

Evaluation of the predictive power of the software PQAi for quality assurance controls in VMAT treatments

Nicolas Bizot^{1,2}, Jean-Marc Nigoul², Rashid Oozeer³

¹ *Master 2 Ingénierie pour la Santé et le Médicament, Option Physique médicale, Grenoble, FRANCE*

² *Service de radiothérapie, AP-HM - hôpital de La Timone, Marseille, FRANCE*

³ *Directeur et Physicien médical, Radiation-Therapy-Consulting, Marseille, FRANCE*

ABSTRACT: Currently, an increasing fraction of radiotherapy (RT) treatments use intensity modulated techniques as the latter allow to deliver a highly conformal dose distribution to the tumor while sparing the surrounding normal tissues. The most striking example being for organ at risk (OAR) with a concave shape [1]. However, as it is well known, such rotational treatments deliver lower doses to a greater volume of surrounding normal tissues. As usual, these complex methods require quantitative verifications in order to assess the integrity of the delivery: the measured and planned dose distributions need to be compared. A new feature compare to classical conformational RT (3DCRT) arises from the necessity to achieve these controls or quality assurance (QA) before the treatments. Indeed, usual in vivo dosimetry is no longer realistic in IMRT due to technical issues related to the beam modulation. Then, the pre-treatment QA are in general made with a dedicated phantom [2,3] and a gamma-analysis is performed. However, these measures are time-consuming (humanly and in machine time) and if the result is not satisfying we need to restart the (inverse) planimetry from the beginning.

From this ascertainment, the society Radiation Therapy Consulting (RTC) had developped a software called PQAi, based on machine learning which allows to evaluate the deliverability of IMRT plans ie to predict whether a plan will pass or fail the QA. In this study, carried out at the university hospital La Timone, we quantitatively evaluate the predictive power of this new software. The software prediction is insured by a first learning phase in which several complexity indices (CI) are computed from the RT plans, in order to determine the degree of complexity of each plans. This complexity is then correlated to the result of the QA to determine which kind of plans are deliverable or not in a second predict phase.

Our focus is on the VMAT (Volumetric-modulated arc therapy) technique [4–7] and we explore the effect of several parameters related to the quality of the prediction. The main one being the learning criteria used for the gamma analysis. We also study the tumor location and the effect of a mirror accelerator.

I Introduction

The QA of IMRT plans can be done in several ways: with 3D phantoms (*Delta*⁴ Phantom, ArcCHECK) like in our case, with planar phantoms (matrix, octavius) or films, using the portal imager (EPID) or with a transmission chamber ¹ (dolphin, IQM, *Delta*⁴ discover). The two last options are less time-consuming as there is no need to position a detector but we still need to carry out the treatment to measure the QA outlining the advantages of a predictive software. Let us mention that another possibility uses the log files [8].

To assess the level of modulation of IMRT plans, different metrics combining several beam parameters have been introduced in the literature. These complexity metrics are either based on the fluence [10–15] or on the geometry [15–19] (aperture based). The software considered in this work combines several of these complexity indices. Each plan is then located in a space with a dimension fixed by the number of complexity indices computed by the software. During the learning part, the software learns how to differentiate plans in this space. This learning phase allows to discriminate plans at the level of the passing (OK) or the failing (KO) plans, the values of gamma passing rate (GPR) or the mean and max gamma values. After the learning, the software is able to provide an advice to determine if a plan passes or not the QA. Depending on this advice we conclude that the plan is: (i) deliverable, (ii) not deliverable and we need to restart the planimetry or (iii) not sure to be deliverable and we need to perform the QA.

It is worth to note that up to our knowledge only few works attempt to predict the result of QA for IMRT plans [20]. However, some of the complexity metrics have already been incorporated into TPS, with the goal of producing plans that are more likely to pass the QA [21, 22]. In all cases, only weak correlations have been found between passing rates and these metrics because only one single metric was considered at a time with the expectation that the latter will explain alone the passing rates. Therefore, an algorithm that integrates different complexity metrics and is able to predict IMRT QA passing rates has yet to be developed.

¹This kind of detectors can be used for pre-treatment QA as well as for in-vivo dosimetry.

II Materials and methods

Two techniques can be differentiated in RT: the 3DCRT and the IMRT. In the first one, the field shape of a beam is determined by the jaws of the accelerator such that only large rectangular fields are available. Moreover, the dose prescription is in one point and the planification method is direct. In the more recent IMRT technique, the spirit is completely different. Indeed, the field shape is determined by the multi-leaf collimator (MLC) which allows to achieve more complex and smaller fields with a better conformation to the tumor volume as well as a better protection of the OAR. The prescription is now related to a volume and it is worth to note that the dosimetry is very different due to the inverse planification process that is, the desired dose distribution is fixed from the beginning and the TPS algorithm tries to find the better way to realise it by varying the different available parameters.

In this study, we will only consider IMRT techniques as the latter require QA controls before the treatment. There are different types of IMRT techniques. First, the static one that includes the step and shoot (*S&S*) mode [23–26] and the sliding window mode [27–29]. These techniques use several beams at fixed angulation of the gantry. In the *S&S* mode, the MLC is motionless for a given beam which have then a fixed field shape to achieve the dose conformation. In the sliding window mode, the field shape is not fixed but varies continuously to achieve the dose conformation. This method is slightly faster than the previous one. The other type of IMRT technique is the VMAT which contrary to static IMRT uses continuous arcs instead of fixed beams. We then need to fix the number of arc, their starting and ending points as well as the angulation of the couch. The VMAT technique is faster in the delivery as compared to static IMRT. In this study, we will only consider the VMAT technique because of its intrinsic larger complexity leading to QA that are less likely to pass. Moreover, static IMRT will certainly be replaced by the faster VMAT technique in the future.

II.1 Optimisation phase

The treatment planning system (TPS) used during the planification process is *Pinnacle*³ (*Philips Medical Systems, USA*) version 16.2 coupled to the Record and verify (*R&V*) system *MOSAIQ* (*Elekta, Sweden*) version 2.64. The optimisation algorithm used

is the DMPO (Direct Machine Parameter Optimisation) for IMRT and SmartArc (based on DMPO) for VMAT. Both of them are based on the collapsed cone [30] method. Furthermore, we have also used a recent module called auto-planning [31,32] that allows in principle to find better solutions to the inverse planning problem and then more complex plans which are less likely to pass the QA. This is an important point as we need a sufficient percentage of failing plans to train the software during the learning phase (see table 4).

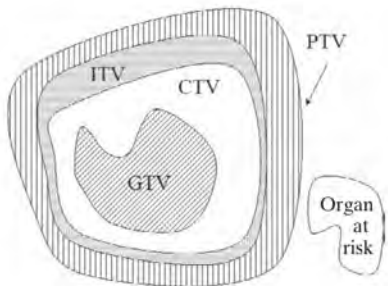


Figure 1: Schematic view of the different target volumes [33]. First, the gross tumor volume (GTV) that corresponds to the palpable or visible extent of the tumor. Then, the clinical target volume (CTV) taking into account the tissue volume that contains a demonstrable GTV and/or sub-clinical microscopic malignant disease. Finally, the planned target volume (PTV) is defined to take into consideration the net effect of all possible geometrical variations of the beams, in order to ensure that the prescribed dose is actually absorbed in the CTV.

As a first step, we intend to redo the planimetry of treated patients² and in order to be as close as possible to the clinical routine we compare our result to the clinical dose-volume histogram (DVH). In addition to respecting the ICRU (International Commission on Radiation Units and Measurements) constraints on the OAR, which are listed in table 1 for the most relevant ones, the new DVH should be equivalent or even better than the old one. For the targets (see Fig. 1 for definitions) that is for the planned target volume (PTV), at least 95% of the volume should receive at least 95% of the prescribed dose. This requirement on the PTV coverage being crucial for tumor control. In addition, the max dose should be less than 107% (hot points) of the prescribed dose. These requirements are outlined in Fig. 10. Once the dosimetry is clinically acceptable, the beam geometry (RT plan) and the dose computation (RT dose) are exported on the scan of the phantom that will be used for the QA. In this way, it will be possible to compare the planned dose with the measured one. Note that the RT plan is also exported to the R&V system in order to deliver

²Reoptimised plans were first performed in static IMRT.

the correct plan.

Location	OAR	Constraints
Pelvic-prostate	Rectum	$V_{50} \leq 50\%$
		$V_{70} \leq 20\%$
		$V_{74} \leq 5\%$
	Bladder	$V_{76} \leq 2\%$
		$V_{60} \leq 50\%$
		$V_{70} \leq 25\%$
H&N	Femoral head	$V_{80} \leq 2\%$
	Hail	$V_{50} \leq 10\%$
		$V_{40} \leq 350cm^3$
H&N	Spinal cord	$D_{max} \leq 40$ Gy
	Spinal cord + 5 mm	$D_{max} \leq 50$ Gy
	Oral cavity	$D_{moy} \leq 30$ Gy
	Mandible	$V_{70} \leq 2\%$
	Parotid	$D_{moy} \leq 26$ Gy
	Temporomandibular joints	$D_{max} \leq 55$ Gy
	Larynx	$D_{moy} \leq 45$ Gy
	Esophagus	$D_{moy} \leq 45$ Gy

Table 1: Main constraints to the OAR applied at the university hospital *La Timone* for the pelvic-prostate and H&N locations, following the ICRU recommendations. Let us remind that an OAR is an organ whose sensitivity to radiation is such that the dose received from a treatment plan may be significant compared with its tolerance, possibly requiring a change in the beam arrangement or a change in the prescribed dose [33]. The above constraints are on the mean dose ($D_{moy} \leq xx$ Gy), the max dose ($D_{max} \leq xx$ Gy) or on the dose to a fraction of the OAR volume ($D_{xx} Gy \leq yy$ %).

Let us now turn to some details about the plans considered in this study. We mostly focus on the prostate alone, pelvic-prostate and head and neck (H&N) tumor locations³. Note that we consider separately the prostate alone and the pelvic-prostate regions as the complexity is in general larger for the second one. This is due to the pelvic nodes that should be treated with larger fields. Typical CT images corresponding to these locations are presented in figures 2 and 12 respectively where the different PTV and the main OAR are outlined in the sagittal, transverse and coronal planes. Typical dose prescriptions at the university hospital *La Timone* are listed in table 2. Let us mention that we have also consid-

Location	Dose per fraction (Gy/fr)	phase 1 (Gy)	phase 2 (Gy)
Prostatic loge	2	66	/
Pelvic-prostate	2	46	34
	1.8	50.4	28.8
H&N	2	50	20
	2	70	/

Table 2: Typical dose prescriptions as a function of the location. The number of fractions is easily determined from the dose per fraction and the total dose.

³ These locations have been chosen according to their complexity. Moreover, the pelvic-prostate and H&N regions were treated in S&S before such that this study allowed to implement their treatment in VMAT for the clinical routine

ered few plans from other locations like the brain, the lung and the abdomen regions. Treatments plans can be done in a sequential way that is in several phases (mainly two) where a boost (larger prescribed dose) is realised in the PTV of the second phase (generally the tumor). Indeed, the PTV of the first phase generally involve regions like the nodes which are only treated in a prophylactic way. Another possibility is an integrated boost ie a treatment in one phase where the boost is realised at the same time. In other words, the PTV with different prescribed dose are treated with a different dose by fraction and then different radiobiological effects. Note also that for simplicity, we have simplified few clinical prescriptions for the pelvic-prostate region usually done in three phases to prescription in two phases. Indeed, the phase one and two with the same PTV have been merged and the phase three became the phase two. Moreover, all phases have been converted to an equivalent dose of 2 Gy/fr. For instance, we have: 25 Gy in 10 fr (2.5 Gy/fr) \cong 28 Gy in 14 fr. This is easily understandable using the quadratic linear model:

$$\frac{D_1}{D_2} = \frac{n_1 d_1}{n_2 d_2} = \frac{\alpha/\beta + d_2}{\alpha/\beta + d_1}, \quad (1)$$

where D_i are the total doses, d_i the dose by fractions and n_i the number of fractions.



Figure 2: Typical CT images of the pelvic-prostate region in the coronal, transverse and sagittal planes respectively. The three relevant PTV are the nodes (dark blue), the seminal vesicles (dark blue, close to the prostate) and the prostate (sky blue). The boosted region, treated in phase 2, only contains the prostate PTV while the nodes and the seminal vesicles are usually just treated in the first phase. The most relevant OAR are the rectum (brown), the bladder (yellow) and the femoral heads (green).

II.2 Quality assurance measurements

In this analysis, we have considered 452 plans in total: 279 of them have been optimised during the internship while 84 correspond to a preliminary work and 89 have been optimised by the physics team for clinical routine. The detail of the number of plans by location and machine is reported in table 3. Note that we keep aside 53 plans for the predict phase and use the others for the learning phase.

Location	Synergy 2	Synergy 1	Total
Total	282	170	452
Prostate alone	48	9	57
Pelvic-prostate	87	15	102
H&N	117	116	233

Table 3: Number of plans used in the study as a function of the location and machine.

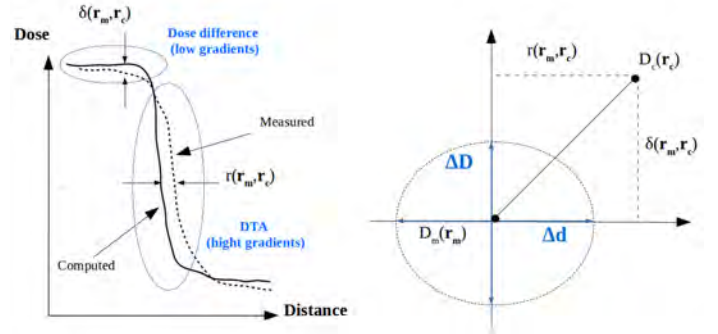


Figure 3: Definition of the dose difference and DTA (left-hand side) and schematic view of the gamma-index computation in two spatial dimensions (right-hand side). The extension to the 3D case is straightforward.

As already mentioned, the measured and planned dose distributions need to be compared to assess the integrity of the delivery of VMAT plans. An objective and quantitative way is provided by the so-called Gamma index [9]. The latter combines both the percentage dose difference and the distance to agreement (DTA). Indeed, in the low gradient regions (see figure 3), the two dose distributions can be directly compared and an agreement condition could be defined as the difference between the measured and calculated doses ie $\delta(\mathbf{r}_m, \mathbf{r}_c) = |D_m(\mathbf{r}_m) - D_c(\mathbf{r}_c)| \leq \Delta D$. However, in the high gradient regions, a small spatial difference between the two distributions induces a large dose difference such that direct dose differences are less relevant and the DTA should be introduced. The DTA is defined as the distance between the measured data point and the nearest point in the calculated dose distribution that exhibits the same dose value ie $r(\mathbf{r}_m, \mathbf{r}_c) = |\mathbf{r}_m - \mathbf{r}_c| \leq \Delta d$. An usual choice being $\Delta D = 3\%$ and $\Delta d = 3 \text{ mm}$. The gamma-index function is defined as follow:

$$\gamma(\mathbf{r}_m) = \min \left[\sqrt{\frac{r^2(\mathbf{r}_m, \mathbf{r}_c)}{\Delta d} + \frac{\delta^2(\mathbf{r}_m, \mathbf{r}_c)}{\Delta D}} \right]. \quad (2)$$

Then, the relevant condition is now defined by $\gamma(\mathbf{r}_m) \leq 1$ ie the acceptable region in the $r - \delta$ plane is located inside the ellipse with axes ΔD and Δd (see figure 3). A gamma passing rate (GPR) larger than 95% ie the percentage of the measured points

having a γ -index smaller than one is required for the plan to be accepted ⁴. Let us remark that the GPR do not completely defines the distribution of measured points such that, it could also be interesting to consider the gamma mean and gamma max values (see figure 14). Moreover, we should separate the local and global gamma analysis which have a different normalisation for the reference dose used to compute the dose difference. In the local analysis that we use in this study, the percentage dose difference is computed from the calculated dose at each measured points such that the tolerance changes with the dose level. On the contrary, in the global analysis the percentage dose difference is the same for each points no matter is the calculated dose and is fixed from the dose max or the prescribed dose. Then, the local gamma analysis is in general more restrictive on the low doses.

The QA are carried out on two mirror accelerators platform synergy beam modulator (*Elekta, Sweden*): the synergy 1 and 2 where the latter is modeled in the TPS while the former is considered as mirror. The measures are achieved with the dedicated *Delta*⁴ (*Scandidos, Sweden*) diode array phantom coupled to the software scandidos. The phantom consists of 1069 p-type silicon diodes in a crossed array (two orthogonal plans) inside a PMMA phantom. Three dimensional dose distribution can be reconstructed from the measured dose in the planes. The percentage of passing and failing plans are summarised in table 4 for two sets of criteria that will be considered in the next. An example of QA result is presented in appendix B. It is worth to note that the QA result ie the OK-KO result, the GPR and the mean and max values are obtained for each prescriptions and arcs while the clinical decision is only taken according to each phases.

Criteria	Synergy 2	Synergy 1
A: 3% - 3mm - 97%	7.09 %	15.79 %
B: 2.5% - 2.5mm - 95%	20.92 %	35.43 %

Table 4: Percentage of failing plans used in the study as a function of the criteria used in the gamma-index analysis as well as the machine (synergy 1 or 2).

The highly conformal dose distribution in VMAT treatments is achieved by varying: (i) the rotation speed of the gantry, (ii) the dose-rate variation and (iii) the position and speed of the MLC leaves. The

⁴Note that we use a threshold of 20% in order to exclude low doses which are subject to larger uncertainties.

deliverability of a treatment ie the result of the QA might be affected by the degree of modulation or in other words by the plan complexity and is due to the mechanical limitations of the accelerator and to the accuracy limitations of the dose calculation in the TPS. This complexity arises from beam modulation as a large number of small and/or irregularly shaped beam segments are needed to achieve high dose conformity. These small segments induce higher dose uncertainties than those used in large fields 3DCRT and necessitate an accurate modelisation of MLCs ie of the leaf transmission and the interleaf leakage.

II.3 Software PQAi

We now turn to a short description of the software functioning (see figure 4). We do not intend to discuss every aspects in detail but rather to clarify the main issues in order to have a sufficiently clear picture in mind. As already mentioned, the software is based on a machine learning algorithm which differentiates plans according to their OK-KO status, the GPR or the gamma mean and max values. The learning phase is insured by the computation of several complexity indices that allow to estimate the plan complexity ie the level of modulation of IMRT treatments.

Fluence-based	Geometry-based
PA, PI, PM [10]	MI [16]
MCS [11, 12]	PIMV [17]

Table 5: List of the main complexity indices used by the software.

The complexity metrics can be classified in two groups, which are based on the fluence or on the geometry (aperture based). The main indices computed by the software are summerised in table 5. For more details on few of them, we report to appendix C. Let us just note that a limitation of fluence-based approaches is their insensitivity to the degeneracy of fluence maps. In other words, many small beams may lead to the same index value than a large beam even though the plan with small beam segments is more complex. On the contrary, aperture-based approaches assess beam complexity by analysing directly the beam aperture.

This is not a simple task to describe the plan complexity using one or even few indices. Indeed, most of them are correlated and it is not obvious to extract one relevant combination because the latter would depend on the plans used in the learning phase. As a consequence, the software compute a large number

of complexity indices and combines them in order to learn how to differentiate them in a space where the dimension is fixed by the number of complexity indices.

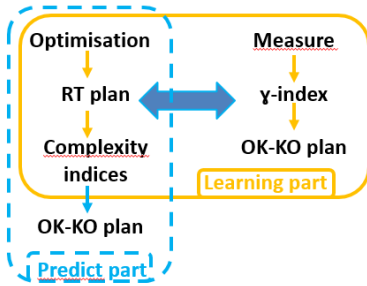


Figure 4: Schematic view of the software functioning.

The software functioning is divided into two parts: the learning and predict part (see figure 4). Among other things, the learning depends on the quality of the modelisation in the TPS, the complexity of the plans obtained during the optimisation phase, the phantom used for the QA that is the response of the detector and of course the accelerator. As we will explore in section III the software prediction may be based on the OK-KO status, the GPR, the mean gamma or the max gamma of the plans. Therefore, the plans used in the learning are implemented in the software with all of these parameters. Note that the GPR, the mean and max gamma are given for each arcs and prescriptions (phases) while the OK-KO status is only for each prescriptions (not for the whole plan). We visualise the data according to the criteria A or B as all of the above parameters depend on this choice. The PCA decomposition gives the better plan to separate the data (see figures 5, 6 and 7). We then have four different learning models according to the OK-KO status, the GPR, the gamma mean or the max.

		QA result	
		OK	KO
Software prediction	OK	True-positive (TP) $P_{TP} = \frac{TP}{TP + FN} = 1 - \alpha$	False-positive (FP) $P_{FP} = \beta$
	KO	False-negative (FN) $P_{FN} = \alpha$	True-negative (TN) $P_{TN} = \frac{TN}{TN + FP} = 1 - \beta$

Table 6: Definition of the false-positive (FP) and false-negative (FN) predictions of the software with their probabilities.

To appreciate the validity of the prediction, we can compute the sensitivity P_{TP} and the specificity P_{TN} which are defined in table 6. Ideally, these probabilities should be equal to one that is, the probabilities of the type 1 and type 2 errors α and β should tend

to zero which is unfortunately never the case in statistics. The precision of the prediction is related to the number of "false-positive" (FP) ie a plan that is predicted by the software to pass the required criteria but fails the QA. Indeed, this is a very dangerous case that should be avoid as the plan is expected to be deliverable but it is not. A "false-negative" (FN) is less problematic as we redo the planimetry even if the plan is deliverable.

For the model based on the OK-KO status, a ROC analysis (sensibility vs specificity curve) is performed in order to determine the appropriate complexity threshold value ie the number of desired FP. We note that the uncertainty of the ROC curve (cf figure 5) is reduced when the number of plans increase outlining the fact that a large number of plans is needed for a correct learning. Otherwise, for models based on GPR, gamma mean or max, the leave one out (LOO) method ⁵ allows to appreciate the robustness of the prediction (see figures 6 and 7) by giving the predicted versus the measured value with an interval.

Finally, having a learning model at disposal, the predict part gives advices (predictions) according to the OK-KO status, the GPR, the gamma mean or the gamma max. The predicted parameter is the same than the one used in the learning model and the advice is presented as a mean value and an interval at 95% (see section III). This advice helps to take a clinical decision. For instance, we can choose to exclude plans with a mean gamma larger than 0.5 or a max gamma larger than 2. Note that the threshold of FP is fixed to 5%, according to the ROC curve, but it can be reduced if necessary.

III Results

Let us now turn to the results of our study. Having a sufficient number of plans ie a large statistics, we expect to bring a quantitative answer to several questions. Indeed, in order to explore the predictive power of the software, we will consider several parameters that may influence the result of a QA prediction ie the learning quality. The main one being the criteria used during the learning phase to discriminate plans (criteria A or B). This is the first step to build a predictive model based on the QA data as a different criteria may better differentiate the plan complexity. Note that the learning criteria is more restrictive than

⁵This cross validation method is used for continuous parameters in opposition to the ROC analysis

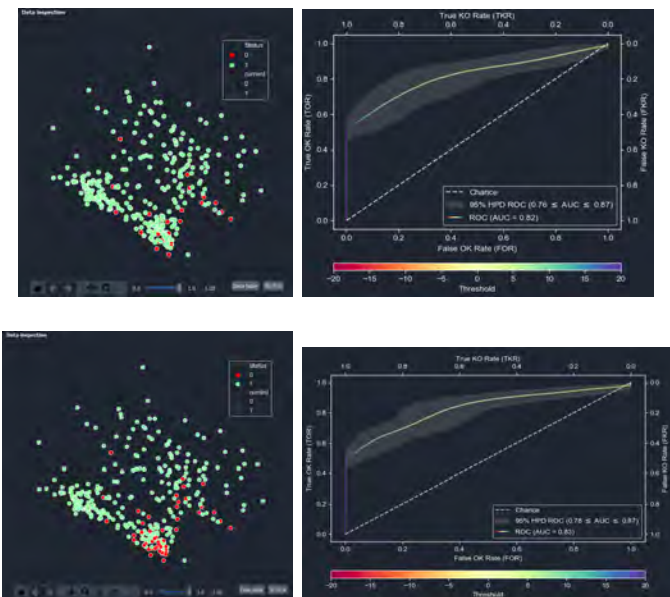


Figure 5: PCA decomposition of the data based on OK-KO status for the criteria A and B (upper and lower left-hand side respectively) as well as the corresponding ROC curves (right-hand side).

the clinical agreement criteria. More precisely, we increase the passing rate (criteria A) or decrease the percent dose difference and DTA (criteria B). Then, we will also consider the effect of the tumor location and the impact of the accelerator (synergy 1 or 2) used for the QA.

III.1 Learning criteria

Let us start by considering the effect of the learning criteria. To that end, we focus on the synergy 2 as it will become more obvious in the next section. We construct two kind of learning models according to the criteria A and B (see figures 5, 6 and 7). As explained before, for these criteria, we build four different models based on the OK-KO status, the GPR and the mean or max gamma. In the ROC analysis for the OK-KO status (cf figure 5), we found an area under the curve (AUC) of 0.82 and 0.83 for the criteria A and B respectively. Then, the model quality is a bit better for criteria B⁶. Note that the uncertainties on the AUC are also given by the software as it can be seen in the figure. Similarly, the cross validation curves⁷ seem to be better in general for model B. Indeed, the associated R^2 -score which allows to assess

⁶A more restrictive criteria of 2.5%-2.5mm-97% confirms this tendency with an AUC of 0.84.

⁷Note that the learning have been done according to the results per fractions but it could be interesting in the future to also consider the results per beams.

the model quality (a lower R^2 -score means a better model quality) is similar for two criteria A and B for the GPR model while it is better for the gamma mean and max models (see table 7). Note that the GPR may not give a proper account of the gamma distribution (cf figure 14) as the latter could be centered on zero or on one and still leads to the same GPR value. Then, the gamma mean and max values could be of great help to take a decision in the predict part, as explained below. We conclude that a more restrictive set of criteria allows to better differentiate the plan complexity (more KO plans and larger differences in the QA results between plans). Then the learning models with criteria B will be used in the next.

Model	Criteria A	Criteria B
GPR	0.39 (0.22,0.51)	0.40 (0.18,0.54)
Gmean	0.42 (0.31,0.50)	0.38 (0.24,0.48)
Gmax	0.35 (0.22, 0.44)	0.33 (0.20,0.43)

Table 7: List of the R^2 -score associated to the GPR, gamma mean and max models with criteria A and B. The brackets correspond to the interval at 95%.

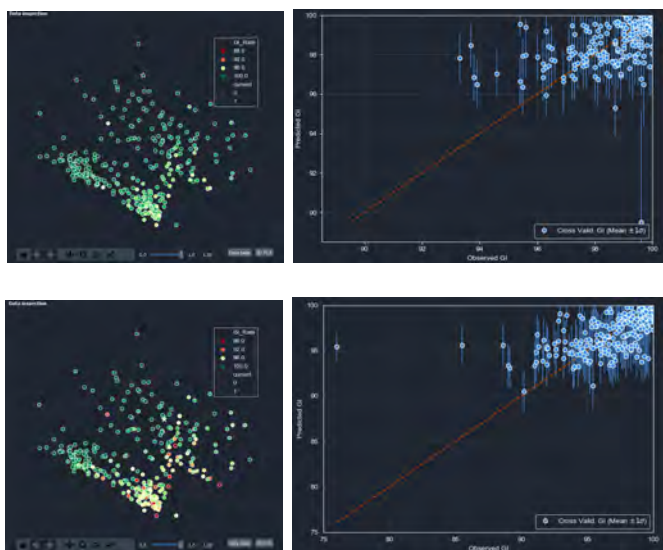


Figure 6: Same as in figure 5 but for the GPR model. Instead of a ROC analysis, the robustness of the model is now evaluated with a cross validation curve.

Let us now turn to the predict part to appreciate the validity of the learning model. We have 53 plans, not involved in the learning part, that can be used to test the predictive power of the software. Following the advices given by the latter, we can determinate for instance, with the OK-KO status model, the number of FP and FN. Unfortunately, these results are not yet available and we hope that they will arrive soon. Then, to give an example of what kind of analysis

can be done with the test plans in the predict part, we present two prototypes of figure. First, the prediction result associated to the gamma mean model is in figure 8 where we can see the measured value versus the prediction (with its interval) ie the software advice. From this figure, we will be able to extract the percentage of the test plans that are compatible with the diagonal line, which provides a good confidence on the predictive power of the PQAi software. Finally, we will use the GPR model for the figure 8 where we can see the prediction of the GPR (with its interval) for each tested plan. If the lowest part of the interval is larger than 95%, we accept the plan otherwise we refuse it. From this figure, we will be able to determine the percentage reduction of the QA.

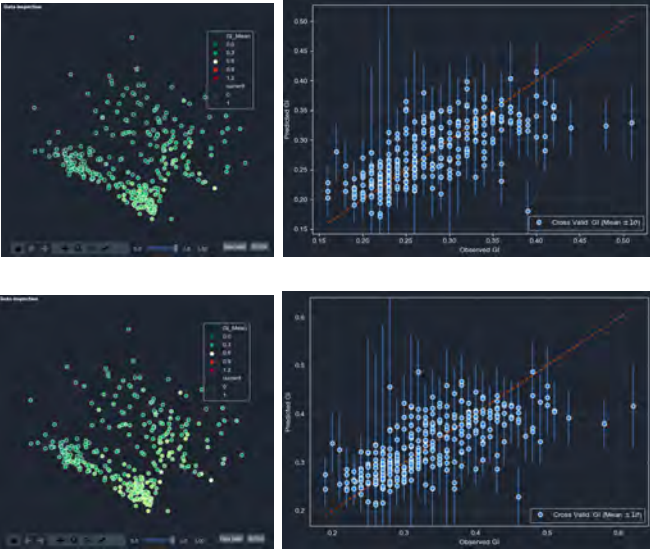


Figure 7: Same as in figure 6 but for the mean gamma model.

III.2 Tumor location and mirror accelerators

We now consider the tumor location and the machine used for the QA as an influence parameter of the learning phase. Following the results of the previous section, we only consider the criteria B now. The tumor location is related in some way to the plan complexity and exploring several tumor locations allows to expect a larger variety of plan complexity that will increase the efficiency of the learning phase. Note that we pursue a preliminary study which have only considered the pelvic-prostate region (cf section II.2). We have considered several set of plans according to a particular location and then constructed the corresponding learning models. We conclude that the

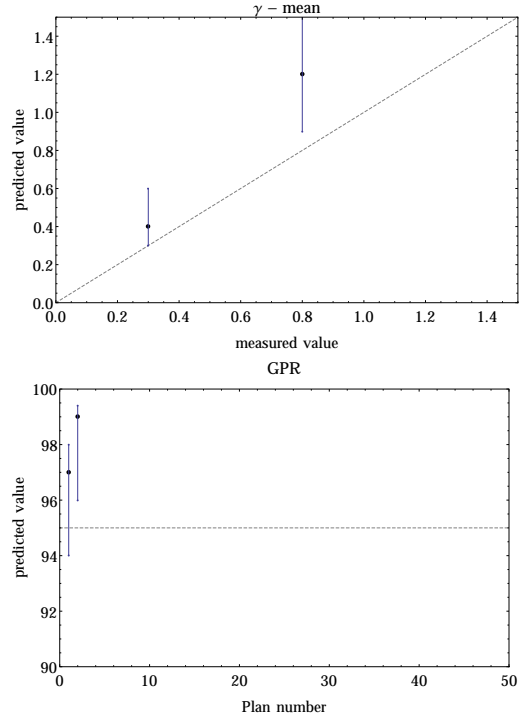


Figure 8: Prototype figures for the predicted γ -mean value versus the measured one (top) and predicted GPR value with its interval (software advice) (bottom) for the test plans.

choice of the tumor location is not really probant as the data points are in general located in the same region. Indeed, the important thing is the degree of modulation which can be similar for different tumor locations. However, as expected we found that the degree of modulation and then the quality of the learning model is smaller for the prostate alone. Moreover, in some cases, the percentage of failing plans is too small which also affect the prediction.

The last influence parameter that we consider is the machine used for the QA. Indeed, we can also ask to what extend the prediction is affected if some of the QA are done with a mirror accelerator because the two accelerators are slightly different while the modulation in the TPS remains the same for both of them. Then, as for the location, we select the data according to the accelerator and construct three models: one with the plans associated with the synergy 1 and 2, one associated with the synergy 2 only and another with the synergy 1 only. As expected, we find a better result for the synergy 2 only model because it is the one modeled in the TPS. In the ROC analysis for instance, the AUC is respectively 0.83, 0.78 and 0.65 using the synergy 2 only, both machines or the synergy 1 only (see figure 9).

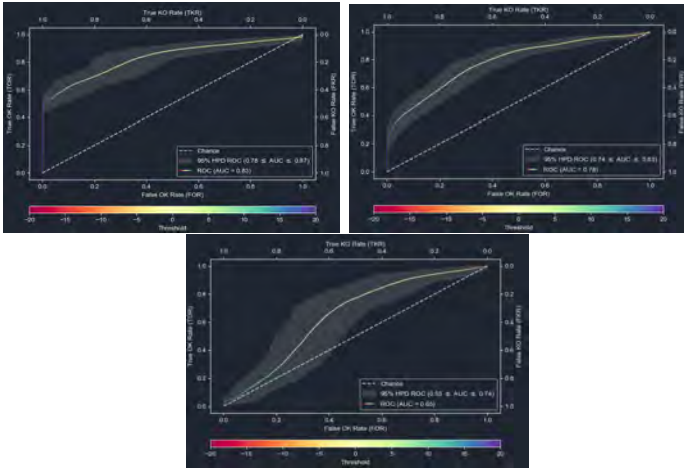


Figure 9: ROC curves for the synergy 2 only (top left), for both machines (top right) and for the synergy 1 only (bottom) models with criteria B.

IV Conclusion and outlooks

In this analysis, we have studied the predictive power of a new software called PQAi which is based on machine learning and allows to predict the results of QA controls in VMAT treatments.

We have explored several parameters that may influence the learning process of the software like the learning criteria, the tumor location and the machine used for the QA. We found that more restrictive learning criteria are needed to better differentiate the data which result in a more efficient predictive model. Indeed, using the test plans and following the advices that arise from the predict part of the software, we will be able (when data will be available) to determine the percentage reduction of the number of QA that should be carried out for criteria A and B respectively. For the tumor location, we found no substantive effect up to the inherent complexity of each anatomical region. For instance the robustness of the model is similar for the pelvic-prostate and H&N regions while it is lower for the prostate alone. Finally, for the machine used for the QA, as expected we found that the prediction is better for the synergy 2 which is the one modeled in the TPS.

It is worth to remind that the software PQAi is still in a development phase such that many points could be improved. Up to now, the results of the QA are entered manually in the software (possibilities of error and time-consuming) but it will possibly change in the future with a direct extraction of the data from the software carrying out the QA. Moreover, the number of plans needed for a sufficiently good prediction

model could be interesting to study as increasing the number will of course increase the quality of the prediction. Such possibility, to vary the number of plans in order to refine the learning model, is not yet implemented in the software and we had not enough time to study it. In the spirit of having a more precise prediction, a more restrictive criteria of 2%-2 mm could be interesting to analyse and we will certainly explore it in the future.

As an outlook of this study, one possible future for the software could be to implement it into a TPS in order to directly select plans that are more likely to be deliverable. Furthermore, we should mention that such predictive QA tool may also be used in a near future for adaptative RT (new plan adapted for each days of treatment). Furthermore, PQAi could also be used for a continuous improvement of practices by gradually increasing the plan complexity and quality still having QA that pass the clinical criteria.

	QA reduction	OK (accepted)	KO (replanned)
Criteria A	65.4%	44.2%	21.2%
Criteria B	55.8%	40.4%	15.4%
Criteria C	55.8%	32.7%	23.1%

Table 8: Percentage of QA reduction, accepted (OK) and replanned (KO) plans.

Aknowledgements

First, I would like to thank Jean-Marc Nigoul and Rashid Oozeer that gave me the opportunity to do this internship and for their help during these six months. Many thanks Jean-Marc for all the discussions we had, I sincerely had the impression to be part of the physics team, especially when we start to use auto-planning.

I also would like to thank the dosimetrists Manu, Laetia and Sam because they were always patient and available to answer my questions (even the stupid ones) and they taught me a lot about how to realise a dosimetry in practice. Thanks to them, I already know a lot of little things that will be very usefull in the future. Obviously, I also have to thank them for their huge patience because it requires a really long time to carry out the QA measures for almost 300 plans.

I also need to thank all the physicists because they never hesitate to share their experience with me. In particular, I would like to thank Emeline for giving me the opportunity to visit the interventional imaging service (even if it was very early in the morning), Bardia to visit the nuclear medicine service and Jean-Marc to visit the gamma-knife service. I was very privileged to discover the profession of medical physicist in this way.

Finally, I have to thank all the DQPRM students, Adrien and Mehdi which did'nt leave me sing alone during our karaoke with Bardia (except Fanny and Estelle because their favorite hobby was to choose very "complicated" songs for us without telling us something).

A Isodose contours and Dose-Volume Histograms

In this appendix, we compile few figures in order to illustrate the optimisation process. Figure 10 shows an example of isodose contours ie the line of equal dose, for a typical H&N plan. This plan contains three different PTV: the nodes, a lymphadenopathy and the tumor . The prescribed doses for the latter are respectively 56, 63 and 70 Gy. Following the ICRU recommendations, the relevant isodoses show in the figure then correspond to 95% of these prescribed doses that is 53.2, 59.85 and 66.5 Gy.

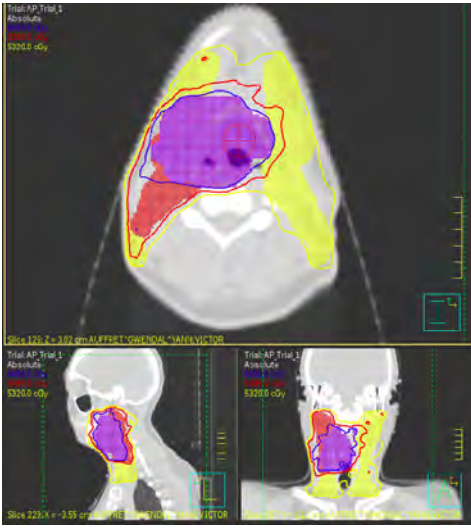


Figure 10: Isodose contours for a typical H&N plan.

Figure 11 gives an example of DVH for the above plan that is the fraction (y-axis) of the PTV and OAR that receive a give dose (x-axis). Such DVH is used to check that a plan respect or not the ICRU constraints on the targets and OAR. Finally, figures 12 show the typical PTV and the main OAR for the H&N region.



Figure 11: Example of DVH for the H&N location corresponding to the same plan than in figure 10.

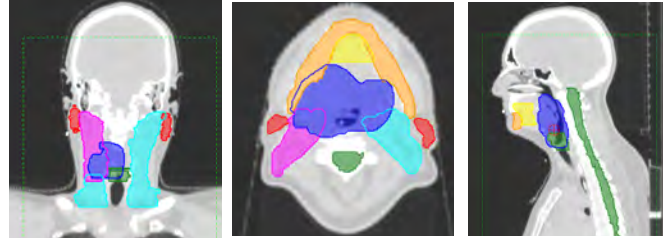


Figure 12: Same as in figure 2 but for the H&N region. Three different PTV are outlined: the nodes (sky blue), the lymphadenopathy (purple) and the tumor (dark blue). Also visible are the main OAR: the mandible (orange), the parotids (red), the oral cavity (yellow), the larynx (dark green) and the spinal cord (brown).

B Example of QA results

In this appendix, we shortly present an example of QA results obtained during the phase of measurement. It is worth to note that the QA measurements are done separately for each arcs of all phases while the GPR condition corresponds to the total dose in a phase. Indeed, despite the phase may fullfill the GPR condition, several arcs may have a GPR smaller than 95%. Figure 13 shows the details of the local gamma analysis with the distributions of the percentage dose difference and DTA. Finally, figure 14 shows the distribution of γ -index from which we extract the GPR as well as the gamma mean and gamma max values.

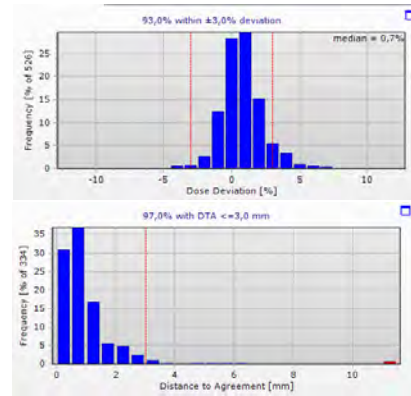


Figure 13: Distribution of measured points with a given dose difference as compare to the planned dose (left-hand side) and same thing for a given DTA (right-hand side)

C Complexity indices

In this appendix, we shortly present three complexity indices used in this analysis.

- **Number of monitor units (MU)**

A commonly used indicator of plan complexity is the total plan MU ie the sum of the MUs from

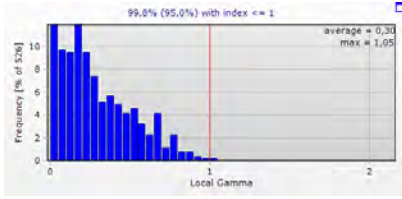


Figure 14: Gamma analysis result with the number of measured points with a given gamma-index. Also shown is the percentage of points with a gamma-index smaller than one (GPR) according to the usual condition (dose difference $\leq 3\%$ and DTA ≤ 3 mm).

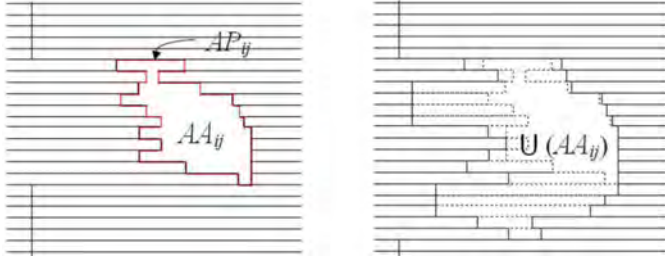


Figure 15: Example of a beam aperture j of beam i . The perimeter of the latter being AP_{ij} while the area is AA_{ij} . The quantity $U(AA_{ij})$ is the union area of all apertures of beam i .

all beams [10]. A larger number of MU could be the sign of a larger complexity of a plan. However, this quick measure of beam modulation does not provide any distinction among the different forms of plan complexity: small beam segments, irregularly shaped, narrow beam segments and large number of beam segments.

- **Modulation index (MI)**

The modulation index [7] is the best known fluence-based complexity metric. It uses the fact that modulation increases as dose conformity increases. Then, MI measures the variations of photon fluence between neighboring pixels of a given beam.

- **Plan average beam area (PA)**

We now turn to a geometry-based index. For the j^{th} aperture A_{ij} of beam i (see Fig. 15), the aperture area (AA) is computed as the total area of all MLC openings within A_{ij} . The beam area can be computed by summing over all aperture j with the proper ponderation of the MU as follow:

$$BA_i = \frac{\sum_j MU_{ij} \cdot AA_{ij}}{MU_i}, \quad (3)$$

where MU_{ij} and MU_i are the MU of segment A_{ij} . Finally, the plan area (PA) is obtained [10] by summing over all beams i , again with the proper

ponderation of the MU:

$$PA = \frac{\sum_i BA_i \cdot MU_i}{MU_p}, \quad (4)$$

where MU_p is the total MU in the plan. A smaller PA indicates a larger plan complexity as it generally corresponds to smaller fields. Similarly, we can compute the plan average beam irregularity (PI) using the aperture perimeter (AP) defined in figure 15.

References

- [1] D. Azria, N. Ailleres, C. Llacer-Moscardo, M. Hay, J. Dubois and P. Fenoglietto, "Conformal intensity modulated radiation therapy for localized prostate cancer: Toward a new standard," *Cancer Radiother* **Vol. 13** (2009) .
- [2] B. Mijnheer, "Guidelines for the verification of IMRT," *ESTRO booklet 9* (2008) .
- [3] G. A. Ezzell, J. M. Galvin, D. Low, J. R. Palta, I. Rosen, M. B. Sharpe et al., "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee," *Med. Phys.* **Vol. 30** (2003) .
- [4] C. X. Yu, "Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy," *Phys. Med. Biol.* **Vol. 40** (1995) .
- [5] S. Crooks, X. Wu and C. T. et al., "Aperture modulated arc therapy," *Phys. Med. Biol.* **Vol. 48** (2003) .
- [6] C. X. Yu and G. Tang, "Intensity-modulated arc therapy: principles, technologies and clinical implementation," *Phys. Med. Biol.* **Vol. 56** (2011) .
- [7] S. Webb and D. McQuaid, "Some considerations concerning volume-modulated arc therapy: a stepping stone towards a general theory," *Phys. Med. Biol.* **Vol. 54** (2009) .
- [8] N. Childress, Q. Chen and Y. Rong, "Parallel/opposed: IMRT QA using treatment log files is superior to conventional measurement-based method," *J. Appl. Clin. Med. Phys.* **Vol. 1** (2015) .
- [9] D. A. Low, W. B. Harms, S. Mutic and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical Physics* **Vol. 25, No. 5** (1998) .
- [10] W. Du, S. H. Cho, X. Zhang, K. E. Hoffman and R. J. Kudchadker, "Quantification of beam complexity in intensity-modulated radiation therapy treatment plans," *Medical Physics* **Vol. 41, No. 2** (2014) .
- [11] A. L. McNiven, M. B. Sharpe and T. G. Purdie, "A new metric for assessing IMRT modulation complexity and plan deliverability," *Medical Physics* **Vol. 37, No. 2** (2010) .
- [12] J. M. Park, S.-Y. Park, J. C. Hyoungnyoun Kim, Jin Ho Kim and S.-J. Ye, "Modulation indices for volumetric modulated arc therapy," *Phys. Med. Biol.* **Vol. 59** (2014) .

- [13] C. K. McGarry, C. D. Chinneck, M. M. O'Toole, J. M. O'Sullivan, K. M. Prise and A. R. Hounsell, "Assessing software upgrades, plan properties and patient geometry using intensity modulated radiation therapy (IMRT) complexity metrics," *Medical Physics* **Vol. 38**, No. 4 (2011) .
- [14] M. Nauta, J. E. Villarreal-Barajas and M. Tambasco, "Fractal analysis for assessing the level of modulation of IMRT fields," *Med. Phys.* **Vol. 38**, No. 3 (2011) .
- [15] S. B. Crowe, T. Kairn, N. Middlebrook, B. Sutherland, J. K. B. Hill, C. M. Langton et al., "Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results," *Phys. Med. Biol.* **Vol. 60** (2015) .
- [16] S. Webb, "Use of a quantitative index of beam modulation to characterize dose conformality: illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy," *Phys. Med. Biol.* **Vol. 48** (2003) .
- [17] M. M. Coselmon, J. M. Moran, J. D. Radawski and B. A. Fraass, "Improving IMRT delivery efficiency using intensity limits during inverse planning," *Med. Phys.* **Vol. 32**, No. 5 (2005) .
- [18] K. C. Younge, M. M. Matuszak, J. M. Moran, D. L. McShan, B. A. Fraass and D. A. Roberts, "Penalization of aperture complexity in inversely planned volumetric modulated arc therapy," *Med. Phys.* **Vol. 39** (2012) .
- [19] S. B. Crowe, T. Kairn, J. Kenny, R. T. Knight, B. Hill, C. M. Langton et al., "Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results," *Australas. Phys. Eng. Sci. Med.* **Vol. 37** (2014) .
- [20] G. Valdes, R. Scheuermann, C. Y. Hung, A. Olszanski, M. Bellerive and T. D. Solberg, "A mathematical framework for virtual IMRT QA using machine learning," *Medical Physics* **Vol. 43**, No. 7 (2016) .
- [21] J. Llacer, T. D. Solberg and C. Promberger, "Comparative behaviour of the dynamically penalized likelihood algorithm in inverse radiation therapy planning," *Phys. Med. Biol.* **Vol. 46** (2001) .
- [22] S. Spirou and C. Chui, "A gradient inverse planning algorithm with dose-volume constraints," *Med. Phys.* **Vol. 25** (1998) .
- [23] J. Galvin, X. Chen and R. Smith, "Combining multileaf fields to modulate fluence distributions," *Int J Radiat Oncol Biol Phys.* **Vol. 27** (1993) .
- [24] C. X. Yu, "Design considerations for the sides of multileaf collimator leaves," *Int J Radiat Oncol Biol Phys.* **Vol. 28** (1994) .
- [25] J. Stein, T. Bortfeld, B. Dorschel and W. Schlegel, "Dynamic X-ray compensation for conformal radiotherapy by means of multi-leaf collimation," *Radiotherapy and Oncology* **Vol. 32** (1994) .
- [26] M. Keller-Reichenbecher, T. Bortfeld, S. Levegrun, J. Stein, K. Preiser and W. Schlegel, "Intensity modulation with the « step and shoot » technique using a commercial MLC : a planning study," *Int J Radiat Oncol Biol Phys.* **Vol. 45** (1999) .
- [27] D. Convery and S. Webb, "Generation of discrete beam-intensity modulation by dynamic multileaf collimation under minimum leaf separation constraints," *Phys. Med. Biol.* **Vol. 43** (1998) .
- [28] S. Spirou and C. Chui, "Generation of arbitrary intensity profiles by dynamic jaws or multileaf collimators," *Medical Physics* **Vol. 21** (1994) .
- [29] C. Chui, M. Chan, E. Yorke, S. Spirou and C. Ling, "Delivery of intensity-modulated radiation therapy with a conventional multileaf collimator: comparison of dynamic and segmental methods," *Medical Physics* **Vol. 28** (1994) .
- [30] A. Ahnesjö, "Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media," *Med. Phys.* **Vol. 16** (1989) .
- [31] D. Gintz, K. Latifi, J. Caudell, B. Nelms, G. Zhanga, E. Moros et al., "Initial evaluation of automated treatment planning software," *Journal of Applied Clinical Medical Physics* **Vol. 17**, No. 3 (2016) .
- [32] I. Hazell, K. Bzdusek, P. Kumar, C. R. Hansen, A. Bertelsen, J. G. Eriksen et al., "Automatic planning of head and neck treatment plans," *Journal of Applied Clinical Medical Physics* **Vol. 17**, No. 1 (2016) .
- [33] E. Podgorsak and al, "Radiation Oncology Physics: A Handbook for Teachers and Students," *IAEA* (2005) .