerance level when more planes were evaluated. The psuedo-3D gamma score alone would not detect the 1 mm positioning and 3% MU errors with 3%/3 mm criteria and 95% tolerance level. The 2%/2 mm criteria could not be applied using the film dosimetry method as established in our work because the 2% criterion is below the uncertainty of relative film dosimetry (k = 2).

The pseudo-3D OCTAVIUS 4D method did not detect the 1 mm positioning error at all. It did not detect the 3% MU error with 3%/3 mm criteria and 95% tolernace level. It was able to detect the 3 mm MLC positioning error, when more planes were evaualted, with 3%/3 mm criteria and 95% tolernace level. The pseudo-3D gamma score alone did not detect any of the errors with 3%/3 mm and 95% tolerance level. The psuedo-3D gamma score from the OCTAVIUS system did detect the 3 mm positioning error and 3% MU error with 2%/2 mm criteria and 95% tolerance level. It also detected these errors with the 2%/2 mm criteria and 95% tolerance when more planes were evaluated in 2D (plane-by-plane).

Generally speaking, 2D gamma analysis (field-by-field or plane-by-plane) can give false positive results (plan fails tolerance even though it should not) because it is likely to find a more suitable dose point with 3D gamma analysis in 3D space. But this depends on the particular properties of measured data. Field-by-field approach can give false negative results (plan passes tolerance even though it should not) because it does not take into account potential errors comming from rotational plan delivery or incorrect dose calculation of the TPS in 3D. On the other hand, field-by-field approach can also give false positive results (plan fails tolerance even though it should pass) because small errors can cancel out in the 3D space when the whole treatment plan is delivered, so they are not relevant for the patient. Measurements in different directions in the 3D space (transverse, sagittal, coronal) can give more information about the dose distribution and its agreement with predictions. Each direction gives different information, actually. For example, the transverse plane is at all times parallel to the beam and each transverse plane is associated with potential leaf positioning errors coming from only one leaf pair. On the other hand, transverse plane is the one where patient CT data are available with good precision in the TPS. Results might be affected by detector directional dependence, the direction of MLC leaf travel and gantry rotation, as well as TPS algorithm for dose calculation on CT data in the 3D grid.

From our findings and taking into account the above mentioned issues, it seems that it is more appropriate to use a pseudo-3D method, at least when new complex techniques are introduced in the clinic. Our pseudo-3D film method was used for example during commissioning of two TrueBeam

linacs and dosimetric equipment for patient plan verification. The same conclusion is drawn in several other works using different approaches, above all in a monography based on recent findings [136]; this topic is very often addressed in literature. But the author of the dissertation has not found a similar study in literature that would compare radiochromic film dosimetry in 2D and 3D, where differences in gamma results are caused solely by considering the extra dimension (and not by inherent dosimetric system differences). The overall pseudo-3D gamma score should not serve as the routine pass/fail decision criterion for patient-specific verification, unless the gamma score threshold for passing and the input criteria of gamma analysis are very carefully estabslished for a given dosimetric system. It is recommended, based on our findings, to analyse individual 2D planes in the 3D space at least before appropriate action limits are established for IMRT verification in the clinic. Field-by-field verification as the only applied method is not recommended. Gamma analysis is a complex tool dependent on many input parameters and field-by-field verification with 2D gamma analysis might not give the right answers.

Our work also points out several observations regarding gamma analysis specifically with radiochromic film dosimetry. Relative gamma analysis applied to dose distributions measured with film is sensitive to the selection of a specific dose normalization point, such as the maximum dose, because noise remains on film even after smoothing. Normalization strategies should be explored and the user should be aware of their significance when using film dosimetry for verification of IMRT plans. Preferably, the normalization method should be to a percentile, such as the 75th percentile as used in this work. If the 3D verification method as described in this work is implemented, dose distributions that belong to one data set (for example transverse, sagittal and coronal dose planes for one treatment plan) should always be measured in one session and with the same batch of films in order to minimize uncertainties in the process. Film dosimetry as implemented in this work is not appropriate for measuring planes in the out-of-field area, i.e. planes with doses below 0.5 Gy. The film dosimetry method implemented in this work cannot be used with dose difference criterion tighter than 3% because the relative uncertainty of relative dose estimation with film is 1.85% (k = 1) for a dose of 2 Gy (and rises for lower doses).

4.3 Overall discussion

Detailed discussion of results, their meaning and comparison with literature was already done in chapters 4.1 and 4.2. Chapter 4.3 aims at bringing all results together and explaining the novelty of our work and contribution to the field of IMRT plan verification.

The strategy of our work was to find a dosimetric method that could be applied both in 2D and in 3D, so that our results of 2D and 3D IMRT plan verification would directly be comparable to each other. We also aimed at using equipment that is usually available at hospitals, so that any site could apply our findings or start using our method in routine practice. The chosen method was therefore radiochromic film dosimetry in a simple water equivalent slab phantom, because radiochromic film is used for a number of applications [11-18].

The first step was to establish a methodology for film dosimetry that could easily be implementable at our working site, would suit our experimental goals and at the same time would be simple enough to apply in routine clinical practice. Under our particular experimental conditions and for 2 Gy on film, we achieved an uncertainty of 1.85% (k = 1) for relative dose determination with EBT3 radiochromic film (without covariance terms that would further decrease the uncertainty). The

relative uncertainty of absolute dose determination with film for a dose of 2 Gy was 2.28% (k = 1), where a large contribution to this value was the uncertainty of absolute dose determination with an ionization chamber (1.33% for k = 1). For relative gamma analysis, the uncertainty of relative dose determination is relevant. This is well within uncertainties published elsewhere [25, 47, 104]. The only drawback that would prevent our method from being used directly at hospitals is the fact that films were processed in MATLAB. This software is not normally available at hospitals. On the other hand, using MATLAB enabled automation of data evaluation which saved a lot of time. Similarly, this could be performed in GNU Octave or Python, which are available to hospitals.

The second step was benchmarking our method against currently available and clincially used method. In our experimental conditions, a rather older version of 2D inoization chamber array was available (PTW Seven29) and in use for routine clinical IMRT verification for the Siemens Artiste linac. Also a rather older version of VeriSoft (v. 3.1) was used, where the user cannot adjust many parameters and therefore it is not very convenient for experimental purposes. From another point of view, however, this is actually an advantage. Most published papers focus on newly released dosimetric systems at the time. In older papers [105, 107], it was the PTW Seven29 and either radiographic EDR2 film or radiochromic EBT2 film. Other recently published works [137, 138] focus on OCTAVIUS 1500 or OCTAVIUS 1000 SRS detector from PTW compared to EBT3 radiochromic film. Thus, our work brings quite unique results combining EBT3 film and PTW Seven29, which is still in use at hospitals, while EBT2 or EDR2 film is not (or very rarely). Our results might be interesting for clinics using the same combination of equipment and were published in our paper [139]. Moreover, our finding is in contrast with previous published works [106, 108] on a similar topic. We proved that radiochromic EBT3 film can be an excellent tool for IMRT verification, unlike the mentioned authors. The PTW Seven29 array was replaced with the OCTAVIUS 1500 detector and a newer version of VeriSoft (v. 8.0) at the Thomayer Hospital and the rotating cylindrical phantom OCTAVIUS became available together with the Varian TrueBeam linacs.

The final step was to perform the actual 3D measurements and compare them to 2D results. This was done for both clinical plans and error-induced plans. The question was whether 3D gamma analysis is more appropriate than 2D gamma analysis, whether 3D global plan dose verification is needed instead of 2D field-by-field verification to reveal relevant errors in IMRT clincial plans and also, whether it is sufficient to measure just onle single plane in the 3D space when the whole treatment plan is delivered to the dosimetric system. A number of previous works has been published on this topic (cited throughout this study). But it needs to be pointed out that none of the previously published works that the author is aware of made a direct comparison of 2D and 3D data. None of the works used the very same measured data set for 2D and 3D evaluation and none of the works used at the same time the same software code where all parameters are user-adjustable. This is not possible with commercially available equipment. So this is a novelty of our work.

In comparison to previously made attempts of using radiochromic or radiographic film for pseudo-3D verification of IMRT and VMAT techniques, as mentioned in chapter 2.3, the advantage of our method is above all the type of phantom used. All of the works [17, 83-87] used a custom-made cylindrical phantom with specific gaps designed for film placement, usually in a spiral manner. Our method uses an IMRT cube phantom which is already available at most clinics. For the previous attempts mentioned in literature, additional steps are sometimes required, such as Monte Carlo simulation [84] or modification of the gamma analysis concept [85]. Some of the works are not even

meant for 3D gamma evaluation [17] or cover a very small portion of the 3D space [83]. In case of suspicious TPS dose calculation results in the 3D space, as shown in Figure 3-12, where calculated dose profiles differ in a sagittal and in a transverse plane, our geometry of film placement into the phantom can be more helpful compared to most of the previous works [17, 83-87] where film is arranged in a spiral manner. Because treatment planning systems and their algorithms are more often based on cartesian coordinates than on polar or any other type of coordinates. Incorrect TPS dose calculation is thus one of the problems that could be revealed with our film method.

The answer to the above mentioned questions can be summarized as follows: Field-by-field verificatin of IMRT treatment plans does not give the user enough information to accept or reject the treatment plan (similar conclusion to [140] using MapCHECK by Sun Nuclear Corporation, Melbourne, FL, USA). The whole plan should be measured in a 3D manner, i.e. several 2D planes in several directions should be evaluated, additionally a 3D gamma evaluation can be performed (similar conclusion to [141] using Presage and film). A simple tolerance limit for the 3D gamma score value can be applied as a pass/fail criterion only when the level of the limit is chosen carefully and inspection of individual planes in the 3D space is recommended. This is an ideal scenario, probably not applicable in routine practice for all treatment plans. But the author believes that this should be done at least when a new treatment technique or a new TPS algorithm is set up at the hospital.

Comparison of our results to literature is problematic in that a range of conclusions can be found, from finding no correlation between 2D and 3D gamma analysis [8, 142], to finding some [9, 143] or strong correlation [7, 144] or even rejecting clinical gamma analysis as a suitable method to predict relevant dose errors [94, 140]. Conclusions usually depend on methods and equipment used. As reported by Hussein et al. [144] in a recent review and in accordance with the demonstrations presented here, there are many differences between the 2D and 3D gamma calculation. Therefore, it is necessary to report a number of parameters when gamma analysis is addressed in published literature and individual papers are not usually comparable. Pulliam et al. [95] have stressed out the importance of having the same system for 2D and 3D comparisons because otherwise the comparison is affected by the systems‘ inherent properties. When using artificial data without measurement (TPS and Monte Carlo calculated dose distributions), these authors conclude that 3D gamma analysis always gives higher pass rates than 2D gamma analysis. But this is only natural, it is the mathematical property of gamma analysis. In real world, there are usually measured data and their format also affects the results of gamma analysis. Some authors, however, report the same conclusions on measured data [141, 145].

Our pseudo-3D method was used at the Thomayer Hospital in Prague when IMRT step-and-shoot technique was commissioned for Siemens Artiste, together with the PTW Seven29 for 2D measurements. It was used again for benchmarking a new detector (PTW OCTAVIUS 1500) when it was purchased for IMRT plan verification. It was also used when new TrueBeam linacs were commissioned (capable of IMAT) and the OCTAVIUS 4D system was purchased. It proved to be a useful benchmarking method in clinical practice. It was also used several times for individual plan verification when some sort of problem occured with the commercial verification equipment in use.

Using the Siemens Artiste linear accelerator with a step-and-shoot IMRT mode together with the treatment planning system XiO 4.80 makes this study one of the rare examples of this experimental setup. Siemens Artiste was the last model of Siemens linear accelerators, released

before the year 2010, and the number of these linac models in the world is relatively low, but they are still in use (which can be declared by ongoing research – e.g. [146-148] and many more). Therefore, the newer EBT3 film is rarely used for this type of linac in the literature. For example, film dosimetry for IMRT techniques can be sensitive to inter-leaf leakage specific to linac models (small non-zero doses on film). Siemens Artiste does not have jaws in the X direction, only MLC. Siemens Artiste also has a specific behaviour during the first few monitor units, being very stable and reproducible with its step-and-shoot IMRT technique. Siemens Artiste is capable of delivering even 0.1 MU. A part of segments that are delivered clinically within one treatment plan typically has less than 2-3 MU.

In comparison with commercially available detection systems, our pseudo-3D method has the advantage of very good in-plane resolution (0.35 mm compared to several milimeters for the best electronic devices and comparable to the 0.34 mm resolution of the newly released portal imager aSi-1200 by Varian) and no need of software algorithms for recalculation into 3D. Thus, it does not require commissioning of the system in terms of beam data acquisition and it is not dependent on particular algorithm (the implementation of which is not usually known to the end user in full). Another advantage of our pseudo-3D method is that no additional (and costly) resources need to be purchased by the hospital. The only extra cost is the film material. For hospitals, it might be more convenient to buy another box of film for a relatively low cost several times per year than to buy a very expensive equipment for which a tender must be held.

Regarding the workload, it is similar for our radiochromic film method and commercially available solutions based on electronic detectors. All commercial solutions need some sort of commissioning, while film dosimetry might be already an established method at the hospital. Electronic detectors might require recalibration before each use, while radiochromic film can be calibrated once per batch. On the other hand, commercial solutions require one delivery per treatment plan while our pseudo-3D method requires three deliveries per treatment plan. Some extra time is needed for scanning with the film method, but 15 planes can be scanned in 15-20 minutes. Data processing and gamma analysis are automated. Even a large number of patient plans can be processed at once (unlike in most types of commercial software), the time needed for 3D gamma analysis evaluation depends on the particular hardware available. Optimizing the code in terms of computation time was not the priority of our work. There is a relatively long time needed for film response stabilization before scanning, up to 24 hours (in our case 72 hours – for practical reasons, because it was more convenient to irradiate films on Friday and evaluate them on Monday). However, hospital personnel is not involved in the process during this period. Compared to gel dosimetry, the film method is more suitable and less time consuming for routine use in the clinic. Our pseudo-3D method might be efficient in the following situations: establishing new treatment techniques, commissioning new detection equipment, dosimetry audits and individual treatment plan verification in questionable cases.

The drawback of our study is the low number of patients used. However, for each patient there was a relatively big amount of data obtained, i.e. 135 2D planes and 58 fields available for comparison in some evaluations. Due to the material and time at the linac available for our experiments, we chose to look at the problem from different perspectives rather than to get a statistically significant amount of data for all of the comparisons. Where possible, statistical evaluation was performed (when individual fields or planes from the 3D space were taken into account, the amount of data was

sufficient). But certain parts of the study, especially section 4.1, 4.2.3 and 4.2.4, would require more data to confirm the conclusions. On the other hand, in patient-specific plan verification, each and every individual plan matters because it is delivered to the patient.

An interesting paper reporting seven real-life examples of failure in clinical dosimetry of intensity modulated beams has been published [68]. It shows different types of errors that occurred in different types of linear accelerators and treatment planning systems that could not be properly identified with routinely used modern detection and verification methods. Only a more in-depth analysis with advanced methods, focused on the particular type of problem, could reveal what was wrong. The seven cases included incorrect TPS settings for leaf-end modelling; failure to account for the tongue and groove effect in the TPS; dose gradient errors in the TPS due to volume-averaged dose profiles entered into the beam model; inherent dose gradient errors in the TPS algorithm (for which the vendor was responsible) or TPS underestimation of dose for narrow MLC segments in a complex VMAT plan. These errors led, for example, to a loss of target dose coverage of 5.5 % or to local dose deviations of up to 31.5 %, which is already clinically significant. This paper points out the need for accurate dosimetry and verification in the modern era of radiotherapy.

Gamma analysis, being dependent on so many aspects while hospitals use it to obtain a simple pass/fail decision criterion, has been and will always be topical, regardless the number of publications that are already available. This is supported by the cited publications from recent years up to today (even for example [149] on EPID dosimetry). As Tomas Kron states in his recent review [150], it is surprisingly difficult to assess if clinically relevant errors can be detected by gamma analysis in various scenarios. There is always work to be done.

5 Conclusions

We developed a pseudo-3D dosimetric method for verification of IMRT techniques that is based on radiochromic EBT3 film and a water equivalent slab phantom. Data processing and gamma analysis (2D and 3D) were performed in MATLAB. The advantages of our pseudo-3D method compared to commercial solutions are: better in-plane resolution (0.35 mm compared to several mm for electronic devices), water equivalence of film, no need of recalculation into 3D (the method is based only on measured data in the 3D space), relatively low cost and automation of data processing (a large number of patient plans can be evaluated at once). The disadvantages of the method are: 3D dose distribution is obtained in 3 subsequent irradiations, extra time is needed for stabilization of film response, the film method is not applicable to doses below 0.5 Gy. The method was used clinically at the Thomayer Hospital for IMRT commissioning and commercial detector benchmarking and proved to be a useful tool. It can be readily used at hospitals because radiochromic film and water equivalent slab phantoms are usually available. The expected use of our method is: commissionig of new techniques and detectors, dosimetry audits, patient-specific verification for questionable cases.

Another big advantage of the method is the possibility to compare 2D and 3D evaluation directly on the same measured data and with the same code (where all parameters are user-adjustable). This is not possible with commercial electronic detectors, so comparison of 2D and 3D gamma analysis in literature is affected by the inherent differences of the 2D and 3D systems.

Based on our pseudo-3D method, it seems that field-by-field verification of IMRT techniques does not give enough information to approve or reject a treatment plan. This was proven by a statistical test performed in chapter 4.2.2, which showed that field-by-field gamma scores were higher than plane-by-plane gamma scores when the whole treatment plan was delivered at once with its original gantry angles. This conlcusion is also based on comparison of field-by-field and pseudo-3D verification for error-induced plans in chapter 4.2.3. One of the error-induced plans would be delivered clinically if the desicion was based on field-by-field verification results only. Global plan dose evaluation is thus recommended because of the interplay of individual fields‘ errors.

Evaluation of only one 2D plane in the 3D space is not sufficient because none of the tested clinical or error-induced plans passed the tolerance of a 90% (95%) gamma score in all of the measured planes, as shown in chapters 3.3 and 4.2. Problematic TPS dose calculation might be revealed if 2D dose distributions in different directions (transverse, sagittal, coronal) are checked at the same place in the 3D space. Different plane directions are affected by different sources of error both during dose calculation and plan delivery (patient CT data set, MLC leaf travel, gantry rotation, etc.).

A 3D gamma score value should not be used as a pass/fail decision criterion in routine clinical practice unless tolerance limits are very carefully established for a given irradiation technique and verification system. In chapter 4.2.4, the clinically relevant errors in the VMAT plan were not revealed by the 3D gamma score value alone with clinically used criteria (3%/3 mm, gamma score > 95%) neither with the pseudo-3D film method, nor with the pseudo-3D OCTAVIUS 4D method. Ideally, it is recommended to complement the 3D gamma score value with inspection of several 2D planes in the 3D space, also in terms of gamma index distributions and absolute dose values, because this can give the user additional valuable information. This is at least recommended for commissioning of new irradiation techniques or treatment planning systems.

Detailed guidelines for general clinical practice cannot be given because gamma analysis is dependent on a large number of parameters including the particular system used (type of accelerator, treatment planning system and hardware and software verification tools). However, chapter 3.1 can serve as a practical guide for establishing film dosimetry in the clinic.

Our film dosimetry method is based on and optimized for EBT3 radiochromic film and EPSON Perfection V700 Photo scanner. We use 16 dose points between 0 Gy and 5.5 Gy for calibration and a 4th order polynomial function to convert scanner values to dose values. We use only the red channel because it turned out that the multichannel approach does not improve the results with our particular scanner. We do not subtract blank film prior to irradiation because it turned out that this does not improve the results either. The lateral response artifact and film-to-light distance have little effect on the results obtained with our particular scanner. The uncertainty of absolute dose determination with film according to our protocol is 2.28% (k = 1). The uncertainty is dose-dependent and the reported value is valid for 2 Gy on film. The uncertainty for relative applications (for a dose of 2 Gy on film) is 1.85% (k = 1) and 3.70% (k = 2). Therefore, relative gamma analysis with 3%/3 mm criteria can be performed but a 2% or 1% dose difference criterion cannot be used. Film dosimetry as implemented in this work is not appropriate for measuring planes in the out-of-field area, i.e. planes with doses below 0.5 Gy, because the uncertainty rises unacceptably for low doses. Noise remains on film even after smoothing and normalization to maximum dose therefore should not be used when performing relative gamma analysis with film data.

For Siemens Artiste and step-and-shoot IMRT field-by-field verification, our EBT3 film method gave higher gamma scores than the clinically used method (PTW Seven29 and VeriSoft v. 3.1). This was proven by a statistical test in chapter 4.2.2. For Varian TrueBeam and a clinical VMAT plan, the pseudo-3D film method seemed to give similar results to the clinically used method (PTW OCTAVIUS 4D phantom, OCTAVIUS 1500 array and VeriSoft v. 3.1) but there was not enough data to confirm this by a statistical test. For Varian TrueBeam and error-induced VMAT plans, the pseudo-3D film method revealed all introduced errors with 3%/3 mm criteria and a gamma score limit of 95% if 2D plane-by-plane evaluation was considered. The clinically used pseudo-3D method (PTW OCTAVIUS 4D phantom, OCTAVIUS 1500 array and VeriSoft v. 3.1) only revealed the 3 mm leaf positioning error with these criteria and 2D plane-by-plane approach. Lowering the criteria to 2%/2 mm helped to reveal the 3% MU error but not the 1 mm leaf positioning error.

Author contribution statement

All the work was done by the author.

Acknowledgements

I would like to thank my supervisors for their time, patience and work invested in reading the texts. I thank my colleagues in the Thomayer Hospital in Prague, especially Alžběta Lépesová, who assisted in some of the film irradiation sessions. I would also like to thank Vladimír Dufek from the Czech National Radiation Protection Institute for valuable advice on film dosimetry and to Simona Horová and Jiří Hanuš for their advice on MATLAB scripting language. Acknowledgement also belongs to all secondary school students who participated somehow in my research.

References

1. Low, D. A., Harms, W. B., Mutic, S. and Purdy, J. A. A technique for the quantitative evaluation of dose distributions. *Medical Physics*. 1998, **25**(5), 656-661. DOI: 10.1118/1.598248.

2. Godart, J., Korevaar, E. W., Visser, R., Wauben, D. J. L. and Van't Veld, A. A. Reconstruction of high-resolution 3D dose from matrix measurements: error detection capability of the COMPASS correction kernel method. *Physics in Medicine and Biology*. 2011, **56**(15), 5029-5043. DOI: 10.1088/0031-9155/56/15/023.

3. Heilemann, G., Poppe, B. and Laub, W. On the sensitivity of common gamma-index evaluation methods to MLC misalignments in Rapidarc quality assurance. *Medical Physics*. 2013, **40**(3), 031702. DOI: 10.1118/1.4789580.

4. Stasi, M., Bresciani, S., Miranti, A., Maggio, A., Sapino, V. and Gabriele, P. Pretreatment patient-specific IMRT quality assurance: A correlation study between gamma index and patient clinical dose volume histogram. *Medical Physics*. 2012, **39**(12), 7626. DOI: 10.1118/1.4767763.

5. Zhen, H., Nelms, B. E. and Tomé, W. A. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Medical Physics*. 2011, **38**(10), 5477. DOI: 10.1118/1.3633904.

6. Watanabe Y., Nakaguchi Y. 3D evaluation of 3DVH program using BANG3 polymer gel dosimeter. *Medical Physics*. 2013, **40**(8), 082101. DOI: 10.1118/1.4813301.

7. Zhang, D., Wang, B., Zhang, G., Ma, C. and Deng, X. Comparison of 3D and 2D gamma passing rate criteria for detection sensitivity to IMRT delivery errors. *Journal of Applied Clinical Medical Physics*. 2018, **19**(4), 230-238. DOI: 10.1002/acm2.12389.

8. Kim, J., Choi, C. H., Wu, H. G., Kim, J. H., Kim, K. and Park, J. M. Correlation analysis between 2D and quasi-3D gamma evaluations for both intensity-modulated radiation therapy and volumetric modulated arc therapy. *Oncotarget*. 2017, **8**(3). DOI: 10.18632/oncotarget.

9. Wu, C., Hosier, K. E., Beck, K. E., et al. On using 3D γ-analysis for IMRT and VMAT pretreatment plan QA. *Medical Physics*. 2012, **39**(6Part1), 3051-3059. DOI: 10.1118/1.4711755.

10. Stevens, S., Dvorak, P., Spevacek, V., Pilarova, K., Bray-Parry,M., Gesner J. and Richmond, A. An assessment of a 3D EPID-based dosimetry system using conventional two- and three-dimensional detectors for VMAT. *Physica Medica*. 2018, **45**, 25-34. DOI:10.1016/j.ejmp.2017.11.014.

11. Natanasabapathi, G. and Bisht, R. K. Verification of Gamma Knife extend system based fractionated treatment planning using EBT2 film. *Medical Physics*. 2013, **40**(12), 122104. DOI: 10.1118/1.4832138.

12. Yamauchi, M., Tominaga, T., Nakamura, O., Ueda, R. and Hoshi, M. GAFChromic film dosimetry with a flatbed color scanner for Leksell Gamma Knife therapy. *Medical Physics*. 2004, **31**(5), 1243-1248. DOI: 10.1118/1.1712393.

13. Blanck, O., Masi, L., Damme, M. C., Hildebrandt, G., Dunst, J., Siebert, F. A., Poppinga, D. and Poppe, B. Film-based delivery quality assurance for robotic radiosurgery: Commissioning and validation. *Physica Medica*. 2015, **31**(5), 476-483. DOI: 10.1016/j.ejmp.2015.05.001.

14. Chiu-Tsao, S. T., Duckworth, T. L., Patel, N. S., Pisch, J. and Harrison, L. B. Verification of Ir-192 near source dosimetry using GAFCHROMIC ® film. *Medical Physics*. 2004, **31**(2), 201-207. DOI: 10.1118/1.1637733.

15. Mourtada, F., Soares, C. G., Seltzer, S. M., Bergstrom, P. M., Fernández-Verea, J. M., Asenjo, J. and Lott, S H. Dosimetry characterization of a ^{32}P source wire used for intravascular brachytherapy with automated stepping. *Medical Physics*. 2003, **30**(5), 959-971. DOI: 10.1118/1.1567832.

16. Cheung, T., Butson, M. J. and Yu, P. K. N. Multilayer Gafchromic film detectors for breast skin dose determination in vivo. *Physics in Medicine and Biology*. 2002, **47**(2), N31-N37. DOI: 10.1088/0031-9155/47/2/401.

17. Grams, M. P. and de los Santos, L. E. F. Design and clinical use of a rotational phantom for dosimetric verification of IMRT/VMAT treatments. *Physica Medica*. 2018, **50**, 59-65. DOI: 10.1016/j.ejmp.2018.05.019.

18. Tessonnier, T., Dorenlot, A. and Nomikossoff, N. Patient quality controls with gafchromic EBT3 for tomotherapy HI-ART 2: Analysis of the calibration's condition to achieve an absolute dosimetry. *Physica Medica*. 2013, **29**. DOI: 10.1016/j.ejmp.2013.08.138.

19. Crijns, W., Maes, F., van der Heide, U. A. and van den Heuvel, F. Calibrating page sized Gafchromic EBT3 films, *Medical Physics*. 2013, **40**(1), 012102. DOI: 10.1118/1.4771960.

20. Niroomand-Rad, A., Blackwell, C. R., Coursey, B. M., Gall, K. P., Galvin, J. M., McLaughlin, W. L., Meigooni, A. S., Nath, R., Rodgers, J. E. and Soares, C. G. Radiochromic film dosimetry: Recommendations of AAPM Radiation Therapy Committee Task Group 55. American Association of Physicists in Medicine. *Medical Physics*. 1998, **25**(11), 2093–2115. DOI: 10.1118/1.598407.

21. Devic, S., Seuntjens, J., Sham, E., Podgorsak, E. B., Schmidtlein, C. R., Kirov, A. S. and Soares, C. G. Precise radiochromic film dosimetry using a flat-bed document scanner. *Medical Physics*. 2005, **32**(7), 2245–2253. DOI: 10.1118/1.1929253.

22. Borca, V. C., Pasquino, M., Russo, G., Grosso, P., Cante, D., Sciacero, P., Girelli, G., La Porta, M. R. and Tofanil, S. Dosimetric characterization and use of GAFCHROMIC EBT3 film for IMRT dose verification. *Journal of Applied Clinical Medical Physics*. 2013, **14**(2), 158-171. DOI: 10.1120/jacmp.v14i2.4111.

23. Khachonkham, S., Dreindl, R., Heilemann, G., Lechner, W., Fuchs, H., Paplmans, H., Georg, D. and Kuess, P. *Characteristic of EBT-XD and EBT3 radiochromic film dosimetry for photon and proton beams*. 2018, **63**(6). DOI:10.1088/1361-6560/aab1ee.

24. Miura, H., Ozawa, S., Hosono, F., Sumida, N., Okazue, T., Yamada, K. and Nagata, Y. Gafchromic EBT-XD film: Dosimetry characterization in high-dose, volumetric-modulated arc therapy. *Journal of Applied Clinical Medical Physics*. 2016, **17**(6), 312-322. DOI:10.1120/jacmp.v17i6.6281.

25. Devic, S., Tomic, N. and Lewis, D. Reference radiochromic film dosimetry: review of technical aspects. *Physica Medica*. 2016, **32**(4), 541-56. DOI: 10.1016/j.ejmp.2016.02.008.

26. Huet, C., Moignier, C., Fontaine, J. and Clairand, I. Characterization of the gafchromic EBT3 films for dose distribution measurements in stereotactic radiotherapy. *Radiation Measurements*. 2014, **71**, 364-368. DOI: 10.1016/j.radmeas.2014.05.020.

27. Reinhardt, S., Hillbrand, M., Wilkens, J. J. and Assmann, W. Comparison of Gafchromic EBT2 and EBT3 films for clinical photon and proton beams. *Medical Physics*. 2012, **39**(8), 5257-5262. DOI: 10.1118/1.4737890.

28. Niroomand-Rad, A., Chiu-Tsao, S. T., Grams, M. P., et al. Report of AAPM Task Group 235 Radiochromic Film Dosimetry: An Update to TG-55. *Medical Physics*. 2020, **47**(12), 5986-6025. DOI:10.1002/mp.14497.

29. Santos, T., Ventura, T. and do Carmo Lopes, M. A review on radiochromic film dosimetry for dose verification in high energy photon beams. *Radiation Physics and Chemistry*. 2021, **179**. DOI:10.1016/j.radphyschem.2020.109217.

30. Massillon-JL, G., Chiu-Tsao, S. T., Domingo-Munoz, I. and Chan, M. F. Energy Dependence of the New Gafchromic EBT3 Film: Dose Response Curves for 50 KV, 6 and 15 MV X-Ray Beams. *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology*. 2012, **01**(02), 60-65. DOI: 10.4236/ijmpcero.2012.12008.

31. Micke, A., Lewis, D. F. and Yu, X. Multichannel film dosimetry with nonuniformity correction. *Medical Physics*. 2011, **38**(5), 2523-2534. DOI: 10.1118/1.3576105.

32. Mendez, I., Peterlin, P., Hudej, R., Strojnik, A. and Casar, B. On multichannel film dosimetry with channel-independent perturbations. *Medical Physics*. 2014, **41**(1), 011705. DOI: 10.1118/1.4845095.

33. García-Garduno, O.A., Lárraga-Gutiérrez, J.M., Rodríguez-Villafuerte, M., Martínez-Dávalos A. and Rivera-Montalvo, T. Effect of correction methods of radiochromic EBT2 films on the accuracy of IMRT QA. *Applied Radiation and Isotopes*. 2016, **107**, 121-126. DOI:10.1016/j.apradiso.2015.09.016.

34. Mayer, R. R., MA, F. Chen, Y., Miller, R. I., Belard, A., mcdonough, J. and O'Connel, J. J. Enhanced dosimetry procedures and assessment for EBT2 radiochromic film. *Medical Physics*. 2012, **39**(4), 2147-2155. DOI:10.1118/1.369410.

35. Marrazzo, L., Zani, M., Pallotta, S., Arilli, C., Casati, M., Compagnucci, A., Talamonti, C. and Bucciolini, M. GafChromic® EBT3 films for patient specific IMRT QA using a multichannel approach. *Physica Medica*. 2015, **31**(8), 1035-1042. DOI:10.1016/j.ejmp.2015.08.010.

36. Méndez, I. Model selection for radiochromic film dosimetry. *Physics in Medicine and Biology*. 2015, **60**(10), 4089-4104. DOI:10.1088/0031-9155/60/10/4089.

37. Dufek, V., Horakova, I. and Koniarova, I. COMPARISON OF DIFFERENT TECHNIQUES FOR EVALUATION OF DOSE DISTRIBUTIONS IN RADIOTHERAPY USING RADIOCHROMIC EBT3 FILMS. *Radiation Protection Dosimetry*. 2019, **186**(2-3), 357-361. DOI:10.1093/rpd/ncz231.

38. Méndez, I., Hartman, V., Hudej, R., Strojnik, A. and Casar, B. Gafchromic EBT2 film dosimetry in reflection mode with a novel plan-based calibration method. *Medical Physics*. 2013, **40**(1). DOI:10.1118/1.4772075.

39. Méndez, I., Polšak, A., Hudej, R. and Casar, B. The Multigaussian method: a new approach to mitigating spatial heterogeneities with multichannel radiochromic film dosimetry. *Physics in Medicine and Biology*. 2018, **63**(17), 175013. DOI:10.1088/1361-6560/aad9c1.

40. Pérez Azorín, J. F., Ramos García, L. I. and Martí-Climent, J. M. A method for multichannel dosimetry with EBT3 radiochromic films. *Medical Physics*. 2014, **41**(6Part1), 62101. DOI:10.1118/1.4871622.

41. Vera-Sánchez, J. A., Ruiz-Morales, C. a González-López, A. Monte Carlo uncertainty analysis of dose estimates in radiochromic film dosimetry with single-channel and multichannel algorithms. *Physica Medica*. 2018, **47**, 23-33. DOI:10.1016/j.ejmp.2018.02.006.

42. Palmer, A. L., Dimitriadis, A., Nisbet, A. and Clark, C.H. Evaluation of Gafchromic EBT-XD film, with comparison to EBT3 film, and application in high dose radiotherapy verification. *Physics in Medicine and Biology*. 2015, **60**(22), 8741-8752. DOI:10.1088/0031-9155/60/22/8741.

43. Martišíková, M., Ackermann, B. and Jäkel, O. Analysis of uncertainties in Gafchromic® EBT film dosimetry of photon beams. *Physics in Medicine and Biology*. 2008, **53**(24), 7013-7027. DOI: 10.1088/0031-9155/53/24/001.

44. Hartmann, B., Martišíková, M. and Jäkel, O. Technical Note: Homogeneity of Gafchromic ® EBT2 film. *Medical Physics*. 2010, **37**(4), 1753-1756. DOI: 10.1118/1.3368601.

45. Sorriaux, J., Kacperek, A., Rossomme, S., Lee, J. A., Bertrand, D., Vynckier, S. and Sterpin, E. Evaluation of Gafchromic® EBT3 films characteristics in therapy photon, electron and proton beams. *Physica Medica*. 2013, **29**(6), 599-606. DOI: 10.1016/j.ejmp.2012.10.001.

46. Bouchard, H., Lacroix, F., Beaudoin, G., Carrier, J. F. and Kawrakow, I. On the characterization and uncertainty analysis of radiochromic film dosimetry. *Medical Physics*. 2009, **36**(6Part1), 1931-1946. DOI: 10.1118/1.3121488.

47. Saur, S. and Frengen, J. GafChromic EBT film dosimetry with flatbed CCD scanner: a novel background correction method and full dose uncertainty analysis. *Medical Physics*. 2008, **35**(7), 3094–101. DOI: 10.1118/1.2938522.

48. González-López, A., Vera-Sánchez, J. A. and Ruiz-Morales, C. The incidence of the different sources of noise on the uncertainty in radiochromic film dosimetry using single channel and multichannel methods. *Physics in Medicine and Biology*. 2017, **62**(22), N525-N536. DOI:10.1088/1361-6560/aa8f74.

49. Pócza, T., Zongor, Z., Melles-Bencsik, B., Tatai-Szabó, D. Z., Major, T. and Pesznyák, C. Comparison of three film analysis softwares using EBT2 and EBT3 films in radiotherapy. *Radiology and Oncology*. 2020, **54**(4), 505-512. DOI:10.2478/raon-2020-0049.

50. *Radiochromic.com* [online]. 2022 [cit. 2022-09-12]. Dostupné z: https://radiochromic.com/

51. Vera Sánchez, J. A, Ruiz Morales, C. and González López, A. Characterization of noise and digitizer response variability in radiochromic film dosimetry. Impact on treatment verification. *Physica Medica*. 2016, **32**(9), 1167-1174. DOI:10.1016/j.ejmp.2016.08.019.

52. Huang, J. Y., Pulliam, K. B., Mckenzie, E. M., Followill, D. S. and Kry, S. F. Effects of spatial resolution and noise on gamma analysis for IMRT QA. *Journal of Applied Clinical Medical Physics*. 2014, **15**(4), 93-104. DOI:10.1120/jacmp.v15i4.4690.

53. Miguel, F. J. S., Clemente-Gutiérrez, F. and Pérez-Vara, C. Analysis of different procedures for absolute dosimetry with EBT3 radiochromic film. *Biomedical Physics and Engineering Express*. 2018, **4**(6), 65008. DOI:10.1088/2057-1976/aae1cd.

54. Lárraga-Gutiérrez, J. M., García-Garduno, O. A., Trevino-Palacios, C. and Herrera-González, J. A.. Evaluation of a LED-based flatbed document scanner for radiochromic film dosimetry in transmission mode. *Physica Medica*. 2018, **47**, 86-91. DOI:10.1016/j.ejmp.2018.02.010.

55. Shameem, T., Bennie, N., Butson, M. and Thwaites, D. A comparison between EPSON V700 and EPSON V800 scanners for film dosimetry. *Physical and Engineering Sciences in Medicine*. 2020, **43**(1), 205-212. DOI:10.1007/s13246-019-00837-3.

56. Schoenfeld, A. A., Wieker, S., Harder, D. and Poppe, B. The origin of the flatbed scanner artifacts in radiochromic film dosimetry—key experiments and theoretical descriptions. *Physics in Medicine and Biology*. 2016, **61**(21), 7704-7724. DOI:10.1088/0031-9155/61/21/7704.

57. Van Battum, L. J., Huizenga, H., Verdaasdonik, R. M. and Heukelom, S. How flatbed scanners upset accurate film dosimetry. *Physics in Medicine and Biology*. 2016, **61**(2), 625-649. DOI:10.1088/0031-9155/61/2/625.

58. Ferreira, B. C., Lopes, M. C. and Capela, M. Evaluation of an Epson flatbed scanner to read Gafchromic EBT films for radiation dosimetry. *Physics in Medicine and Biology*. 2009, **54**(4), 1073-1085. DOI:10.1088/0031-9155/54/4/017.

59. Lewis, D. and Chan, M. F. Correcting lateral response artifacts from flatbed scanners for radiochromic film dosimetry. *Medical Physics*. 2015, **42**(1), 416-429. DOI:10.1118/1.4903758.

60. Pérez Azorin, J. F., Garcia, L. I. R., Ozcoidi, D. M. and Almansa, J. F. Polarized dosimetry method for Gafchromic EBT3. *Physica Medica*. 2016, **32**(8), 972-980. DOI:10.1016/j.ejmp.2016.06.013.

61. Leiws, D. and Devic, S. Correcting scan-to-scan response variability for a radiochromic film-based reference dosimetry system. *Medical Physics*. 2015, **42**(10), 5692-5701. DOI:10.1118/1.4929563.

62. Calvo-Ortega, J. F., Pozo, M., Moragues, S. and Casals, J. Fast protocol for radiochromic film dosimetry using a cloud computing web application. *Physica Medica*. 2017, **39**, 1-8. DOI:10.1016/j.ejmp.2017.05.072.

63. Lewis, D., MICKE, A., Yu, X. and Chan, M. F. An efficient protocol for radiochromic film dosimetry combining calibration and measurement in a single scan. *Medical Physics*. 2012, **39**(10), 6339-6350. DOI:10.1118/1.4754797.

64. Maeyama, T., Ishida, Y., Kudo, Y., Fukasaku, K., Ishikawa, K. L. and Fukunishi, N. Polymer gel dosimeter with AQUAJOINT ® as hydrogel matrix. *Radiation Physics and Chemistry*. 2018, **146**, 121-125. DOI:10.1016/j.radphyschem.2018.01.014.

65. Spěváček, V., Pilařová, K., Končeková, J. and Konček, O. The influence of antioxidant THPC on the properties of polymer gel dosimeter. *Physics in Medicine and Biology*. 2014, **59**(17), 5141-5161. DOI:10.1088/0031-9155/59/17/5141.

66. Lazzaroni, S., Liosi, G.M., Mariani, M. and Dondi, D. An innovative Fe3 selective ligand for Fricke-gel dosimeter. *Radiation Physics and Chemistry*. 2020, **171**. DOI:10.1016/j.radphyschem.2020.108733.

67. Khezerloo, D., Nedaie, H. A., Takavar, A., et al. PRESAGE® as a solid 3-D radiation dosimeter: A review article. *Radiation Physics and Chemistry*. 2017, **141**, 88-97. DOI:10.1016/j.radphyschem.2017.06.002.

68. Nelms, B. E., Chan, M. F., Jarry, G., Lemire, M., Lowden, J., Hampton, C. and Feygelman, V. Evaluating IMRT and VMAT dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Medical Physics*. 2013, **40**(11), 111722. DOI: 10.1118/1.4826166.

69. Feygelman, V., Zhang, G., Stevens, C. and Nelms, B. E. Evaluation of a new VMAT QA device, or the "X" and "O" array geometries. *Journal of Applied Clinical Medical Physics,* 2011, **12** (2), 146-168. DOI:10.1120/jacmp.v12i2.3346.

70. Neilson, C., Klein, M., Barnett, R. and Yartsev, S. Delivery quality assurance with ArcCHECK. *Medical Dosimetry*. 2013, **38**(1), 77-80. DOI: 10.1016/j.meddos.2012.07.004.

71. Sun Nuclear Corporation (2019) *ArcCHECK® & 3DVH®* [pamphlet]. Retrieved from https://www.sunnuclear.com/documents/datasheets/arccheck3dvh.pdf (14 August 2019).

72. Nelms, B. E., Opp, D., Robison, J., Wolf, T. K., Zhang, G., Moros, E. and Feygelman, V. VMAT QA: Measurement-guided 4D dose reconstruction on a patient. *Medical Physics*. 2012, **39**(7Part1), 4228-4238. DOI: 10.1118/1.4729709.

73. Allagaier, B., Schule, E. and Wurfel, J. Dose reconstruction in the OCTAVIUS 4D phantom and in the patient without using dose information from the TPS. *PTW*. 2013. D913.200.06/00.

74. Urso, P., Lorusso, R., Marzoli, L., Corletto, D., Imperiale, P., Pepe, A. and Bianchi, L. Practical application of Octavius ® -4D: Characteristics and criticalities for IMRT and VMAT verification. *Journal of Applied Clinical Medical Physics*. 2018, **19**(5), 517-524. DOI: 10.1002/acm2.12412.

75. PTW Freiburg (2019) *Radiation Medicine QA Solutions* [brochure]. Retrieved from https://www.ptwdosimetry.com/en/products/OCTAVIUS-4d/ (14 August 2019).

76. Hauri, P., Verlaan, S., Graydon, Ahnen, L., Klöck, S. and Lang, S. Clinical evaluation of an anatomy-based patient specific quality assurance system. *Journal of Applied Clinical Medical Physics*. 2014, **15**(2), 181-190. DOI: 10.1120/jacmp.v15i2.4647.

77. ScandiDos (2019) *Delta^4 Phantom+* [pamphlet]. Retrieved from https://delta4family.com/products (14 August 2019).

78. Hu, Y., Ruan, C., Nguyen, A., Duggar, W., Mobit, P., Rajaguru, P., He, R. and Yang, C. SU-E-T-44: Phantom 3D Dose Calculation and Anatomy Based DVH Evaluation On VMAT Patient QA Using the Newest Version of Delta4 Dosimetry System. *Medical Physics*. 2013, **40**(6Part11), 213-213. DOI: 10.1118/1.4814479.

79. Sadagopan, R., Bencomo, J. A., Martin, R. L., Nilsson, G., Matzen, T. and Balter, P. A. Characterization and clinical evaluation of a novel IMRT quality assurance system. *Journal of Applied Clinical Medical Physics*. 2009, **10**(2), 104-119. DOI: 10.1120/jacmp.v10i2.2928.

80. Ahnesjö, A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Medical Physics*. 1989, **16**(4), 577-592. DOI: 10.1118/1.596360.

81. Valve, A., Keyriläinen, J. and Kulmala, J. Compass model-based quality assurance for stereotactic VMAT treatment plans. *Physica Medica*. 2017, **44**, 42-50. DOI: 10.1016/j.ejmp.2017.11.009.

82. Nichol, L. E. A. Introduction to Dosimetry Check [presentation]. *2nd UK and Ireland Dosimetry Check User Meeting Symposium*. Retrieved from http://dosimetrycheck.com/papers_and_patents.htm (14 August 2019).

83. McCaw, T. J., Micka, J. A. and Dewerd, L. A. Development and characterization of a three-dimensional radiochromic film stack dosimeter for megavoltage photon beam dosimetry. *Medical Physics*. 2014, **41**(5), 052104. DOI: 10.1118/1.4871781.

84. Barbeiro, A. R., Ureba, A., Baeza, J. A., Linares, R., Jimenez-Ortega, E., Mateos, J. C., Velazques, S. and Plaza, A. L. SU-E-T-644: QuAArC: A 3D VMAT QA System Based On Radiochromic Film and Monte Carlo Simulation of Log Files. In: *Medical Physics*. 2015, **42**(6Part22), 3484-3484. DOI: 10.1118/1.4925007.

85. Park, J. Y., Lee, J. W., Choi, K. S., Lee, J. S., Kim, J. H., Homg, S. and Suh, T. S. Development of a novel quality assurance system based on rolled-up and rolled-out radiochromic films in volumetric modulated arc therapy. *Medical Physics*. 2011, **38**(12), 6688-6696. DOI: 10.1118/1.3659706.

86. Richardson, S. L., Tomé, W. A., Orton, N. P., McNutt, T. R. and Paliwal, B. R. IMRT delivery verification using a spiral phantom. *Medical Physics*. 2003, **30**(9), 2553-2558. DOI: 10.1118/1.1603965.

87. Paliwal, B., Tomé, W., Richardson, S. and Mackie, T. R. A spiral phantom for IMRT and tomotherapy treatment delivery verification. *Medical Physics*. 2000, **27**(11), 2503-2507. DOI: 10.1118/1.1319523.

88. Robar, J. L. and B. G. Clark. A practical technique for verification of three-dimensional conformal dose distributions in stereotactic radiosurgery. *Medical Physics*. 2000, **27**(5), 978-987. DOI: 10.1118/1.598962.

89. Gueli, A. M., Mannino, G., Troja, S.O. et al. 3D dosimetry on Ru-106 plaque for ocular melanoma treatments. *Radiation Measurements*. 2011, **46**(12), 2014-2019. DOI: 10.1016/j.radmeas.2011.07.032.

90. Troja, S. O., Egger, E., Francescon, P., Gueli, A. M., Kacperek, A., Coco, M., Musmeci, R. and Pedalino, A. 2D and 3D dose distribution determination in proton beam radiotherapy with GafChromic film detectors. *Technology and Health Care*. 2000, **8**(3), pp. 155-164.

91. Kim, J., Yoon, M., Kim, S., Shin, D., Lee, S. B., Lim, Y. K., Kim, D. W. and Park, S. Y. Three-dimensional radiochromic film dosimetry of proton clinical beams using a gafchromic EBT2 film array. *Radiation Protection Dosimetry*. 2012, **151**(2), 272-277. DOI: 10.1093/rpd/ncs010.

92. Gueli, A. M., De Vincolis, R., Kacperek A. and Troja, S. O. An approach to 3D dose mapping using gafchromic® film. *Radiation Protection Dosimetry*. 2005, **115**(1-4), 616-622. DOI: 10.1093/rpd/nci255.

93. Rangel, A., Palte, G. and Dunscombe, P. The sensitivity of patient specific IMRT QC to systematic MLC leaf bank offset errors. *Medical Physics*. 2010, **37**(7), 3862–3867. DOI: 10.1118/1.3453576.

94. Nelms, B., Zhen, H., Tomé W. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Medical Physics*. 2011, **38**(2), 1037–1044. DOI: 10.1118/1.3544657.

95. Pulliam, K. B., Huang, J. Y., Howell, R. M., Followill, D., Bosca, R., O'Daniel, J. and Kry, S. F. Comparison of 2D and 3D gamma analyses. *Medical Physics*. 2014, **41**(2), 021710-. DOI: 10.1118/1.4860195.

96. Xing, A., Arumugam, S., Deshpande, S., George, A., Vial, P., Holloway, L. and Goozee, G. Evaluation of 3D Gamma index calculation implemented in two commercial dosimetry systems. *Journal of Physics: Conference Series*. 2015, **573**(012054), 1-4. DOI: 10.1088/1742-6596/573/1/012054.

97. Howard, M. E., Herman, M. G., Grams, M. P. and Leal, A. Methodology for radiochromic film analysis using FilmQA Pro and ImageJ. *PLOS ONE*. 2020, **15**(5). DOI:10.1371/journal.pone.0233562.

98. Hermida-López, M., Lüdemann, L., Flühs, A. and Brualla, L. Technical Note: Influence of the phantom material on the absorbed-dose energy dependence of the EBT3 radiochromic film for photons in the energy range 3 keV-18 MeV. *Medical Physics*. 2014, **41**(11). DOI:10.1118/1.4898598.

99. Ataei, G., Rezaei, M., Gorji, K. E., Banaei, A., Goushbolagh, N.A., Farhood, B., Bagheri, M. and Firouzjah, R. A. Evaluation of dose rate and photon energy dependence of gafchromic EBT3 film irradiating with 6 MV and Co-60 photon beams. *Journal of Medical Signals and Sensors*. 2019, **9**(3). DOI:10.4103/jmss.JMSS_45_18.

100. Kairn, T., Hardcastle, N., Kenny, J., Meldrum, R., Timé, W. A. and Aland, T. EBT2 radiochromic film for quality assurance of complex IMRT treatments of the prostate: micro-collimated IMRT, RapidArc, and TomoTherapy. *Australasian Physical & Engineering Sciences in Medicine*. 2011, **34**(3), 333-343. DOI:10.1007/s13246-011-0087-z.

101. Mathot, M., Sobczak, S. and Hoornaert, M.-T. Gafchromic film dosimetry: Four years experience using FilmQA Pro software and Epson flatbed scanners. *Physica Medica*. 2014, **30**(8), 871-877. DOI:10.1016/j.ejmp.2014.06.043.

102. González-López, A., Vera-Sánchez, J. A. and Ruiz-Morales, C. Technical Note: Statistical dependences between channels in radiochromic film readings. Implications in multichannel dosimetry. *Medical Physics*. 2016, **43**(5), 2194-2199. DOI:10.1118/1.4945273.

103. Almady, B., Wesolowska, P., Santos, T. and Izewska, J. EP-1511: Gamma analysis. *Radiotherapy and Oncology*. 2016, **119**, S698-S699. DOI:10.1016/S0167-8140(16)32761-X.

104. Marroquin, E. Y. L., Herrera González, J. A., Camacho López, M. A., Villarreal Barajas, J. E. and García-Garduno, O. A. Evaluation of the uncertainty in an EBT3 film dosimetry system utilizing net optical density. *Journal of Applied Clinical Medical Physics*. 2016, **17**(5), 466-481. DOI:10.1120/jacmp.v17i5.6262.

105. Llinares, A. U., Fernández, A. S., Pain, E. A., Caballero, L. J. R. and Oquendo, M. I. EP-1319 OCTAVIUS PHANTOM AND SEVEN29 2D ARRAY IN IMRT PATIENT VERIFICATION. TECHNICAL AND CLINICAL ORGANIZATION ASPECTS. *Radiotherapy and Oncology*. 2012, **103**, S500-S501. DOI: 10.1016/S0167-8140(12)71652-3.

106. Nalbant, N., Donmez, K. N. and Bilge, H. Pre-Treatment Dose Verification of Imrt Using Gafchromic Ebt3 Film and 2DArray. *Journal of Nuclear Medicine & Radiation Therapy*. 2014, **182**(5). DOI: 10.4172/2155-9619.1000182.

107. Hussein, M., Rowshanfarzad, P., Ebert, M. A., Nisbet, A. and Clark, C. H. A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems. *Radiotherapy and Oncology*. 2013, **109**(3), 370-376. DOI: 10.1016/j.radonc.2013.08.048.

108. Onishi, Y., Nakayama, S., Watanabe, S., Kaneshige, S., Monzen, H., Matsumoto, K., Shintani, N. and Kamomae, T. Comparison of dose accuracy between film and two-dimensional detectors in intensity-modulated radiation therapy. *Journal of the Korean Physical Society*. 2015, **67**(1), 89-95. DOI: 10.3938/jkps.67.89.

109. Marrazzo, L., Arilli, C., Pasler, M., et al. Real-time beam monitoring for error detection in IMRT plans and impact on dose-volume histograms. *Strahlentherapie und Onkologie*. 2018, **194**(3), 243-254. DOI:10.1007/s00066-017-1245-3.

110. Depuydt, T., Van Esch, A. and Huyskens, D. P. A quantitative evaluation of IMRT dose distributions: refinement and clinical assessment of the gamma evaluation. *Radiotherapy and Oncology*. 2002, **62**(3), 309-319.

111. Scanditronix Wellhöfer (2019) *I'mRT* [pamphlet]. Retrieved from http://www.rosalina.in/IBA-Dosimetry/I%27mRT.pdf (14 August 2019).

112. Low, D. A. and Dempsey, J. F. Evaluation of the gamma dose distribution comparison method. *Medical Physics*. 2003, **30**(9), 2455-2464. DOI: 10.1118/1.1598711.

113. Arumugam, S., Xing, A., Goozee, G. and Holloway, L. Detecting VMAT delivery errors: A study on the sensitivity of the ArcCHECK-3D electronic dosimeter. *Journal of Physics: Conference Series*. 2013, **444**. DOI:10.1088/1742-6596/444/1/012019.

114. Steers, J. M. and Fraass, B. A. IMRT QA and gamma comparisons: The impact of detector geometry, spatial sampling, and delivery technique on gamma comparison sensitivity. *Medical Physics*. 2021, **48**(9), 5367-5381. DOI:10.1002/mp.14997.

115. Woon, W. A., Ravindran, P. B., Ekayanake, V. S., Lim, Y. F. and Khalid, J. A study on the effect of detector resolution on gamma index passing rate for VMAT and IMRT QA. *Journal of Applied Clinical Medical Physics*. 2018, **19**(2), 230-248. DOI:10.1002/acm2.12285.

116. Amin, A. E., Sallam, A. M., El-Sayed, E. M., El-Sayed, S. M. and Salem, N. R. Validation of the Octavius 4D measuring system in verifying advanced external beams radiotherapy techniques. *Egyptian Journal of Biomedical Engineering and Biophysics*. 2018, 0-0. DOI:10.21608/ejbbe.2018.2295.1010.

117. Mijnheer, B., Jomehzadeh, A., González, P., et al. Error detection during VMAT delivery using EPID-based 3D transit dosimetry. *Physica Medica*. 2018, **54**, 137-145. DOI:10.1016/j.ejmp.2018.10.005.

118. Pogson, E. M., Arumugam, S., Hansen, C. R., et al. Comparison of multi-institutional pre-treatment verification for VMAT of nasopharynx with delivery errors. *Physica Medica*. 2018, **53**, 25-31. DOI:10.1016/j.ejmp.2018.07.007.

119. Hussein, M., Adams, E. J., Jordan, T. J., Clark, C. H. and Nisbet, A. A critical evaluation of the PTW 2D-ARRAY Seven29 and OCTAVIUS II phantom for IMRT and VMAT verification. *Journal of Applied Clinical Medical Physics*. 2013, **14**(6), 274-292. DOI: 10.1120/jacmp.v14i6.4460.

120. Thongsawad, S., Srisatit, S. and Fuangrod, T. Error detection sensitivity test using complex errors on three patient-specific VMAT QA systems. *Journal of Physics: Conference Series*. 2019, **1285**(1), 012030. DOI:10.1088/1742-6596/1285/1/012030.

121. Netherlands Commission on Radiation Dosimetry. *Code of Practice for the Quality Assurance and Control for Intensity Modulated Radiotherapy: Report 22*. Delft, The Netherlands, 2013. DOI: 10.25030/ncs-022.

122. Netherlands Commission on Radiation Dosimetry. *Code of Practice for the Quality Assurance and Control for Volumetric Modulated Arc Therapy: Report 24*. Delft, The Netherlands, 2015. DOI: 10.25030/ncs-024.

123. Chandraraj, V., Stathakis, S., Maniskam, R., Esquivel, C., Supe, S. S. and Papanikolaou, N. Comparison of four commercial devices for RapidArc and sliding window IMRT QA. *Journal of Applied Clinical Medical Physics*. 2011, **12**(2), 338-349. DOI: 10.1120/jacmp.v12i2.3367.

124. Stathakis, S., Gutiérrez, A. N., Esquivel, C., Mavroidis, P., Papanikolaou, N. and Myers, P. Evaluation of PTW Seven29 for tomotherapy patient-specific quality assurance and comparison with ScandiDos Delta 4. *Journal of Medical Physics*. 2012, **37**(2). DOI: 10.4103/0971-6203.94741.

125. Moreno, R. D., Venencia, D., Garrigo, E. and Pipman, Y. A method to enhance spatial resolution of a 2D ion chamber array for quality control of MLC. *Journal of Applied Clinical Medical Physics*. 2011, **12**(4), 63-73. DOI: 10.1120/jacmp.v12i4.3456.

126. Wiezorek, T., Banz, N., Schwedas, M., Scheithauer, M., Salz, H., Georg, D. and Wendt, T. G. Dosimetric Quality Assurance for Intensity–Modulated Radiotherapy. *Strahlentherapie und Onkologie*. 2005, **181**(7), 468-474. DOI: 10.1007/s00066-005-1381-z.

127. Van Esch, A., Huyskens, D. P., Behrens, C. F. et al. Implementing RapidArc into clinical routine: A comprehensive program from machine QA to TPS validation and patient QA. *Medical Physics*. 2011, **38**(9), 5146-5166. DOI: 10.1118/1.3622672.

128. Fredh, A., Scherman, J. B., Fog, L. S. and Munck Af Rosenschöld, P. Patient QA systems for rotational radiation therapy: A comparative experimental study with intentional errors. *Medical Physics*. 2013, **40**(3), 031716. DOI: 10.1118/1.4788645.

129. GARCIA-VICENTE, Feliciano, Virginia FERNANDEZ, Rocio BERMUDEZ, Alberto GOMEZ, Leopoldo PEREZ, Almudena ZAPATERO and Juan J. TORRES. Sensitivity of a helical diode array device to delivery errors in IMRT treatment and establishment of tolerance level for pretreatment QA. *Journal of Applied Clinical Medical Physics,* 2012, **13**(1), 111-123. DOI:10.1120/jacmp.v13i1.3660.

130. Al Sa'd, M., Graham, J., Liney, G. P. and Moore, C. J. Quantitative comparison of 3D and 2.5D gamma analysis: introducing gamma angle histograms. *Physics in Medicine and Biology*. 2013, **58**(8), 2597-2608. DOI: 10.1088/0031-9155/58/8/2597.

131. Kruse, J J. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. *Medical Physics*. 2010, **37**(6), 2516-2524. DOI: 10.1118/1.3425781.

132. Taylor, M. S. Comment on "On the insensitivity of single field planar dosimetry to IMRT inaccuracies" [Med. Phys. 37, 2516–2524 (2010)]. *Medical Physics*. 2010, **37**(7), 3880. DOI: 10.1118/1.3456928.

133. Kruse, J. J. Response to "Comment on 'On the insensitivity of single field planar dosimetry to IMRT inaccuracies' " [Med. Phys. 37, 3880 (2010)]. *Medical Physics*. 2010, **37**(7), 3881. DOI: https://doi.org/10.1118/1.3460816.

134. Budgell, G. Comment on "On the insensitivity of single field planar dosimetry to IMRT inaccuracies" [Med. Phys. 37, 2516–2524 (2010)]. *Medical Physics*. 2010, **37**(12), 6497. DOI: 10.1118/1.3514142.

135. Budgell, G. J., Perrin, B. A., Mott, J. H. L., Fairfoul, J. and Mackay, R. I. Quantitative analysis of patient-specific dosimetric IMRT verification. *Physics in Medicine and Biology*. 2005, **50**(1), 103-119. DOI: 10.1088/0031-9155/50/1/009.

136. Mijnheer, B. *Clinical 3D dosimetry in modern radiation therapy*. Boca Raton, 2017. Imaging in medical diagnosis and therapy, 28. ISBN 978-148-2252-217.

137. Stelljes, T. S., Harmeyer, A., Reuter, J., Looe, H. K., Chofor, N., Harder, D. and Poppe, B. Dosimetric characteristics of the novel 2D ionization chamber array OCTAVIUS Detector 1500. *Medical Physics*. 2015, **42**(4), 1528-1537. DOI: 10.1118/1.4914151.

138. Yoosuf, A. B. M., Jeevanandam, P., Whitten, G., Workman, G. and McGarry, C. K. Verification of high-dose-rate brachytherapy treatment planning dose distribution using liquid-filled ionization chamber array. *Journal of Contemporary Brachytherapy*. 2018, **10**(2), 142-154. DOI: 10.5114/jcb.2018.75599.

139. Hanušová, T., Horáková, I. and Irena Koniarová, I. IMRT plan verification with EBT2 and EBT3 films compared to PTW 2D-ARRAY Seven29. *Radiation Physics and Chemistry*. 2017, **140**, 365-369. DOI: 10.1016/j.radphyschem.2017.02.041.

140. Carrasco, P., Jornet, N., Latorre, A., Eudaldo, T., Ruiz, A. and Ribas, M. 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Medical Physics*. 2012, **39**(8), 5040-5049. DOI: 10.1118/1.4736949.

141. Lafratta, R., Ibbott, G., Adamovics, J. and Followill, D. Comparison of 2D and 3D Gamma calculations for an IMRT QA phantom. *Journal of Physics: Conference Series*. 2015, **573**, 012055. DOI: 10.1088/1742-6596/573/1/012055.

142. Rajasekaran, D., Jeevanandam, P., Sukumar, P., Ranganathan, A., Johnjothi, S. and Nagarajan, V. A study on correlation between 2D and 3D gamma evaluation metrics in patient-specific quality assurance for VMAT. *Medical Dosimetry*. 2014, **39**(4), 300-308. DOI: 10.1016/j.meddos.2014.05.002.

143. Jin, X., Yan, H., Han, C., Zhou, Y., Yi, J. and Xie, C. Correlation between gamma index passing rate and clinical dosimetric difference for pre-treatment 2D and 3D volumetric modulated arc therapy dosimetric verification. *The British Journal of Radiology*. 2015, **88**(1047). DOI: 10.1259/bjr.20140577.

144. Hussein, M., Clark, C. H. and Nisbet, A. Challenges in calculation of the gamma index in radiotherapy – Towards good practice. *Physica Medica*. 2017, **36**, 1-11. DOI: 10.1016/j.ejmp.2017.03.001.

145. Kim, K. H., Kim, D. S., Kim, T. H., Kang, S. H., Cho, M. S. and Suh, T. S. The influence of the IMRT QA set-up error on the 2D and 3D gamma evaluation method as obtained by using Monte Carlo simulations. *Journal of the Korean Physical Society*. 2015, **67**(10), 1859-1867. DOI: 10.3938/jkps.67.1859.

146. Sadrollahi, A., Nuesken, F., Licht, N., Rübe, C., Dzierma, Y. and Zhang, Q. Monte-Carlo simulation of the Siemens Artiste linear accelerator flat 6 MV and flattening-filter-free 7 MV beam line. *PLOS ONE*. 2019, **14**(1). DOI: 10.1371/journal.pone.0210069.

147. Khaledi, N., Dabaghi, M., Sardari, D., Samiei, F., Ahamdabad, F. G., Jahanfarnia, G., Saadi, M. K. and Wang, X. Investigation of photoneutron production by Siemens artiste linac: A Monte Carlo Study. *Radiation Physics and Chemistry*. 2018, **153**, 98-103. DOI: 10.1016/j.radphyschem.2018.06.006.

148. Turgul, T. and Erogul, O. Determination of initial electron parameters by means of Monte Carlo simulations for the Siemens Artiste Linac 6 MV photon beam. *Reports of Practical Oncology and Radiotherapy*. 2019, **24**(4), 331-337. DOI: 10.1016/j.rpor.2019.05.002.

149. Alharthi, T., Arumugam, S., Vial, P., Holloway, L. and Thwaites, D. EPID sensitivity to delivery errors for pre-treatment verification of lung SBRT VMAT plans. *Physica Medica*. 2019, **59**, 37-46. DOI: 10.1016/j.ejmp.2019.02.007.

150. Kron, T., Lehmann, J. and Greer, P. B. Dosimetry of ionising radiation in modern radiation oncology. *Physics in Medicine and Biology*. 2016, **61**(14), R167-R205. DOI: 10.1088/0031-9155/61/14/R167.

Appendix

Core MATLAB scripts created in this work

The following is a list of the most important MATLAB scripts, in alphabetical order, created in this work for various tasks that have been described throughout the text and gives a brief description of each script's funcionality. It does not aim to be a software documentation or anything of that kind.

blankfilm.m – subtracts blank film from irradiated film and saves the differencial image

calcurve.m – fits calibration curve to calibration dose points

chi2ind.m – performs statistical chi square test of independence

dose.m – converts scanner values to dose values with an appropriate calibration curve

findcross.m – finds cross sections of different 2D planes in the 3D space, draws profile plots and quantifies differences

gamma2Dfbyf.m – processes input data (measured and predicted field dose in isocentric coronal plane), performs 2D field-by-field gamma analysis, shows and saves results

gamma2DGauss.m – adds gaussian noise to film dose matrices and performs 2D gamma analysis repeatedly to simulate propagation of uncertainties to gamma score

gamma3DglobalSiemens.m – processes input data (15 film palnes and 15 XiO predicted planes in ASCII), builds 3D data cubes, performs 3D gamma analysis – global normalization, shows and saves results (gamma index maps, gamma scores, gamma index images, dose plane images, profile images, gamma histograms)

gamma3DglobalTrueBeam.m - processes input data (15 film palnes and 15 Eclipse predicted planes in DICOM), builds 3D data cubes, performs 3D gamma analysis – global normalization, shows and saves results (gamma index maps, gamma scores, gamma index images, dose plane images, profile images, gamma histograms)

gamma3DlocalSiemens.m - processes input data (15 film palnes and 15 XiO predicted planes in ASCII), builds 3D data cubes, performs 3D gamma analysis – local normalization, shows and saves results (gamma index maps, gamma scores, gamma index images, dose plane images, profile images, gamma histograms)

gamma3DlocalTrueBeam.m - processes input data (15 film palnes and 15 Eclipse predicted planes in DICOM), builds 3D data cubes, performs 3D gamma analysis – local normalization, shows and saves results (gamma index maps, gamma scores, gamma index images, dose plane images, profile images, gamma histograms)

gamma2Drel.m – function that is called for 2D gamma analysis

gama3Drel.m – function that is called for 3D gamma analysis

hist.m – creates histograms from already saved data, optimizes graphical display and saves them as images

interscan.m – performs 2D gamma analysis on 10 consecutive scans of the same film taking the first one as reference and saves results