

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

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

¹Master 2 Ingénierie pour la Santé et le médicament, Option physique médicale, Grenoble

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

³Director and Medical Physicist, Radiation-Therapy-Consulting, Marseille

Introduction

Currently, an increasing fraction of radiotherapy treatments use intensity modulated techniques (IMRT) as the latter allow to deliver a highly conformal dose distribution to the tumor while sparing the 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 realise these controls or *quality assurance* (QA) before the treatments. Indeed, in vivo dosimetry is no longer realistic in IMRT due to technical issues related to the leaf speed. Then, the QA are in general made with a dedicated phantom.

However, these measures are *time-consuming* and if the result is not satisfying one needs to restart the (inverse) planimetry from the beginning. From this ascertainment, the society Radiation Therapy Consulting (RTC) had developped a software, in collaboration with the CEA, based on *machine learning* which allows to evaluate different IMRT plans by computing a passing rate ie the probability that a plan pass the QA. The main goal of this work being to *quantitatively* evaluate the predictive power of this learning method.

Materials and Methods

Materials :

- Two mirror accelerators (machines A and B) platform synergy beam modulator (Elekta, Sweeden)
- TPS Pinnacle³ version 16.2 with the module auto-planning (Philips, USA)
- Delta⁴ diode array phatom (Scandidos, Sweeden)
- R&V MOSAIQ (Elekta, Sweeden) version 2.64

Methods :

- Optimisation:** Realise the dosimetry of patients in VMAT (Volumetric-Modulated Arc Therapy) for several tumor locations.
- Measure:** Control the deliverability of these plans with the Delta⁴ phatom.
- Learning part:** Implement the data in the software ie the RT plan and the QA result (gamma passing rate, γ -mean and γ -max).

- Predict part:** Test the robustness of the prediction. For each tested plan, the software compute a passing probability from several *complexity indices*. Depending on this probability one conclude that a plan is:
 - deliverable without further check
 - not deliverable and one needs to restart the dosimetry
 - not sure to be deliverable and one needs to realise the QA

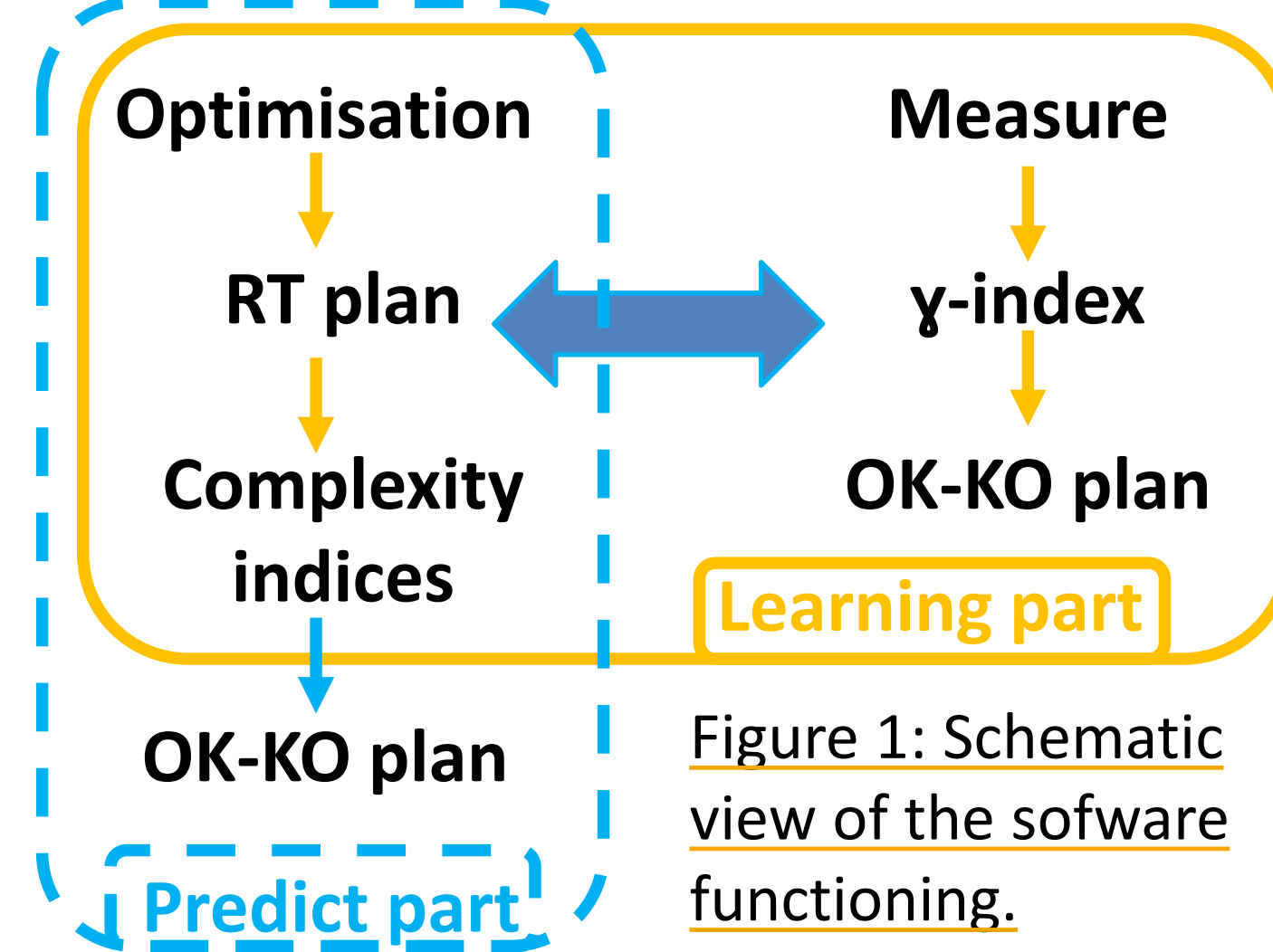


Figure 1: Schematic view of the software functioning.

Gamma-index:

The *deliverability* of a plan is *quantitatively* evaluated using the so-called gamma index [1] which combines the DTA (*Distance to agreement*) and the *dose difference* to compare the calculated and measured dose distributions.

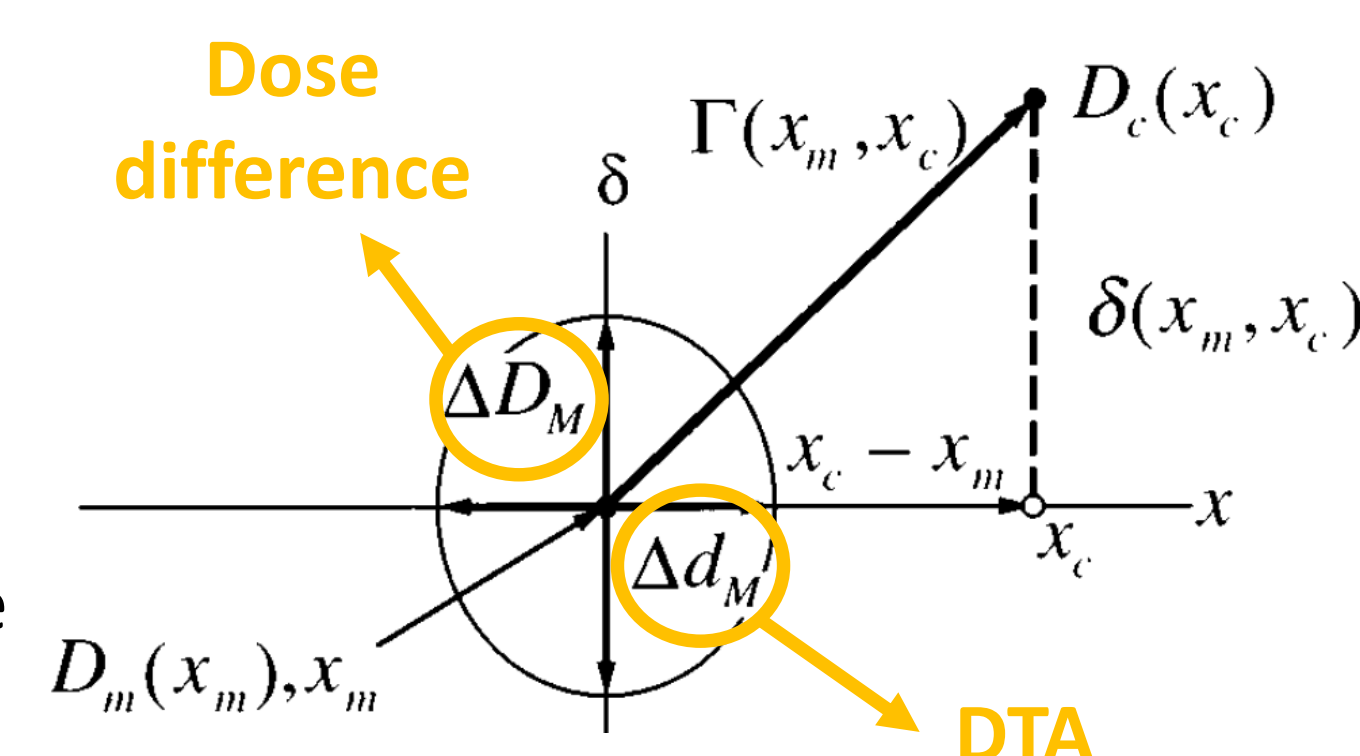


Figure 2: Two-dimensional geometric representation of the gamma-index.

Complexity indices:

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* (eg number of UM) or on the *geometry* (aperture based like the area or the perimeter of beam segments).

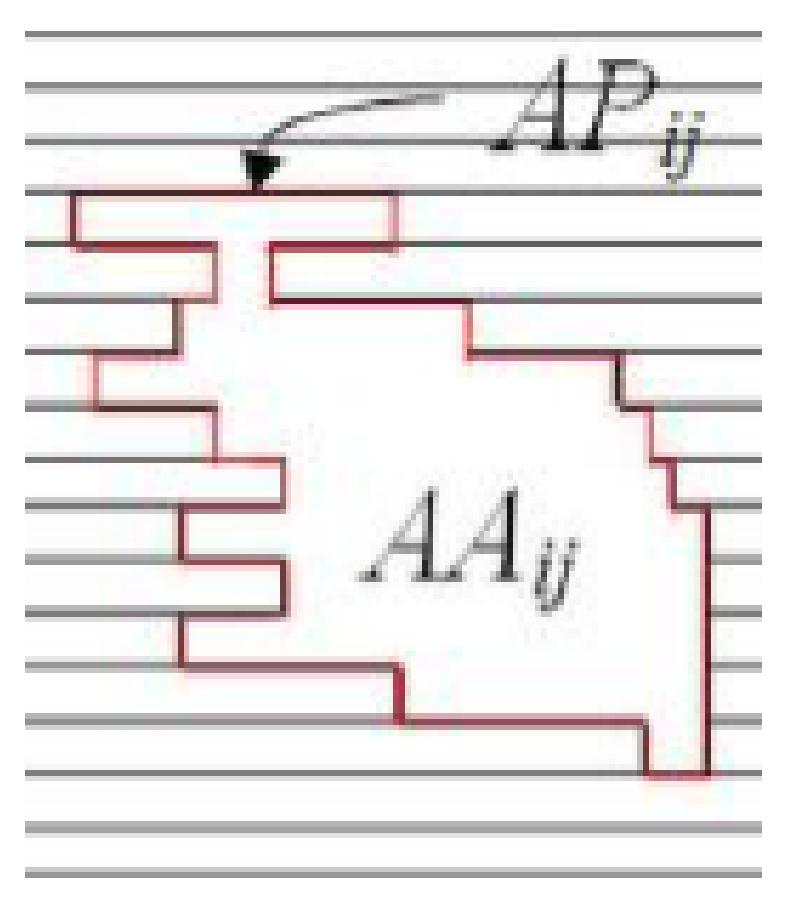


Figure 3: Illustration of the beam aperture.

[1] D. A. Low et al, A technique for the quantitative evaluation of dose distributions, Medical Physics (1998)

Preliminary results

Up to now, the study is still in a *preliminary phase*:

- Few plans remain to be optimised.
- Most QA remain to be done.
- First data have been implemented in the software which is still in its final stage of development.

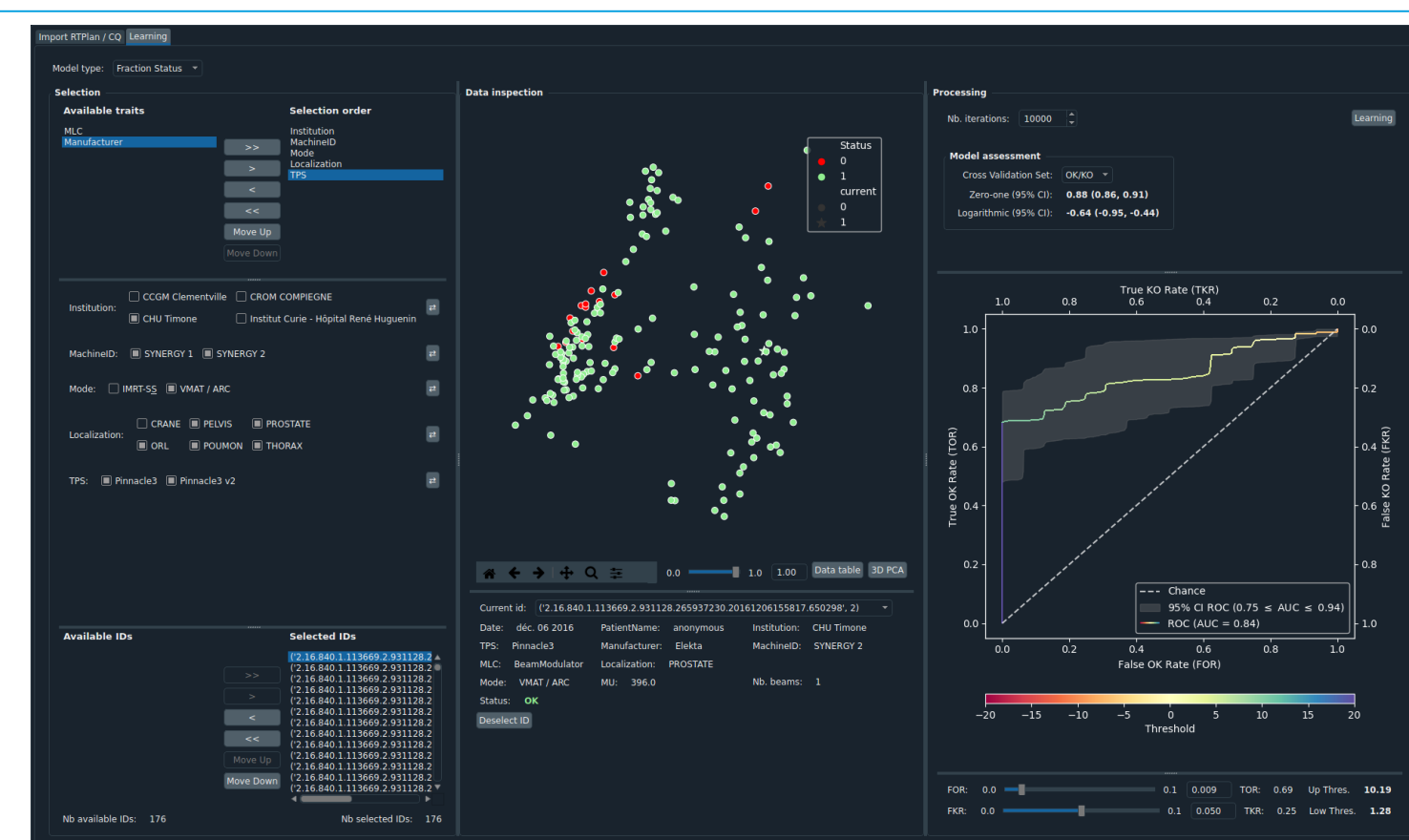


Figure 4: Screenshot of the software interface.

- A large statistics is expected: the preliminary study included 84 plans.
- Plans with large fields are considered:
 - Prostate (1 arc), pelvic-prostate (2 arcs), H&N (2 arcs)
 Only pelvic-prostate region was explored in the preliminary study.
- The H&N location has the largest number of failing QA ie the highest complexity **New features!**
- A fair comparison between machines A and B is possible as several redundant measures are available.

Criterion	1	2
Prostate machine A	3,70%	3,70%
H&N machine A	10,64%	23,40%
Prostate Machines A&B	3,70%	3,70%
H&N Machines A&B	12,50%	28,57%

Table 1: Percentage of failing plans (KO) for the two criteria and machines.

- Local gamma analysis with 20% threshold.
- All plans pass the usual criterion:
 - Dose difference $\leq 3\%$
 - DTA $\leq 3\text{mm}$
 - γ -passing rate $\leq 95\%$
- Sufficiently large number of failing plans needed for the learning phase ~10-20%
 - Increase of the plan complexity during the optimization phase + H&N location
- New criteria to test:
 - Criterion 1: 3% - 3mm - 97%
 - Criterion 2: 2,5% - 2,5mm - 95%

Location	Plans	QA machine A	QA machine B	QA machines A&B
Pelvic-Prostate	54	27	5	5
H&N	171	47	31	25
Brain	20	6	7	6
Lung	4	2	1	1
Total	252	85	47	40

Table 2: Number of optimised plans and QA realised with the machines A and/or B for several tumor locations.

Outlooks

Having a sufficient number of plans ie a *large statistics*, we expect to bring a quantitative answer to several questions:

- The **number of plans** needed to get a sufficiently precise prediction for the QA (no "false-positive" ie a plan that is predicted by the software to pass the required criteria but fails the QA).
- The importance of the **tumor location**. Indeed, the latter is related in some way to the plan complexity. Exploring several tumor locations allow to expect a larger variety of plan complexity that will increase the efficiency of the learning phase. In this spirit, we focus in this work on pelvic-prostate, H&N, brain and lung locations.
- To what extent the prediction is affected if some of the QA are done with a mirror accelerator (machine A and B). Indeed, the two accelerators are slightly different while the modelisation in the TPS remains the same.
- The effect of changing the **passing criteria** to modify the learning part of the software ie the OK and KO plans (even if the criteria remain the same to determine if a plan is deliverable or not). For instance, one can increase the passing rate (criterion 1) or decrease the percent dose difference and DTA (criterion 2).