Doctoral thesis: opponent's review

Institution: FJFI ČVUT v Praze

Author: Ing. Tereza Hanušová

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

Main objectives of the review:

- (format, clarity, language and style)
- recency and relevance of the subject of research
- methods
- goals of the thesis met?
- results and scientific contribution

FORMAT, CLARITY, LANGUAGE AND STYLE

The thesis is written in English - its standard is overall good, i.e. understandable, although it would require corrections to reach scientific journal level.

In my opinion, the abstract should reflect better a common standard and the thesis content. The 1st paragraph is rather a general introduction to the topic. The example sentences…

- *Compared to gel dosimetry, our method is more precise and less demanding.*

  Note: as far I understand, comparison with gel dosimetry was not main objective of the thesis so I do not see why to comment on this in abstract.

- *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.*

…in my opinion and based on experience, do not belong to abstract. Instead I would expect and prefer more description of goals, methods and results.

Some results are presented twice, in my opinion, without the need for it. For example differences between gamma score for 15 planes in tables (e.g. Table 4-26) and in the graph again (e.g. Figure 4-23)

Some formulations strike the eyes as one would not expect them in a scientific work, for example:

- *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%).*

- *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*
- in context of gamma analysis using established terms of LOCAL and GLOBAL normalization in a different sense for purposes of this thesis is confusing although carefully explained and reminded within the whole text. LOCAL/GLOBAL normalization normally reffers to normalization of dose difference to calculate gamma index. In this thesis it reffers to normalization of absolute dose distributions to obtain relative ones.
- I suspect author uses '*false positive*' and '*false negative*' in the manner opposite than is common. I understand that 'false positive result' means that truth is the positive result

  e.g. page 73: *false positive results (= plan failing tolerance even though delivered dose distribution in patient is acceptable)*

With regards to the research objectives I would prefer better organization of chapters – I found very difficult to find and/or distinguish relevant information regarding, e.g., benchmarking film dosimetry, benchmarking gamma algorithm implementation, testing sensitivity to errors using films (field-by-field, plane-by-plane, pseudo-3D and 2D gamma calculation, pseudo-3D and 3D gamma calculation, film + Omnipro) then comparing to electronic detectors and their own evaluation SW, IMRT, VMAT, etc.

RECENCY OF THE SUBJECT OF RESEARCH

Yes. Approaches to plan specific dosimetry QA including their implementation in clinics and setting operational decision criteria based on available equipment and methods still remain a topic for research. Particularly a question of sensitivity of various approaches and equipment to errors is certainly relevant and probably always will be. Also impact of orientation of films to radiation source is certainly an interesting question.

METHODS (AND RESULTS PRESENTATION/INTERPRETATION)

...please see separate attachment at the end of this review

GOALS OF THE THESIS MET?

Goals of the thesis are formulated in the Introduction, i.e.

One goal of this study is to find a dosimetric method … should provide measured doses in the whole 3D space with resolution at least comparable to current 2D electronic detectors. Moreover, the method should be accessible to all clinics using existing resources.

Yes, the author implemented EBT3 film dosimetry for measurement of relative dose distributions in combination of transverse, coronal and sagital planes in a dedicated homogeneous phantom. This includes film processing as well as gamma evaluation scripts in MATLAB. Although whether the method as presented would be accessible to all clinics using existing resources is – in my opinion -

questionable. Author presents film evaluations using commercial SW but as far I understand, film processing remains with MATLAB which cannot be considered as a widely accessible tool. Author is aware of this and suggests using alternative digital data processing platforms such as Python, etc. I am afraid that in such case the work to be done would be comparable to work done in this thesis so using outcomes of the thesis to ease their job I do not see straightforward.

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…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

Yes, the author compared same set of ('pseudo-3D', i.e. multiple 2D planes) measured data with the reference using own 2D and 3D gamma analysis. Although based on methods chosen the 3D application was limited to areas of planes intersections so total volume where 3D approach had a chance to bring a different result was rather limited as for most of evaluated points there was no difference between 2D and 3D especially for locally normalized film planes applied to both approaches. Global normalization of 'pseudo-3D' dose matrix that would move the method more towards 3D brought rather confusing results. With regards to '*whether a 3D method should be used*', it depends on whether the question stands as '2D vs 3D gamma calculation' or '2D vs (pseudo-)3D measurements with either gamma calculation' and whether '2D' applies to '2D evaluation of pseudo-3D data' or 'single-plane' or 'field-by-field' but yes, I agree sampling 3D space with more planes with different orientations towards the source has more chance to detect potential errors, and I also agree that – especially for modern standard dose delivery, i.e. VMAT – whole 3D dose should be analysed.

RESULTS AND SCIENTIFIC CONTRIBUTION

Details of implementing EBT3 film dosimetry may help others doing the same. Reading presentations of measurements and mutual comparisons including explanations of possible reasons of difference certainly may help others to make a better picture about all aspects and problems involved when establishing own local methods to address plan specific dosimetric QA and also during linac and TPS commissioning.

On other hand, I am not sure that the main original products of the work, i.e. MATLAB scripts for film processing and gamma calculations are easy to share. Similarly I think about the other (side-) products such as workflows to handle various HW and SW applied, there is probably not enough details to follow.

Regarding main outcome/recommendation to utilize rather 3D (dosimetry) approach for commissioning or investigating when routine method fails, I too remain unsure, especially because I do not feel arguments supported by results presented are strong enough to justify this recommendation – also for low coverage of potential influencing factors (6 or 9 patients/treatment plans, only MLC errors tested for IMRT – 3 fields, 2 patients/treatment plans, and also limited

coverage of possible VMAT delivery errors – 1 patient/treatment plan, 1 and 3 mm MLC bank shift and total MU) including limiting gamma score pass/fail threshold (arbitrarily and locally chosen 90% for one system and 95% for another) or considering other than 3%|3mm basic gamma parameters.

CONCLUSION

Despite number of questions and comments I find this work interesting and useful and therefore **I recommend this work for adoption as a doctoral thesis.** Main reason is large volume of complex work done with large variety of resources and careful application of statistical and uncertainty analysis. I also consider author's publication activity showing that she is capable to perform and publish scientific work.

Prague, 8.3.2023

Ing. Pavel Dvorak, Ph.D.

ATTACHMENT: METHODS (AND RESULTS PRESENTATION/INTERPRETATION)

(main 14 questions are highlighted in bold font)

Theoretical background provides review of radiochromic film dosimetry, 3D dosimetry systems and film-based approximations to 3D dosimetry. With regards to the thesis goals, a very important section is *2.4 Correlations between 2D and 3D dosimetry published in the literature*. Here I am missing a clear differentiation between '*2D vs 3D gamma evaluation*' and '*2D and 3D gamma pass rates*'. This is important for the goals of the thesis as the author's main focus on interpreting/representing gamma evaluation results is the pass rate (*gamma score* used within the thesis). In my opinion, and author is aware of this, these are not the same things – there are many ways how to represent/interpret result of gamma evaluation, both 2D or 3D.

Chapter 3 – Material and methods describes implementation of EBT3 film dosimetry and its validation/benchmarking by comparing result with results obtained using alternative electronic dosimeters. Using author's own words: '*The aim is to prove that film performs well enough to be used for our further experiments.*'

This benchmarking was done by comparing measured doses with TPS calculations. But TPS calculation is ultimately also a subject of dosimetric verification. Or is it only machine performance? Why measured data from various detectors were not compared mutually to separate the problem of benchmarking dose measurement from the problem of benchmarking analysis? The author herself is well aware of challenges of comparing gamma analysis implementations.

Another benchmarking example: 2D gamma calculation script in MATLAB as presented in *Chapter 3.3.3 2D gamma analysis in MATLAB*

This was done by comparing gamma calculation of 15 film planes – 2D gamma in MATLAB vs same measured doses in OmniPro I'mRT SW. Differences in gamma scores per each plane are presented in Figure 3-9 and Table 3-8. These differences are huge.

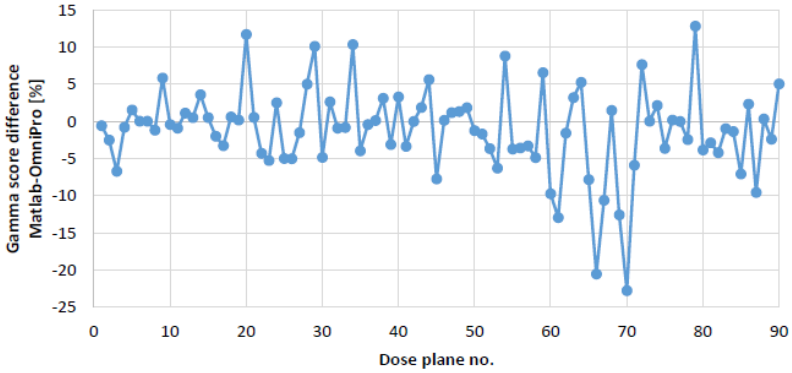

Figure 3-9: 2D gamma score difference between MATLAB and OmniPro I'mRT for all 2D planes (transverse, sagittal, coronal) for all patients in the I'mRT Phantom.

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

The author points probably correctly that major reason is using different gamma evaluation parameters, especially normalization and ROI. However, what is this for? The author applied statistical analysis to gamma scores calculated by both SW and found that the difference is not statistically significant: Page 56: *The differences in gamma scores between OmniPro I'mRT and MATLAB in Table 3-8 are not statistically significant.* **So the outcome should be that the 2D gamma using own MATLAB script has been validated by commercial SW?** I dont think this is correct conslusion as I believe that everyone using gamma analysis to compare dose distributions would agree that differences in gamma scores up to 10 or even 20% are large with regards to the analysis objectives. If I want to compare two gamma algorithm implementations I must keep as many influencing parameters as possible constant. Normalization of dose difference (being compared with preset tolerance, e.g. 3%) is probably the most important one. This is definitely possible to control using both MATLAB and OmniPro I'mRT.

Thesis/Page 53: *Another important reason was the comparison of results to OmniPro I'mRT version 1.7. ... 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.*

This is incorrect. All dose distributions are 'discrete'. '*Digital gamma*' option in OmniPro I'mRT includes interpolation between points changing sign of dose difference when searching the minimum gamma. This leads to more accurate pass/fraction compared to their standard '*Gamma*' but for the cost of not having direct gamma values. However, if the main focus of research is gamma score, this is probably option to consider.

Regarding the principle of the dosimetry method described in *Chap. 3.3.2 Pseudo-3D measurement with film*. As far I understand, (film) 15 planes are considered to sample whole 3D (space) matrix. Their position and orientation is determined by the phantom and its orientation. Citing from the thesis, page 51... *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.* OK, there is some uncertainty in film planes position. One of the main features of the proposed methods, the 3D gamma calculations (applied to pseudo-3D measurements) are based on analyzing 3D matrices where reference (TPS) is constructed again as 'pseudo-3D', i.e. empty 3D matrix with voxels filled at positions corresponding to expected film planes positions. This means that application of 3D gamma calculation differs from the 2D (plane-by-plane) calculation only in regions of planes intersections where orthogonal plane data expand search space for gamma calculation. **Why wasn't full 3D dose matrix considered for reference doses exported from TPS?** This would make the 3D gamma calculation much more meaningful. In addition, considering the above mentioned film planes positioning uncertainty, one could say it is a must. This is not even extra revolutionary approach, e.g. commercial SNC's ArcCheck applies this principle to its shell-like measurements as the author herself describes on page 20-21.

Focusing on, in the Introduction, declared main goals of the thesis, i.e.

...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

the most relevant results are probably presented in *Chapter 4.2.1 Pseudo-3D verification results*, *Chapter 4.2.3 Comparison of field-by-field and pseudo-3D verification for error-induced IMRT plans and 4.2.4 Comparison of the pseudo-3D approach with film and OCTAVIUS 4D for an error-induced VMAT plan.* For example...

Page 74, Table 4-7

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										

- If I understand well then both 2D and 3D sets of results apply to same measurements with same normalizations (75percentil of dose of individual plane) so the only difference is 2D vs 3D gamma calculation script. **How is it possible that in some cases 3D gamma score is lower than 2D gamma score, e.g. Patient 3/transverse plane/-2cm or Patient 2/coronal plane/-2 ?**
- **The reason for missing planes for 3D gamma evaluation is that they do not interesect with any other so the gamma result is the same as with the 2D calculation?**
- for example, Patient 2/sagittal plane/isocentric shows much lower pass rate than other planes, only slightly above 60%. The author does not consider this suspicious result? And 20% she does, as comments elsewhere? **Isn't mistake in measurement, processing or analysis most likely reason? Was the measurement repeated?**
- Patient 3/sagittal plane/-2...3D result 28.6% but 2D result, which should be lowe from principle reasons, result is 81.6%? **Is this same measurement, processing and gamma parameters applied so the only difference is intended 2D vs 3D gamma code?**
- **on page 73 the author comments result of Patient 4/sagittal plane/-4cm...17.8%...**

  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).

  Is this plane so unique (very asymmetric) that this very limiting aspect does not apply to other planes, also 4cm from the isocenter, for the same plan?

PATIENT 9 – VMAT

- film/pseudo-3D/2D gamma

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.

	2D gamma score [%]													
	EBT3 film pseudo-3D method							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	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%

So, the original plan gamma score for the plane 5 is 27.8, with the 1mm error 53.6 and with 3mm 42.3. Isn't it weird to have results with error introduced plans better than with the original plan without deliberate error? Especially when investigating sensitivity of the method to reveal errors? Unsurprisingly, results with OmniPro are much closer to expected – almost no difference for 1mm error, probably because the gamma parameters (3%|3mm) are well above the difference (1mm), some difference for next two errors but consistent for all planes. **Also results for other planes measured with film show rather unexpected trends – bigger error larger gamma score (e.g. planes 13 and 15), etc. Were these measurements**

repeated and confirmed? If there is such large uncertainty of the gamma score, does it make sense to consider application of this implementation of film dosimetry for these purposes at all?

...here is the author's interpretation:

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]).

...so, only the plane 5 is so far from the PTV that the dose is so low that even relative dosimetry is so uncertain that cannot be used? And the other planes were all close to PTV? Was the shape pf PTV/dose distribution so asymmetric? With VMAT technique? **Please explain why only plane 5 is so special that cannot be considered.**

Does author's interpretation mean that to use gamma score and film dosimetry, only errors reflected in differences above 4.5% of gamma score can be considered significant? In my opinion this is lots...considering application that I cannot say anything for gamma scores within range of 85.5% and 94.5% for example threshold 90%? Please comment...

For above questions and arguments I have doubts about the author's interpretation...

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.

...especially interpretation that lower gamma scores for films mean higher sensitivity to error detection.

- film/pseudo-3D/3D gamma

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%

What is a reason for higher gamma score in plane 13 for 3mm error than for the 1mm error? Is it because uncertainty of film dosimetry implementation reflects in such high uncertainty of gamma scores? **Again, isn't such high uncertainty disqualifying the method as a whole?**

Thesis/Page 47: 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. Isn't such uncertainty comparable with declared uncertainty of implemented film dosimetry? If so, why does this level of uncertainty allows significantly more consistent pass rates for the Octavius system as presented in Table 4-26?

- Related parts of the thesis summary:

For a VMAT plan, the pseudo-3D film dosimetry method was able to detect 1 mm and 3 mm MLC leaf positioning errors ... as well as 3% MU error ...with 3%/3 mm criteria and 95% tolerance level when more planes were evaluated...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).

Presented successfull error detection is based on lower number of film planes with gamma score below 95%. How many planes should not pass 95% to declar overal result 'error detected'? Gamma scores of some planes are higher than scores for the original error free plan. This relates to Table 4-26.

The same measurements with 3D gamma evaluation, Table 4-28, showed 1/9 planes failed 1mm error plan, 2/9 planes failed 3mm error plan and 1/9 planes failer 3% MU error plan. Making final decision whether plan is OK or not using this approach introduces yet another degree of freedom to the problem (number of passing planes). **Is such method presentable as sensitive to errors in principle? With the final outcome crucially dependent on orbitrary chosen pass/fail threshold (e.g. 95%) and even more with minimum fraction of total number of planes below this threshold?**

...and next

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).

So, the electronic system would detect all errors should it had lower basic tolerances? Is it really surprising outcome when errors in delivery of 3% and 3mm, respectively, were tested using 3%|3mm basic tolerance levels so they would appear clearly with lower tolerances of 2%|2mm? **Isn't rather this, i.e. adjusting basic gamma tolerances, more appropriate approach to set sensitivity of any dosimetry method based on gamma?**