

UCLouvain

MASTER THESIS

Multistage flagging system for online adaptive proton therapy

Author:
Maxime BERNARD
NOMA:
59261700

Supervisor:
Pr. Edmond STERPIN

*A thesis submitted in fulfillment of the requirements
for the degree of Master in Physics [120]*

in the

Institute of Experimental and Clinical Research
Research pole: Molecular Imaging, Radiotherapy & Oncology

Faculty of Science - School of Physics

Academic year: 2019-2020

“The surest way to corrupt a youth is to instruct him to hold in higher esteem those who think alike than those who think differently.”

Friedrich Wilhelm Nietzsche

UCLouvain

Abstract

Faculty of Science

Research pole: Molecular Imaging, Radiotherapy & Oncology

Master in Physics [120]

Multistage flagging system for online adaptive proton therapy

by Maxime BERNARD

Lately the high clinical potential of proton therapy, coupled with the appearance of new technologies such as pencil beam scanning and the mounting of imagery tools in the very midst of the treatment room increased the interest around its development.

This particular cancer treatment modality takes advantage of the singular way the protons deposit their energy throughout the matter: the Bragg peak. This very narrow dose profile peak allows very little dose deposition before the peak and close to no dose after. But as efficient as it is, this energy deposition is also very sensitive to density variations and if we want to know exactly where the peak will be located we have to develop methods to account for these variations. One of them consists in an online checking for these variations by imaging the patient anatomy in the tumor area right before his daily dose delivery to make sure it still corresponds to the images taken for the treatment planning. If there is a mismatch we can choose to adapt his treatment to ensure a good dose conformity and a safe dose delivery. This is what is called online adaptive proton therapy.

That is in that context that our research question was raised. How can we determine if a patient needs a treatment plan adaptation or not? Of course we could compute the dose distribution every time the patient lay on the table for its treatment but the dose distribution is way too long to compute. In fact, the more we wait for approval, the more likely the patient is to move, leading then to new positioning errors. We therefore need a flagging system working in the order of the second. We decided this system would work on 3 different stages of increasing complexity. The first stage, and the fastest one, would be based on the difference evaluation between the reference scanner and the daily scanner using Hounsfield unit (HU) matrices. The second stage would consist in computing the difference between WEPL maps from both reference and daily scanner. And finally, the third and last stage, which is also the most time consuming one would be the computation of the dose distribution.

For this purpose, using a pool of 13 patients, we tried to find metrics that could serve as adaptation indicators. We started with the dose distribution based indicator as we wanted to use it as a reference to compare our future metrics with. We found the homogeneity index to be performing very well and we decided to use it. We then came to the ΔHU based indicators and we found that peak signal to noise ratio and the mean squared error could be used as a first way to filter out patients not needing a treatment plan adaptation. For the WEPL based indicators, the percentile 90 performed very satisfyingly, allowing us to discriminate with large margins both patients to adapt and patients not needing adaptation. Our study showed us that a multistage flagging system for online adaptive proton therapy was feasible even though we could improve the robustness of our conclusions if we had more patients.

Acknowledgements

Ce mémoire n'aurait probablement jamais vu le jour sans les conseils et le soutien apportés par certaines personnes.

Je tenais dès lors à remercier tout particulièrement mon promoteur, le Professeur Edmond Sterpin. Pour sa confiance d'une part, pour m'avoir permis de travailler sur ce projet mais aussi pour ses nombreuses relectures du manuscrit et pour ses précieuses explications qui ont permis de faire de ce travail ce qu'il est aujourd'hui.

Je voudrais également remercier Elena Lucia Borderias Villarroel, Gregory Buti et Umair Javaid pour leur aide précieuse, leur disponibilité et leur accueil chaleureux au sein de leur équipe cette année.

Par ailleurs je voulais profiter de l'occasion pour exprimer toute ma reconnaissance à mes parents et à mes frères sans qui je ne serais pas là où j'en suis aujourd'hui.

Et pour finir, je tiens à remercier Eléonore pour son appui durant ces deux dernières années et ses relectures attentives.

Contents

Abstract	v
Acknowledgements	vii
1 Introduction	1
1.1 Motivation of the study	1
1.2 State of the art in cancer treatment	3
1.3 Radiotherapy workflow	8
1.4 Management of uncertainties in proton therapy	10
1.4.1 Robust optimization	11
1.4.2 Adaptive proton therapy	12
2 Physics notions	15
2.1 Interactions of protons with matter	15
2.1.1 Inelastic interaction with atomic electrons	15
2.1.2 Inelastic interactions with the nucleus	16
2.1.3 Elastic interactions with the nucleus	16
2.2 Energy loss rate	17
2.3 Range and energy straggling	19
2.4 Multiple Coulomb scattering	20
2.5 Hounsfield Units and WEPL	21
2.5.1 Hounsfield Units	21
2.5.2 WEPL	22
3 Material and methods	23
3.1 Material	23
3.1.1 Patient data	23
3.1.2 Treatment planning software	23
3.2 Methodology	25
3.2.1 Dose distribution indicators	26
3.2.2 HU based indicators	27
3.2.3 WEPL based indicators	30
4 Results	31
4.1 Metrics	31
4.1.1 Dose distribution based indicators	31
4.1.2 Hounsfield units metrics	35
4.1.3 ΔHU results discussion	38
4.1.4 Water equivalent path length metrics	39
4.1.5 WEPL metrics results discussion	44
4.2 Indicators	49
4.2.1 First approach	49
4.2.2 Second approach	53

Chapter 1

Introduction

1.1 Motivation of the study

Today, cancer is treated with three main strategies: surgery, radiotherapy and systemic therapy using drugs (chemotherapy, immunotherapy...). Radiotherapy, which we will focus on in the course of this study, is used in about 50% of cases.

Lately the high clinical potential of an advanced radiotherapy modality called proton therapy, coupled with the appearance of new technologies such as pencil beam scanning and the embedment of imagery tools in the very midst of the treatment room raised the level of interest around its development. Proton therapy takes advantage of the peculiar characteristics of the protons when interacting with matter, the concentration of the dose at the end of their path called the Bragg-Peak.

This Bragg-Peak enables accurate irradiation of the tumor volumes. While in conventional radiotherapy, using high energy X-rays, the all path through which the photons travel is irradiated, protons only deposit reduced dose before the tumor and most of the dose in the peak with no dose deposition after the tumor, enabling improved target conformity and OAR (Organs at risk) sparing. But because of its finite path and the high dose concentration at its end, proton therapy also suffers from a very high sensitivity of its protons to geometrical and anatomical uncertainties.

When a patient comes to the clinic to undergo proton therapy, the first thing we do is mapping the patient anatomy in the tumor area in order to plan his treatment. This alone can take up to a few days in a clinical setting as it needs a physician approval, and uncertainties management using robust optimization. Because of the sensitivity of the protons to density variations, any anatomical changes in the patient occurring between the day of the planning and the daily dose delivery can potentially lead to unwanted dosimetric and clinical effects such as bad tumor coverage or excessive dose received in wrong areas inducing healthy tissue complications.

To avoid such complications and wrong behavior of the dose distribution, there are two paths we can take. Either we can make the treatment plan more robust hence increasing the overall dose or we can take daily images and adapt the treatment plan if needed. Robust optimization aims at incorporating the uncertainties due to anatomical changes directly in the plan optimization by considering all possible scenarios. But covering the all range of possible changes weakens the advantages from proton physics thus prejudicing the treatment optimality. The other option available is the adaptive proton therapy. In this case repeated patient images are acquired throughout the therapy and the treatment plan is adapted if comparison with the planning CT images shows significant anatomical changes.

While the decision to adapt the treatment or not is made, one need to keep in mind that the patient is waiting on the treatment table. It means that the decision has to be as fast as possible. In fact, if the decision to adapt might take some time, the actual adaptation will take at least that much time and we want to avoid any unnecessary stress for the patient and prevent any movement while the process occurs. It could be said that re-optimization procedure works just fine and that we could do it systematically for every patient, but it takes a lot of time. And in a clinical framework, with limited human resources, we want to optimize the treatment duration.

The purpose of this study is to find a fast and automatic way to detect which patient will need an adaptation and which will not. To this end, we will first compare dose distributions in planning and daily CT's to extract from the patients pool the ones we know for sure need a plan adaptation. When it is done, we will search for a reliable dose distribution based indicator that would predict the same patients to adapt. We could stop there and rely only on this indicator as it would be highly reliable but dose distribution based indicators are also longer to compute. That is why we will then focus on finding fastest ways to predict the adaptation necessity using on the one hand daily CT's comparison based on Hounsfield Units and on the other hand CT's comparison using WEPL maps. In fact we will see that CT images and, with more accuracy, the WEPL can be used as surrogate to predict dose distortion.

From these comparisons we will try to find indicators that would give us the same results we obtained with dose distributions but in a much faster manner. To find these so-called indicators we will work with different metrics, study each of them and their behavior and see if any could predict with accuracy the patients needing a treatment plan adaptation. The system we want to develop needs to be both fast and highly secure. It can allow a small number of plan adaptation for patients which did not need one but we don't want to miss a single patient that would have needed that treatment plan adaptation.

1.2 State of the art in cancer treatment

In the human body, cell divisions happen all the time and are controlled by different biological systems. Sometimes, DNA mutations occur due to environmental factors, viruses, replication errors, exposure to ionizing radiations. . . and can affect those systems. Most of the time, these mutations are fixed but if it has not been detected or has not been repaired properly, it can cause cells to get out of the division regulation, hence starting to multiply in an uncontrolled way leading to the formation of a tumor.

Two main categories of treatment exist[1]:

- Systematic treatments
 - Chemotherapy
 - Immunotherapy
 - Metabolic radiotherapy
- Loco-regional tumor control
 - Surgery
 - External radiotherapy
 - Brachytherapy

In our case, we will focus on external radiotherapy, but one can note that patients are usually treated using a combination of these modalities. Radiotherapy is involved in around 50% of treatments and its goal is mainly to secure loco-regional control.

The basic mechanism of radiotherapy is the use of ionizing radiations to treat cancer. These ionizing radiations are radiations carrying enough energy to free electrons from atoms or molecules. DNA exposed to this kind of radiations will undergo direct and indirect actions from the radiations. While direct actions will straight away damage the DNA, indirect actions will lead to the creation of free radicals that, in turn, will impair the DNA. Those damages will impact the intra-cellular mechanisms potentially leading to the death of the cell.

The radiations we use during therapy affect both normal tissue cells and cancer cells. However, normal cells have efficient machinery for DNA repair that better preserve the cellular integrity while, in cancer cells, this mechanism is flawed and residual errors persist after reparation. This opens a therapeutic window. It can be better understood using a tumor control probability over prescribed dose graph. The dose is a measure of the mean energy imparted($d\bar{\epsilon}$) by ionizing radiations to a material of mass dm expressed in Gray (Gy):

$$D = \frac{d\bar{\epsilon}}{dm}. \quad (1.1)$$

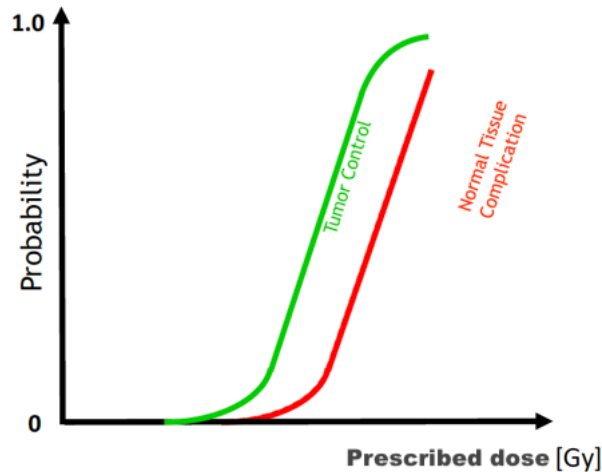


FIGURE 1.1: Tumor control probability (green) and Normal tissue complication probability (red) vs the prescribed dose. *Source: [1]*

On the graph 1.1 we can see two curves. The green one describes the growth of the tumor control probability when we increase the prescribed dose and the red one the normal tissue complication induced by this same dose. We can observe that the more dose we deliver to the tumor, the more we increase the probability of controlling it. But in the meantime, we also increase the probability to see normal tissue complications. We therefore want both of these curves the furthest away from each other possible.

The core issue in radiotherapy is to find the best trade-off between the tumor control and the sparing of the organs at risk. We need to find the answer to how to deliver the highest dose to the tumor volume while at the same time delivering the lowest dose to healthy tissues. A number of options are available including [1]:

- Improved ballistics
- Better choice of particles
- Differential radiobiological effects
- Radiosensitization

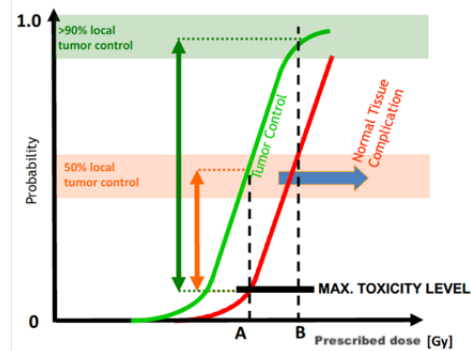


FIGURE 1.2: *Source: [1]*

The graph 1.2 highlights the importance of moving both curves away from each other in order to be able to increase the tumor control probability (TCP) without impacting the normal tissue complication probability (NTCP) too much.

An important first step toward ballistic improvement was made when developing better dose delivery using dose sculpting, a technology allowing a better dose conformation to the tumor.

For this purpose the beams are shaped to the tumor form using custom-made collimator and then using multi-leaf collimator.

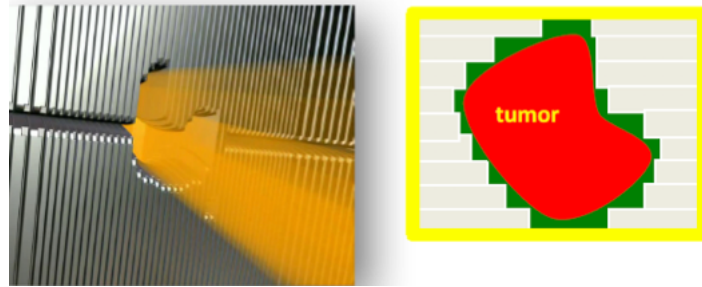


FIGURE 1.3: Multi-leaf collimator used for optimal tumor contour conformation. Source: [1]

We further improved the conformation of the dose to the tumor structure using intensity modulated beams hence reducing the dose received by the organs at risk (OAR). The OAR are the most radio-sensitive organs which we absolutely want to spare from the radiations.

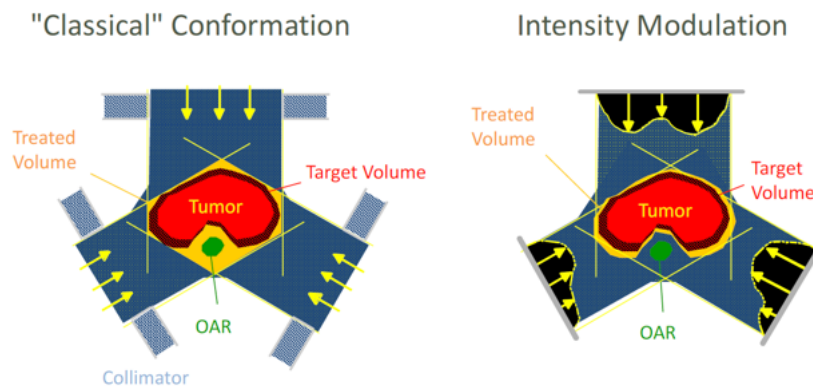


FIGURE 1.4: Intensity modulated beams enable yet a better dose conformation to the tumor volume allowing sparing of OAR. Source: [1]

The next meaningful step improving the efficiency of radiotherapy resides in a better choice of particle.

As it was mentioned earlier, proton therapy wants to take advantage of the Bragg peak, the high dose concentration the protons deposit at the end of their path through the matter. It is this improved dose deposition compared to the one from X-rays that makes the proton therapy more appealing. The underlying physical reasons for the differences in these profiles will be covered in the chapter 2.

Here is a typical dose profile for a proton showing the depth deposition as a function of the depth in water:

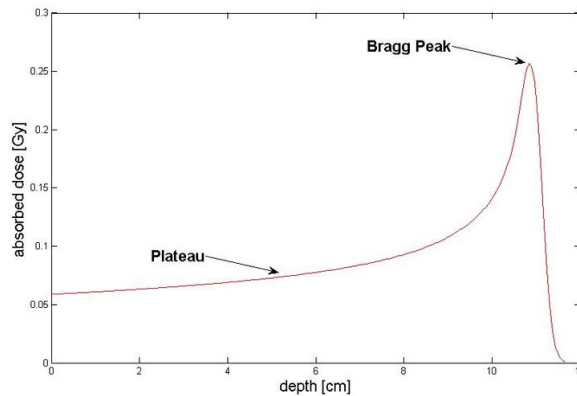
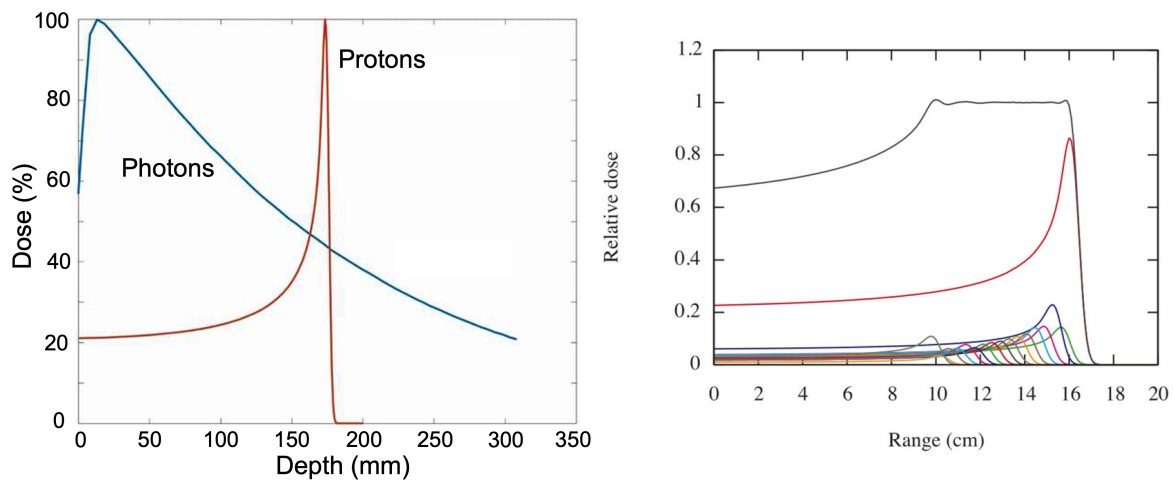


FIGURE 1.5: Description of a 130 MeV proton dose deposition as a function of the depth travelled in water [2]

The depth at which the peak is located depends on the initial energy of the protons. For the same material, the more energetic they are the deeper the peak will be. But a single Bragg peak is too narrow to cover the all volume of the tumor. That is why we will need multiple Bragg peaks to cover the all depth in which the tumor is located, therefore creating a spread-out Bragg peak (SOBP) by modulating the initial energy of the protons:



(A) Dose profil comparison for photons and protons in water. (B) Accumulation of different dose profiles forming a so-called spread-out Bragg peak.

FIGURE 1.6: [3]

The graphs at Fig 1.6 highlight the potential of protons for strongly enhancing cancer treatment. The Graph (A) as well as the figure 1.7 emphasize the difference in dose profile between photons and protons.

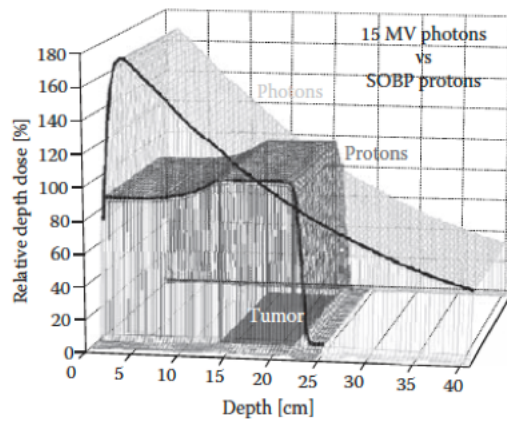


FIGURE 1.7: 3D comparison of SOBP and photons dose profiles

The clinical potential can also be illustrated on anatomical images (CT scanner), for instance in Fig 1.8 for a craniospinal irradiation:

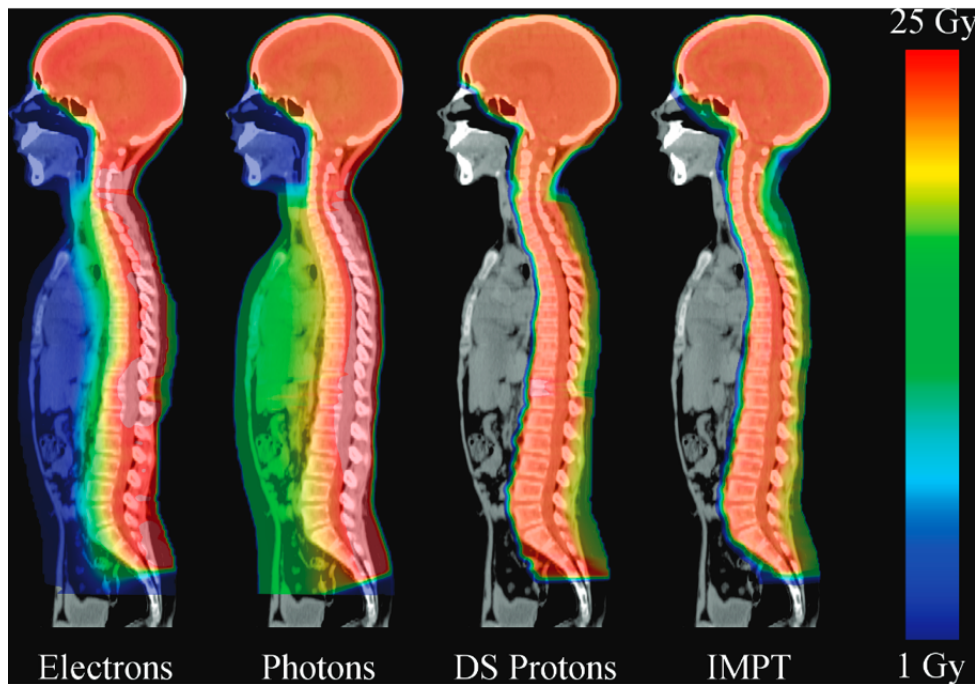


FIGURE 1.8: Treatment delivery modality comparison. From left to right: Electrons, photons, double scattered protons and Intensity modulated proton therapy. Source: [4]

On this image, 4 modalities are depicted: electrons, conventional radiotherapy using X-rays and the last two, using protons, are double scattering proton therapy and intensity modulated proton therapy. On both images showing the dose distribution for protons, the benefits regarding the sparing of OAR and the dose conformation become obvious. This is important because patients often suffer from secondary effects of the irradiations underwent by the healthy tissues during the treatment and it could enable us to reduce such effects. It is particularly helpful when treating children because they are more radio-sensitives and are at risk to develop radio induced cancers because of their growth and long life expectancy post-exposure [5].

Today, proton therapy is mostly used in pediatric cancers and in very complex cases where the organs at risk are very close to the tumor like in the case of the eyes or the skull base.

1.3 Radiotherapy workflow

To ensure the optimization of the treatment, namely the maximum dose the tumor while sparing the best we can the organs at risk, and a safe treatment delivery, a comprehensive workflow must be followed:

- Imaging
- Contouring and prescription
- Treatment optimization
- Treatment verification
- Treatment delivery

First of all, images have to be taken to build an anatomical map of the patient. For this purpose, we usually use 3D-CT imaging and IRM. 4D-CT can also be needed, for instance, when the tumor is located in the lungs. Sometimes these CT images are also coupled with a PET Scan which offers a metabolic imagery on top of the anatomic one which is very useful to better determine the position of the tumor.

In Leuven (Belgium), a proton therapy center has just been achieved and is equipped with a Proteus One from IBA (Ion beam applications)[6]. The Proteus One combines the proton therapy tools with a CBCT (Cone beam computed tomography) integrated in the gantry which allows online patient position corrections. It also includes an on rails CT in the same room. This configuration is unique in the world.

When the images are ready, a physician uses them to contour the area of the tumor (the zone to treat) and all the organs at risk (areas we want to spare). When the tumor volume is contoured, many aspects have to be taken into account. In fact, if we want to ensure loco-regional control, all the tumor cells must have been irradiated with the appropriate dose which means the contours have to encompass the visible volume of the tumor. Tumor cells can also be found at the microscopic level in supposedly healthy tissues. Tissues at significant risk of microscopic presence are typically included in the target contours. Finally during the treatment itself, the geometric location of the tumor macroscopic volume and the suspected affected tissues may not be right where we expected them to be during the planning. All the above considerations have been standardized in the contour nomenclature.

Nowadays, three types of contours are used and are defined accordingly:

- GTV (Gross Tumor Volume): Encompass all the Macroscopic, clinically demonstrable, tumor volume.
- CTV (Clinical Target Volume): Encompass the GTV and the suspected tumor cells at the microscopic level.
- PTV (Planning Target Volume): Circumscribes the CTV and the added margins taking into account the uncertainties related to the location of the tumor during the treatment.

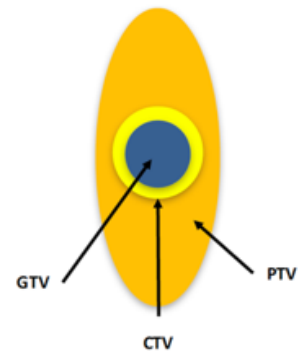


FIGURE 1.9:
Source:[1]

When the contours are well defined by the physician, he also determines dose prescriptions for the tumor volume and fixes dose limitations for the different OAR. It is an important step because the physician and the physicist will have to determine the best trade-off between sparing the OAR and assuring the best dose delivery to the tumor.

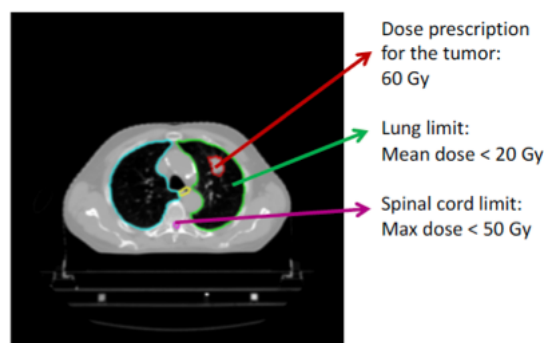


FIGURE 1.10: Establishment of the dose prescriptions and OAR limits using contours

The contour we will focus on for this study is the CTV(T+LN) which correspond to the combination of the CTV for the main tumor (T) and the CTV for the nodes (LN).

After the contours definition comes the treatment optimization in which the medical physicist will plan a treatment that delivers an optimal dose distribution and meets best the prescriptions.

Using DVH's as in 1.11, the physician will check the conformity of the planned treatment and approve it or not. A DVH is a dose-volume histogram that allows us to display information about the dose received by a certain volume for every OAR and the target volume. In this case we have cumulative DVH as opposed to differential DVH.

If the treatment plan is approved, then we can carry on to the delivery of the treatment. Depending on the type of cancer and the choices made by the physician, the treatment will be delivered in a particular pace called fractionation. The conventional fractionation consists in 5 fractions of 1.8 to 2Gy per week on a 6 to 7 weeks periods. Other fractionations exist such as hypofractionation and hyperfractionation but will not be covered in this research.

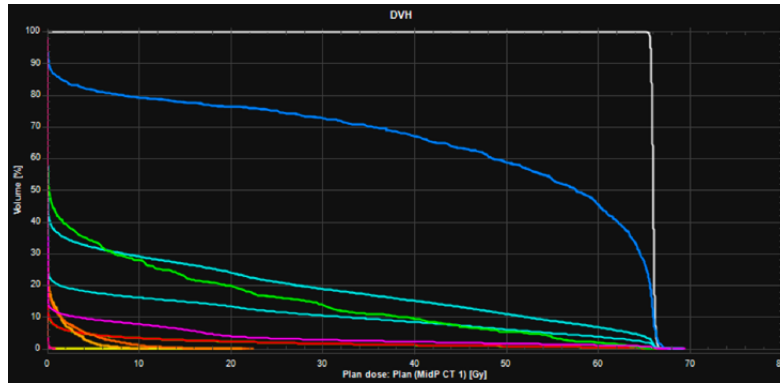


FIGURE 1.11: Dose volume histograms for different organs at risk and for the CTV(T+LN) (white)

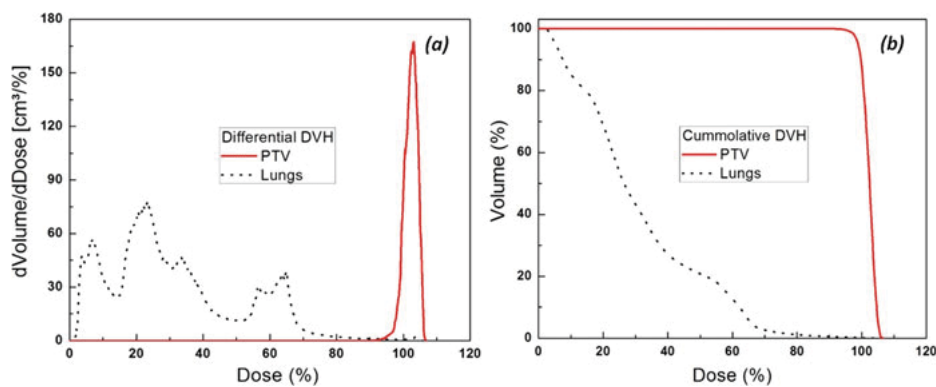


FIGURE 1.12: Schematics of (a) differential DVHs and (b) cumulative DVHs for the PTV and lungs in a phase-2 mediastinum treatment plan [7]

1.4 Management of uncertainties in proton therapy

As it was mentioned earlier, the protons feature important advantages but their path is very sensitive to density variations in the anatomy of the patient. The control of the proton range in matter is difficult because it necessitates the knowledge of the exact material density the protons are going through. While it is not a big issue for photons as we can see from the dose distribution curve 1.6, we can imagine that, for protons, a slight offset of the Bragg peak could potentially lead to quite catastrophic consequences.

On figure 1.13 is displayed an example of the planned dose on the reference image CT's compared to the planned dose on the daily CT 35 days after the first images were taken. We can observe that the tumor has shrunk throughout the therapy leading to major anatomical changes and inducing a degraded dose which will not cover the tumor properly and will also harm healthy tissues beyond it.

The uncertainties we face are not only originating from internal anatomical changes throughout the treatment. They can arise from different sources. They can come from changes in the patient position on the table day by day inducing an overall CTV position variation. Changes can also occur inside the patient due to different digestive phases, patient weight loss along the treatment or an imperfect conversion of imaging data into quantities needed to compute the dose distribution. All these factors cause uncertainties on the ranges of the protons and therefore on the dose distribution initially planned and they must be accounted

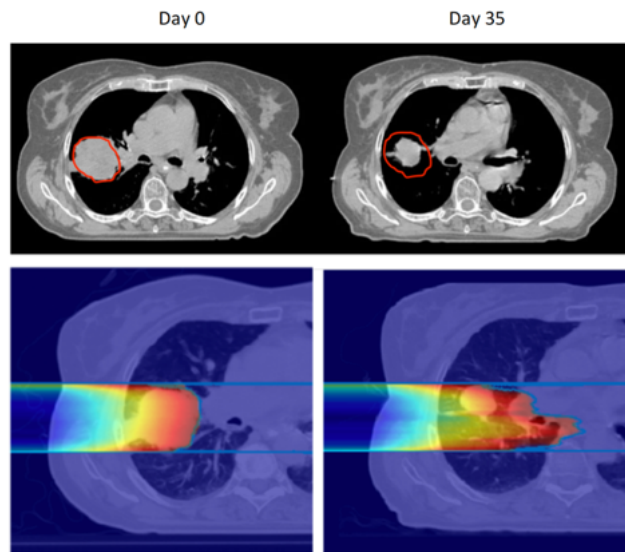


FIGURE 1.13: Comparison of the planned dose distribution with the same planned dose 35 days later on a changed anatomy caused by the size reduction of the tumor. *Source: [1]*

for in order to ensure the conformity of the dose with the planned treatment. We say that the treatment needs to be robust against those uncertainties.

1.4.1 Robust optimization

We will not go into details here but more about about this strategy can be found in some of the articles we consulted [8] [9].

This first strategy consists in including the uncertainties previously mentioned directly into the optimization process. It is based on the simulation of many different scenarios which each corresponds to different uncertainties realizations.

A direct approach would consist in the simulation of all the different kinds of uncertainties and in the combination of the different scenarios. But this leads to an enormous number of scenarios possible. In fact it has been calculated [10] that for a lung tumor, if we take into account the geometrical systematic errors from breathing and the physical errors due to the overestimation and underestimation of protons penetration depth on a ten phases CT scan, we already have 1470 different scenarios. And it is by considering the same errors for every beam orientations which will not be the case.

We see here that research has yet to be done in order to find a way to generate a number of scenarios more manageable for an optimizer because such a large number would take too much time to compute.

Moreover, the more robust we want to be, the more we will increase the dose to compensate for the uncertainties. And the more we increase the dose the further we stray from the benefits provided by protons.

It is in that context that treatment plan adaptation can make a real difference.

1.4.2 Adaptive proton therapy

Adaptive proton therapy, as opposed to robust optimization which aim at compensate for all uncertainties beforehand in the optimization process, will correct the treatment plan on a case-by-case bases and, if necessary, will allow a re-adaptation of the treatment plan the day of the treatment.

After the treatment planning by the physician and the physicist, the patient will receive its daily treatment at the clinic. The first step is always to acquire new CT images of the patient. We then perform a rigid registration of those images on the reference images for position alignment. On these new registered images, we can evaluate the planned dose to check any dose distribution mismatch. The dose on the registered CT's is called the degraded dose. From the comparison between the reference dose and the degraded one and from the comparison of the images, a decision has to be made. If we consider the changes do not modify the treatment plan to much we move on to the delivery. If too many anatomical changes occurred, the treatment can be stopped and a new plan has to be computed and optimized to fit the daily anatomy of the patient before moving on to the treatment again.

But it takes a lot of time and we cannot afford it in the context of an online process. We cannot go through all the planification workflow while the patient is on the table. We need a simplified, fast but still reliable method. One of this method is called "Dose restoration".

The principle of this method is to simply copy the contour from the reference CT images and to apply it to the daily images with the rigid registration field from the registration applied on the images. We then proceed to the optimization of the plan based on the copied contour. The daily contours being the same as the reference ones, they are fast to acquire and no check from a physician is needed. It is therefore possible to make the online procedure automatic.

Following is the description of an adaptive proton therapy workflow:

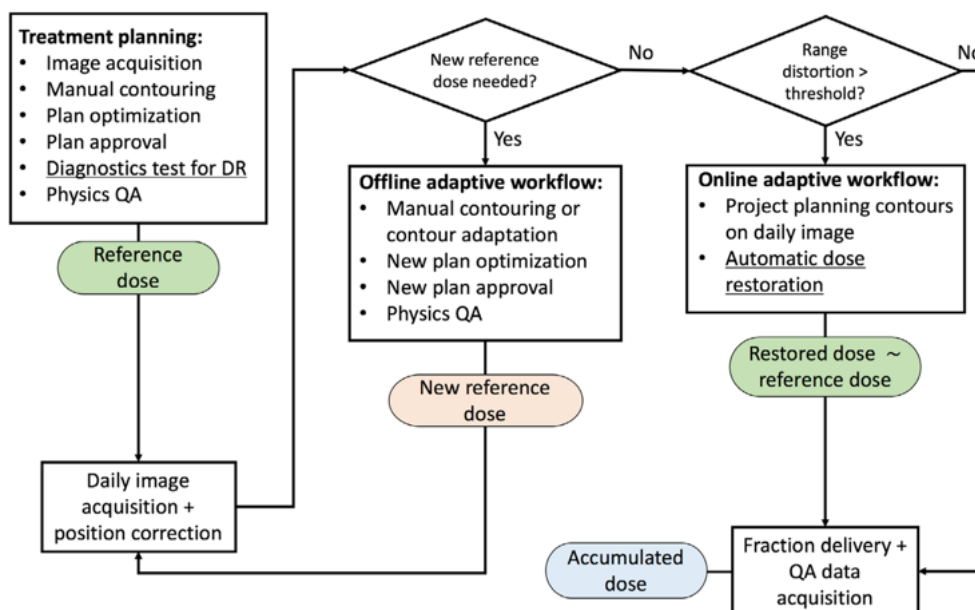


FIGURE 1.14: Proton therapy workflow including online and offline adaptation workflow

The online adaptive proton therapy procedure will be based on the evaluation of anatomical differences according to three different stages of increasing complexities and therefore of increasing computation time. We would like to be able to flag the patients needing treatment plan adaptation the fastest we can.

The first steps of the adaptive proton therapy are always the daily imagery of the patient anatomy and the rigid registration of these acquired images on the reference CT images. Then, the different stages of difference evaluation will be as follow:

1. This first stage will consist in the difference evaluation of both reference and daily CT images converted in Hounsfield Units. More details about these units in the next chapter. This is the fastest of the three stages but also the less reliable.
2. The second stage will be based on the computation of the water equivalent path length (WEPL). The WEPL map differences are longer to compute but are also more reliable as they are based on the stopping power. These new concepts will be covered in more details in the next chapter.
3. And finally, the last stage is the one with the highest level of complexity, hence the more reliable and the longest to compute. It is based on the comparison of the planned and the degraded dose. It is the best test we can do and it will be therefore used as a reference to compare the other indicators.

Based on these different evaluation stages, we want to be able to build a triggering system which would tell us when the anatomical changes are too important hence an online adaptation is necessary. The more time the patient stays on the table the more likely he is to move, cough... The adaptation itself already taking time, we want the flagging/triggering system to be as fast as possible. But one has to be careful not to fall into excesses and re-adapt when it is not necessary. Adaptive proton therapy is not always necessary and has not to be triggered if the daily anatomy fits the reference images. It is also important to note that both methods, robust optimization and adaptive PT, are not necessarily antagonist. Ideally, we would like to find a trade-off between the robust optimization parameters and a certain level of adaptation to use the full potential of proton therapy.

The aim of this study is to find thresholds and flagging indicators for every stage that would trigger an automatic online adaptive procedure when necessary. For this purpose scripts written with Python will be used for the first stage and Matlab scripts for the second stage. For the highest level of complexity, we will use the software REGGUI which was partly designed within the UCLouvain which lets us visualize CT slices, contours and dose distributions among many other things. The next chapter will focus on the physical phenomenon governing what was covered in the current chapter. Then the research methodology will be described in the third chapter and our results will be displayed in chapter 4 to finally finish with the conclusion and future prospects in the chapter 5.

Chapter 2

Physics notions

In this chapter we will focus on the physics of proton therapy and how we can explain the protons behavior reviewed in the previous chapter. We will start out with a first section about the different interactions of protons with matter. We will then continue with the concept of stopping power, range and energy straggling and finally finish with the multiple coulomb scattering and different important notions we need before moving onto the rest of the research.

2.1 Interactions of protons with matter

2.1.1 Inelastic interaction with atomic electrons

This interaction is the one we are the most interested in because it is the most predominant in our range of energies (3 to 300MeV) and with the material used in the clinical framework. It will only cause small modifications of the proton trajectory but it is through this process that the proton will lose most of its energy inducing atomic excitation and ionisation.

When an electron is extracted from its shell by a proton, if enough energy was transferred to it, the electron can in turn ionize the medium it travels through creating what is called δ -rays. We can distinguish two types of interactions.

The **soft** collisions can be described as collisions occurring when fast particles pass near an atom at a relatively large distance, that is with an impact parameter b greater than the classical radius of the atom a ($b \gg a$). In this case, the coulomb force is exerted between the proton and the atom as a whole. We have excitation or ionization of the valence electrons and only a small amount of energy is transferred (order of the eV). This type of collision is the most frequent one. Then we have the so called **hard** collisions which occur when the particles involved come closer to the atom ($b \sim a$). In this case we consider the interaction happening only with a single electron. The electrons at stake here are the one producing the δ -rays. These electrons will be ejected with a maximum kinetic energy equal to:

$$T'_{max} = 2m_0c^2\left(\beta^2 - \frac{1}{\beta^2}\right), \quad (2.1)$$

which can be simplified as:

$$T'_{max} \approx \frac{T_p}{500}. \quad (2.2)$$

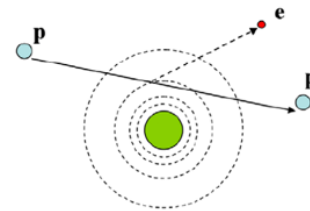


FIGURE 2.1: Source: [1]

With T'_{max} being the maximum kinetic energy (MeV) transferred by the proton to the electron and T_p the kinetic energy of the proton (The equation 2.1 is only valid for kinetic energy $< m_0c^2$). Hard collisions occur more seldomly than soft collisions, but the energy transferred is much larger, resulting in a comparable contribution in the total energy transferred. The energy loss is provided by the stopping power which will be covered in the following section.

2.1.2 Inelastic interactions with the nucleus

This interaction leads to the loss of the initial proton with production of neutrons, protons or heavier particles such as deuteron and a recoil nucleus as well as an activation of the system through γ -rays production. Though it is less frequent than the other types of interaction, this process will have a significant impact in terms of radioprotection.

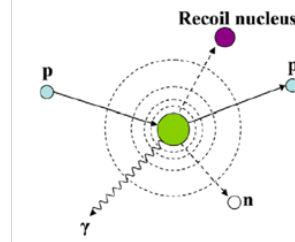


FIGURE 2.2: Source: [1]

2.1.3 Elastic interactions with the nucleus

This interaction happens when a proton passes close to the atomic nucleus and experience a repulsive coulombic interaction which, due to the large mass of the nucleus, deflects the proton from its original trajectory. This is the interaction that cause the most important deviations but with a negligible loss of energy. Fortunately, the total cross section (2.4) also decreases rapidly with the energy of the initial proton meaning that at the energies we are working at, only a few protons will undergo large deflections.

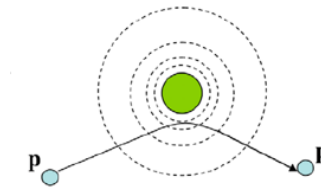


FIGURE 2.3: Source: [1]

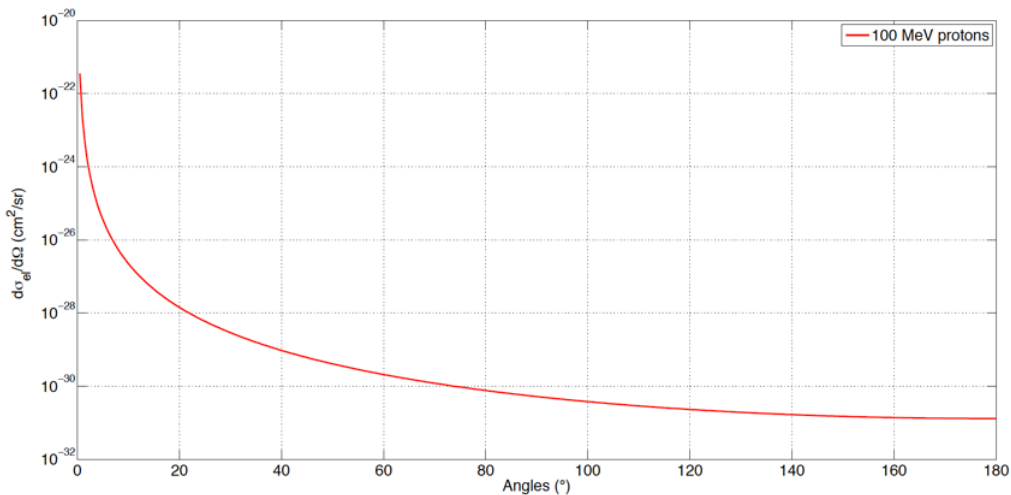


FIGURE 2.4: Decreasing of the Differential cross section for electromagnetic elastic scattering for a 100 MeV proton [11].

Finally, protons can technically lose energy through Bremsstrahlung process, but it only starts to have an impact on the energy loss when the protons reach relativistic energies which is not the case in the energy range we work with.

2.2 Energy loss rate

The energy loss rate, also called stopping power, describes the occurrence of very frequent small losses of energy along the path of any charged particles due to their interactions with matter. It is defined as the quotient of dE and dx where E is the mean energy loss and x the covered distance.

$$S = -\frac{dE}{dx} (\text{MeV/cm}). \quad (2.3)$$

We can also express the energy loss rate by normalizing by the mass density. We get what is called the mass stopping power:

$$\frac{S}{\rho} = -\frac{dE}{\rho dx} (\text{MeV cm}^2 \text{g}^{-1}). \quad (2.4)$$

A more complete definition of the stopping power is given by the Bethe (1930) and Bloch (1933) equation which include quantum effects:

$$-\frac{dE}{\rho dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{z^2}{\beta^2} \left[\ln \left(\frac{2m_e c^2 \gamma^2 \beta^2}{I} \right) - \beta^2 - \frac{\delta}{2} - \frac{C}{Z} \right], \quad (2.5)$$

where

$\frac{dE}{dx}$ is the linear stopping power (MeV/cm),

ρ is the density of the material (g/cm^3),

$\frac{dE}{\rho dx}$ is the total mass stopping power ($\text{MeV}/((\text{g/cm}^2))$),

N_A is the Avogadro Number,

z_p is the atomic number of the proton (=1),

A is the relative atomic mass,

e is the elementary charge,

Z is the atomic number of the absorbing material r_e is the classical electron radius,

m_e is the electron rest mass,

c is the speed of light in the vacuum,

β is the ratio $\frac{v}{c}$,

γ is the Lorentz factor,

I is the mean excitation potential of the absorbing material,

δ is a density correction that account for the shielding of the outer electrons by the inner ones which will results in a reduction in energy loss in the high energies,

C is a shell correction factor which will only be important for low energies namely when the energy of the particle at stake is close to the energy of the atomic electrons

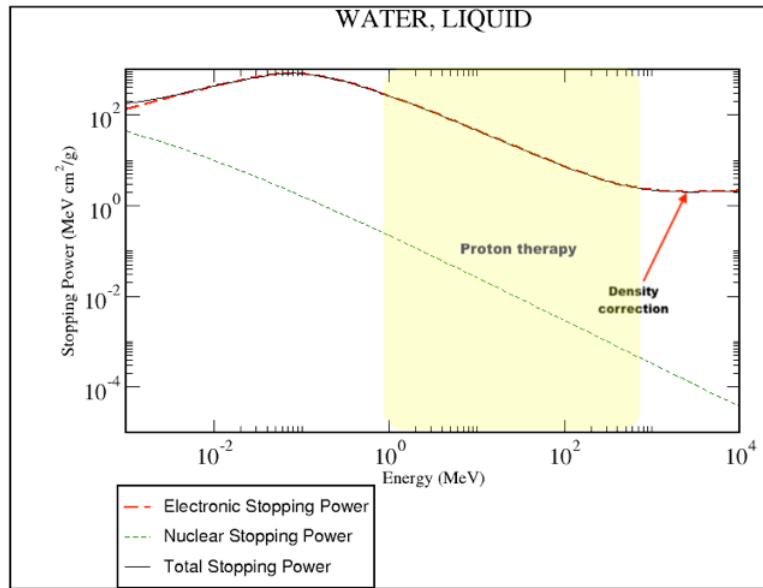


FIGURE 2.5: Stopping power of protons in water as a function of the energy. Computed with P-Star [12]

On the graph 2.5, we displayed 3 curves. In red the electronic mass stopping power also called the collision mass stopping power which is due to the inelastic collisions with the bound atomic electrons, in green the nuclear mass stopping power resulting from the elastic scattering and finally in black the total mass stopping power. We also highlighted in yellow the range of energies used in proton therapy. We see that the proton beam energies usually used in radiotherapy fall short of the relativistic region. In fact, the relativistic effects occur when we reach energies close to the proton rest mass of 938 MeV which is not our case. It means we can neglect the density correction term δ .

The Bethe-Bloch equation allows us to very well depict how the characteristic of the particle dictates the behavior of its energy loss rate. For instance we see that at energy levels used in the clinical framework, the mass stopping power decreased as the square of the speed [13]. We can also notate that from the dependence of the stopping power to the density of electrons in the material ($N_A \rho \frac{Z}{A}$) and to I^{-1} , it is clear that the energy loss rate of protons in the human body will be mainly dictated by the material density they cross. Moreover, the dependence with β makes the energy loss rate very sensitive to the proton velocity. In the light of the above we can already understand how the Bragg-peak form. As it exists a direct link between the dose and the stopping power described as:

$$D_{med} = \Phi_p \left(\frac{S}{\rho} \right)_{p,med}, \quad (2.6)$$

where Φ_p is the local fluence of protons. We could also consider the fluence of electrons which were transferred energy by the protons but it is not necessary for a good computation of the dose.

When the protons will enter the material, they will start deposit a dose directly proportional to their stopping power which will depend on their initial energy. As they will cross the material, the protons will lose energy thus increasing their stopping power as we saw from the graph 2.5. The stopping power will then increase as the protons lose energy thus increasing the dose deposited in the material in an exponential manner leading to the formation of the so-called Bragg-peak.

2.3 Range and energy straggling

The range of a proton beam of a certain energy in a given medium is defined as the depth at which half the protons from the beam will come to rest. The range is an average quantity as small variations in the energy loss occur due to the individual behavior of each proton. This effect is called the range straggling.

In the previous section we considered the energy loss rate to happen in a continuous and smooth manner, but it is because we assumed it to be an average phenomena and we neglected the energy loss rate of individual protons. In most cases it gives us a good first order approximation but the accumulation of many small variations can lead to an energy straggling as for the range.

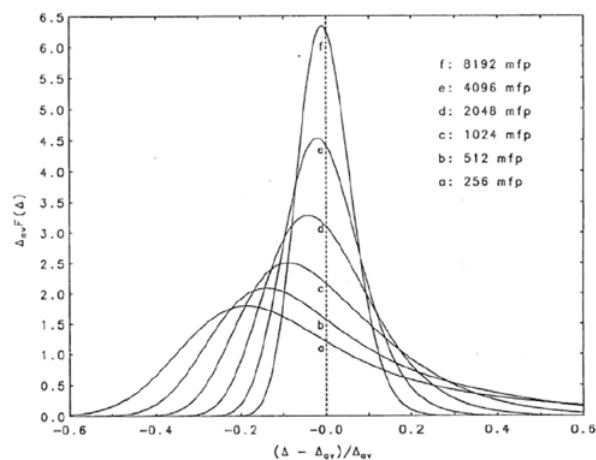


FIGURE 2.6: Energy loss PDFs are plotted for various thicknesses of water absorbers, where the thickness is expressed in units of mean free path (mfp). For visual clarity, the energy-loss PDF's have been scaled on both the abscissa and ordinate. The single event energy loss is expressed as a fraction of the mean energy lost in the entire absorber thickness, or $(\Delta - \Delta_{av})/\Delta_{av}$. Each PDF was scaled so that the integral over all energy-loss values yields unit value. For thin absorbers (curves *a-e*), the PDFs are broader and asymmetric and are modeled with the Valisov or Landau theories. For thick absorbers (curve *f*), the PDFs are symmetric and well-approximated with Bohr's theory (1915), i.e. a Gaussian distribution (Caption reproduced from [1])

The graph 2.6 represents different energy loss rate probability density functions for protons through water absorbers of different thicknesses (express here in units of mean free path) bringing out the energy straggling.

2.4 Multiple Coulomb scattering

Coulomb scattering occurs when a proton passes close enough to a nucleus to feel the repulsive force from its positive charge. We call the combined effect of all scattering event the multiple-Coulomb scattering (MCS). Even the smallest changes in the proton trajectories can have substantial effects and it needs to be accounted for when computing dose distributions.

The main consequence of this MSC is a broadening of the beam. This broadening being larger the deeper we are in the material as we can see on the following graph:

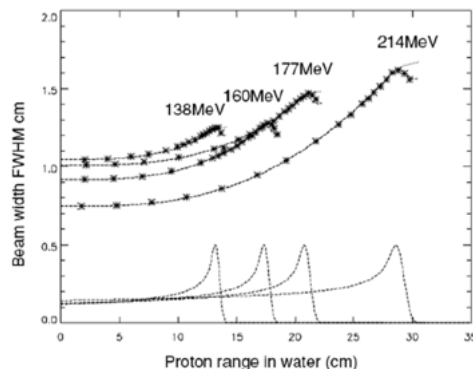


FIGURE 2.7: Broadening of the beam width in water due to multiple Coulomb scattering (Source: [14])

An other unwanted effect directly linked to the MSC is the range mixing effect. In fact we know that the protons range is very affected by the material density. As the beam broadened, part of the protons beam could start going through a different density material next to the one it was planned to go leading to different path length for protons in the same beam hence leading to the building of a secondary Bragg Peak.

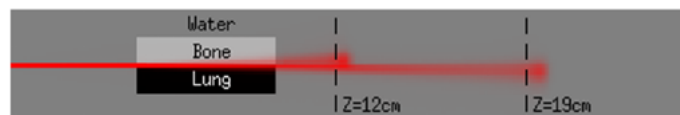


FIGURE 2.8: Dose deposited by a 150 MeV proton pencil beam into a heterogeneous phantom composed of lung, water and bone materials. (Caption reproduced from [1])

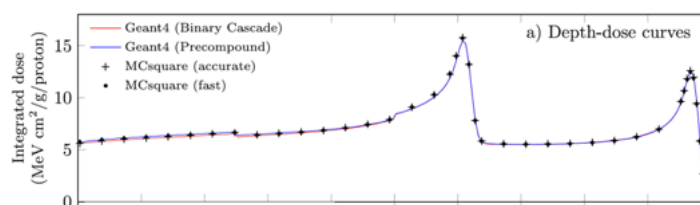


FIGURE 2.9: (a) Integrated depth-dose curves for a 150 MeV proton pencil beam simulated in the heterogeneous phantom of Fig 2.8 using both nuclear models of GEANT4 and both configurations of MCsquare. (Caption reproduced from [1])

2.5 Hounsfield Units and WEPL

2.5.1 Hounsfield Units

When doing a scanner, the attenuation coefficient of X-rays within the tissue is used to produce a gray leveled image. In fact the density of the tissue is directly proportional to the X-rays attenuation. And Hounsfield units can be used to translate this attenuation:

$$HU = 1000 \times \frac{\mu_x - \mu_{water}}{\mu_{water}} \quad (2.7)$$

Where μ_x is the linear attenuation coefficient of the element and μ_{water} the linear attenuation coefficient of water. Using Hounsfield unit and fixing the value for water to 0 we can now scale every voxel (3D pixel) of the CT from -1000 for the air to +1000 for the densest bones.

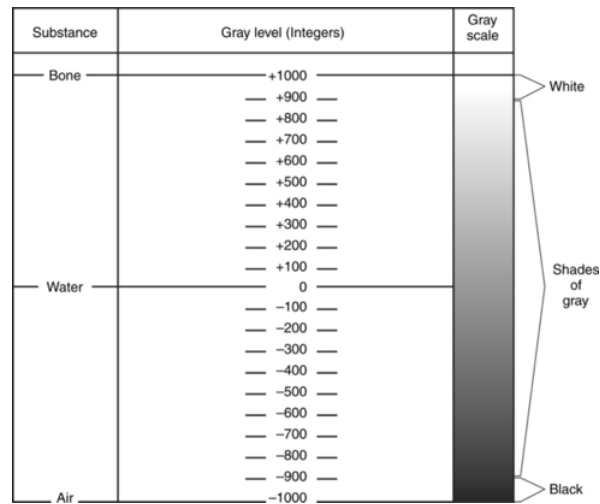


FIGURE 2.10: Matching of the gray scale to HU units corresponding to different density material in the human body (Source: [15])

We will see later in the Chapter 3 that we can actually convert the CT images using the gray scale value of each voxel of the 3D image, usually a $512 \times 512 \times 1XX$ matrix, with the following linear transformation:

$$HU_{value} = Pixel_{value} \times slope + intercept \quad (2.8)$$

2.5.2 WEPL

We usually do our computations in water phantoms considering tissues are radiologically close enough to water to be comparable. But the human body is an inhomogeneous medium and therefore distortions of the beam will occur when the density is less than that of water, as in lungs and other cavities, or when the density is more important than that of water, as in bones for instance. Hence proton beams of the same energy will have different geometric paths in the body depending on the density of the material they cross.

The Water Equivalent Path Length can then be used to approximately account for this density variation effect. Approximately because the attenuation of protons at a point after an absorber of density ρ_m is assumed to be the same as for a scaled thickness t_w of water where $t_w = t_m \frac{\rho_m}{\rho_w} \frac{S_m}{S_w}$.

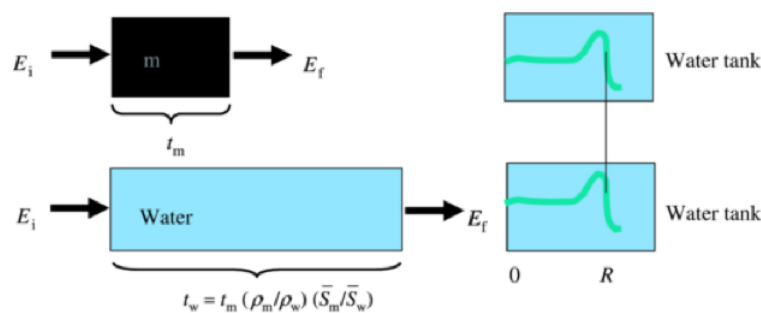


FIGURE 2.11: Water equivalent path length to obtain the same peak in a water tank as if it crossed a denser material of thickness t_m [16].

The WEPL can therefore be defined as the water-equivalent tissue thickness that would attenuate a proton beam in the same way the actual tissue (bones, . . .) would from the source to the calculation point.

Chapter 3

Material and methods

In the first part of this chapter we will introduce the different specific softwares and the files used for the study. In a second part we will describe the data we have and the all reflection process we will go through while conducting this research.

3.1 Material

3.1.1 Patient data

First, a word about the data we are in possession of. We have access to Dicom files from 14 different patients who have been treated for lung cancers within the university hospital Saint-Luc in Bruxelles. The Dicom (Digital imaging and communications in medicine) files are the most commonly used for storing and sharing medical images and can contain a lot of information from the images themselves to the type of scanner used, from the personal information of the patient to the service where the images have been taken. In our case the Dicom have been anonymize to protect the patients identities.

For every patient, we have the reference CT images (the MidpCT1) and a first repeated CT which we call MidpCT2. We also have a second repeated CT (MidpCT3) for every patient except the patient number 9. For all our patients we also have the planned dose distribution files and the recomputed (degraded) dose on both repeated CT. All the dose files contain the data for three beams from three different angles except for the patient 4 which was treated using only beams from two directions.

3.1.2 Treatment planning software

The main software used in this study is the RayStation. The RayStation is a clinical treatment planning system even though our version is a research only system. It comes with many features like multi-criteria optimization, ultrafast computation speed and 4D adaptive radiation therapy. This is the software that helped plan the treatment for every patient and from which the Dicom files we use were uploaded.

Here is what we can see when we open a patient file in the RayStation:

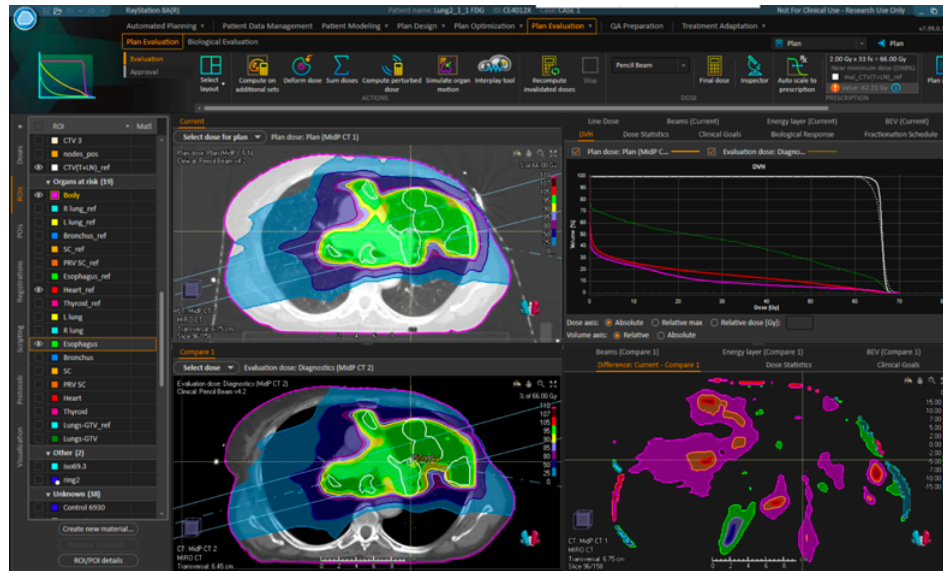


FIGURE 3.1: Actual RayStation display with different possible layouts. Here from top left to bottom right: MidpCT1 with planned dose, DVH of the CTV(T+LN) and some OAR, MidPCT2 with degraded dose, Dose distribution difference.

In this case the plan has already been computed, therefore we can compare both reference and repeated CT images and check for dose distribution differences and see the DVH for every OAR and target volumes that have been drawn.

Another important software we use is REGGUI. REGGUI is a Matlab based interface featuring many different tools such as registration methods, filtering methods, dose volume histogram computation and many others. In our case, as we will see in the next section, we will mainly use it for CT slices visualization and for its powerful rigid registration tool.

The REGGUI interface appears as follows:

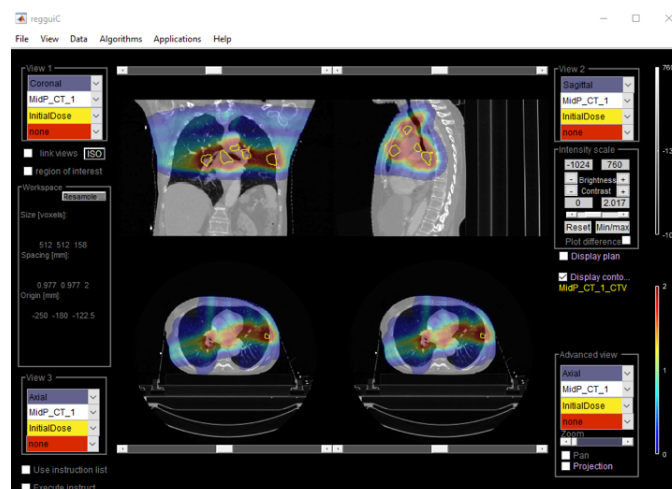


FIGURE 3.2: REGGUI display. From top left to bottom right: Coronal section, Sagittal section and two axial sections. All sections with the target volume contours (yellow) and the associated dose distribution.

We have a layout displaying sagittal, coronal and axial section where we can scroll through the different CT slices on which was added the dose and the contours computed with the RayStation.

REGGUI will also be used directly from MatLab and not from the GUI for WEPL computations for instance. Other scripts were also written using Python which I was more comfortable with.

3.2 Methodology

It should be recalled that for every patient we have a reference MidpCT which we call MidpCT1 and two repeated MidpCT which we call MidpCT2 and MidpCT3.

The first thing we have to do to be able to work on our images is to register them. In fact, we know that rigid registration will help us compensate for the uncertainties we already mentioned earlier like patient positioning and thus allow us to better compare both CT's but this is not the only thing the registration is necessary for. We also need to take into account the fact that the scanner grid from the reference CT scan and the daily scan will not be the same. This will not have any impact on the axial section which will always be 512*512 but the sagittal and coronal section will have different sizes in different repeated CT's. The REGGUI rigid registration tool will help us make sure the images from both CT's are properly aligned before computing anything on them.

Once this step is done, we can upload the contour file from the MidpCT2 and apply the same registration on them. All the contours are contained in a RTSTRUCT file and when it is uploaded, we can choose amongst all the contours that have been drew. The contour that interest us in the context of this study, as we mentioned earlier, is the CTV(T+LN). Having done that, we can then upload the dose file which is the degraded dose computed with the RayStation. We also apply the same registration field on this dose matrix to make sure its size is the same as the CT and the contour size.

The image processing can be summarized as follow:

1. Upload reference CT (MipCT1) images on REGGUI
2. Upload daily CT (MidpCT2) images on REGGUI
 - Rigid registration on the reference CT
3. Upload MidpCT2 CTV(T+LN) contours
 - Apply rigid registration field on the contours
4. Upload MidpCT2 degraded dose
 - Apply rigid registration field on the dose
5. Download all registered files

This process is repeated for the first repeated CT (MidpCT2) and the second repeated CT (MidPCT3) and for every single patient. Except for the first patient which we have had difficulties with. We unfortunately had to remove it from the patients pool.

Now that the data are ready and in the right format to be analyzed, we are able to start the core study. To help specify the different stage we will go through while conducting the study we will use the patient 6 as an example for its important dose distribution differences.

3.2.1 Dose distribution indicators

The first thing to do is to identify the patients which actually need a treatment plan adaptation. In other words, the dose distribution indicators will help us build a set of patients which will be our reference. In the next sections, the indicators should be able to reproduce the same results. In order to do that, we will use the concept of $D_{95\%}$, $D_{50\%}$ and $D_{5\%}$ metrics which are respectively the dose receive by 95, 50 and 5 percent of the volume. Here we will restrain our observation to the target volume. In fact, the metrics on the target volume are more sensitives. Moreover, the CTV is a high priority target and we want to focus on it.

Here is what the DVH for the patient 6 looks like:

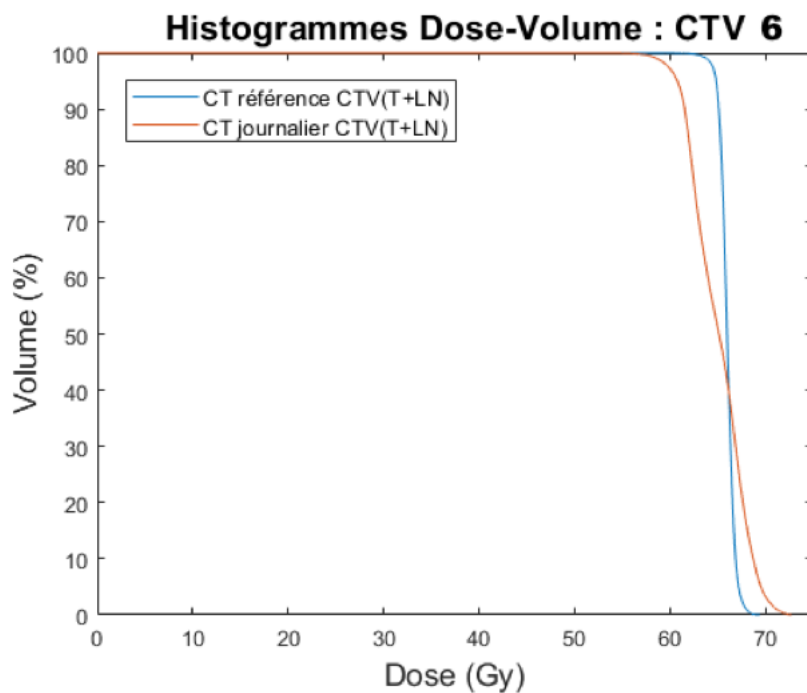


FIGURE 3.3: Dose volume histogram on the CTV(T+LN) for the reference CT (Blue) and the dailt CT (Orange)

On this graph we can observe both DVH, the blue for the CTV(T+LN) on the reference CT and the orange for the daily CT. We can clearly see that by applying the planned treatment on the day of the first delivery, we have an overdosing of about 10% of the volume and an underdosing of almost 50% of the volume. From this DVH we can extract data such as the $D_{95\%}$ or the $D_{5\%}$.

One of the indicators we will test is the homogeneity index. This is an index that can allow us to analyze the uniformity of the dose in the CTV. It is relevant because we want the dose distribution to be the most homogeneous in the target volume, that is, we do not want to see high dose and low dose regions appearing. This index can also be seen as an indicator of the DVH steepness. In fact, the steeper the DVH for a certain volume is, the more homogeneous the dose inside this volume will be. It is defined as such:

$$HI = \frac{D_{95\%} - D_{5\%}}{D_{50\%}} \quad (3.1)$$

We will also test other metrics which are built from the difference of the D_{pres} and the $D_{95\%}$ and the difference of the D_{pres} and the $D_{5\%}$. The difference between the $D_{95\%}$ of the reference CT and the $D_{95\%}$ of both repeated CT's will also be explored as possible solutions for good indicators. As these dose distribution-based indicators are the more reliable but also the most complex and time-consuming ones to compute, their purpose will mainly be to serve as reference to compare the HU and WEPL based indicators we will find in the rest of the study.

Another approach with those dose distribution-based indicators is to find a relation between them and the probability of re-adaptation. We suggest the use of a sigmoid as a relation between the indicators values and the probabilities. We made the hypothesis that the sigmoid could best describe the way human think the probability of adaptation. We would then like to be able to reproduce the same sigmoid parameters with the Hu and WEPL indicators.

3.2.2 HU based indicators

The first stage of evaluation, the less complex and the fastest one, is the comparison of the CT scanner based on the Hounsfield Unit difference. For this purpose we will first have to convert the images from the Dicom files from Gray leveled pixels to HU using the linear transformation we already mentioned. The transformation works as follow:

$$HU_{value} = Pixel_{value} \times slope + intercept \quad (3.2)$$

The slope and the intercept can be found in the Dicom file:

(0020, 0032) Image Position (Patient)	DS: ['-250', '-180', '-94.5']
(0020, 0037) Image Orientation (Patient)	DS: ['1', '0', '0', '0', '1', '0']
(0020, 0052) Frame of Reference UID	UI: 1.2.840.113704.1.111.6656.1388670304.5
(0020, 0060) Laterality	CS: ''
(0020, 1040) Position Reference Indicator	LO: ''
(0020, 1041) Slice Location	DS: "-94.50"
(0020, 4000) Image Comments	LT: ''
(0028, 0002) Samples per Pixel	US: 1
(0028, 0004) Photometric Interpretation	CS: 'MONOCHROME2'
(0028, 0010) Rows	US: 512
(0028, 0011) Columns	US: 512
(0028, 0030) Pixel Spacing	DS: ['0.9765625', '0.9765625']
(0028, 0100) Bits Allocated	US: 16
(0028, 0101) Bits Stored	US: 12
(0028, 0102) High Bit	US: 11
(0028, 0103) Pixel Representation	US: 0
(0028, 1050) Window Center	DS: ['-0600', '-0600']
(0028, 1051) Window Width	DS: ['01600', '01600']
(0028, 1052) Rescale Intercept	DS: "-1024"
(0028, 1053) Rescale Slope	DS: "1"

FIGURE 3.4: Snippet of a dicom file where the rescale slope and intercept can be found

From the CT slice we have converted to HU, we can build a 3D matrix representing the all 3D anatomy of the patient. The idea now is to take the difference of the reference and daily 3D matrix and find metrics we could use to build good indicators of anatomy variations.

But before this, to preserve our self from too many errors, we will restrain the 3D matrices to the area where the target volume has been defined.

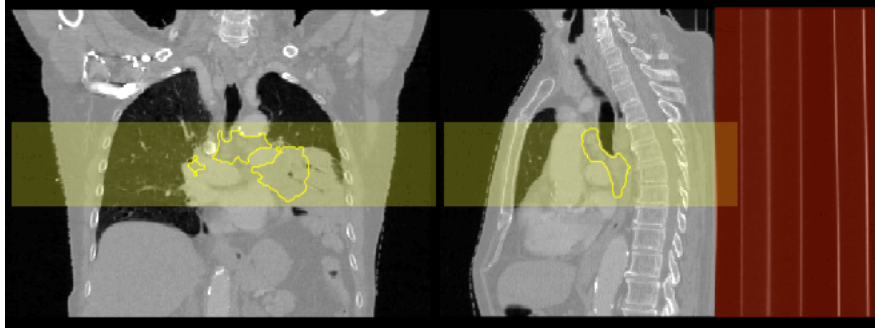


FIGURE 3.5: CT scan from REGGUI where we focus on the slices where the target volume contours appear.

The yellow contours represent the CTV(T+LN) has the physician defined it. As the beams we will use to treat the tumor will be perpendicular to the table, we don't need to consider the changes that could have happened above and below the tumor area. This will allow more accurate indicators but also less computation time. We still need to consider the anatomy in the beam path though because changes before the tumor will influence their penetration.

We will also envisage to crop the table from the images (In red on the previous image) to avoid it to influence our metrics when it does not have any impact on the anatomy. Based on this we will be able to compute the HU difference matrix for every patient and for both repeated CT's.

$$\Delta CT(HU) = CT_{ref}(HU) - CT_{daily}(HU) \quad (3.3)$$

And from the difference we computed we will test metrics from which we would like to build indicators.

The metrics that we have selected are the following:

- The mean squared error (MSE)

The MSE is usually used in image processing to assess for image compression quality.

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2 \quad (3.4)$$

The lower the MSE value, the lower the error hence the more similar the images are.

- The percentile 85/90

The percentile can be understood as such. The percentile 85 for instance will give us the value under which 85% of the observations can be found. In our case it will provide us with the difference value under which 85% of the data can be found.

- The mean absolute error

This metrics works as the MSE but is less sensitive to large errors and it can be useful when comparing pixel to pixel.

- The peak signal to noise ratio

Similarly to MSE, the PSNR is used to compare image compression quality. It represents a measure of the peak error.

$$PSNR = 20 \cdot \log_{10}(Max_I) - 10 \cdot \log_{10}(MSE) \quad (3.5)$$

Where Max_I is the maximum value for a pixel found in the reference CT. The higher the PSNR the more similar the images are.

All the metrics and resulting graphs will be computed using Python.

Comparison of the indicators we will get from our metrics with the results we will have from the dose distribution based indicators will allow us to determine threshold and see if these HU indicators can already reliably predict which patient needs a treatment plan adaptation.

Let us remind that a good indicator can tolerate false positive (FP) but never a false negative (FN). A false positive is when our indicators flag a plan as needing to be re-adapted when it did not actually need to and a false negative is when our indicators tell us we can securely carry on the treatment when we should have re-adapted the treatment.

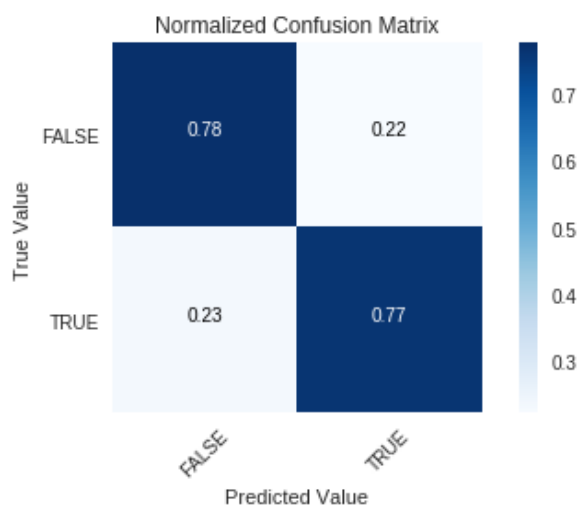


FIGURE 3.6: Normalized confusion matrix generated with the library Sklearn from Python.

The different outcomes possible can be represented as what is called a confusion matrix where on the X-axis are represented the predicted values and on the Y-axis are the true values.

We have 4 different possibilities:

1. Predicted False when it is actually False : True negative (TN)
2. Predicted True when it is actually False : False positive (FP)
3. Predicted False when it is actually True : False negative (FN)
4. Predicted True when it is actually True : True positive (TP)

If our indicator triggers an adaptation for a patient who did not need one, the only drawback will be on the clinical workflow. However if our indicator tells us we can move one to the treatment delivery when we should have been doing a re-planning of the treatment, it will have much more prejudicial effects.

3.2.3 WEPL based indicators

The second stage of evaluation is a bit more complex than HU matrices differences because it takes into account stopping power of the material the beams will cross. For this purpose, MatLab will be used as REGGUI possesses powerful WEPL-related computation tools.

The first step of the code will be to unpack the useful data from the Dicom files that is the CT images, the contours CTV(T+LN) from both reference and the repeated CT. Finally the planned dose and the distorted dose are loaded. It is to notate that every beam has his own dose file. It will allow us to weight them according to the dose they each deliver compare to the total dose. Every patient has been treated using three beams from different incidence angles except for the patient 4 which has been treated only using two beams.

Using all the information from the patient position to the beam angles, we will compute a WEPL map for the reference CT and a WEPL map for the repeated CT for every beam angle. This is done using a build-in function from REGGUI called "WEPL computation".

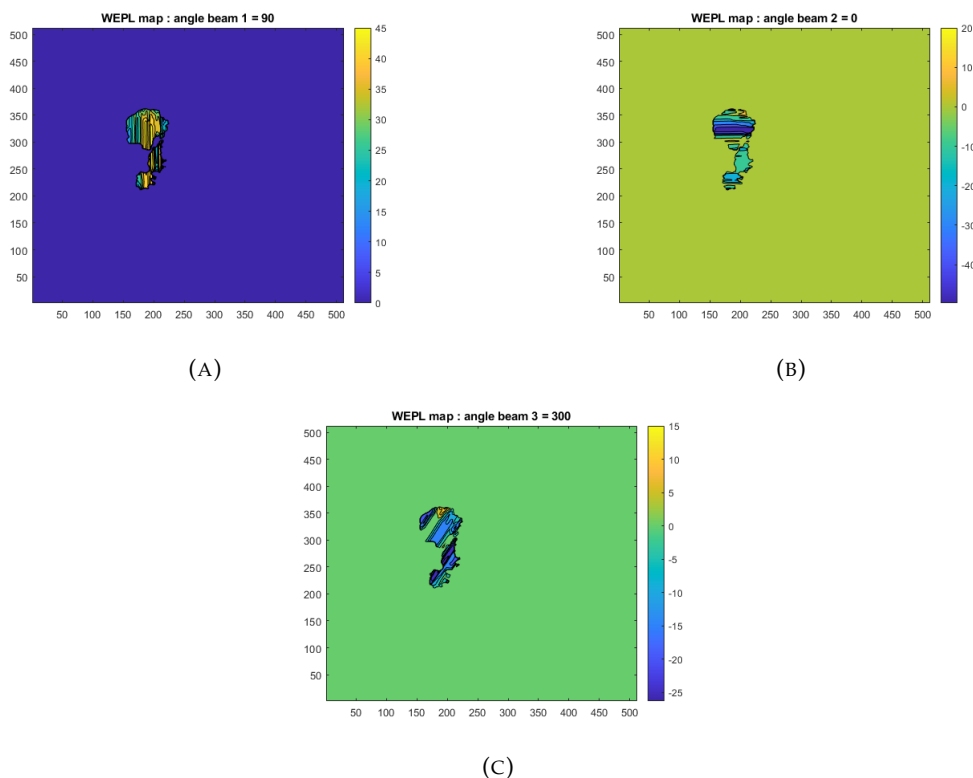


FIGURE 3.7: WEPL map for the three different beam angles: a) 90° b) 0° c) 300° . The difference scale unit on the left is the mm.

We will then compute the differences between those WEPL map for every angle and try to infer any useful information from this difference.

Chapter 4

Results

4.1 Metrics

In this section we will focus on finding metrics that could allow us to build indicators properly determining whether or not we need to adapt a treatment plan for a patient. As seen in the previous section, many metrics exist. For every type of metric, we will first aim our attention on exposing all the results we gathered. And in a second part, we will discuss those results and try to interpret them. For each metric, the aim is to find the best threshold, that is, the threshold that allows absolutely no false negative with a minimum of false positive. We also have to take into account that this threshold must be the same for every repeated CT.

4.1.1 Dose distribution based indicators

As mentioned in the chapter 2 the dose distribution is the most accurate way to tell if a re-planning is necessary or not but it is also the most time consuming one. That is why we will start by working with it and try to find the best and the most reliable indicator and then use it to find a less time consuming indicators giving us the same results.

The first indicator that was tested was the ΔHI which tells us about the homogeneity difference in the target(CTV) between two different CT's (See chapter 2 for more details). To that end we used MATLAB with the dedicated functions from REGGUI and found the following results:

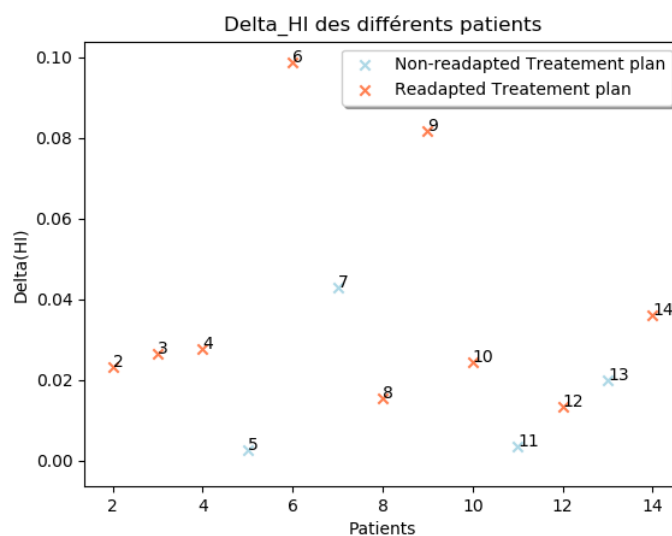


FIGURE 4.1: MidpCT2 ΔHI for the 13 Patients using REGGUI

In the figure 4.1 we colored in **red** the patients who needed a re-adapted treatment plan and in **blue** the ones who did not need it. This distinction between these two cases was based on the work of Julie Vuadens who defined it according to her clinical considerations [17].

We can see on fig 4.1 that we cannot distinguish the non-readapted treatment plan (blue) from the readapted treatment plan (red). Either our HI values or the choice of plan to adapt is wrong. An other possibility would simply be that this indicator is not fit for adaptation flagging.

We first decided to verify our HI values and to recompute them all. And to make sure we did not do anything wrong, we did the computations using two different methods.

The first computation of the ΔHI was made using directly the RayStation which is supposed to give us the most accurate results. We encoded all the $D_{95\%}$, $D_{5\%}$ and $D_{50\%}$ values in an excel sheet for computation. The second method used MatLab and the integrated functions from REGGUI to compute the ΔHI . Both methods gave us the same results so we were confident about the accuracy of our HI values.

It was then the choice of treatment plan to adapt that was to be verified.

The choice on the criteria for the adaptation or not of a plan is totally up to us. And the ones we based our decision on were the following:

$$D_{95\%} > 95\%D_{prescribe}$$

$$D_{5\%} < 105\%D_{prescribe}$$

Where $D_{prescribe} = 66Gy$

Which means that on the one hand we want no less than 95% of the target volume to receive at least 62.7Gy and the other hand we do not want more than 5% of the volume to be irradiated with 69.3Gy or more. (Est ce bien juste?)

We therefore plotted the absolute difference between the $D_{prescribe}$ and the $D_{95\%}$ and the absolute difference between the $D_{prescribe}$ and the $D_{5\%}$ for both repeated CT's.

Following are the results we got for the MidpCT2 which is the first repeated CT:

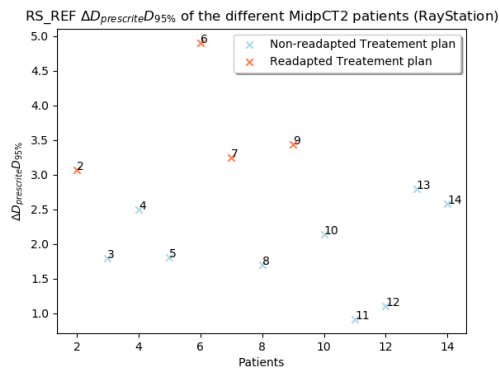


FIGURE 4.2: Difference between the $D_{prescribe}$ and the $D_{95\%}$ for the MidpCT2

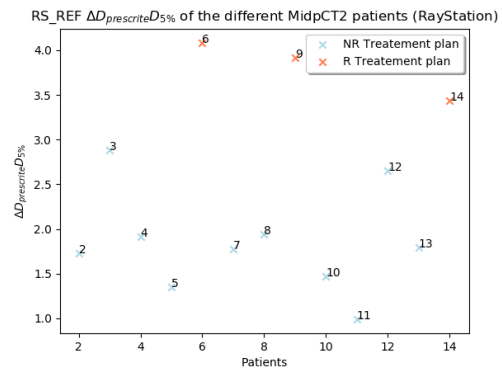


FIGURE 4.3: Difference between the $D_{prescribe}$ and the $D_{5\%}$ for the MidpCT2

Keeping in mind that any difference exceeding (or too close to) $\Delta D_{pres}D_{95\%} = 3.3$ would mean that we did not meet the requirements for qualifying as a good plan, we decided, from the $D_{95\%}$, that the patients to adapt were the patients 2,6,7 and 9 and, from the $D_{5\%}$, the patients 6, 9 and 14. All together, it means that for the first repeated CT, the patients 2,6,7,9 and 14 would require a treatment adaptation. We will make a reservation to the decision to adapt patient 2 though. In fact, we choose to adapt it in the first place but its $\Delta D_{pres}D_{95\%}$ value is 3.07 which is actually underneath the threshold 3.3. The patient 2 is in what we will call the rest of the chapter a "grey zone".

And following are the results we got for the MidpCT3 namely the second repeated CT:

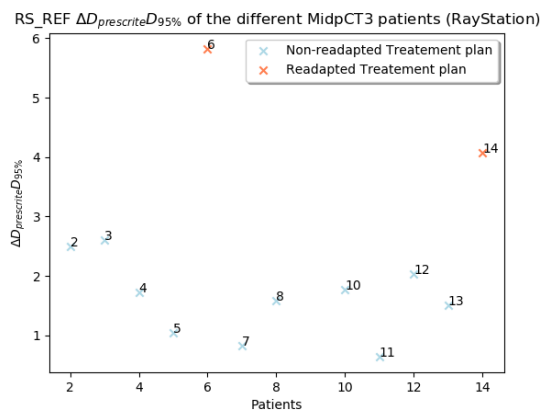


FIGURE 4.4: Difference between the $D_{prescribe}$ and the $D_{95\%}$ for the MidpCT3

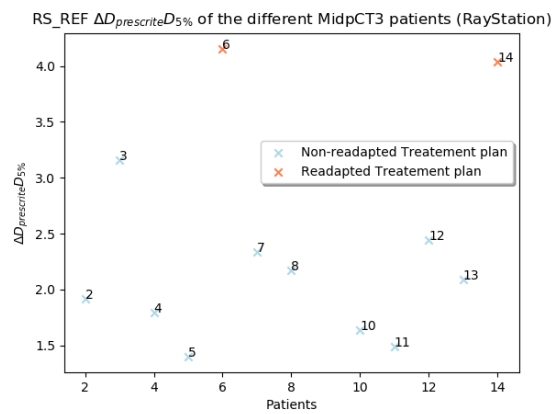


FIGURE 4.5: Difference between the $D_{prescribe}$ and the $D_{5\%}$ for the MidpCT3

From the fig 4.8 and fig 4.7 and with the same criteria used for the MidpCT2, we conclude that only patients 6 and 14 needed to have a treatment plan adaptation.

Now that we decided which patient to adapt we can plot the ΔHI we found earlier to see if it is more relevant now with the patients to adapt we just found.

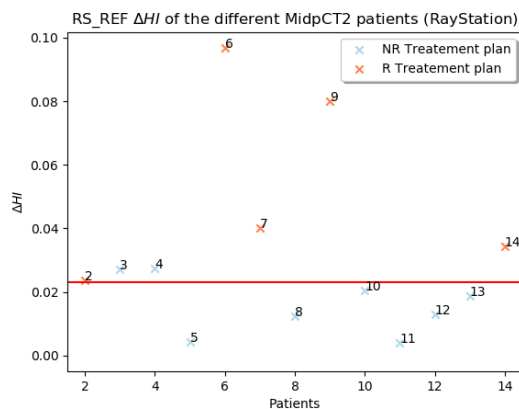


FIGURE 4.6: MidpCT2 ΔHI for the 13 patients with a threshold placed at $\Delta HI = 0.023$

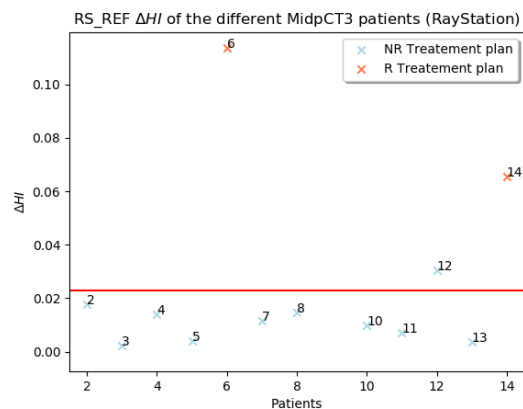


FIGURE 4.7: MidpCT3 ΔHI for the 13 patients with a threshold placed at $\Delta HI = 0.023$

We see that fixing a $\Delta HI = 0.023$, we get all our patients who need a re-planning over the red line with only two false positives for the patients 3 and 4 from the MidpCT2 and one false positive for the patient 12 from the MidpCT3 which is acceptable.

ΔHI is therefore a satisfying dosimetric indicator but we still wanted to try other indicators to make sure none are better. The $\Delta D_{95\%}$ was suggested as a possible indicator and here are the results for both MidpCT we have:

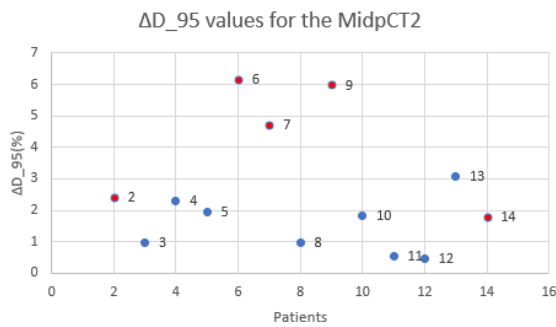


FIGURE 4.8: $\Delta D_{95\%}$ for the 13 MidpCT2 patients

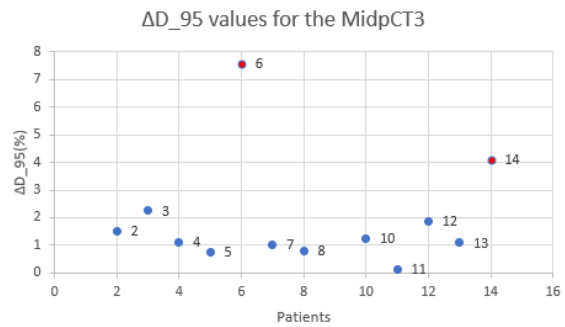


FIGURE 4.9: $\Delta D_{95\%}$ for the 13 MidpCT3 patients

As we can observe, this indicator gives good results for the MidpCT3 but is less efficient for the MidpCT2. In fact it suggests that we should adapt patient 4,5,10 and 11 which is 2 more false positives than with the ΔHI .

The indicators $D_{pres} - D_{95\%}$ and $D_{ref} - D_{95\%}$ were also investigated but were not able to properly predict the right patients needing a treatment plan adaptation.

We therefore decided to keep the ΔHI as our best dose distribution based indicator. It will serves us as a base to compare our futur results.

4.1.2 Hounsfield units metrics

Now that we determined which patients needed a treatment adaptation and which patient did not, we can search for other reliable indicators which can lead us to the same conclusion as the ΔHI but this time in a much less time consuming manner. Remember: the ΔHI is a solid indicator but require a software like the Raystation with dose calculations to be computed.

We extracted the HU matrices from the DICOM files as it was mentioned in the Chapter 3 and used them to compute different metrics on the HU difference matrices between the reference and repeated CT's.

Here are the results we got for different metrics:

The mean squared error

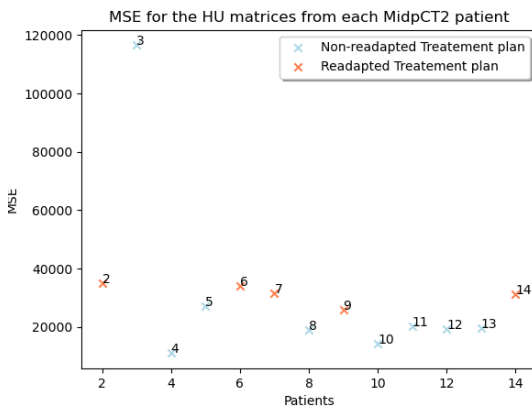


FIGURE 4.10: MSE values for the ΔHU matrices from the MidpCT2.

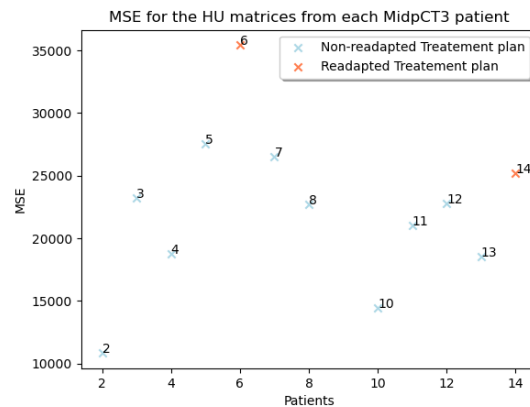


FIGURE 4.11: MSE values for the ΔHU matrices from the MidpCT3.

The mean absolute error

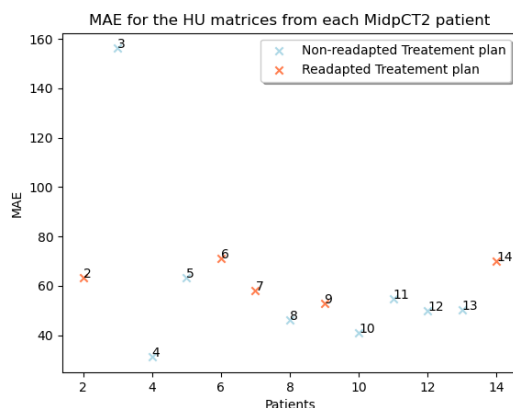


FIGURE 4.12: Mean absolute error values for the ΔHU matrices from the MidpCT2.

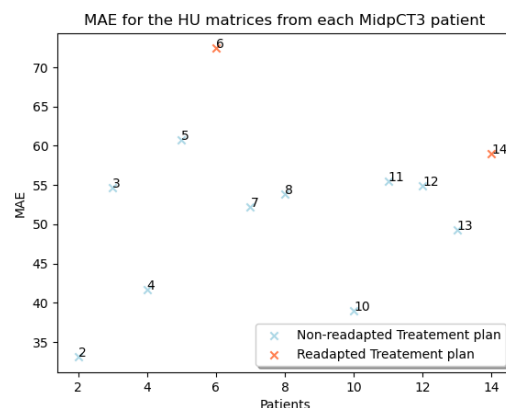


FIGURE 4.13: Mean absolute error values for the ΔHU matrices from the MidpCT3.

The percentile 85

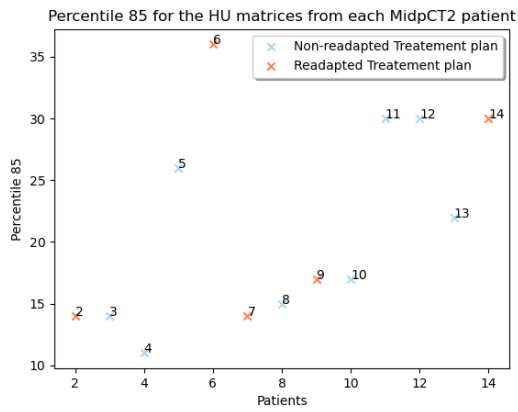


FIGURE 4.14: Percentile 85 values for the ΔHU matrices from the MidpCT2.

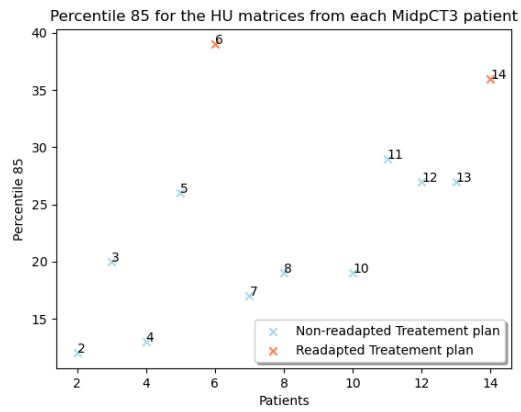


FIGURE 4.15: Percentile 85 values for the ΔHU matrices from the MidpCT3.

The percentile 90

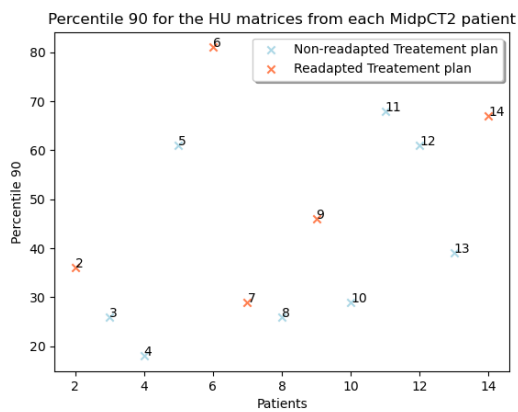


FIGURE 4.16: Percentile 90 values for the ΔHU matrices from the MidpCT2.

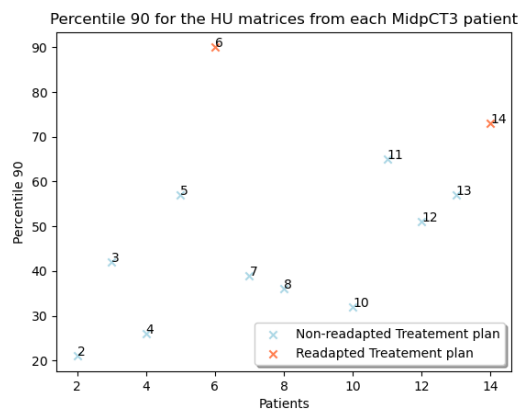


FIGURE 4.17: Percentile 90 values for the ΔHU matrices from the MidpCT3.

The peak signal to noise ratio

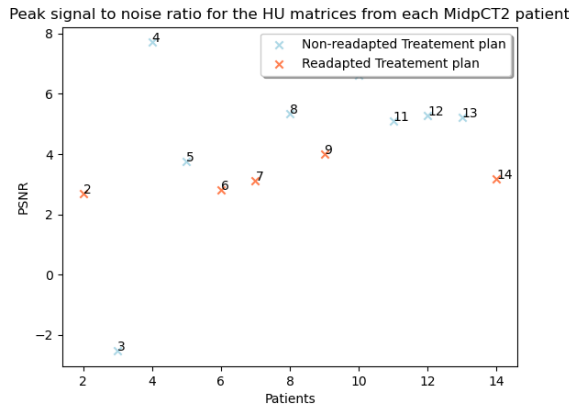


FIGURE 4.18: PSNR values for the ΔHU matrices from the MidpCT2.

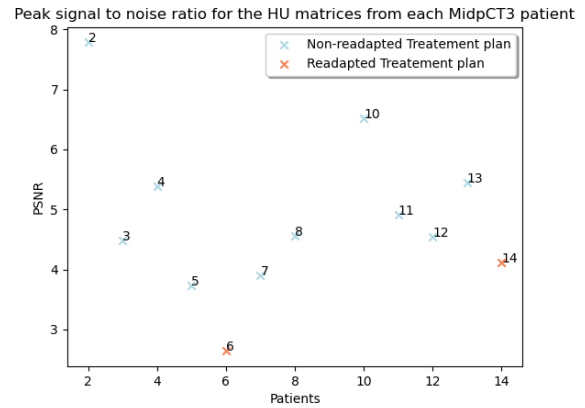


FIGURE 4.19: PSNR values for the ΔHU matrices from the MidpCT3.

To further explore those metrics, we decided to see if it existed any correlation between the metrics and the ΔHI values. In order to do that we will plot the metrics as a function of the ΔHI and use a linear function to fit the data and evaluate the correlation using the Pearson's correlation coefficient R [18]. The relation between both set of data seemed linear and that is why we chose that kind of regression .

Here are the correlations found for the Mean squared errors for both MidpCT:

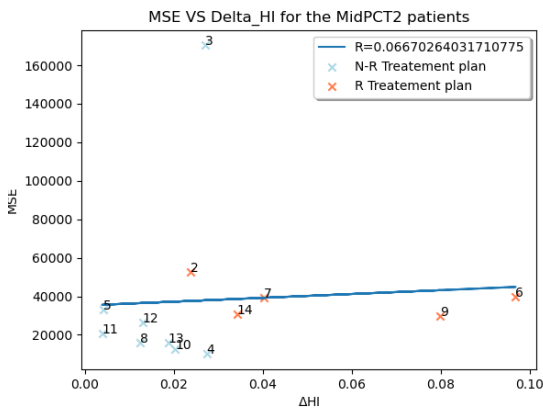


FIGURE 4.20: ΔHU MSE vs ΔHI for the MidpCT2 patients with linear regression.

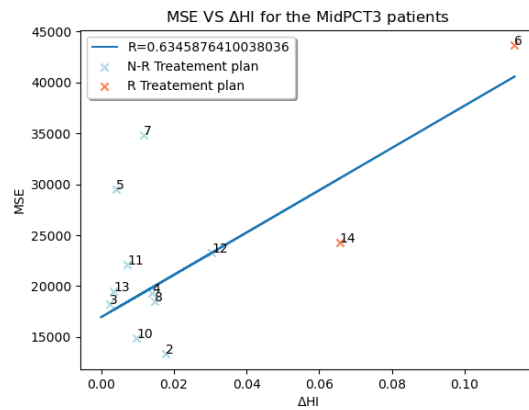


FIGURE 4.21: ΔHU MSE vs ΔHI for the MidpCT3 patients with linear regression.

And here are the correlations found for the peak signal to noise ratio for both MidpCT:

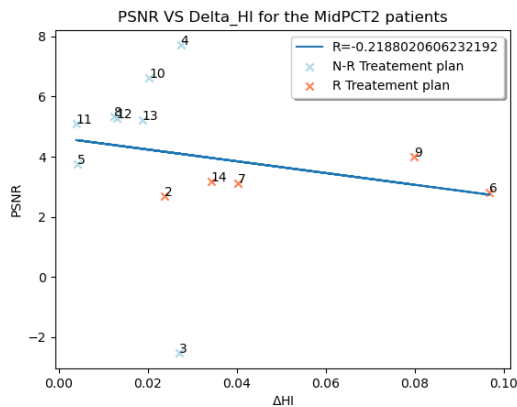


FIGURE 4.22: ΔHU PSNR vs ΔHI for the MidpCT2 patients with linear regression.

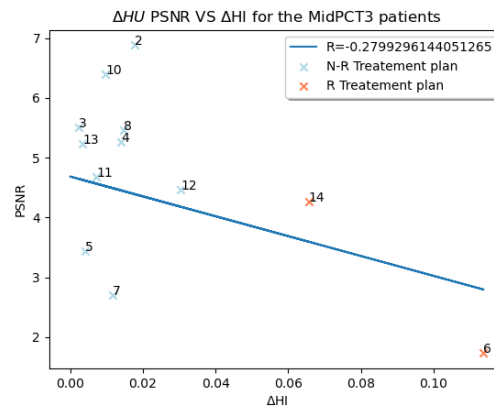


FIGURE 4.23: ΔHU PSNR vs ΔHI for the MidpCT3 patients with linear regression.

4.1.3 ΔHU results discussion

The mean squared error

For the MidpCT2, if we draw a straight line at $MSE = 25000$ we have all the patients needing to be adapted above that line with patients 3 and 5 as false positives but a few false positive does not make a metric flawed. In fact, most of the time we will actually have false positives. We can draw the exact same line on the MidpCT3 graph and we will have patients 5,6,7 and 14 above it. Again the patients to re-adapt are well above the line and we have two false positive with patients 5 and 7.

The mean absolute error

The mean absolute error seems to perform quite well too but does not improve the distinction between adaptation or not. If we draw a straight line at $MSE = 55$, we have one more false positive in the MidpCT2 pool with the patient 11 while we have one less with the MidpCT3. The MAE will not be kept as a possible option for a good metric as it offers nothing more than the MSE.

The percentile 85

Even if the percentile 85 seem to be a good metric for the MidpCT3, we see that for the MidpCT2, all the patients are randomly distributed and no clear distinction can be drawn between both possible cases. We therefore will not keep the percentile 85 as an option.

The percentile 90

The same conclusion as for the percentile 85 is drawn for the percentile 90. We will not keep the percentile 90 as an possible metric.

The peak signal to noise ratio

Let's remind here that, for the PSNR, the higher it is, the better. With that in mind, it seems to be a good metric as we can draw a straight line at $PSNR = 4.2$ and all the patients to adapt will be under this line whether it is for the MidpCT2 or the MidpCT3. Again we will have some false positive but PSNR still seems to be a reliable metric.

For now, concerning the ΔHU , we will keep the MSE and the PSNR as good options to build an HU based indicator.

Correlation discussion

We computed the linear regression for the MSE and the PSNR as it was the metrics which seemed the most reliable.

Regarding the MSE, we observe a very bad correlation coefficient for the MidpCT2 ($R = 0.07$) and it is probably partly due to the fact that the patient 3 which is way off the chart. The MidpCT3 R-value is much better with $R = 0.63$.

Regarding the PSNR, we do not have very great correlation coefficients values for both repeated CT with $R = -0.22$ for the MidpCT2 and $R = -0.28$ for the MidpCT3.

In summary, if we take into account the parameter "correlation" we cannot say that any of the metrics we considered for the ΔHU fulfills our expectations and allows us to think we could build a good indicator out of it. But if we remove the correlation from the equation we see that both MSE and PSNR allow us to properly discriminate the patient we want to adapt from the ones that do not need an adaptation.

4.1.4 Water equivalent path length metrics

In this section we try to find a metric that would reliably predict which patient need an treatment plan adaptation based on the WEPL map differences computed with the integrated REGGUI functions on MatLab.

The percentile 85

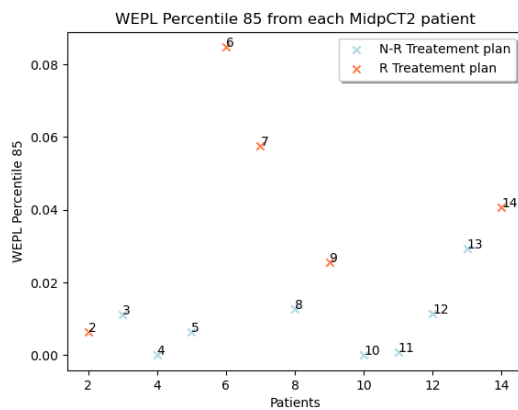


FIGURE 4.24: Percentile 85 for the WEPL Map differences from the MidpCT2.

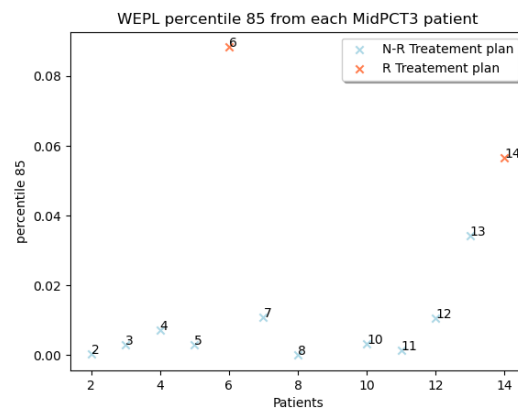


FIGURE 4.25: Percentile 85 for the WEPL Map differences from the MidpCT3.

The percentile 90

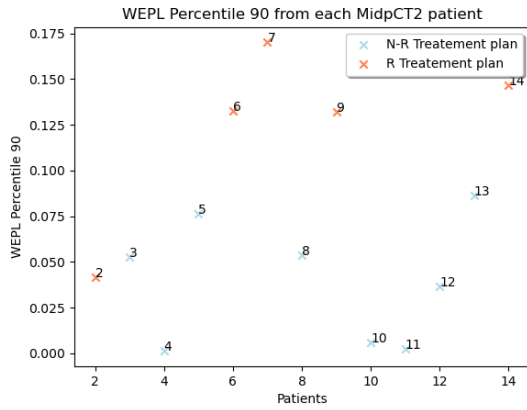


FIGURE 4.26: Percentile 90 for the WEPL Map difference from the MidpCT2.

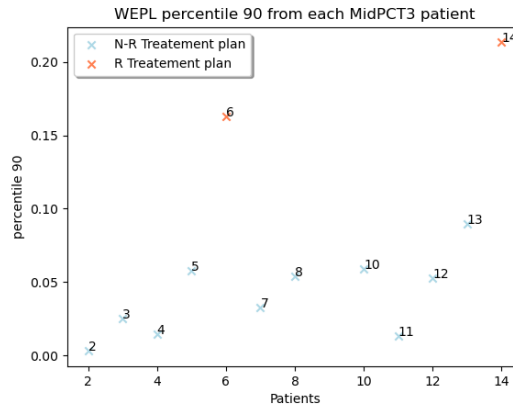


FIGURE 4.27: Percentile 90 for the WEPL Map difference from the MidpCT3.

The percentile 95

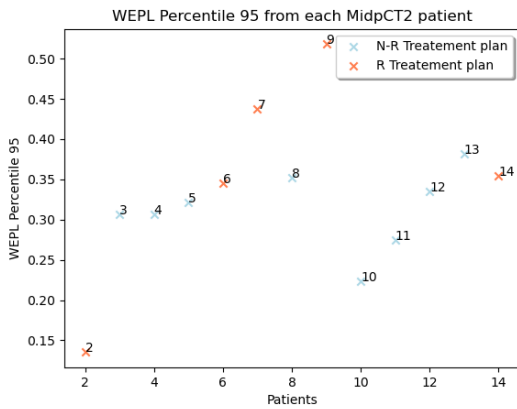


FIGURE 4.28: Percentile 95 for the WEPL Map difference from the MidpCT2.

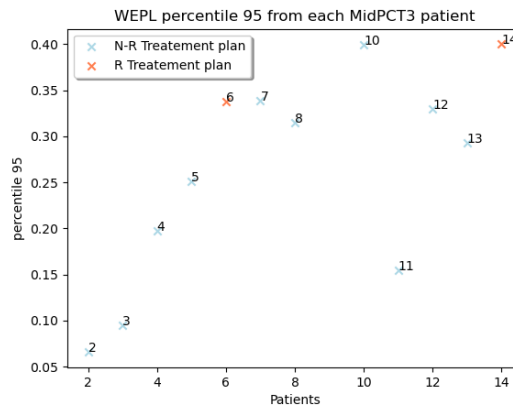


FIGURE 4.29: Percentile 95 for the WEPL Map difference from the MidpCT3.

$\Delta WEPL = 15\text{mm}$

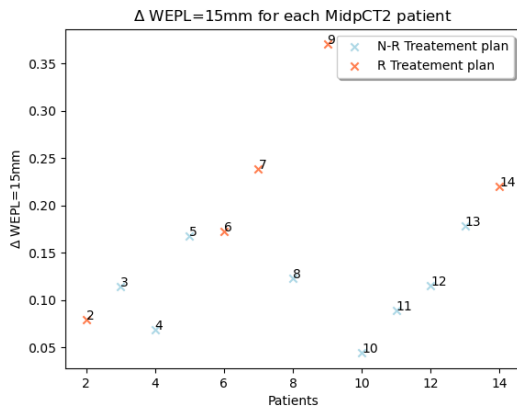


FIGURE 4.30: $\Delta WEPL = 15\text{mm}$ for the WEPL Map difference from the MidpCT2.

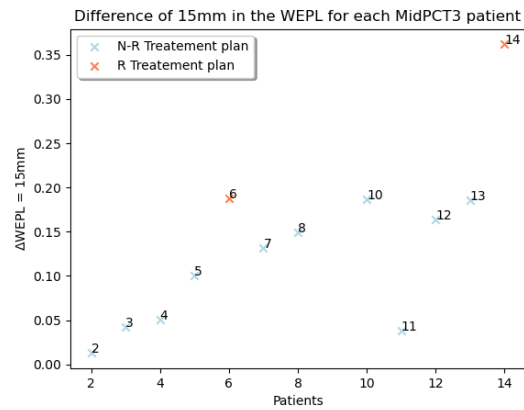


FIGURE 4.31: $\Delta WEPL = 15\text{mm}$ for the WEPL Map difference from the MidpCT3.

The mean squared error

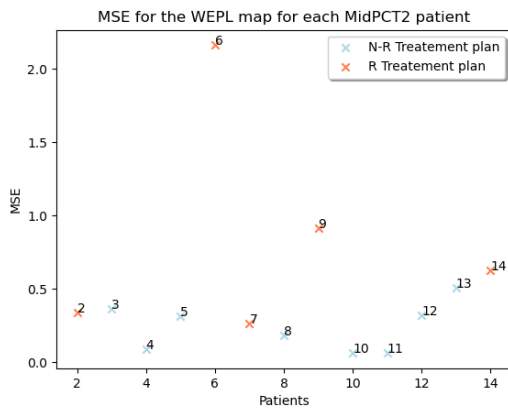


FIGURE 4.32: Mean squared error for the WEPL Map differences from the MidpCT2.

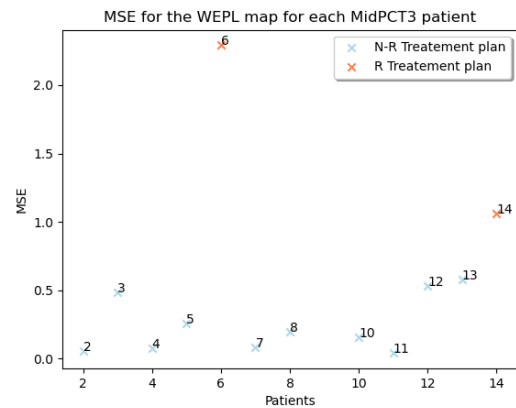


FIGURE 4.33: Mean squared error for the WEPL Map differences from the MidpCT3.

The peak signal to noise ratio

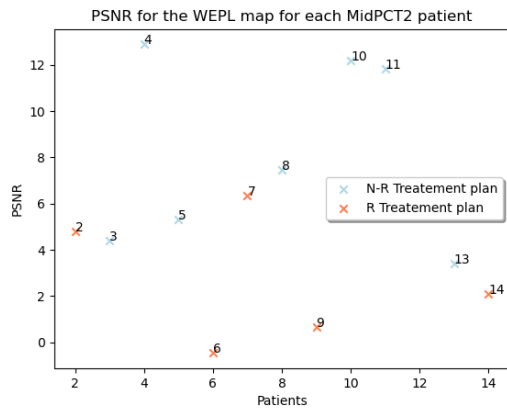


FIGURE 4.34: Peak signal over noise ratio for the WEPL Map difference from the MidpCT2.

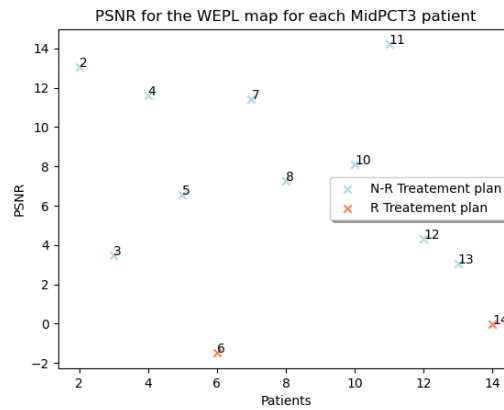


FIGURE 4.35: Peak signal over noise ratio for the WEPL Map difference from the MidpCT3.

To further analyze the different metrics reviewed, we will also plot them as a function of ΔHI to highlight any correlation between both variables. We analyzed all metrics except the percentile 95 which did not satisfy our expectation as it will be reviewed .

Here are the correlations found for the Mean squared errors for both MidpCT:

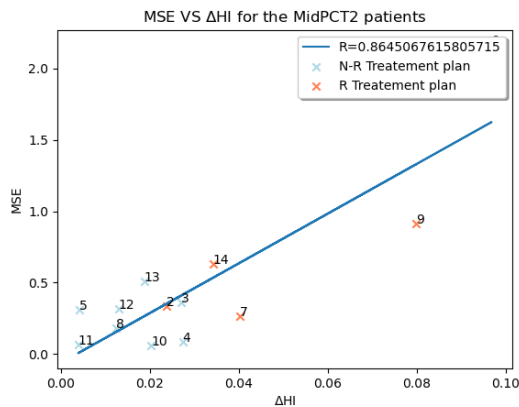


FIGURE 4.36: WEPL MSE vs ΔHI for the MidpCT2 patients with linear regression.

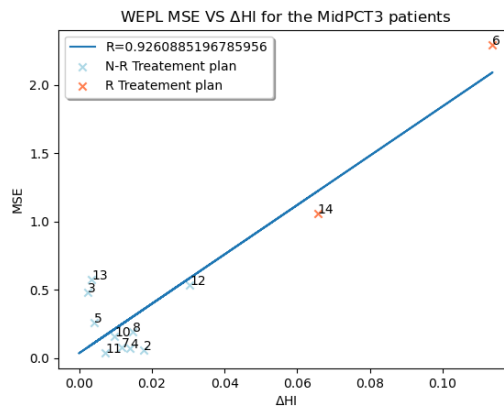


FIGURE 4.37: WEPL MSE vs ΔHI for the MidpCT3 patients with linear regression.

Here are the correlations found for the percentile 85 for both MidpCT:

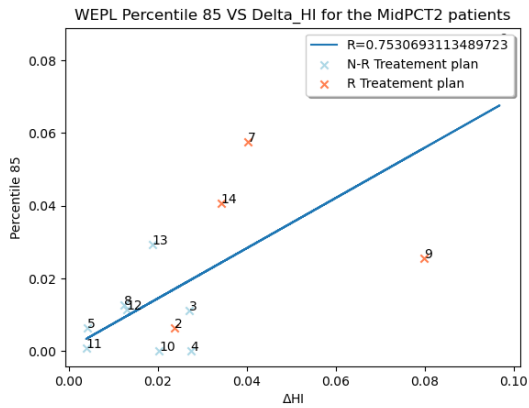


FIGURE 4.38: WEPL percentile 85 vs ΔHI for the MidpCT2 patients with linear regression.

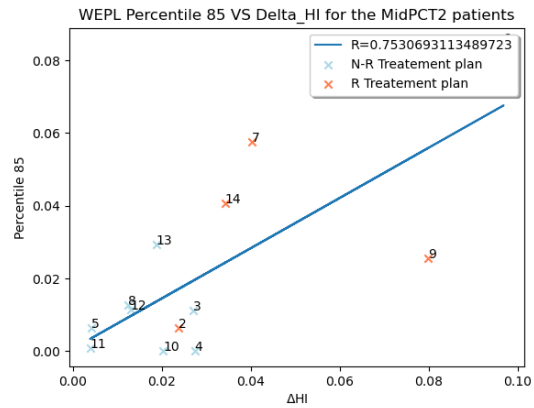


FIGURE 4.39: WEPL percentile 85 vs ΔHI for the MidpCT3 patients with linear regression.

Here are the correlations found for the percentile 90 for both MidpCT:

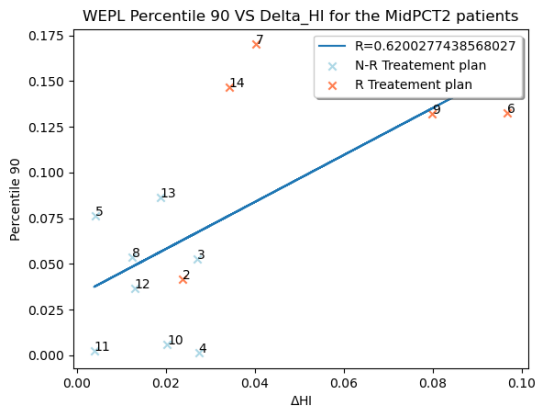


FIGURE 4.40: WEPL percentile 90 vs ΔHI for the MidpCT2 patients with linear regression.

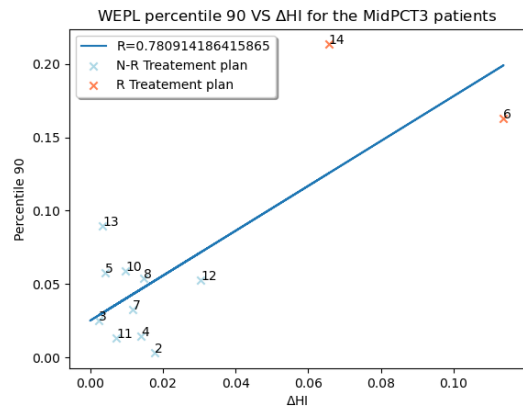


FIGURE 4.41: WEPL percentile 90 vs ΔHI for the MidpCT3 patients with linear regression.

Following are the correlations found for the $\Delta WEPL = 15\text{mm}$ for both MidpCT:

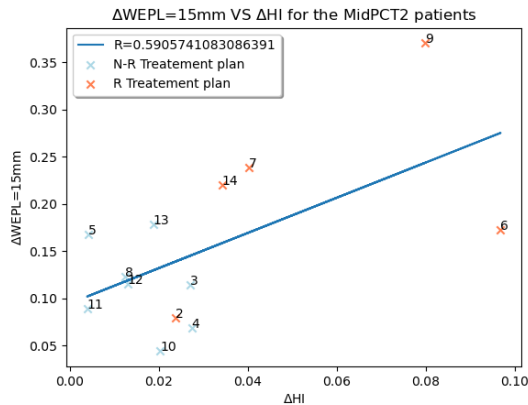


FIGURE 4.42: $\Delta WEPL = 15\text{mm}$ vs ΔHI for the MidpCT2 patients with linear regression.

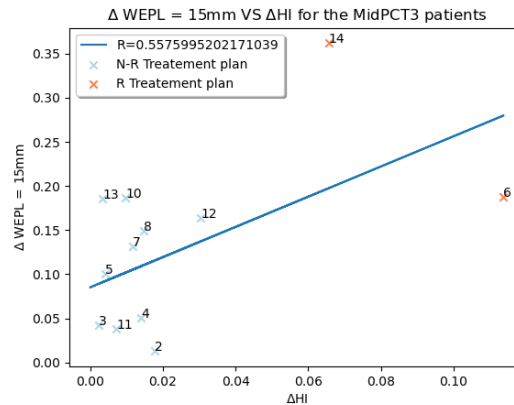


FIGURE 4.43: $\Delta WEPL = 15\text{mm}$ vs ΔHI for the MidpCT3 patients with linear regression.

And finally, here are the correlations found for the peak signal to noise ratio for both MidpCT:

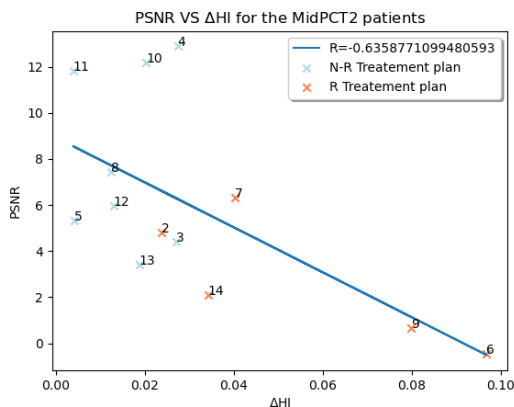


FIGURE 4.44: PSNR vs ΔHI for the MidpCT2 patients with linear regression.

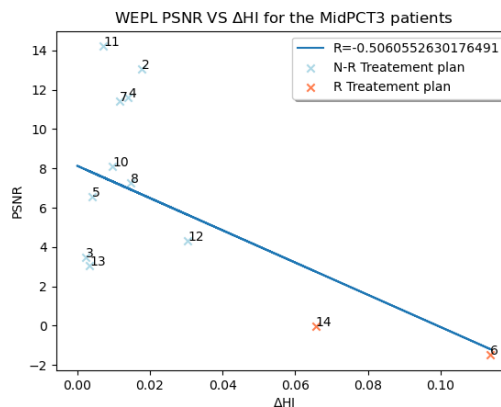


FIGURE 4.45: PSNR vs ΔHI for the MidpCT3 patients with linear regression.

4.1.5 WEPL metrics results discussion

The percentile 85

The graphs shows very good results and especially for the MidpCT3. The only black spot we could raise here is for the MidpCT2. In fact we have the patient 2 which we decided to adapt that have a small percentile 85 value indicating we will not need to adapt its treatment plan. That is what we would have defined as a false negative and we want to avoid it at all cost. But one has to remember that in the section 4.1.1 we made a reservation to the decision to adapt patient 2 and it can explain the results we have here for its percentile 85 value.

The percentile 90

We draw here the exact same conclusions as for the percentile 85 but adding that patient 2 can be seen as false negative if we set a threshold around $percentile_{90} = 0.1$ but can also be properly flagged if the threshold is set lower, at values near **0.03**. The same threshold slide could have been applied for the percentile 85.

The percentile 95

The percentile 95 as opposed to the percentile 85 and 90 will not be withheld as a possible metric. In fact, we cannot draw any conclusion from the results it gives us except for the fact that it seems useless.

$\Delta WEPL = 15mm$

For the $\Delta WEPL$, it seems we have a good metric if we disregard the patient 2. In fact, if we consider that the patient 2 does not actually need to be adapted, the $\Delta WEPL$ set at **0.15** gives us two false positive for the MidpCT2 and 4 false positive for the MidpCT3. We will then chose to keep it as a usable metric.

The mean squared error

Here we can tell that the mean squared error can potentially perform well if a threshold is set at **0.25** but it would induce the adaptation of 4 more patients in both MidpCT2 and MidpCT3 which is less efficient than the other metrics considered as valid so far.

The peak signal to noise ratio

In the case of the PSNR, if we set the threshold at **6.5**, we can distinguish the treatment plan to adapt from the one not to adapt with 4 false positive for the MidpCT2 (patients 3,5,12 and 13) and 3 false positive for the MidpCT3 (patients 3,12 and 13). It is then a valid metric even if it is not the most efficient one.

Correlation discussion

Here again we decided to compute the trend line between the metric values and the ΔHI values seeking for any correlation.

First, in the case of the mean squared error, we find a very strong correlation between our variables with $R = 0.86$ for the MidpCT2 and $R = 0.93$ for the MidpCT3.

For the percentile 85, we observe a strong correlation between the variables with $R = 0.75$ for the MidpCT2 and $R = 0.91$ for the MidpCT3.

In the case of percentile 90 we have a good regression with $R = 0.62$ for the MidpCT2 and $R = 0.78$ for the MidpCT3. The correlation is slightly worse compared to the percentile 85 but still good.

Again, if we consider the patient 2 as wrongly classified and as sitting in a grey zone, we see a nice clustering of both specific cases. Here the correlation seems weaker than in the previous cases with $R = 0.56$ for the MidpCT2 and $R = 0.59$ for the MidpCT3. Here the correlation seems similar to the $\Delta WEPL$ with $R = -0.64$ for the MidpCT2 and $R = -0.51$ for the MidpCT3.

All the metric values are summarized and brought together into the following tables. A first table for the MidpCT2 and a second for the MidpCT3 with the patients considered to be needing a treatment plan adaptation highlighted in red.

MidPCT2

<i>Patients</i>	WEPL					
	Percentile 85	Percentile 90	Percentile 95	15 mm	MSE	PSNR
1	NaN	NaN	NaN	NaN	NaN	NaN
2	0.006	0.041	0.135	0.079	0.336	4.809
3	0.011	0.052	0.307	0.114	0.364	4.418
4	0	0.001	0.306	0.069	0.086	12.888
5	0.006	0.076	0.321	0.168	0.312	5.312
6	0.085	0.133	0.345	0.172	2.163	-0.467
7	0.058	0.170	0.437	0.239	0.265	6.335
8	0.013	0.054	0.352	0.123	0.183	7.449
9	0.026	0.132	0.518	0.370	0.914	0.669
10	0	0.006	0.224	0.044	0.061	12.205
11	8.88E-04	0.002	0.275	0.089	0.066	11.828
12	0.012	0.036	0.335	0.115	0.317	5.951
13	0.029	0.087	0.382	0.179	0.506	3.397
14	0.040	0.147	0.354	0.219	0.628	2.105

TABLE 4.1: WEPL metrics for the MidpCT2 patients. The red highlighting is used for the patients initially chosen to be adapted.

<i>Patients</i>	ΔHU				
	Percentile 85	Percentile 90	MSE	MAE	PSNR
1	NaN	NaN	NaN	NaN	NaN
2	14	36	34998.015	63.270	2.690
3	14	26	116517.639	156.154	-2.533
4	11	18	11019.522	31.196	7.709
5	26	61	27296.819	63.187	3.769
6	36	81	34078.674	71.207	2.806
7	14	29	31640.318	58.025	3.128
8	15	26	19043.869	46.379	5.333
9	17	46	25854.773	52.793	4.005
10	17	29	14139.641	40.938	6.626
11	30	68	20102.642	54.665	5.098
12	30	61	19232.987	49.989	5.290
13	22	39	19483.642	50.337	5.234
14	30	67	31226.043	69.971	3.186

TABLE 4.2: ΔHU metrics for the MidpCT2 patients. The red highlighting is used for the patients initially chosen to be adapted.

MidPCT3	WEPL					
	Percentile 85	Percentile 90	Percentile 95	15 mm	MSE	PSNR
<i>Patients</i>						
1	NaN	NaN	NaN	NaN	NaN	NaN
2	2.95E-04	0.003	0.066	0.012	0.057	13.068
3	0.003	0.025	0.095	0.042	0.482	3.457
4	0	0.014	0.197	0.051	0.074	11.604
5	0.007	0.058	0.251	0.100	0.258	6.585
6	0.088	0.163	0.338	0.188	2.292	-1.496
7	0.003	0.033	0.339	0.132	0.079	11.409
8	0.010	0.053	0.315	0.149	0.193	7.24
9	NaN	NaN	NaN	NaN	NaN	NaN
10	0.003	0.059	0.399	0.187	0.156	8.114
11	0.001	0.013	0.154	0.038	0.039	14.208
12	0.011	0.052	0.329	0.164	0.532	4.315
13	0.034	0.089	0.293	0.185	0.578	3.066
14	0.057	0.213	0.399	0.362	1.059	-0.023

TABLE 4.3: WEPL metrics for the MidpCT3 patients. The red highlighting is used for the patients initially chosen to be adapted.

MidPCT3	ΔHU				
	Percentile 85	Percentile 90	MSE	MAE	PSNR
<i>Patients</i>					
1	NaN	NaN	NaN	NaN	NaN
2	12	21	10825.972	33.086	7.786
3	20	42	23188.999	54.677	4.478
4	13	26	18777.046	41.671	5.394
5	26	57	27528.651	60.742	3.733
6	39	90	35413.636	72.416	2.639
7	17	39	26509.745	52.259	3.897
8	19	36	22738.709	53.828	4.563
9	NaN	NaN	NaN	NaN	NaN
10	19	32	14473.182	38.943	6.525
11	29	65	21004.218	55.484	4.908
12	27	51	22806.645	54.935	4.550
13	27	57	18529.119	49.241	5.452
14	36	73	25168.925	59.043	4.122

TABLE 4.4: ΔHU metrics for the MidpCT3 patients. The red highlighting is used for the patients initially chosen to be adapted.

Based on these tables and on the graphs previously analysed one by one, we defined threshold for every metrics meant to best distinguish the patient we will re-adapt from the ones we will not. Of course, these threshold need to be the same for every repeated CT's. We have not bothered setting threshold for the ΔHU percentile metrics because of their non-reliability.

Here are the thresholds we set:

WEPL

	Percentile 85	Percentile 90	Percentile 95	15 mm	MSE	PSNR
<i>Threshold</i>	0.02	0.1	0.34	0.15	0.25	7

ΔHU

	MSE	MAE	PSNR
<i>Threshold</i>	25000	55	4.2

TABLE 4.5: Table displaying all thresholds set for each metrics with both WEPL and ΔHU

4.2 Indicators

In this section, using the data we have about the different metrics we gathered in the previous section, we will try to build indicators which would give us a good prediction of which patients need to see their treatments re-adapted and which do not. We actually would like to find a way to translate metric values to adaptation probabilities. To do so, we investigated two different approaches. The first approach will use very strict criteria and separate the patients between three different classes: adapted, grey zone and not adapted and make use of ramp functions. The second approach will consist in a splitting of the patient pool in only two classes: adapted or not. In this case we will use logistic functions instead of ramp functions as in the first approach.

In both cases we will use confusion matrices to analyse the probabilities and compare the predictions obtained with the actual adaptation and non-adaptation cases to help us choose which indicators perform the best. We would like to remind that the final goal is to build a multistage flagging system. Therefore, we would like to find indicators that would give good predictions at every stage.

4.2.1 First approach

In a first instance, we suggested linking metric values to adaptation probabilities by simply stating that our thresholds (table 4.5) corresponded to a probability of adaptation of 50%. Following this, the further above this threshold the value is, the more likely we are to adapt the treatment plan with of course an upper limit of 100%. In the same way, the further underneath this threshold the value is, the less likely we are to adapt. This rule was reversed for the PSNR as we know that the bigger the PSNR value is, the better. The resulting function is what is called a ramp function, that is, a linear function between two plateau at 0% and 100%.

For instance, for the MSE WEPL metric, here is how we computed the adaptation probability:

$$Probability = \frac{MSE_Value}{Threshold} \times 50 \implies Probability = \frac{0.336}{0.25} \times 50 = 67.28\% \quad (4.1)$$

And we repeated the process for every patient to obtain the following graph:

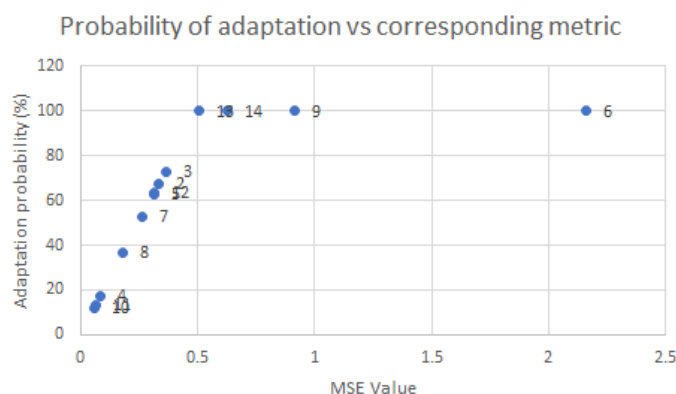


FIGURE 4.46: Ramp function representing the increase in adaptation probability with the increase of the MSE value

We chose this particular type of function because it is what seemed the most intuitive when it comes to associate metric values to probabilities.

The following tables give us the probability of adaptation given by every metric for every patient first for the WEPL and then for the ΔHU . Seeking for visibility, we will only displayed results for the MidpCT2:

<i>Patient</i>	WEPL: adaptation probabilities (%)				
	Percentile 85	Percentile 90	15 mm	MSE	PSNR
1					
2	16	20.75	26.4	67.28	63.01
3	28	26.45	38.13333333	72.94	66.02
4	0	0.6	22.83333333	17.26	0.86
5	16	38.1	56	62.58	59.14
6	100	66.4	57.43333333	100	100
7	100	85.05	79.5	52.92	51.27
8	31.75	26.85	41.06666667	36.58	42.698
9	63.75	66.15	100	100	94.85
10	0	2.95	14.66666667	12.12	6.12
11	2.21945	1.09	29.6	13.14	9.02
12	28.75	18.2	38.36666667	63.36	54.225
13	73.5	43.25	59.5	100	73.87
14	100	73.35	73.3	100	83.805

TABLE 4.6: Table displaying the adaptation probabilities associated with the metrics studied for the WEPL. The red highlighting is used for the patients initially chosen to be adapted.

<i>Patient</i>	ΔHU: daptation probabilities (%)		
	MSE	MAE	PSNR
1			
2	69.99603087	57.51845996	67.971
3	100	100	100
4	22.03904343	28.35992784	8.224
5	54.59363761	57.44285878	55.122
6	68.15734815	64.733692	66.595
7	63.28063675	52.75046638	62.757
8	38.08773877	42.16231544	36.508
9	51.7095458	47.993521	52.317
10	28.27928288	37.21663794	21.114
11	40.20528482	49.69503833	39.306
12	38.46597321	45.44426849	37.019
13	38.96728404	45.76068785	37.689
14	62.45208653	63.610395	62.075

TABLE 4.7: Table displaying the adaptation probabilities associated with the metrics studied for the ΔHU . The red highlighting is used for the patients initially chosen to be adapted.

We decided that a patient needed to be re-adapted if the probability was above 65% and did not need to below 35%. In between we have a grey zone where it might be safer to re-adapt the treatment.

The first observation we can already make here is that, using a ramp function, none of the metrics seem to be able to predict with high confidence which patients need a treatment plan adaptation. In fact for all three ΔHU metrics, the patient 9 is always classified in the Grey zone. But it could still be accepted as a patient in the grey zone would be adapted as we do not want to take any risk. What we really want to avoid here is a metric that would tell us a patient does not need an adaptation when it was actually necessary. Let's remind here that, for the MidpCT2, we want patients 6,7,9 and 14 to be flagged and the patient 2 to be in a grey zone. On the other hand, for the WEPL metrics, we make similar observations. However, we want to highlight that percentile 85, percentile 90 and $\Delta WEPL = 15mm$ all classify patient 2 as clear for treatment when we want patient 2 to be in the grey zone. But even MSE and PSNR have a weak spot with patient 7 which is classified in the grey zone when we would have prefer it to be flagged as to be adapted.

Given the following observations about single metric flagging, we tried to find metric combinations that would be able to predict the patients to adapt with more accuracy. We started with the double metrics combinations. Doing so, we found no combination of ΔHU metrics that could make a proper prediction. But we could actually find combinations of WEPL metrics that did it convincingly:

Patients	Double metrics combinations(%)		
	WEPL_perc85+MSE	WEPL_perc90+MSE	WEPL_15mm+MSE
2	41.64	44.015	46.84
3	50.47	49.695	55.537
4	8.63	8.93	20.047
5	39.29	50.34	59.29
6	100	83.2	78.717
7	76.46	68.985	66.21
8	34.165	31.715	38.823
9	81.875	83.075	100
10	6.06	7.535	13.393
11	7.679	7.115	21.37
12	46.055	40.78	50.863
13	86.75	71.625	79.75
14	100	86.675	86.65

TABLE 4.8: Table displaying the adaptation probabilities associated with different double metrics combinations studied. Green is used for probabilities < 35%, yellow for the grey zone: 35-65% and red for the probabilities > 65%.

The distinction between good combinations and bad ones was done using confusion matrices :

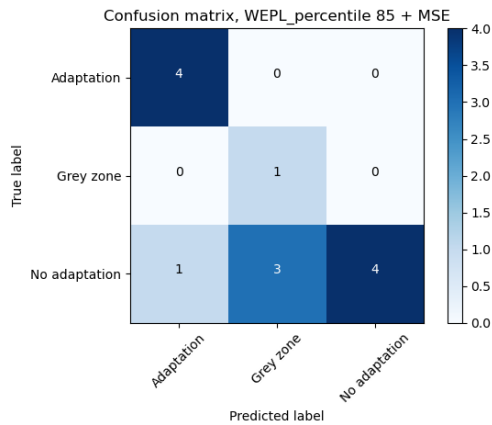


FIGURE 4.47: Confusion matrix for the combination of WEPL percentile 85 and WEPL MSE.

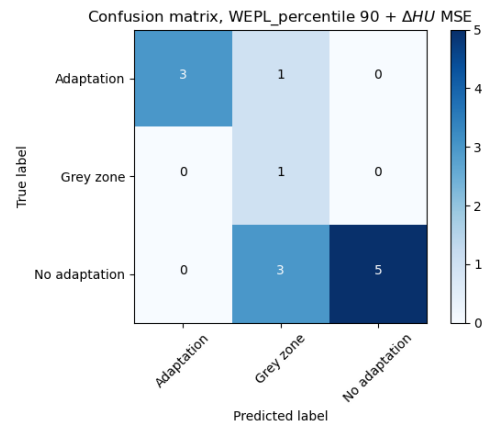


FIGURE 4.48: Confusion matrix for the combination of WEPL percentile 90 and ΔHU MSE.

The figure 4.47 represents the indicator found in the first column of the table 4.8. It shows that no patient truly needing adaptation have been predicted as belonging to the grey zone or to the no adaptation pool.

This is not the case for the other indicator in figure 4.48, combining WEPL and ΔHU metrics, where one patient known as needing to be adapted was classified in the grey zone. As we already said, it would not be a major issue as we would still adapt a patient in the grey zone but it helps us sort the indicator.

This procedure using confusion matrices was used for every single metrics and double metric combinations and allowed us to separate the best metrics/indicators from the others.

Proceeding this way, we found three ways of combining metrics that gave us correct results, meaning that patient 6,7,9 and 14 were predicted as to be adapted and the patient 2 was in a grey zone. The combinations only involved WEPL metrics. We will also note that patient 13 is always flagged for an adaptation process but will count as a false positive.

These exact same WEPL metrics combinations were applied to the MidpCT3 and gave us the right patient predictions with one false positive for two of them and two false positives for one of them, increasing thus our trust in those combinations reliability.

4.2.2 Second approach

For the second approach we decided to focus on the single metric indicators.

Until now, we worked with ramp functions as it seemed the be most intuitive manner we could use to define adaptation probabilities. Indeed, it feels natural for a us to linearly link metric values to adaptation probabilities. But it might not be the actual best way of doing so and we therefore search in the literature for a function that could be more accurate. For this purpose we investigated the use of logistic functions. In fact, logistic functions are particularly useful in the case of dichotomous variables [19] and it could be convenient here as we have two possible outcomes: adaptation or not. An other important advantage of the logistic function is that it is continuous and differentiable. For this approach we will consider the patient 2 as not to be adapted and we will set the adaptation threshold at 50%.

The idea was the following:

- We will first do what we did in the previous section, that is, plot the probability of adaptation as a function of the specific metric values to obtain a ramp function.
- We now fit a logistic function on the ramp function and extract the sigmoid parameters.
- We inject our metrics values in this logistic function to obtain adjusted adaptation probability values.
- We do exactly the same for the ΔHI values to obtain the adaptation probabilities as a function of them.
- We can now plot both probability functions on different axes to see if any correlation between them can be found as we expect a linear correlation to be existing.

The first three steps of this procedure can be represented as we do on the following figure:

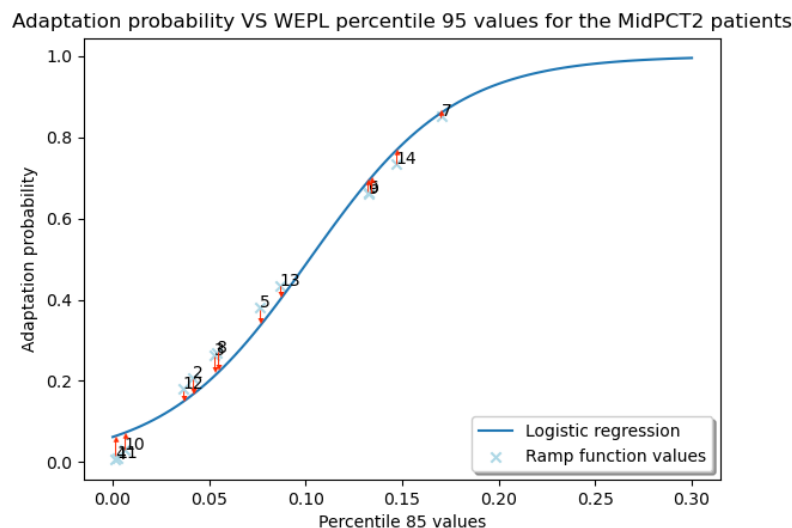


FIGURE 4.49: This graph represents the three first steps that were described above. First, the ramp function depicting the different adaptation probabilities was plotted. On this ramp function was then computed a logistic regression. Using this logistic regression, we were able to adjust the adaptation probabilities for each percentile 85 value.

We choose the WEPL percentile 85 in this example but it also applies to the other metrics. Here is the adaptation probability as a function of the ΔHI values for the MidpCT2 patients:

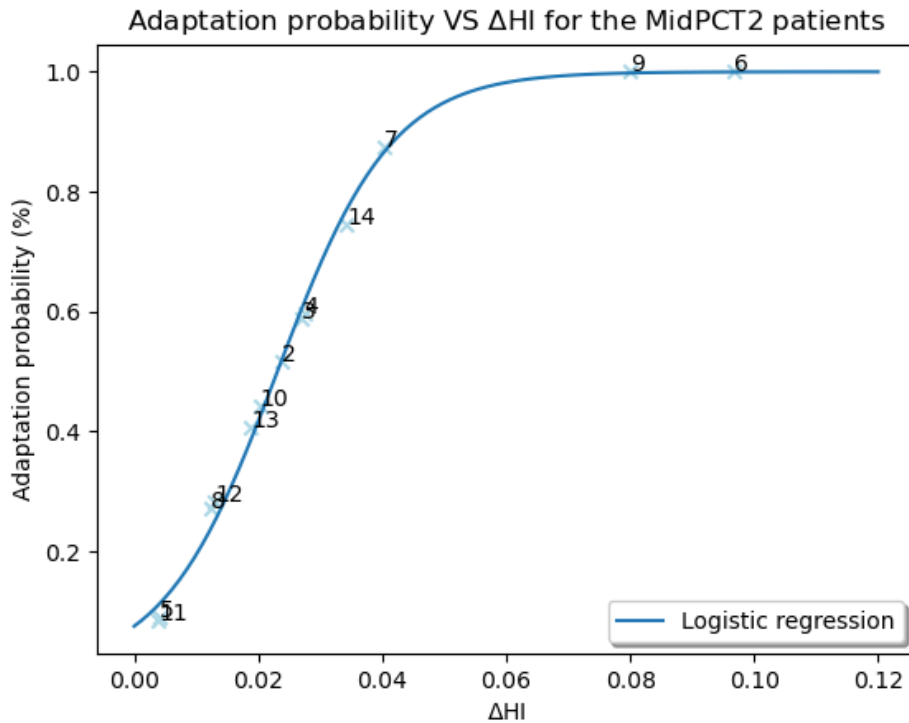


FIGURE 4.50: Logistic regression on the adaptation probability values as a function of ΔHI obtained using the method mentioned at 4.2.1 considering a 50% adaptation probability corresponded to a ΔHI value of 0.023.

Fig 4.50 represent the ramp function for the ΔHI values on which has been plotted a logistic regression. The ΔHI was systematically used because, as it was mentioned in the chapter 3 and earlier in this chapter, it relies on the description of the dose distribution homogeneity. Therefore it is used as a reference to compare our newly found probabilities.

After computing it for the ΔHI , we did it for both ΔHU metrics we decided were satisfying enough, namely the MSE and the PSNR.

Here are displayed the correlations between our metric adaptation probabilities and the ΔHI adaptation probabilities:

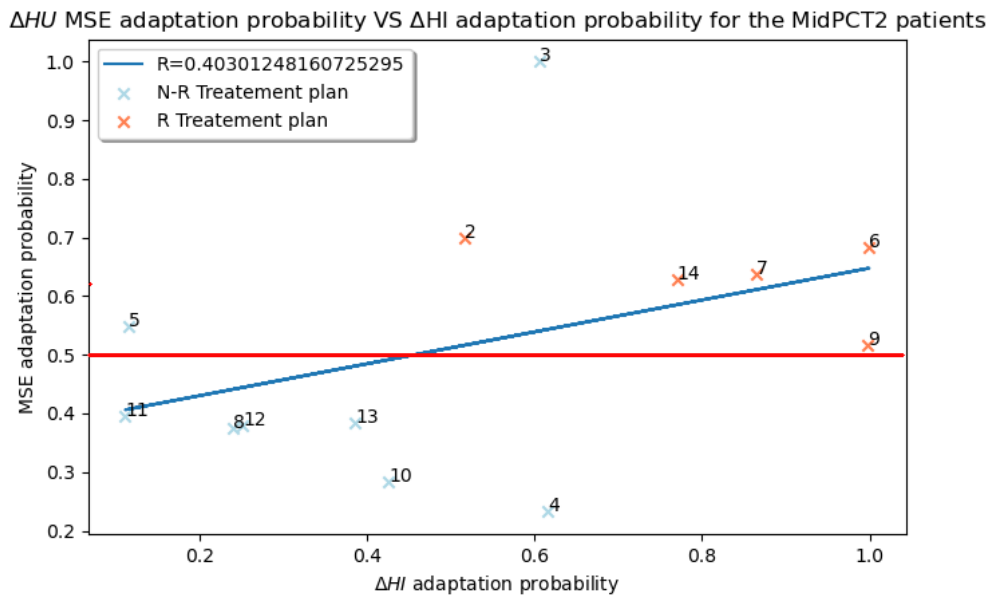


FIGURE 4.51: Plot of the ΔHU MSE adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

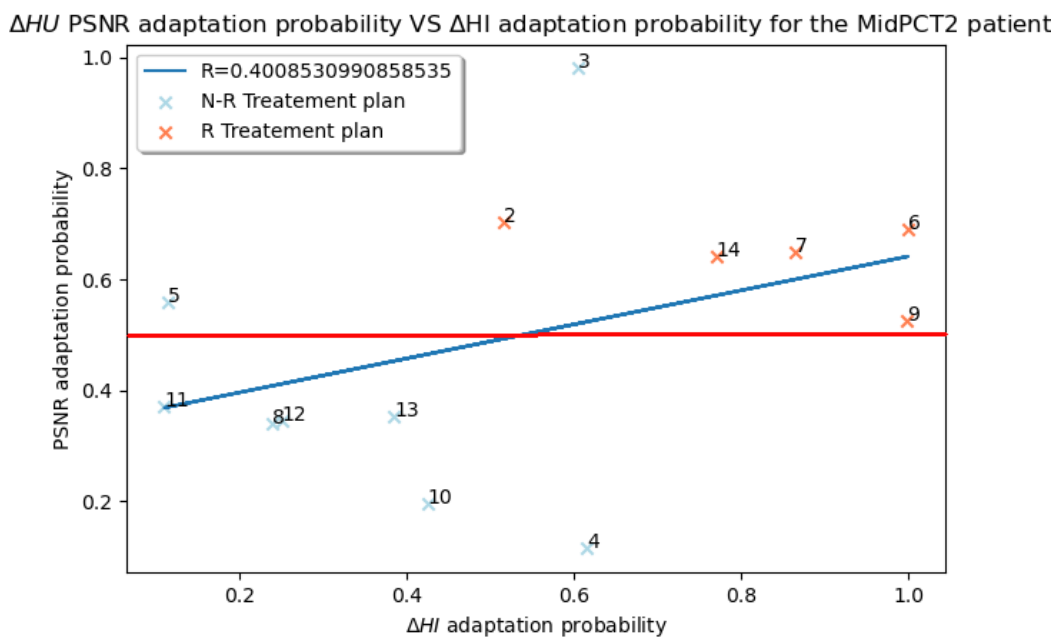


FIGURE 4.52: Plot of the ΔHU PSNR adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

Both ΔHU metrics seem to be performing well. The only thing we could hold against both of them is that we would have prefer patient 9 to be better discriminated and not having an

adaptation probability so close to 50%.

To better characterise those metrics, we decided to compute their sensitivity and specificity using confusion matrices:

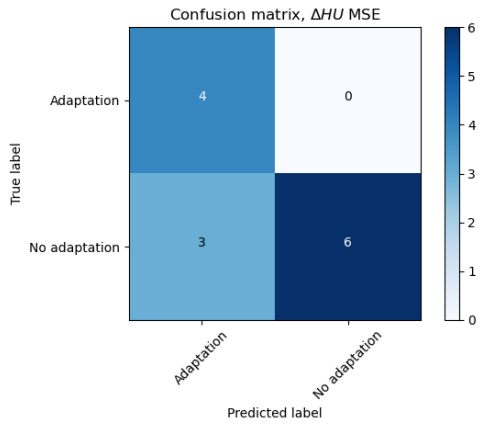


FIGURE 4.53: Confusion matrix for the MidpCT2 ΔHU MSE.

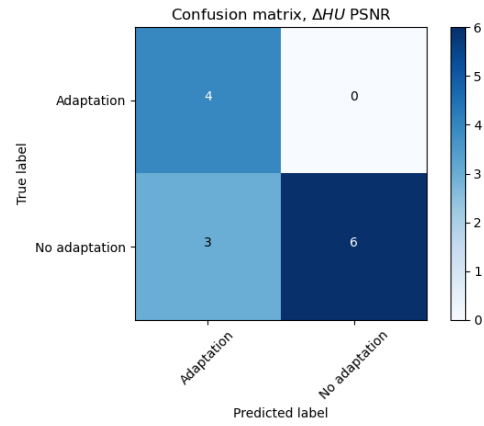


FIGURE 4.54: Confusion matrix for the MidpCT2 ΔHU PSNR.

We remind here that the sensitivity is the ratio

$$\frac{TP}{TP + FN} \quad (4.2)$$

and the specificity is the ratio

$$\frac{TN}{TN + FP} \quad (4.3)$$

where

TP are the true positives

TN are the true negatives

FN are the false negatives

FP are the false positives.

Both metrics seem to equally well perform with a sensitivity equal to 1 and a specificity of $\frac{2}{3}$.

When the ΔHI metrics were done, we computed the WEPL metrics adaptation probabilities and also plotted them as a function of the ΔHI adaptation probabilities:

WEPL percentile 85 adaptation probability VS ΔHI adaptation probability for the MidPCT2 patients

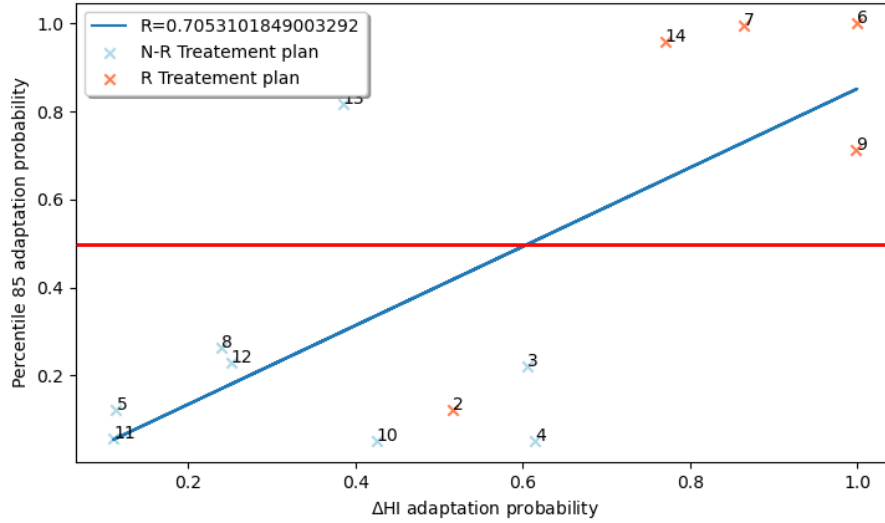


FIGURE 4.55: Plot of the percentile 85 adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

WEPL percentile 90 adaptation probability VS ΔHI adaptation probability for the MidPCT2 patients

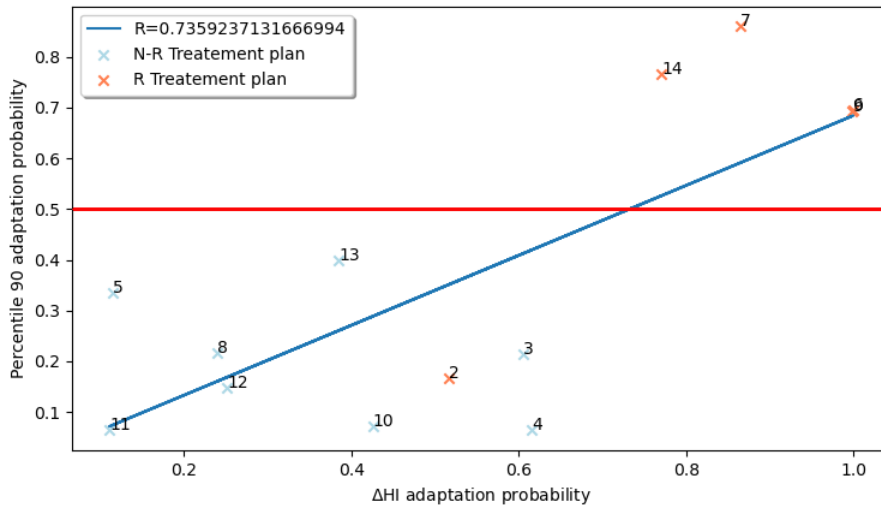


FIGURE 4.56: Plot of the percentile 90 adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

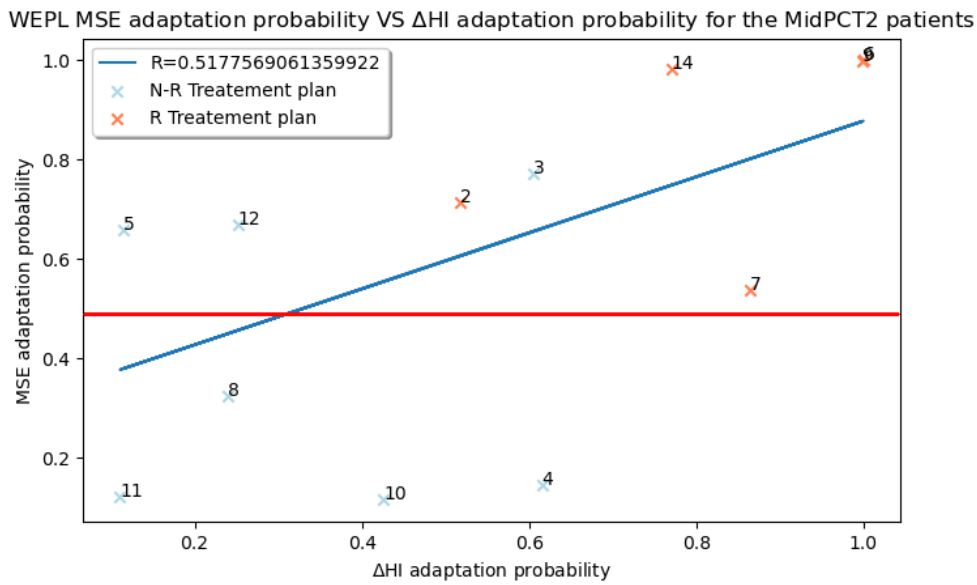


FIGURE 4.57: Plot of the MSE adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

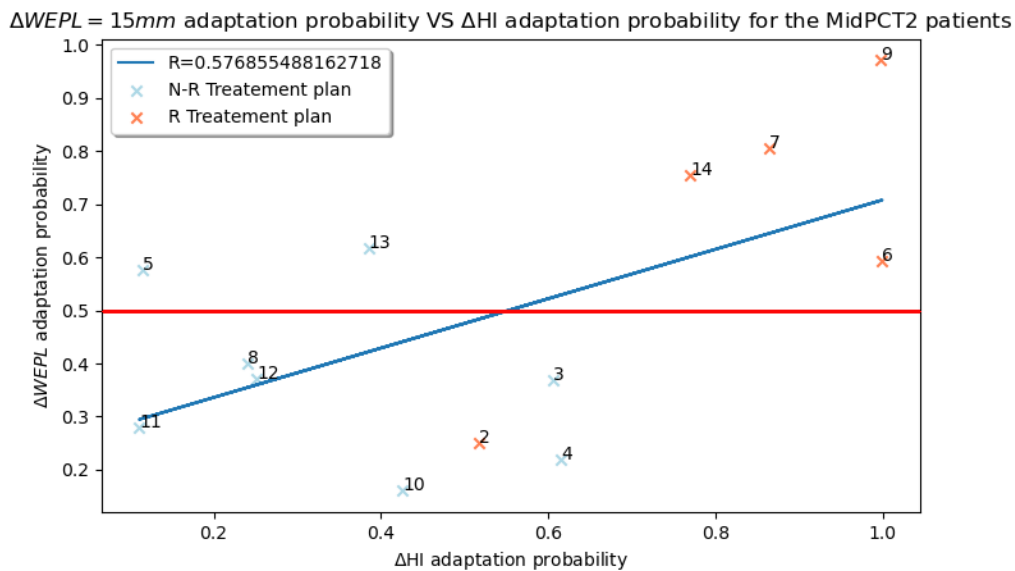


FIGURE 4.58: Plot of the $\Delta WEPL = 15mm$ adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

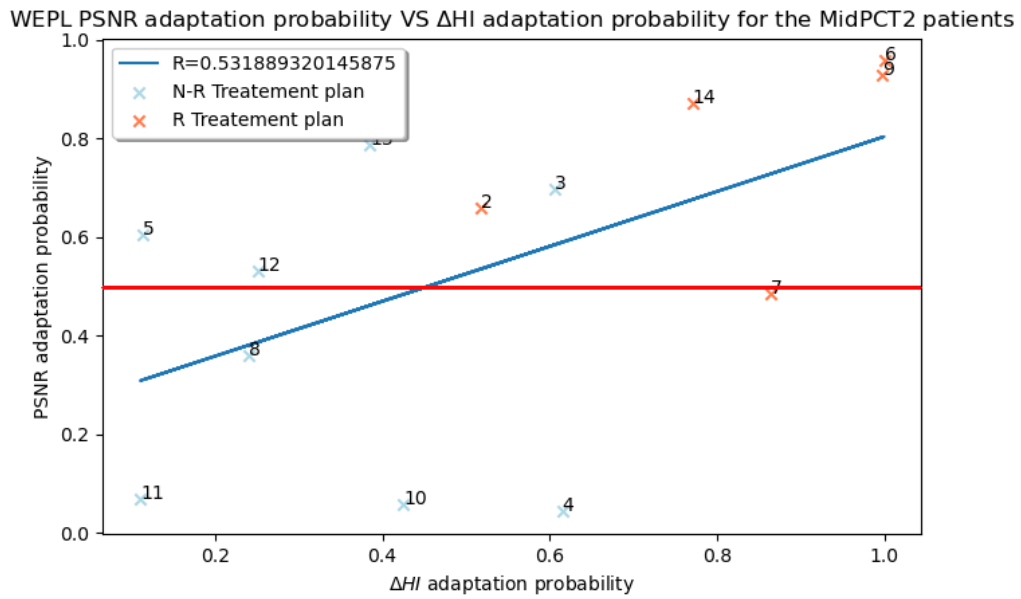


FIGURE 4.59: Plot of the PSNR adaptation probabilities as a function of the ΔHI adaptation probabilities. A linear regression has also been plotted to study the correlation between both set of data.

All 5 metrics seem to perform quite satisfyingly. However some of them seem better than the others.

To be able to better compare them we also computed each confusion matrix:

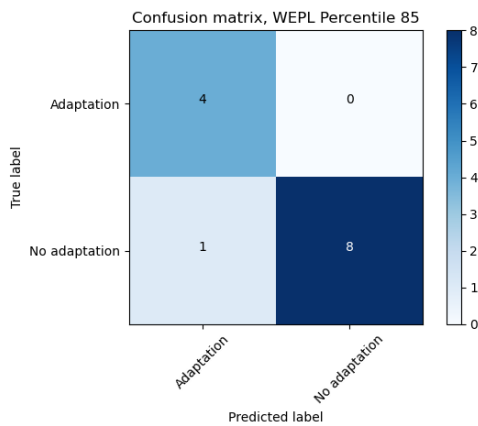


FIGURE 4.60: Confusion matrix for the MidpCT2 WEPL Percentile 85.

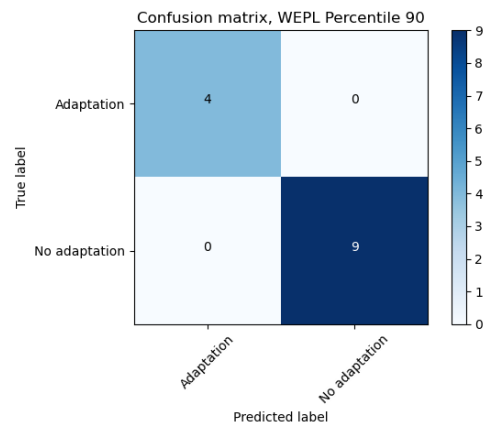


FIGURE 4.61: Confusion matrix for the MidpCT2 WEPL Percentile 90.

Both Percentile metrics perform very well and especially the Percentile 85 which have a sensitivity and a specificity equal to 1.

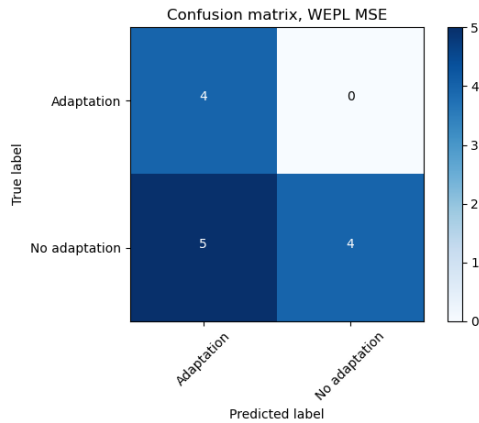


FIGURE 4.62: Confusion matrix for the MidpCT2 WEPL MSE.

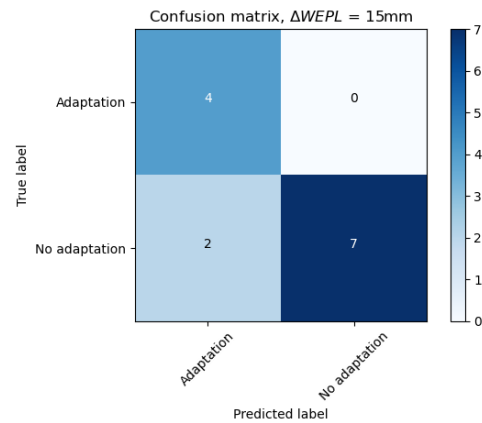


FIGURE 4.63: Confusion matrix for the MidpCT2 $\Delta WEPL = 15\text{mm}$.

The WEPL MSE confusion matrix tells us this metric has a sensitivity of **1** and a specificity of $\frac{4}{9}$. The $\Delta WEPL = 15\text{mm}$ on the other hand perform better with a sensitivity still equal to **1** and a specificity of $\frac{7}{9}$.

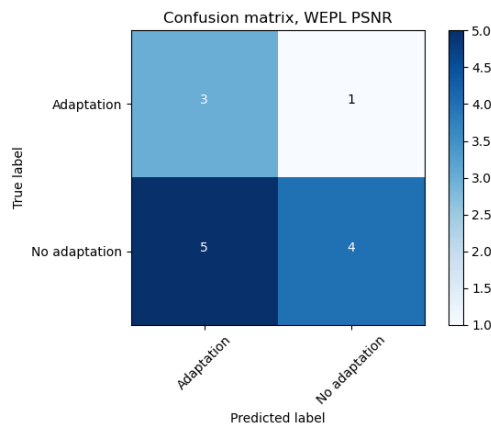


FIGURE 4.64: Confusion matrix for the MidpCT2 WEPL PSNR.

The WEPL PSNR is the only metric with a sensitivity $\neq 1$. In fact, it includes a false negative. Its sensitivity is then equal to $\frac{3}{4}$. We will exclude it from the potential useful metrics.

Chapter 5

Conclusion

The late technology improvements and the recent cost reduction increased the interest in the proton therapy development. As part of this study, we focused on the adaptive proton therapy, a modality that allows us to adapt a patient treatment plan if any detrimental modifications occurred in the patient anatomy between the day of the planning and the daily dose delivery.

In an effort to enable online adaptive proton therapy, I was given the task to study a multi-stage flagging system that would allow an online adaptive proton therapy triggering. The first stage of this increasing complexity triggering system uses ΔHU based indicators. The second stage uses WEPL maps and the third and last stage uses a dose distribution based indicator which was our reference throughout the research. This flagging system needs to be fast and highly reliable. To comply with the reliability, we used dose distribution computed with a dedicated cutting-edge software as a reference to compare our results. The ΔHU and WEPL based indicators on the other hand are meant to allow an important computation time reduction while delivering the same results as the dose distribution based indicator.

When analysing every metric on each of the three stages we already observe encouraging results. In fact most of them already allowed us to make a more or less good distinction between the patients to adapt and the ones not to adapt. Only a few metrics had to be ruled out straight away for their ineffectiveness. But metric values were not enough to build a multistage flagging system and we needed a way to translate those metric values in adaptation probabilities.

In the course of this study, we first consider the use of metric combinations as possible indicators. We did so because, at first sight, we did not think we would be able to extract any useful information from single metric indicators. In fact, we first set our adaptation flagging threshold at 65% of probability and not 50%. This caused all the single metrics, whether it was for the ΔHU or the WEPL, to seem irrelevant when analysed with the help of confusion matrices. Therefore we sought for combinations of our metrics that would comply with our strict criteria and we found three of them. All three combinations uses WEPL MSE either with WEPL percentile 85, WEPL percentile 90 or $\Delta WEPL=15\text{mm}$. But as this approach ruled out all the ΔHU metrics, it did not allow the setting of a multistage flagging system and we decided to focus on a different approach.

An other approach, where we lowered our adaptation threshold to 50% and used an other translation method to get our adaptation probability values, allowed us to suggest a way to establish a multistage flagging system using single metric indicators. The first stage of this system would consist in a computation of the MSE or the PSNR as both metrics are providing almost the same results. This first stage already allows us to rule out some patients from the possibility of an adaptation. The second stage would use the calculation of the WEPL

percentile 90. This WEPL metric performed particularly well and allows a separation between adaptation and non-adaptation with large margins. At this level, if a patient has been flagged as "to be adapted" in both stages, we can state with a high level of confidence that the patient really needs a treatment plan adaptation. In a similar way, if a patient has not been flagged in any of the stages, we can claim with trust that they do not need the adaptation. If a patient happens to be flagged in one of the stages but not in both, we would then still trigger the online adaptation to ensure a secure daily dose delivery. The third and last stage, which is the dose distribution based indicator also provide excellent results. But as it relies on dose distribution computation, which take too much time to be worked out, we only used it as a reference to evaluate the confidence we could put in the indicators from the first two stages.

Despite the promising results brought by this last approach, using single metric indicators, we must not forget that they are statistically weak due to the small number of patients available. In fact we only had 13 patients to work with and we would have needed much more patients than that if we wanted to be certain our results are truly reliable. Another important thing I want to bring the attention to is the choice of our metric thresholds. In fact, we based all of our probability computations on these thresholds as we suggested they represented a 50% adaptation probability. These thresholds are "user dependant" and it would be interesting to find a more accurate and consistent way to define our thresholds. For instance, we could use machine learning with support-vector machines to try finding the best thresholds in a more dependable way [20].

However, and despite what is said above, this work already gives a good insight of where we should investigate with more care when we will have the data needed for a real statistical study. We showed the feasibility of a multistage flagging system and proved its pertinence.

Bibliography

- [1] Edmond Sterpin. Lecture notes in dosimétrie en radiothérapie et contrôle de qualité, 2018-2019.
- [2] Solange Gagnebin. *Experimental determination of absorbed dose to water in a scanned proton beam using a water calorimeter and an ionization chamber*. PhD thesis, 05 2010.
- [3] Ugo Amaldi, Manjit Dosanjh, Jacques Balosso, Jens Overgaard, and Brita Sørensen. A facility for tumour therapy and biomedical research in south-eastern europe. 09 2019.
- [4] Camilla Stokkevåg, Grete May Engeseth, Kristian Ytre-Hauge, Dieter Röhrich, Odd Odland, Ludvig Muren, Marianne Brydøy, Liv Hysing, Artur Szostak, Matthew Palmer, and Jorgen Petersen. Estimated risk of radiation-induced cancer following paediatric cranio-spinal irradiation with electron, photon and proton therapy. *Acta oncologica (Stockholm, Sweden)*, 53:1–10, 07 2014.
- [5] KR Kutanzi, A Lumen, I Koturbash, and IR Miousse. Pediatric Exposures to Ionizing Radiation: Carcinogenic Considerations. *International journal of environmental research and public health*, 13, oct 2016.
- [6] À propos du centre de protonthérapie. <https://www.uzleuven.be/fr/centre-de-protontherapie/propos-du-centre-de-protontherapie>, 2020. [Online; accessed 29-May-2020].
- [7] Amjad Hussain and Wazir Muhammad. *Treatment Planning in Radiation Therapy*, pages 63–129. 11 2017.
- [8] Kinga Bernatowicz, Xavier Geets, Ana Barragan, Guillaume Janssens, Kevin Souris, and Edmond Sterpin. Feasibility of online IMPT adaptation using fast, automatic and robust dose restoration. *Physics in Medicine & Biology*, 63, Apr 2018.
- [9] Jan Unkelbach, Markus Alber, Mark Bangert, Rasmus Bokrantz, Timothy C Y Chan, Joseph O Deasy, Albin Fredriksson, Bram L Gorissen, Marcel van Herk, Wei Liu, Houra Mahmoudzadeh, Omid Nohadani, Jeffrey V Siebers, Marnix Witte, and Huijun Xu. Robust radiotherapy planning. *Physics in Medicine & Biology*, 63, nov 2018.
- [10] E. Sterpin, A. Barragan, K. Souris, and J.A. Lee. Planification de traitement robuste en protonthérapie. *Cancer/Radiothérapie*, 20(6):523 – 529, 2016. 27e Congrès national de la Société française de radiothérapie oncologique.
- [11] C. Champion M. Karamitros M.A. Bernal Z. Francis V. Ivantchenko S.B. Lee J.I. Shin S. Incerti H.N. Tran, Z. El Bitar. Modeling proton and alpha elastic scattering in liquid water in geant4-dna. 01 2015.
- [12] Stopping power and range tables for protons. <https://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html>, 2020. [Online; accessed 29-May-2020].
- [13] J.C Rosenwald P Mayles, A Nahum. *Handbook of Radiotherapy Physics: Theory and Practice*. CRC Press, 2007.

- [14] Wayne D Newhauser and Rui Zhang. The physics of proton therapy. *Physics in Medicine and Biology*, 60(8):R155–R209, mar 2015.
- [15] Radiology key. <https://radiologykey.com/computed-tomography-4/>, 2020. [Online; accessed 29-May-2020].
- [16] Rui Zhang and Wayne D Newhauser. Calculation of water equivalent thickness of materials of arbitrary density, elemental composition and thickness in proton beam irradiation. *Physics in Medicine & Biology*, 54(6):1383–1395, 02 2009.
- [17] Julie Vuadens. Optimisation d’indicateurs de déclenchement de procédure de proton-thérapie adaptative en ligne. Master’s thesis, Uclouvain, June 2019.
- [18] M M Mukaka. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi*, 24(3):69 – 71, 2012.
- [19] Beatrice A Golomb Steven C Bagley, Halbert White. Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*, 54(10):979 – 985, 2001.
- [20] Corentin Guyot. *Plasmonic nanocomposites embedding gold and silver nanoparticles: in situ synthesis and local optical properties by spectroscopic imaging ellipsometry*. PhD thesis, University of Mons, Mons, Belgium, 2020.