NTNU
Norwegian University of
Science and Technology

UNIVERSITETSSYKEHUSET NORD-NORGE
DAVVI-NORGGA UNIVERSITEHTABUOHCCEVIESSU

Radiotherapeutic Limits of Acceptance for Errors in Synthetic CT Synthesised from MRI

Master's Thesis

Leif-André B. Jensen

Department of Physics
Norwegian University of Science and Technology
Norway

Submitted
June 10th, 2020

Supervisors
Dr. Kathrine Røe Redalen (NTNU),
Dr. Ola Engelsen (UNN),
and Veronika K. Tømmerås (UNN)

Abstract

CT scans, resulting in electron density maps or mass density maps, are standard input for radiotherapy planning systems. However, MRI provides superior soft tissue contrast. In cases where MRI is required for accurate lesion delineation, both MRI and CT are performed, creating systematic registration complications in the planning images and a more time-consuming workflow. Alleviating these issues, MRI-only techniques are developed where synthetic CT (sCT) images are synthesised from MRI images. However, MRI struggles to separate bone from air, creating significant errors in sCT images. Using dose-volume histogram (DVH) point analysis and a 2 % deviation acceptance criterion, this thesis examines the resulting planning target volume (PTV) dose deviation in radiotherapy planning as a consequence of bone misclassification, relevant in sCT images for MRI-only radiotherapy planning. This is for the purpose of finding the limit of acceptable magnitude of error, described by the bone-area A, dependent on the distance, d, between the centres of the misclassification volume and the PTV. This study finds the limits of acceptance for these errors for three different PTV sizes, S, using several pairs of otherwise equal CT images where a bone volume is replaced by air. Optimised radiotherapy planning is performed on the image containing air and copied to the bone case. The resultant decision formulas provide useful tools for evaluating the radiotherapeutic consequence of misclassification, assuming a volumetric modulated arch therapy (VMAT) treatment plan and a misclassified bone of CT value equal to 800 HU. No statistical significance of the variable S is found.

Sammendrag

CT-skanninger, som resulterer i elektrontetthetskart eller massetetthetskart, er standard inndata for strålebehandlingplanleggingssystemer. Imidlertid gir MR overlegen bløtvevskontrast. I tilfeller der MR er nødvendig for nøyaktig avgrensning av lesjoner, utføres både MR og CT, noe som skaper systematiske registreringskomplikasjoner i planleggingsbildene og en mer tidkrevende arbeidsflyt. For å lindre disse problemene, utvikles bare-MR-teknikker der syntetiske CT-bilder (sCT) blir syntetisert fra MR-bilder. MR sliter imidlertid med å skille bein fra luft, noe som skaper betydelige feil i sCT-bilder. Ved bruk av punktanalyse i dose-volum histogram (DVH) og et 2 %-avvik akseptkriterium, undersøker denne avhandlingen det resulterende planleggingsvolumets (PTV) doseavvik i strålebehandlingplanlegging som en konsekvens av beinfeilklassifisering, relevant i sCT-bilder for bare-MR stråleplanlegging. Dette gjøres for å finne grensen for akseptabel feilstørrelse, beskrevet av beinarealet A, avhengig av avstanden, d, mellom sentrene i feilklassifiseringsvolumet og PTV. Denne studien finner akseptgrensene for disse feilene for tre forskjellige PTV-størrelser, S, ved bruk av flere par av ellers like CT-bilder der et beinvolum er erstattet av luft. Optimalisert stråleplanlegging utføres på bildet som inneholder luft og kopieres til beintilfellet. De resulterende beslutningsformlene er nyttige verktøy for å evaluere den radioterapeutiske konsekvensen av feilklassifisering, forutsatt volumetrisk modulert bueterapi (VMAT) behandlingsplan og et feilklassifisert bein med CT-verdi lik 800 HU. Ingen statistisk signifikans av variabelen S er funnet.

Acknowledgements

This master's thesis was written as a conclusion to the master's degree programme in Applied Physics and Mathematics at the *Norwegian University of Science and Technology* (NTNU) in Trondheim, Norway. This work was performed during the spring of 2020 in cooperation with the *University Hospital of North Norway* (UNN) in Tromsø, Norway.

I would like to give a big thanks to all three of my supervisors; all of whom have made this past year very enjoyable and educational. I would like to thank Kathrine for making academic life interesting again and pointing me in the direction of medical physics. Without Veronika's help I would not have been able to use the complex radiotherapy treatment planning tools necessary to perform these experiments. I am very grateful to Ola for giving me the opportunity to write both my project thesis and master's thesis at UNN, in an environment of great people, helping me in taking the first steps of my career in medical technology. Finally, a special thanks goes out to my friend, André Pedersen, for knowing by heart which statistical test to perform in order to compare two regression lines.

Leif-André B. Jensen
Djupvik, June 2020

Contents

Abbreviations

AAA:	Analytical anisotropic algorithm.
AC:	Alternating current.
CT:	Computed tomography.
DNA:	Deoxyribonucleic acid.
DVH(s):	Dose volume histogram(s).
DTA:	Distance to agreement.
FID:	Free induction decay.
FLASH:	Fast low angle shot.
HU:	Hounsfield unit.
LINAC:	Linear accelerator.
MLC:	Multileaf collimator.
MR:	Magnetic resonance.
MRA:	Magnetic resonance angiography.
MRI:	Magnetic resonance imaging.
MU:	Monitor unit.
OAR(s):	Organ(s) at risk.

PET:	Positron emission tomography.
PETRA:	Pointwise encoding time reduction with radial acquisition.
PR:	Pass rate.
PTV(s):	Planning target volume(s).
QA:	Quality assurance.
RF:	Radiofrequency.
sCT:	Synthetic CT.
SPECT:	Single-photon emission computed tomography.
TE:	Echo time.
TPS:	Treatment planning system.
TR:	Repetition time.
UTE:	Ultrashort echo time.
VMAT:	Volumetric modulated arc therapy.
ZTE:	Zero echo time.

1 Introduction

External beam radiotherapy, which involves exposing diseased human tissue to substantial amounts of deadly, ionising radiation, is a popular form of cancer treatment [1]. This is a carefully planned process, as one strives towards giving the *planning target volume* (PTV) the required dose of radiation while simultaneously sparing the healthy tissue and *organs at risk* (OARs) [2]. Radiotherapy planning and delivery necessitates one or more sets of medical images.

To accurately calculate the dose of radiation deposited in each location in the body, using an *analytical anisotropic algorithm* (AAA), an electron density map is required. This is best obtained through a *computed tomography* (CT) scan [3]. However, CT imaging lacks the soft tissue contrast of nuclear *magnetic resonance* (MR) imaging (MRI) [4]. The superior soft tissue contrast is often necessary in order to accurately delineate a lesion that is surrounded by healthy soft tissue. The use of MRI in radiotherapy treatment planning increased from 6 % of cases to 24 % between 2006 and 2017 in the USA [5]. This requirement of two imaging modalities and two separate scans poses certain issues; like achieving accurate image registration when both internal and external movement might occur between scans, patient scheduling, and economics. In addition, a dose of radiation is given to the patient during the CT scan.

In an attempt to alleviate these complications, *synthetic CT* (sCT) imaging techniques have been developed, where MR images are used to create approximate CT images [6]. Some of these techniques have been found to create sCT images that are clinically acceptable for radiotherapy planning [6][7][8][9]. However, there is one important obstacle to overcome in such techniques; because of the low MR signal given from both bone tissue and air, the two are often difficult to distinguish in MRI [10]. Bone tissue and air have vastly different electron densities [11], meaning misclassification of the two in sCT images is highly undesirable. Solutions to such problems have been proposed and implemented. One is to elastically fit a bone-atlas created from previous patients' anatomy [6]. However, this method struggles when it is exposed to atypical patient anatomy, for example children. Another solution is to utilise *ultrashort echo time* (UTE) MRI sequences to obtain bone tissue MR signal [10]. This method struggles with atypical $T2$ properties in the bone tissue and is still prone to misclassification. Once any bone misclassification occurs in a patient sCT image, the radiographer must be able to evaluate whether there is a need to repeat the scan, possibly with different settings, or use conventional CT imaging. This thesis aims to explore the consequences of bone-to-air misclassification and find under which conditions the misclassification errors are acceptable in *volumetric modulated arc therapy* (VMAT) treatment planning.

Similar work has been performed by this author previously in a project thesis conducted at the *Norwegian University of Science and Technology* (NTNU) and *University Hospital of North Norway* (UNN) in the autumn of 2019 [12]. Using gamma evaluation [13][14], the project thesis found a maximum acceptable size of misclassified bone for VMAT treatment planning, though only for a spherical PTV of constant diameter and at a constant distance from the misclassification volume. However, it was found that the PTV dose deviation between the air and bone case was substantial at that misclassification magnitude. This calls into question the gamma evaluation method's ability to fail cases of too great PTV dose deviation, motivating this thesis to apply a different acceptance criterion using *dose-volume histograms* (DVHs) [15][16]. Additionally, a linear correlation

between the PTV dose deviation and the size of the misclassification volume was observed, motivating the use of linear regression models to find the limits of acceptance in this study. With the exception of the project thesis, and to the best of this author's knowledge, no other study of the radiotherapeutic consequences of bone misclassification has been conducted.

The primary aim of this thesis is to use the aforementioned DVH acceptance criterion to derive equations describing the limit of acceptance for bone misclassification, given by the maximum two-dimensional size of bone, area A, and the minimal distance, d, between the centres of the PTV and misclassification volume. Additionally, the dependency of the PTV size, S, is also tested. Such equations might provide a useful tool for radiographers, radiologists, and/or medical physicists in evaluating whether occurrences of bone misclassification in sCT images are acceptable or if new scans must be made. Gamma evaluation is performed in order to assess its ability to correctly fail grave PTV dose deviations while simultaneously providing a method for evaluating the dose cohesion in the entire irradiation volume.

Because of a difference in techniques for generation of sCT images of the pelvis and head [7], where more effort is put into avoiding bone misclassification in the latter, the radiotherapeutic consequences of an entire skull misclassification are explored.

Finally, the results of these experiments are tested by digitally inducing bone misclassification errors in real sCT images generated by *Siemens'* sCT product [7].

2 Theory

As this master's thesis is inspired by the work done in a similar project thesis by this author, most of the content in this section is adapted from the corresponding section in the project thesis, with minor changes [12]. However, some major changes are: a new section on the topic of cancer (Section 2.1), two new subsections in Section 2.2 explaining some relevant MRI sequences (Section 2.2.3 and Section 2.2.4), an addition to Section 2.3 in order to go into detail on DVHs (Section 2.3.3), a few CT theory additions to Section 2.4 for the purpose of discussing results, an extension of Section 2.5 to include sCT generation of the head and neck area, and a section describing the theory behind the statistical analysis utilised in this thesis (Section 2.8).

2.1 Cancer

A short description of cancer biology is given in this section. The source of the following is the textbook "Radiobiology for the Radiologist" [1].

Cancer is an abnormal growth of cells caused by changes in the genetic material of cells. The mutations result in uncontrolled cell division (proliferation) and failure of programmed cell death. This creates a Darwinian-like biological system where the cancer cells become dominant and create a tumour. Cells might also spread further, for example through blood streams (metastasis). Progression from healthy tissue to tumour tissue is stepwise, with steps occurring randomly through errors in the DNA, or resulting from external factors such as radiation, chemical mutagens, or viral infection. Tumours originate from cells that over time have undergone multiple genetic changes, leading to deactivation of tumour suppressing genes and/or activation of oncogenes. Oncogenes are *gain-of-function* mutations, meaning only one copy of the gene needs to be activated. The tumour suppressor genes are *loss-of-function* mutations, meaning both copies must be lost for the suppressor gene to be deactivated.

In normal cells, the telomere caps at the ends of the chromosomes shorten with each cell division. Once a cell has undergone 40 to 60 cycles, the telomeres are too short to continue dividing and the cell line eventually dies. During cellular division, stem cells and cancer cells activate telomerase, which maintains the telomere length and immortalises the cell. To kill a malignant tumour, the cells must be neutralised by external forces, for example by carefully planned radiotherapy, which will be further explored in Section 2.3.

2.2 MRI

A detailed description of MRI is beyond the scope of this paper. The interested reader is referred to the plethora of literature on the subject. However, a short description covering the concepts relevant to this thesis will be given here. Unless otherwise stated, the bulk of the following information is found in the textbook "MRI in Practice" [4].

2.2.1 Basic MRI

Nuclear magnetic resonance imaging is different from all other medical imaging techniques. MRI is a non-invasive medical imaging technique exploiting protons' quantum mechanical properties in strong magnetic fields. Unlike X-ray imaging, CT, single-photon emission computed tomography (SPECT), and

positron emission tomography (PET), MRI does not give the patient a dose of harmful, ionising radiation. Yet its most important attribute is its superior soft tissue contrast, which is often required to accurately delineate certain tumours. Since its discovery, MRI has developed and improved profoundly; and today it has a wide range of diagnostic and therapeutic applications. Understanding the source of contrast in MRI is important to understand its applications and limitations.

In most cases, protons are the source of an MR signal. MRI utilises the characteristic magnetic dipole moment, $\vec{\mu}$, exhibited by atomic nuclei with an uneven number of nucleons. ^{1}H, consisting of just one proton, is such a nucleus, abundant in water (H_2O), but also in fat and other organic molecules. As human beings consist mostly of water, this is the nucleus most commonly explored in MRI. When exposed to an external magnetic flux density, $\vec{B_0}$, the protons' dipole moments start to precess about $\vec{B_0}$ with a characteristic frequency, ω_0, known as the *Larmor precession frequency*. This is proportional to the magnetic field in the following way:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

where γ is the gyromagnetic ratio. Because of differing electron shielding in differing molecules, the bound protons will experience a slightly different local magnetic field in fat molecules compared to those in water molecules. This results in protons bound to the different molecules obtaining different ω_0-values. Using select sequences, this enables differentiation of the signal originating from fat from the signal originating from water. This is explored further in Section 2.2.2. The protons' magnetic dipole moment, causing the precession about the external magnetic field vector, is due to their *spin*.

Spin is a quantum mechanical property of all nuclei and is an important part of understanding the origin of an MR signal. The protons have spin $\frac{1}{2}$, meaning they can occupy one of two possible energy states: *up*, being aligned parallel to the external magnetic field (low-energy state), and *down*, being aligned anti-parallel to the magnetic field (high-energy state). The ratio of population of the two energy states is given by:

$$\frac{N_{up}}{N_{down}} = e^{\frac{\gamma h B_0}{kT}} \approx 1 + \frac{\gamma h B_0}{kT} \tag{2.2}$$

where h is Planck's constant, k is Boltzmann's constant, and T is the temperature. As slightly more protons align *with* the external magnetic field, a net magnetisation caused by the protons, $\vec{M_0}$, is created. From eq. (2.2) it is apparent that a stronger external magnetic field, and lower temperatures, lead to a larger net magnetisation vector.

A useful signal can be obtained by manipulating $\vec{M_0}$. To be able to measure $\vec{M_0}$, it must be moved away from the direction of the external magnetic field, which is often referred to as the z-direction. This is done by applying a *radiofrequency* (RF) pulse to the precessing protons for a period of time, t. Only protons that precess within the frequencies of the pulse's frequencies are affected. The RF pulse, created by alternating currents in coils, creates an alternating magnetic flux density, $\vec{B_1}$, orthogonal to $\vec{B_0}$. This results in a flip of the magnetisation vector given by:

$$\alpha = \gamma B_1 t \tag{2.3}$$

where α is the flip angle. Now the protons precessing in phase about $\vec{B_0}$ have a net magnetisation component in the orthogonal (xy) plane, $M_\perp$, which is the signal that is recorded due to induced voltages in the receiver. This process is called *excitation*. Because of the torque of the external magnetic field, the magnetisation vector will immediately start to realign itself with B_0 through *relaxation*. Generally, there are two relaxation processes: spin-lattice relaxation and spin-spin relaxation. Spin-lattice relaxation is unrecoverable and is due to the protons' interaction with their surroundings; the result of which is a jump to a lower energy state and realignment with $\vec{B_0}$. Therefore, this is also called *longitudinal relaxation*. $T1$ is defined as the time it takes for the z-component of the net magnetisation to regain $1 - \frac{1}{e} \approx 63\,\%$ of its size. The spin-spin interactions and other interactions, including spin-lattice interactions, can cause the protons to dephase, leading to a loss of signal in the xy-plane. A better name for this is *transverse relaxation*. $T2$ is defined as the time it takes for the xy-component of the net magnetisation to fall to $\frac{1}{e} \approx 37\,\%$ of its original value following the RF pulse. $T1$ and $T2$, measurable because of $\vec{M_0}$, are the tissue properties that create the contrast in MRI. Different tissues exhibit different $T1$ and/or $T2$.

Images in MRI can be *weighted* towards $T1$, $T2$, or both (proton density weighting). This is done by adjusting two operational parameters: echo time (TE) and repetition time (TR), which describe the time from excitation of the slice to the maxima of the following echo and the time between excitations, respectively. These parameters are easiest to understand by discussing the most common MRI sequence: the spin echo (SE) sequence. This involves a 90° excitation pulse, flipping the magnetisation vector completely into the xy-plane, repeated every TR. At time $TE/2$ following the 90° pulse, a 180° pulse is applied. This rephases the dephased protons that are dephased due to inhomogeneities in the magnetic field, creating a maximum signal, or an echo, at time TE, which is recorded. There are many online animations that illustrate this effect nicely [17]. A mathematical description of an MR signal was given by Bloch in 1946. He proposed a set of equations used to calculate the magnetisation, $\vec{M_0}$ [18]. The effect of TE and TR on the MRI signal is clearly shown in the Bloch equations solved for the spin echo sequence:

$$M_\perp(TR, TE) = M_0(1 - e^{-TR/T1})e^{-TE/T2} \tag{2.4}$$

From eq. (2.4) it is apparent that at a comparatively long TR ($TR > T1$), the $T1$ effects largely disappear as $(1 - e^{-TR/T1}) \to 1$. For $T2$, the same is true for short TE ($TE << T2$) as $e^{-TE/T2} \to 1$. Therefore, to get a $T1$-weighted image, short TR and short TE are required, while the opposite is true for $T2$-weighting. A short TE and long TR give proton density-weighting, whereas the opposite case gives too little signal to be useful. TE is considered *short* at about 10 ms and *long* at about 100 ms, while TR is considered *short* at about 500 ms and *long* at about 4000 ms; though what is considered long and short also depends on the strength of the magnetic field and the tissue type.

In the sequences most relevant for this thesis, *gradient* echo imaging is used instead of spin echo imaging. In these sequences, the gradient coils, whose purpose is be explained later, create the echos that are recorded. No 180° degree pulse is applied to create the echo, meaning the dephasing due to field inhomogeneities is not reversed. The $T2$ is therefore not observed; instead the observed signal will be dependent on $T2^* = 1/T2 + 1/T2'$, where $T2'$ is attributable to the magnetic field inhomogeneities. Gradient echo sequences also often utilise flip

angles smaller than 90°, which create a steady state resulting in the following equation for the perpendicular magnetisation:

$$M_{\perp}(TR, TE, \alpha) = M_0 \frac{(1 - e^{-TR/T1})e^{-TE/T2^*}}{(1 - \cos(\alpha) \cdot e^{-TR/T1})} \tag{2.5}$$

It is apparent that the same weighting rules apply here as for eq. (2.4), though with $T2^*$ substituted for $T2$.

A problem in most MR images is that bone tissue in the patient gives very low MR signal intensity. This is because of bone's ultrashort $T2$-properties ($TE >> T2$) [10]. From eq. (2.4), it is then found that $e^{-TE/T2} \to 0$, resulting in approximately no signal arriving from tissue with such $T2$-properties. Bone can be divided into two different tissue types [19]. *Cancellous* bone, or *trabecular* bone, is the softer tissue that is found within the core of mature adult vertebral bones. Such tissue, containing some water, will yield more MR signal than *cortical* bone. Cortical bone is very dense bone tissue containing very little water, yielding close to no signal in normal MR imaging. Air, too, yields a very low MR signal as it contains very few hydrogen atoms, making it difficult to differentiate between bone and air by comparing signal intensity values in MRI, as shown in Figure 2.1. This problem in identifying and separating bone and air is a prominent issue for developers trying to assign a specific tissue-dependent value to a specific voxel, for example in the case of creating sCT images. This is the underlying problem motivating this thesis.

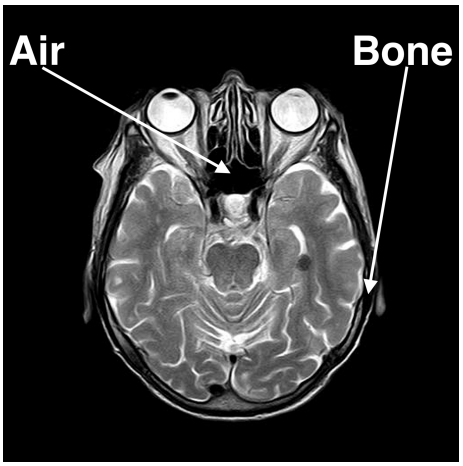


Figure 2.1: MR image [20] illustrating the lack of difference between air and bone. This image has been edited by this author.

Spatial encoding is required to create MR images. The signal's origin must be placed in a coordinate system. In MRI, this is done using gradient coils, which create a magnetic field which strength is a linear function of the position along that direction. Assuming a gradient in the z-direction, G_z, the Larmor precession frequency is affected as follows:

$$\omega = \gamma(B_0 + zG_z) \tag{2.6}$$

As mentioned previously, only protons precessing with frequencies that are contained in the RF pulse are excited by it. Slice selection is made by choosing frequencies that correspond to a certain range of z-values, meaning spatial encoding has been performed in the z-direction. The gradient in this direction is often

called the slice selection gradient, G_{SS}. Two other gradients are applied to spatially encode in the two other directions. These are the phase-encoding gradient, G_{PE}, often in the y-direction/phase-encoding direction, and the readout gradient, G_{RO}, often in the x-direction/frequency-encoding direction. The readout gradient is on while the echo is recorded, while the phase-encoding gradient often changes between each repetition to record only one line in the phase-direction at a time. Over a certain amount of time, the signal, also known as the free induction decay (FID), is recorded and fills one line of *k-space*; which is a spatial frequency domain. The final image is created by performing a two-dimensional Fourier transform of the k-space. In the frequency-encoding direction of k-space, the protons have different Larmor precession frequency dependent on their position. In the phase-encoding direction, the protons are encoded a phase at various spatial locations. However, a change in phase over distance is just another spatial frequency, making the phase-encoding direction more appropriately called the *indirect frequency direction*. Three-dimensional acquisition of k-space is possible (Section 2.5). This is done by using the gradient in the z-direction as a second indirect frequency direction. To summarise; spatial encoding is performed by storing the measured signal in a spatial frequency domain, the position in which is determined by gradient coils applying different magnetic field strengths in different locations.

2.2.2 Dixon method

The ability to quantify the amount of fat and water in each voxel is desirable in MRI for many applications. Fat-protons experience a different local magnetic field than water-protons, resulting in a slightly different Larmor frequency (Section 2.2.1). At 1.5 T, this difference is about 225 Hz [21]. The difference is proportional to the external magnetic field strength, meaning it doubles to about 450 Hz at 3 T. In 1984, wanting to separate the fat-signal, F, from the water-signal, W, W. Thomas Dixon proposed a method to obtain fat-only and water-only images [22]. At the time, implementation was not successful, as perfect B_0-homogeneity was required. But with modern magnet design and shimming techniques (using metals and/or coil currents to create a homogeneous field), working methods were finally implemented in the early 2000's. Dixon's proposed solution was to record the signal using two different echo times: one where the signal from water and fat are in-phase, and another where they are out-of-phase. As a 180° pulse will rewind these phase shifts, this method can only be utilised in gradient echo sequences. At 1.5 T, this method gives echo times of $TE = 4.4$ ms for the in-phase image, IP, and $TE = 2.2$ ms for the out-of-phase image, OP [23]. This is illustrated in Figure 2.2. By simple addition, subtraction, and averaging, water-only and fat-only images are found in the following way:

$$IP = W + F$$
$$OP = W - F$$

$$\frac{1}{2}(IP + OP) = \frac{1}{2}(W + F + W - F) = W$$
$$\frac{1}{2}(IP - OP) = \frac{1}{2}(W + F - W + F) = F$$

This creates four sets of images: in-phase, out-of-phase, water-only, and fat-only. The fat-only and water-only images yield the potential of fat/water-quantification, which is useful in many cases, including the generation of voxel-based sCT, further explored in Section 2.5.

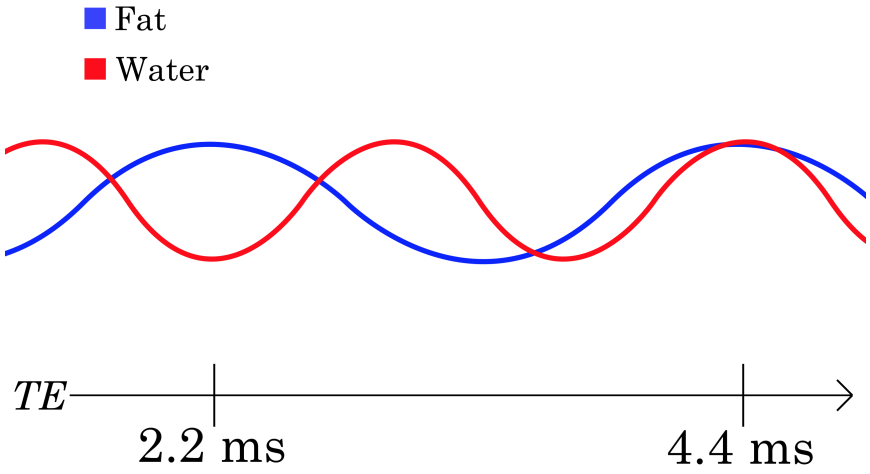


Figure 2.2: Signal of fat (blue) and water (red), at slightly different frequencies ($f_{water} - f_{fat} = 225\,\text{Hz}$), being in-phase at $1/225\,\text{Hz} = 4.4\,\text{ms}$ and out-of-phase at $1/(2 \cdot 225\,\text{Hz}) = 2.2\,\text{ms}$. Illustration created by this author.

2.2.3 Ultrashort/zero echo time imaging

Differentiating bone tissue from air is difficult using the most common types of MRI imaging (Section 2.2.1). The ultrashort relaxation times place high demands upon the MRI hardware. Specifically, the dead time between excitation and signal acquisition must be short enough to be able to detect the signal originating in such tissue. The dead time is necessary in order to ring down the coil with stored RF energy after excitation, before retuning the receive coil for signal acquisition. In an article from 2012, Grodzki et al. [24] wrote that most clinical scanners at the time had a dead time of 40 to 100 µs, but that with some hardware changes some clinical scanners had reported dead times of only 8 µs. Grodzki et al. go on to explain how these improvements made ultrashort echo time (UTE) sequences possible. A short summary follows in the remainder of this subsection.

When dead times have been made short enough to measure a signal coming from tissues such as cortical bone, a couple of popular methods of filling k-space have been developed. Cartesian single point acquisition is one of them. Here, only a single point in k-space is filled up at a time instead of a whole row. Naturally, this yields very long scan times when resolution is high. Another, faster, way of filling k-space, with reasonable resolution, is by filling it radially from the centre, as visualised in this online learning resource [25]. In this UTE sequence, gradients and acquisition start simultaneously once the dead time has passed. However, to avoid image distortions, a zero echo time (ZTE) sequence can be used to fill k-space radially by applying the gradients before the excitation and starting acquisition after the dead time has passed. This leads to a gap in the centre of k-space. The article goes on to propose a solution to filling the centre part of this k-space by using Cartesian single point acquisition. The specific sequence described is the *PETRA* (pointwise encoding time reduction with radial acquisition) sequence. Using such a sequence, even cortical bone

will yield a strong MR signal, which is useful in the generation of sCT images (Section 2.5).

These types of sequences have been used to approximate CT data previously. In combined PET/MR imaging, such sequences help approximate the location of bone (and soft tissue) in the patient. This data is called a *μ-map* and is important for correcting for the γ-photon attenuation within the patient in the PET part of the imaging process. Attenuation and the attenuation coefficient, μ, will be explored further in Section 2.3.1 and Section 2.4, respectively.

2.2.4 Vessel imaging

Because of the flowing blood in blood vessels, the protons in the blood may change slices in the time it takes for the signal to be acquired. This results in a loss of signal within the blood vessel, as illustrated in Figure 2.3. As previously discussed, a loss of signal in a voxel without air may cause problems when trying to differentiate between air, bone, and now blood. Many methods of imaging blood vessels (angiography) in MRI, *magnetic resonance angiography* (MRA), have been created. In this subsection, a *time-of-flight* MRA gradient echo sequence will be presented, illustrated in Figure 2.4.

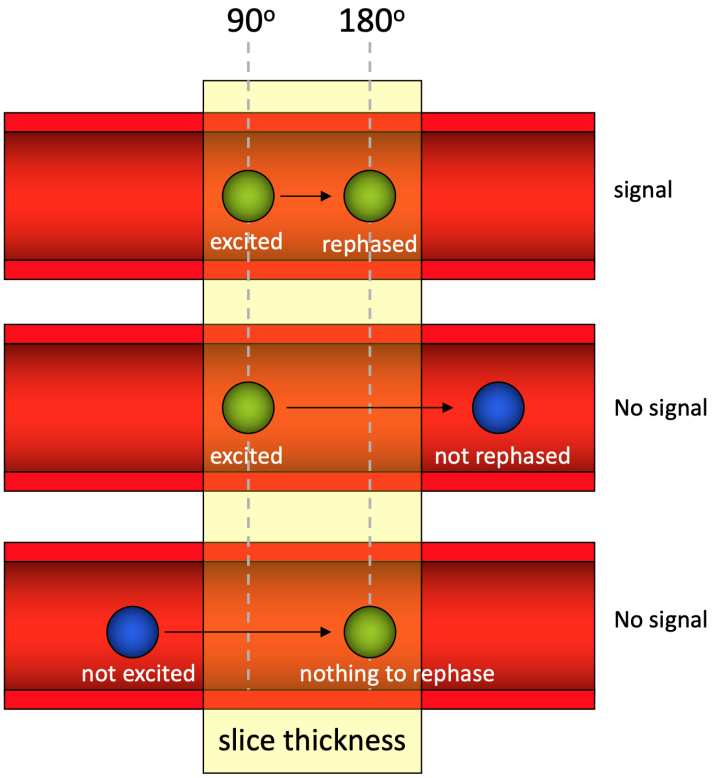


Figure 2.3: Illustration of the problem of flow in a spin echo MRI sequence [26].

Time-of-flight MRA uses the flow of the blood to its advantage. One of the ways to do this is to magnetically, partially *saturate* the stationary protons in the slice [27]. That means lowering the signal of everything in the slice that is not flowing through. This can be done by inducing a magnetisation steady state, M_{SS}, that is significantly lower than M_0 in a slice. This can be achieved in many ways, but a fast way to do it is to use the *FLASH* (fast low angle shot) gradient

echo sequence [28]. This sequence utilises low flip angles, α, and short repetition times, TR. This gives blood time to both enter the slice and its MR signal to be recorded while not letting the longitudinal magnetisation in the stationary protons relax back to M_0. Thus, the blood will give off a comparatively high signal in the image (Figure 2.5) and the blood vessels can be accurately located.

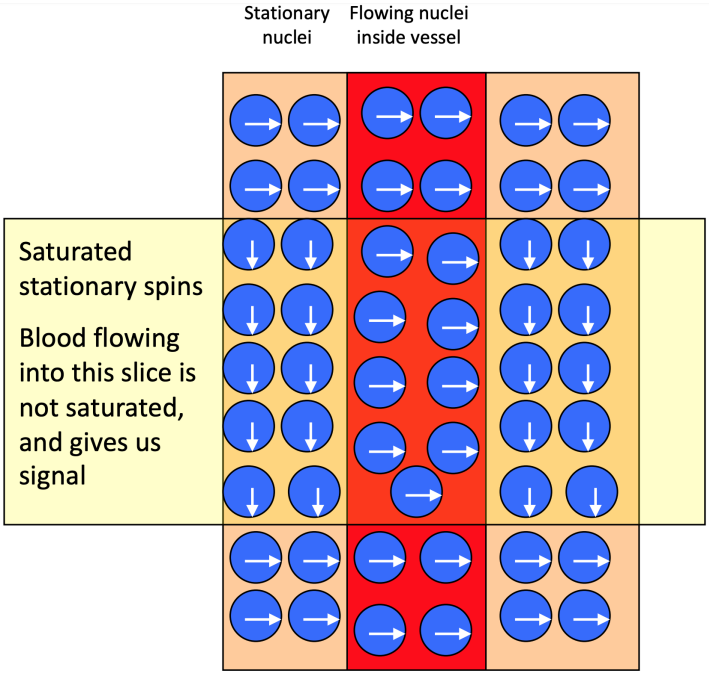


Figure 2.4: Illustration of how a fast gradient echo, time-of-flight MRA sequence provides a relative strong signal from flowing protons [26].

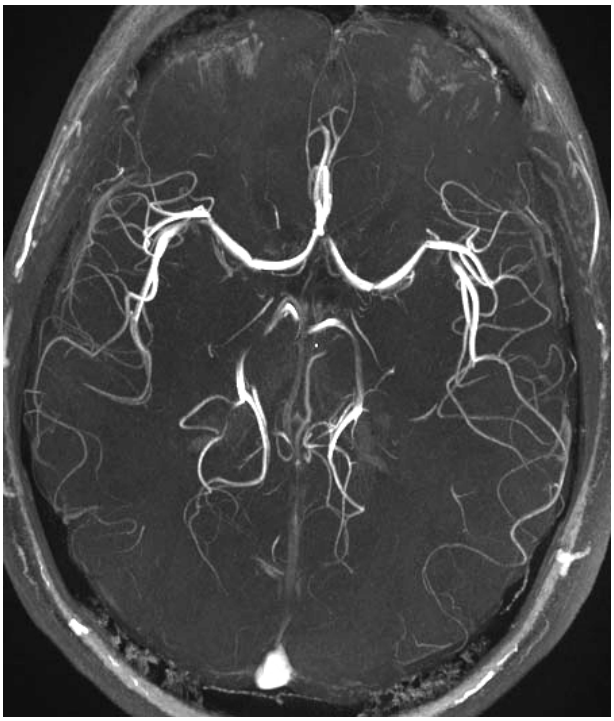


Figure 2.5: Example of MRA showing blood vessels (white) in the brain [29].

2.3 External photon beam radiotherapy

Radiation therapy is widely used to treat cancer (Section 2.1), often in combination with surgery and/or chemotherapy, both curatively and palliatively. The most common radiotherapy method is external beam radiotherapy, which most commonly exposes the patient to ionising photons (X-rays). This external photon beam radiotherapy method will be described in this section.

2.3.1 Ionising photon interaction with biological matter

Ionising radiation is used for cancer treatment because of its interaction with biological tissue, specifically a cell's genetic material, DNA. This interaction is explained thoroughly by Hall and Giaccia [1]. Briefly summarising; the energy of ionising radiation is absorbed by tissue through different interactions. When a photon has interacted with a material, it is said to have been *attenuated*. The interaction may cause damage to the DNA of a cell, either directly, or indirectly through the production of free radicals in the DNA's environment. If the damage is great enough, the cell is unable to repair the damage and consequently dies. These interactions are exploited to kill unwanted, living cells (cancer, Section 2.1) in a patient. This is often done by creating a beam of X-rays with a *linear accelerator* (LINAC). It is named as such because it accelerates electrons at a metal material by applying a certain radiofrequency AC voltage. The interactions of the incoming electrons in the material cause emission of X-rays (bremsstrahlung), with a spectrum of energies. The X-ray energy is given by the voltages in the accelerator. The common range of voltages are from 6 MV to 15 MV [2]. For the purposes of this thesis, the most important takeaway is that ionising radiation is harmful, both for cancer cells and healthy tissue. The harmfulness of the radiation is dependent on the *dose* delivered to the cell, as well as the type of tissue.

The term *dose*, in this case, describes the amount of radiation energy absorbed in matter. The ionising radiation deposits its energy in matter by ionising atoms and molecules in its path, releasing ions and electrons, which in turn transfer their energy to other atoms, molecules, or electrons. In most medical contexts, *Gray* [Gy] is the unit utilised; it is defined as the absorbed energy (in *Joule*, [J]) per mass [kg]. In radiotherapy, the radiation delivered by a LINAC is given in *monitor units* (MU), one of which is equal to a specific dose (often 1 cGy), at a specific depth in water (often 10 cm), at a specific distance to the radiation source (often 100 cm), with a specific beam field size (often 10 x 10 cm), at the particular beam energy. The monitor units are measured in the LINAC head, monitoring the radiation output. The reference conditions, which differ between departments and machines, gives information on how the measurement is performed, and is used to adjust the LINAC to give the correct dose.

2.3.2 Radiation treatment planning

The LINAC has several abilities designed to provide optimal dose to the tumour volume while simultaneously giving minimal dose to the surrounding healthy tissue and OARs. The LINAC is equipped with a beam-shaping component called a *multileaf collimator* (MLC), illustrated in Figure 2.6. The MLC consists of multiple movable rods of a material with a high atomic number that stop the X-rays that are applied, letting only the X-rays headed for the PTV go through [2]. Another way of sparing healthy tissue and OARs is to irradiate

the patient from different angles. The LINAC head is on a gantry that can rotate around the patient on the table, realising the possibility of multiple-angle radiotherapy. One can, for example, rotate the head 90° about the tumour volume between each irradiation, thereby drastically decreasing the dose given to the healthy tissue in front of the tumour, though giving this lessened dose to a larger part of the patient's body. Combining these two abilities by irradiating from several angles and shaping the beam to fit the tumour volume from each angle often gives the best results.

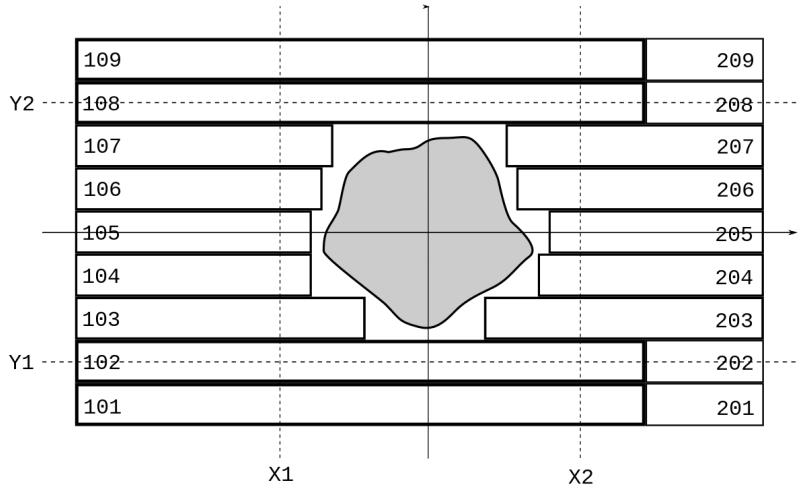
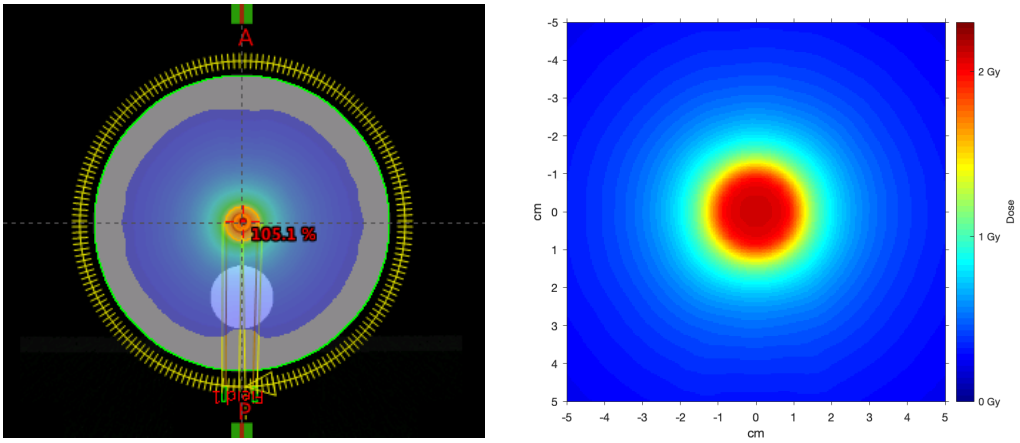


Figure 2.6: MLC shaping beam to PTV [30].

A popular radiation method is the volumetric modulated arc therapy (VMAT) technique. This is a technique where the LINAC continuously delivers radiation while rotating around the patient and shaping the beam, effectively giving the patient a fitted beam of radiation from *every* angle [3]. According to medical physicist and supervisor, Veronika Tømmerås, this is the most common method of radiotherapy treatment at UNN, Tromsø, and is therefore the one used in this thesis.

Independently of the technique used for radiotherapy, every treatment plan requires a calculation of the dose given to every part of the patient. Using a planning CT image, containing essential electron density/mass density information further explored in Section 2.4, and the beam data (angles, energy, duration), an algorithm in the treatment planning system (TPS) computes the doses of radiation within every voxel in the image [3]. Dependent on the calculation algorithm, either electron density or mass density is required. The electron density information is utilised in the TPS relevant for this thesis, and will therefore be the focus in Section 2.4. The result of the calculation is a dose image, or dose matrix, a spatial map of the dose within the patient (Figure 2.7). These calculations, and their resultant dose matrices, are essential for quality assurance (QA) in radiotherapy, for example by gamma evaluation, further explored in Section 2.6. These matrices are also used to calculate *dose-volume histograms* (Section 2.3.3).



(a) Dose matrix overlaid planning CT image in Varian's *Eclipse* (TPS).

(b) Dose matrix only. Zoomed in.

Figure 2.7: Treatment planning images from VMAT treatment plan on a water sphere with a cylindrical bone volume in the axial direction.

2.3.3 Dose-volume histograms

A dose-volume histogram (DVH) simplifies the three-dimensional dose distribution data (3D dose matrix) into a two-dimensional graph. This is done in order to analyse the three-dimensional dose distribution within a specified volume (often PTVs or OARs). In this subsection, the calculation and uses of the DVH will be described. The following information is found in the textbook "Handbook of Radiotherapy Physics" [2].

The most commonly used type of DVH is the *cumulative* DVH. The voxels within the volume of interest are placed into *bins* dependent on their dose. The voxel is placed in a specific bin if the dose in the voxel is higher than a specified threshold dose for the specific bin. For the cumulative DVH, a voxel can be placed in multiple bins. The number of voxels within each bin, often given relative to the total volume, is plotted for the dose, as seen in Figure 2.8.

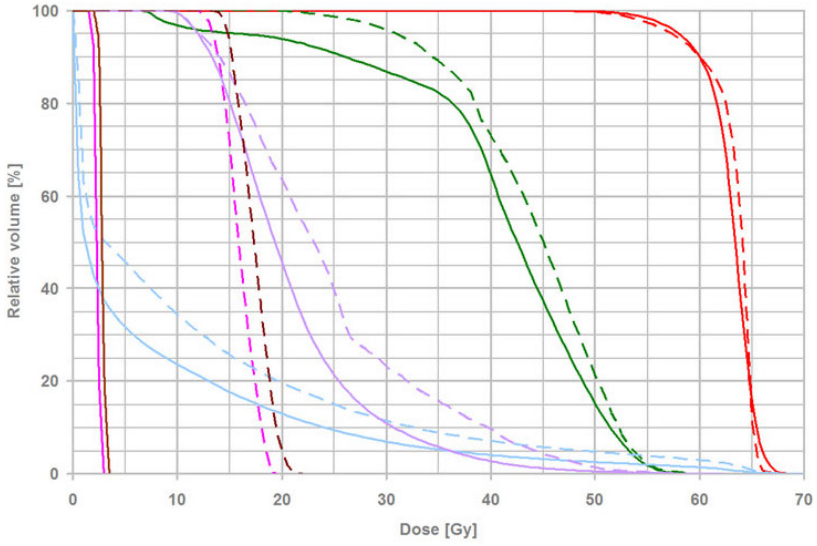


Figure 2.8: Cumulative DVH example, plotting relative volumes for absolute dose [31].

It is desirable for the cumulative DVH of the PTV to resemble a step function, like the red line at the right side of Figure 2.8. This indicates that a large percentage of the volume has a similar dose. The step should occur at the prescription dose. For OARs, a concave appearance is preferred, like the light blue lines in Figure 2.8. In addition to providing a tool for analysis, DVHs are also used in treatment planning by setting DVH goals in the dose optimisation process to shape the curve into a step function (Section 3.2).

The shapes of DVHs are useful for comparing treatment plans, but, in order to have a uniform method of reporting dose in a structure, singular measured values are needed. Such values can be derived from the DVH. The mean dose, D_{mean}, can simply be calculated by dividing the sum of every voxel dose in the volume by the total number of voxels in the volume. Similarly, the median dose, D_{median}, is found by sorting every voxel in the volume. Other measurements are more complex and utilise the DVH directly; a popular choice for such a measurement is the dose, D_{Vref}, which is given by the dose such that the volume receiving a dose higher than D_{Vref} is equal to V_{ref}, where V_{ref} is the relative volume. Here, V_{ref} can be any percentage that is wanted; using the red line example in Figure 2.8, $D_{98\,\%}$ is approximately 55 Gy, while $D_{2\,\%}$ is approximately 67 Gy.

In Figure 2.8, the size of the dose axis (70 Gy) indicates that this DVH describes a full multi-fractional treatment regime. A singular fraction of dose given to a PTV is typically close to 2 Gy. As is the case in this experiment, DVHs can also be calculated for a singular fraction (Section 3.2).

2.4 CT

Spatial electron density information is essential for certain TPS dose calculation algorithms. This information is commonly obtained through CT imaging. In this section, a brief explanation of the basics of CT will be given.

As mentioned, CT provides electron density information on the patient in question [7]. The value in each voxel in the CT image, its CT value, is a measure of its linear attenuation coefficient, μ_x, relative to that of water, μ_w, at specific energies [3]. The linear attenuation coefficient describes the probability of a photon, in this case X-ray, interacting with the material in question. The intensity, I, of a mono-energetic beam that has travelled a distance, L, through a material of attenuation coefficient, μ, at the specific beam energy, is given by:

$$I = I_0 e^{-\mu L} \tag{2.7}$$

where I_0 is the initial beam intensity [32]. The linear attenuation coefficient information is essential to accurately calculate the dose delivered to the patient by external photon beam radiation from a LINAC.

CT scans obtain this information because they utilise X-rays. A fan beam of X-rays passes through the patient, and the transmitted radiation, of intensity I, is detected. Multiple projection views are acquired as the X-ray source rotates around the patient. From these projections, image reconstruction algorithms generate a matrix, where each element is assigned its *Hounsfield unit* (HU), defined as:

$$HU = 1000 \cdot \frac{\mu_x - \mu_w}{\mu_w} \tag{2.8}$$

This defines the CT value of water as 0 HU, while it gives air, with the an assumed zero attenuation, a CT value of -1000 HU. Fat, having lower attenuation than water, has CT values of about -100 HU, while bone, with significantly higher attenuation than water, can have CT values higher than 1000 HU [11], dependent on the bone tissue type. Both of these are dependent on the X-ray energy used in the scan. Hounsfield units at given energies are approximately extrapolated to yield the electron density information that is required to perform accurate dose calculations. This is illustrated in Figure 2.9, where the electron density relative to that in water, *RED*, is plotted for the measured CT values at three different energies. In summary; using Hounsfield units, CT scans utilise X-rays to obtain electron density information. The need for this electron density information complicates a radiotherapy workflow that needs MR images. The use of sCT may provide a solution to these problems.

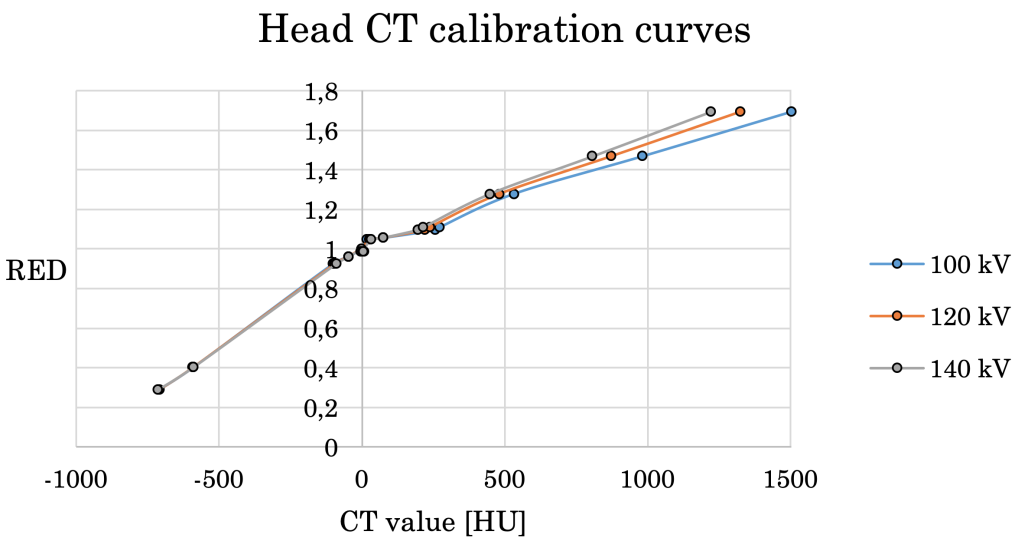


Figure 2.9: Examples of CT calibration curves that provide electron density relative to water, *RED*, as a function of CT value for energies 100 kV, 120 kV, and 140 kV. This plot is from measurements performed at a CT scanner at UNN in 2016.

2.5 Synthetic CT

Because of the aforementioned requirement of electron density information for radiotherapy planning and increasing use of MRI in said planning (increased from 6 % to 24 % of radiotherapy cases in the USA from 2006 to 2017 [5]), sCT has emerged as a possible solution. Obtaining electron density information from MRI would help alleviate many problems related to patient internal and external movement between scans, while also being more cost-efficient. There is an added benefit of lessening the dose to the patient's non-tumour tissue, though this is small compared to the dose delivered by the LINAC. This section will describe the general idea of sCT and delve deeper into clinically approved methods of generating sCT images of the pelvis and brain developed by *Siemens Healthineers* and their clinical partners [7].

The general idea of sCT is to assign CT values to voxels using data from MRI only. Many different ways have been used to do this, often in combination. The following descriptions of these approaches are from Uh et al. [6]. The bulk density approach uses segmentation of different structures and assigns CT

values dependent on what structure it is believed to be (air, bones, soft tissue). This gives an unrealistic sCT image segmented into a low number of HU classes. The atlas-based approach often uses images of several patients whom previously have undergone both MRI and CT scans to elastically deform an atlas to fit the patient in question. This method does not require any extra MR sequences to be performed, but it does struggle with any atypical shapes, where the patient in question's anatomy deviates too much from the patients that were used to create the atlas. The voxel-based approach uses the signal intensity in each voxel in one or several MR images to assign a CT value to the voxel. This requires specialised MR sequences, increasing the scan time. Differentiation of bone tissue and air is difficult due to bone's ultrashort $T2$ (Section 2.2.1). This will be a problem for all of these approaches. To provide physicians and medical physicists with sCT images of the pelvic region, *Siemens Healthineers* and their clinical partners have developed a method using a voxel-based and atlas-based approach.

This sCT method is described in their white paper from 2019 [7]. It uses an MR sequence called T1 VIBE DIXON to classify the soft tissue voxels. This sequence uses the technique described in Section 2.2.2 to obtain water-only and fat-only images, which classify each voxel as either fat or water. Voxels that contain mostly water are given a CT value of 0 HU, while voxels containing mostly fat are given CT value of -75 HU. The air is classified by thresholding. This thresholding may at first classify the bone as air because of the difficulty in distinguishing the two. Therefore, a bone-multi-atlas is used to assign CT values of cortical bone (1170 HU) and cancellous bone (204 HU) to these volumes (Section 3.5). The tissues can be distinguished from each other because of the liquid within the cancellous tissue.

Siemens' method for generating sCT images of the head and neck is more complicated. It is purely a voxel-based method and requires three additional MRI sequences. These are the: T2-SPACE, PETRA, and FLASH Gradient Echo sequences. The T1 VIBE DIXON sequence is run to get the fat and water images.

The T2-SPACE sequence is a three-dimensional $T2$-weighted sequence [33] that is used to visualise the brain anatomy and define the resolution of the sCT. The FLASH gradient echo sequence (Section 2.2.4) is used to ensure blood vessels are not classified as bone, while the PETRA sequence (Section 2.2.3) is used to ensure that bone is not misclassified as air. All images are combined to find the approximate CT value in each voxel.

These approaches have been proved, by using the gamma evaluation technique further explored in Section 2.6, to provide clinically acceptable dose calculations in radiotherapy planning [7].

2.6 Gamma evaluation technique

Gamma evaluation is a method developed in order to compare dose images. It was first proposed by Low et al. [13] as a method of determining whether the measured dose distributions were acceptable compared to calculated dose distributions using certain acceptance criteria. The method is still in clinical use today, both as a QA tool and in research, as it, in addition to assuring quality in treatment, quantifies the effect of propagating errors in radiotherapy treatment planning. This section will look into the gamma evaluation method and its criteria.

The principle of gamma evaluation is straightforward; the dose distribution

in two dose matrices, one reference ($D_r(\mathbf{r})$) and one target ($D_c(\mathbf{r})$), are compared, and the tested points/voxels are given a numerical quality index dependent on their differences. In clinical QA settings, when performing dose measurements in phantoms and comparing with calculated values, the measured dose (the real case) is referred to as the reference matrix, while the calculated matrix is the target matrix. Therefore, in this paper, the real case (bone case) dose matrices are considered as reference, whereas the dose matrices calculated on CT images where bone has been replaced by air (air case) are considered as target matrices. The aforementioned quality index is called the gamma value, γ, and is a measure of disagreement in regions that fail the criteria, while indicating the calculation quality in the regions that pass [34]. The aforementioned criteria are the dose difference, ΔD_M, and distance to agreement (DTA), Δd_M. These are often denoted as $\Delta D_M[\%]/\Delta d_M[\text{mm}]$ in literature. To provide accurate gamma calculations, Δd_M must be sufficiently large in comparison to the voxel/pixel spacing. According to Low et al. [35], Δd_M should be at least three times larger than the resolution, meaning a dose matrix with 0.3906 mm resolution, as is the case in this study (Section 3.2), requires a minimal Δd_M of 1.1718 mm. As is illustrated in Figure 2.10, assuming equal significance of the two criteria, these are represented by an elliptical surface with a reference point, $\mathbf{r}_r$, receiving dose, D_r, in its centre. With $\mathbf{r}_c$ being the compared point, this ellipsoid is defined by:

$$1 = \sqrt{\frac{|\mathbf{r}_r - \mathbf{r}_c|^2}{\Delta d_M^2} + \frac{(D_c(\mathbf{r}_c) - D_r(\mathbf{r}_r))^2}{\Delta D_M^2}} \tag{2.9}$$

For the target distribution, $D_c(\mathbf{r})$, to pass the evaluation, it needs to contain at least one point $(\mathbf{r}_c, D_c)$ within this ellipsoid, meaning it must have at least one point for which:

$$\Gamma_r(\mathbf{r}_c, D_c) \equiv \sqrt{\frac{|\mathbf{r}_r - \mathbf{r}_c|^2}{\Delta d_M^2} + \frac{(D_c(\mathbf{r}_c) - D_r(\mathbf{r}_r))^2}{\Delta D_M^2}} \leq 1 \tag{2.10}$$

The point within $D_c(\mathbf{r})$ with minimal deviation from the reference point is used to define the gamma value, $\gamma(\mathbf{r}_r)$, which is simply defined as the aforementioned minimal deviation. This creates the following pass-fail criterion: when $\gamma(\mathbf{r}_r) \leq 1$, the point passes, while failing when $\gamma(\mathbf{r}_r) > 1$. A complete gamma evaluation is performed by repeating this process for every reference point in the reference matrix. This results in a gamma evaluation image providing important information concerning the level of coherence between two dose distributions (Section 3.3.2).

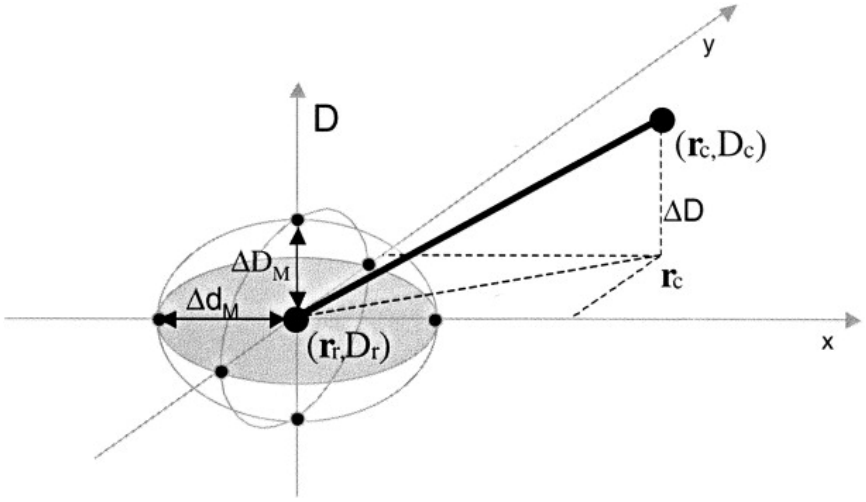


Figure 2.10: The theoretical concept of the gamma evaluation method. Reference and target dose distributions are denoted as $(\mathbf{r}_r, D_r)$ and $(\mathbf{r}_c, D_c)$, respectively. Figure used with permission [34].

Situationally dependent, two main gamma evaluation methods are utilised clinically and in research. These are *local* gamma evaluation and *global* gamma evaluation. In local gamma evaluation, the relative dose difference is calculated relative to the dose in each point, while in global gamma evaluation, the relative dose difference is calculated relative to the maximum dose in the reference matrix. In both cases, it is common to disregard points where the dose is below a certain threshold, often a given percentage of the maximum dose in the dose matrix. This is done because small, inconsequential differences in dose will likely exceed common gamma criteria (1 %/1 mm, 2 %/2 mm, 3 %/3 mm), causing the point to fail without notable ramifications. Using one of these gamma evaluation techniques, with a fitting dose cutoff, one can quantify the coherence of the two dose matrices by comparing passed versus failed points.

Quantification of the coherence of the two matrices is done by performing gamma evaluation and obtaining the *gamma pass rate*. The pass rate is defined as the number of points/voxels in the reference matrix that passed the evaluation divided by the total number of points/voxels that were evaluated. The gamma pass percentage found from this computation is often used in clinics and in research, though differing gamma criteria and clinically acceptable pass rates are employed (Section 2.7).

2.7 Criteria for clinical acceptance

In this section, the theoretical background and previous work considered for the choice of the different acceptance criteria will be explained. The main measurement of interest in this thesis is the DVH measurement D_{mean}. This is because it has been used in similar research previously [15], but also because the project thesis [12] found that the two-dimensional dose through the centre of the PTV deviated profoundly. In the specific scenarios posed in this thesis, where the planning images show air volumes when it is supposed to be volumes of bone, the doses within the PTVs decrease from what is expected, lowering the chances of killing the cancerous tissue because of the unexpected increase in attenuating tissue. However, the gamma evaluation method, which is able to test the entire irradiation volume for discrepancies, will still be used as a secondary test.

In this thesis, the acceptance criterion that is used when measuring D_{mean} deviations is 2 %. In 2016, Sjölin et al. [15] used the 2 % D_{mean}-deviation criterion to test whether the gamma evaluation method could detect errors that they considered unacceptable. Others have used different DVH point measurements to compare dose distributions; Korsholm et al. [16] used the DVH points: D_{median}, $D_{98\,\%}$, and $D_{2\,\%}$, also utilising 2 % as acceptance criterion.

Different gamma criteria and clinically acceptable pass rates are used, both in clinics and in research (Section 2.6). Johnstone et al. [36] point out that the inconsistencies in choice of gamma criteria must be addressed in order to aid method comparisons. Though many use different gamma criteria, most utilise 95 % as the acceptance criterion for the gamma pass rate [14][15], likely because confidence intervals of 95 % are widely used in medical research. At UNN, for radiotherapy QA, 95 % gamma pass rate for a global gamma evaluation, using 2 %/2 mm criteria, and with a 20 % dose cutoff, is considered acceptable. Therefore, these are the values that will be used when performing gamma evaluation in this thesis.

2.8 Statistical analysis

In this section, some of the relevant theory behind the statistical analysis utilised in this thesis is presented. Mostly, this entails the uncertainty of linear regression (Section 2.8.1) and the comparison of linear regression lines (Section 2.8.2).

2.8.1 Linear regression uncertainty

Regression models use data points to estimate a function that describes the dependency of the variables. In a linear regression model, a linear function is created: $y = ax + b$. Here, a is the slope of the line and b is the intersect of the line with the y-axis. However, regression models are imperfect. Uncertainty calculations of regression models are much used and easy to compute. However, this uncertainty calculation assumes error-free measurement of the data points for which the regression was performed. Below follows an explanation of how the uncertainty in the measurement of the data is combined with the uncertainty of the model itself [37].

A common way to calculate the uncertainty of a regression model is through the *standard error of the estimate*:

$$\sigma_{est} = \sqrt{\frac{\sum(y - y_{est})^2}{n - n_p}} \tag{2.11}$$

where y is the measured value at a given point, y_{est} is the value in the same point as estimated by the regression model, n is the number of data points, and n_p is the number of parameters. For a linear model, $n_p = 2$.

This does not describe the uncertainty in the model itself, but estimates the uncertainty in y_{est}. To completely describe the uncertainty in the model, the uncertainty in the slope, a, must be included. The slope is dependent upon the uncertainty of y_{est}, resulting in an uncertainty in a given by:

$$\sigma_a = \sqrt{\frac{\sigma_{est}^2}{\sum_{i=1}^{n}(x_i^2 - \bar{x}^2)}} \tag{2.12}$$

where x_i is the ith value of x and $\bar{x}$ is the mean value of all x.

The resulting function of the linear regression can be expressed as:

$$y_{est} = \bar{y} + a(x - \bar{x}) \tag{2.13}$$

where $\bar{y}$ is the average of the measured values of y.

Since the uncertainty of y_{est} is estimated by σ_{est}, the uncertainty in $\bar{y}$, following normal rules for error propagation [38], is estimated by:

$$\sigma_{\bar{y}} = \frac{\sigma_{est}}{\sqrt{n}} \tag{2.14}$$

The equation for combining uncertainties is given by:

$$\sigma_C^2 = \sigma_A^2 + \sigma_B^2 \tag{2.15}$$

where σ_A and σ_B are independent sources of error. As σ_a and $\sigma_{\bar{y}}$ are independent sources of error, equation (2.15) can be applied to find the uncertainty in the regression model:

$$\sigma_m = \sigma_{est} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\left(\sum_{i=1}^{n} (x_i^2 - \bar{x}^2)\right)^2}} \tag{2.16}$$

Now, other, unrelated uncertainties can be considered. This is simply done by applying equation (2.15) again, resulting in:

$$\sigma = \sqrt{\sigma_m^2 + \sigma_e^2} \tag{2.17}$$

where σ_e is the mentioned unrelated uncertainty.

One such uncertainty might be the instrument precision. If the measurement equipment is very accurate, providing measurements with an accuracy of several decimal points, this uncertainty might be negligible. However, if that is not the case or the data is made up of indirect measurements where error propagation decreases the accuracy in the final measurement, this uncertainty should be included.

Once all relevant uncertainties have been combined, *confidence bands* can be found. Confidence bands create a confidence interval surrounding the regression curve, representing the area within which one is X% confident of finding the regression line. A confidence level of 95 % is widely used. When limited on the number of data points, n, the Student's t-distribution is used to find this band, through the critical value, $t_{\alpha,\nu}$. By simply multiplying each uncertainty with the critical t-value corresponding to the required confidence level and degrees of freedom ($\nu = n - n_p$), the confidence band, CB, is approximated:

$$CB = y_{est} \pm t_{\alpha,\nu}\sigma \tag{2.18}$$

An example of a linear regression line and surrounding confidence band is shown in Figure 3.5 (Section 3.3.1). Because of the dependence of x in σ_a (equation (2.12)), confidence bands of linear functions take an hourglass shape. However, sometimes this is barely visible, as is the case in Figure 3.5. The expected shape can be seen in Figure 4.5 (Section 4.2).

2.8.2 Chow test

When working with multiple different linear regression lines, one might require to calculate whether or not the lines significantly ($p < 0.05$) differ from one another. This can be tested using the Chow test [39].

This method compares two separate regression lines (y_1 and y_2) where a structural difference is known (for example differing PTV sizes, Section 3.1) to a regression line created by combining the data of the two (y_c). The null hypothesis is that the two lines are the same and can be described by the same, combined line.

The Chow test creates a test statistic, C, that follows the Fisher-Snedecor distribution (F-distribution). The critical values in the F-distribution is found through the desired significance level (α) and two separate values of degrees of freedom (ν_1 and ν_2). If the test statistic is not larger than the relevant critical value in the F-distribution, the null hypothesis cannot be rejected. The Chow test provides both the test statistic, C, and the two values of degree of freedom:

$$C = \frac{(S_c - (S_1 + S_2))/k}{((S_1 + S_2)/(N_1 + N_2 - 2k))} \tag{2.19}$$

Here, k is the number of parameters and N_1 and N_2 are the number of data points in the respective data sets. S_C, S_1, and S_2, are the sum of squared residuals from the data, given by:

$$S_z = \sum_{i=1}^{N_z} (y_i - y_z)^2 \tag{2.20}$$

where y_z is the estimated value at the value of x corresponding to y_i. The two values of degrees of freedom are found by: $\nu_1 = k$ and $\nu_2 = N_1 + N_2 - 2k$.

This calculation does not directly provide a specific p-value; it only tells whether or not the test statistic reached its required level of significance. However, software for calculating the p-value, using the test statistic and the two values of ν, does exist (Section 3.3.1).

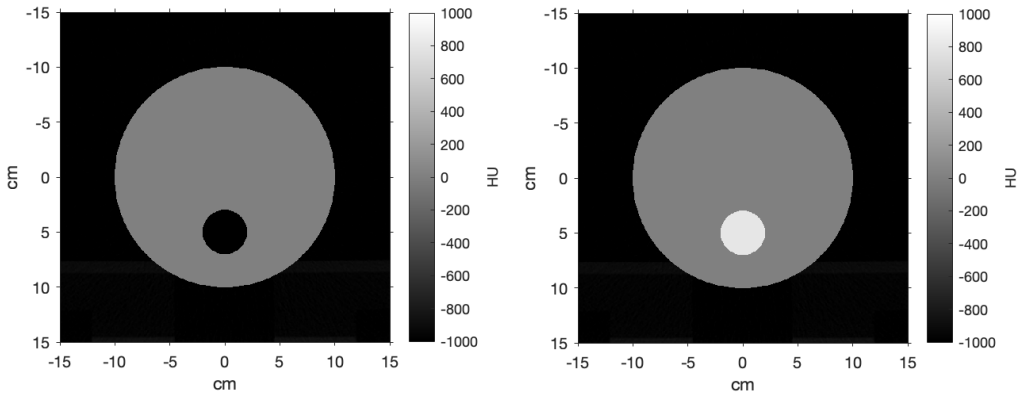
3 Method

In this section, the steps taken in order to calculate the limit of acceptable magnitude of bone misclassification, in relation to misclassification volume size, distance to the tumour centre, and planning target volume size, will be detailed. To simulate cases of bone misclassification, several sets of two conventional CT images were created by editing one original image (Section 3.1). Each set of CT images consisted of one CT image with a bone cylinder volume and another where the bone volume had been replaced by air. For each image containing an air volume, an optimised VMAT treatment plan was created, for differing PTV sizes, and copied to its corresponding CT image containing a bone volume (Section 3.2). The doses were calculated for both and the relevant data were retrieved for analysis. Another similar experiment was performed to test the consequences of misclassifying an entire skull (Section 3.4). Finally, the experimental results were tested in realistic scenarios using authentic sCT images (Section 3.5).

3.1 Acquiring planning CT images

Providing the basis for the treatment planning CT images, a CT scan was performed on a water sphere phantom. The scanner used was the *Siemens Biograph Vision*, made for PET/CT imaging. The scan was performed at 100 kV maximum energy and 119 mA s exposure time product (important point on maximum energy in Section 5). The slice thickness was 1 mm and the plane resolution was 512x512 pixels with a spacing of 0.5859 mm in both directions. This CT scan was the basis for all following CT images in this thesis, with the exception of the sCT images (Section 3.5), ensuring that the CT images were all identical with the exception of the differences of interest.

The base CT image was digitally edited multiple times. Using another Siemens product, *MM RT Image Suite* in *syngo*.via [7], the water sphere phantom was digitally replaced by a slightly larger (20.0 cm diameter) water sphere (0 HU) with its origin in the beam isocentre. The same HU overwriting tool was used to create the cylindrical bone and air volumes within the water sphere, in the axial direction. The value 800 HU was suggested for bone by *syngo*.via, and was therefore chosen. This CT value might represent a realistic misclassification of bone consisting of a larger part cortical bone and a smaller part cancellous bone (Section 2.5). Air is -1000 HU by definition (Section 2.4) and was therefore chosen for the cylindrical volumes in the CT images that represented the misclassification. The cylindrical volume's length was chosen to be long enough to ensure it always fully covered the PTVs in the direction orthogonal to the axial plane. The PTVs were created later, and irradiated with a 0.5 cm margin (Section 3.2). As the largest PTV in this thesis was a sphere of diameter 2.5 cm, a cylindrical length > 3.5 cm was sufficient. Each CT image with a certain sized cylindrical bone volume had a corresponding CT image with an equal sized air volume, as shown in Figure 3.1. 29 such sets were created, with six different misclassification volume sizes, given in two-dimensional area A, and five different distances between the tumour and misclassification volume centres d. The experimental set-up is illustrated in Figure 3.2.



(a) Centre axial slice of water sphere with air cylinder. Representing the bone misclassification case.

(b) Centre axial slice of water sphere with bone cylinder. Representing the real case.

Figure 3.1: A set of otherwise equal CT images with equal sized air/bone volumes. This shows the set at which $A = 12.5\,\mathrm{cm}^2$ and $d = 5.0\,\mathrm{cm}$.

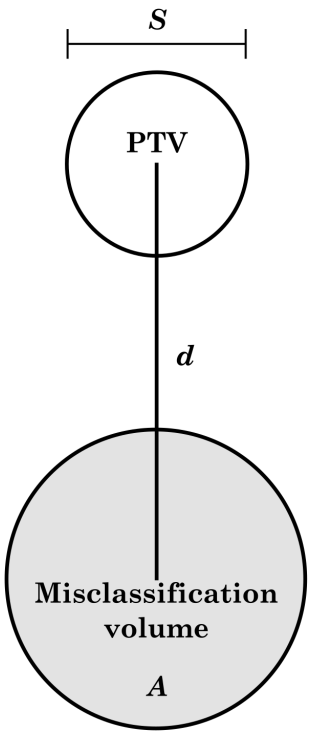


Figure 3.2: Two-dimensional geometric illustration of experimental set-up, with variables S (length of horizontal line and diameter of the PTV), A (area within the grey circle, base area of cylinder), and d (length of vertical line and distance between misclassification volume centre and PTV centre).

The specific values used for A were $2.5\,\mathrm{cm}^2$, $5.0\,\mathrm{cm}^2$, $7.5\,\mathrm{cm}^2$, $10.0\,\mathrm{cm}^2$, $12.5\,\mathrm{cm}^2$, and $15.0\,\mathrm{cm}^2$. The values chosen for d were 3.0 cm, 4.0 cm, 5.0 cm, 6.0 cm, and 7.0 cm. The treatment planning volumes were chosen to be spheres with diameters, S, of 1.5 cm, 2.0 cm, and 2.5 cm, and were to be located in the centre of the water sphere. In clinics, the sizes of PTVs differ greatly. A 2.0 cm diameter PTV was suggested by supervisor and medical physicist Veronika Tømmerås, as it represents a typical, clinical PTV size. The two other values of S

(1.5 cm and 2.5 cm) were chosen as they fit the experimental set-up while approximately resulting in a doubling in volume for each step up in size. More sets of CT images containing more combinations of the variables A and d could have been created. However, the treatment planning and dose calculation processes (Section 3.2) were time consuming. Going through the treatment planning and dose calculation processes once, for a single value of S, took an entire working day. As several repetitions of this process for each combination of S, A, and d, were desired, this number of values for S, A, and d were deemed fitting.

As a cylinder with base area 15.0 cm^2, equalling a radius of 2.19 cm, would leave only 0.6 mm between the edges of the PTV and cylinder at $d = 3.0$ cm, a set of CT images with these values of A and d was not created. This was done in order to create as equal conditions for each experiment as possible, disregarding the variables of interest. At $d = 7.0$ cm there will always be a significant amount of water between the misclassification volume and PTV, meaning the photon path would never go directly from bone/air and into the PTV. For the experiments to be equal, every experiment must have a set-up where the photon path to the PTV through the misclassification volume is, in order: water (0 HU), misclassification volume (-1000 HU or 800 HU), water (0 HU), and PTV (0 HU). The effect of deviating from this by going directly from misclassification volume to PTV was seen in the project thesis experiment [12], where the linear relationship of dose deviation within the PTV with the area A was compromised. For $S = 2.0$ cm, this was also an issue for the 12.5 cm^2 cylinder case at $d = 3.0$ cm (0.1 mm between edges). Additionally, for $S = 2.5$ cm, the 10.0 cm^2 cylinder and 12.5 cm^2 cylinder would overlap the PTV at $d = 3.0$ cm. These experiments were therefore not conducted.

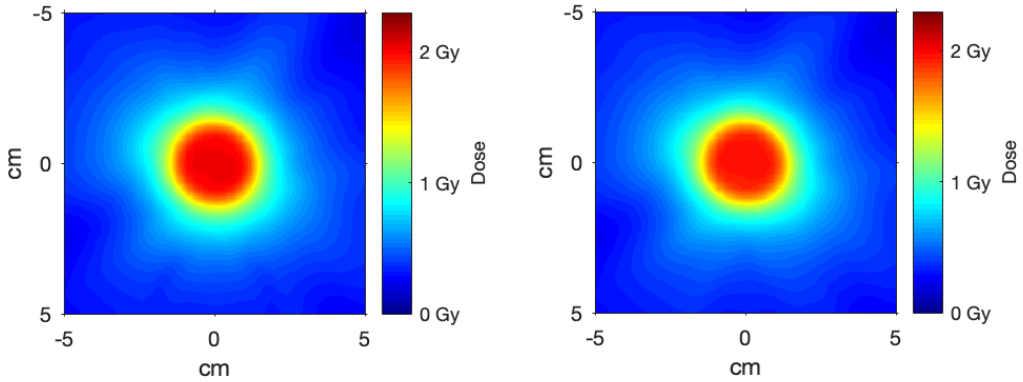
3.2 Creating treatment plan

The 29 sets of planning CT images were imported into a TPS; the Varian external beam planning v 15.6, using the calculation algorithm AAA v 15.6.06. Being the most common treatment plan for external beam radiotherapy, VMAT was chosen (Section 2.3.2). As lower X-ray energies are more prone to attenuation than higher ones, the lowest possible LINAC energy (6 MV) was chosen, providing maximum differences in doses due to the attenuation difference of bone and air.

First; the three different PTVs were created in the contouring tool using a three-dimensional sphere brush to create default accuracy, high risk PTVs, which were placed in the centre of the water sphere and copied to every planning CT. In order to simulate a real bone misclassification case, the original planning and optimisation processes were done using the planning CT images containing air before being copied over to its corresponding bone planning image. In the external beam planning tool, a new plan was created, giving the PTV of choice a dose of 2 Gy, with the patient positioned head first-supine. The arc geometry tool was used to create an optimised VMAT plan with a target margin of 0.50 cm, a fixed collimator angle of 0°, and the recommended, isocentric, full, clockwise rotation. The upper and lower dose boundaries in the optimisation tool, the DVH goals mentioned in Section 2.3.3, were set to 2.03 Gy and 1.97 Gy, respectively, both with 100 % priority. Normal structure resolution (2.5 mm) was chosen before running the optimisation tool in automatic optimisation mode.

After optimisation, the doses were calculated, resulting in dose matrices, as shown in Figure 3.3, and DVH point (D_{mean}) measurements (Appendix A.2).

Plan normalisation was set so 100 % of the dose covers 50 % of the target volume and the MU value was documented (Appendix A.5). The reference condition for the LINAC was 130 MU (Section 2.3.1). Each combination of variables (S, A, and d) was tested three times ($E = \{1,2,3\}$). Most often, the optimisation process resulted in three different MU values, indicating three different plans had been created. When MU values were repeated, meaning the automatic optimisation had created the same plan twice, the optimisation process was repeated until a distinct MU value was found. In some cases, many tries were needed in order to create three different plans. A fourth repetition through the entire experiment would likely have been very time consuming. Some combinations may not have provided a fourth distinct plan at all. Therefore, three distinct plans were deemed enough. When a distinct plan was created, it was copied onto the coinciding planning CT image containing a bone volume, where the treatment beam was perfectly replicated by performing a dose calculation using the identical MU value found in the air case.



(a) Centre axial slice of dose matrix calculated with air volume.

(b) Centre axial slice of dose matrix calculated with bone volume.

Figure 3.3: Calculated dose matrices for the case at which $S = 2.0\,\text{cm}$, $A = 12.5\,\text{cm}^2$, and $d = 5.0\,\text{cm}$.

Two-dimensional dose matrices were exported for gamma analysis (Section 3.3.2). The dose matrices were exported showing absolute, planar plan dose through the isocentre of the PTV, and aligned to show the axial slice, equivalent to what is shown in the planning CT images (Figure 3.1). The dose matrices were 10 x 10 cm and 256 x 256 pixels, resulting in a resolution of approximately 0.3906 mm.

The D_{mean} DVH point measurements (Appendix A.2) were recorded into Microsoft Office Excel (MS Excel) for further analysis (Section 3.3). The values were obtained using the cumulative DVH with absolute doses [Gy] and relative volumes [%].

3.3 Data analysis

As Sjölin et al. [15] used the DVH parameter, D_{mean}, as their gold standard measurement in their correlation analysis to test the gamma evaluation's ability to correctly fail cases of too high dose deviation (>2 %), D_{mean} deviation was utilised as the main measurement of interest in this thesis. This is because a bone misclassification error leading to bone (800 HU) being replaced by air (-1000 HU) in the planning sCT will lead to a volume of significantly lower attenuation, which in turn leads to an underestimate of the amount of radiation required to sufficiently irradiate the PTV. Gamma evaluation, which, unlike PTV-specific DVH measurements, takes all areas of significant dose into consideration, was performed to evaluate the general dose distribution, while also testing whether the pass rate of two-dimensional gamma evaluation accurately highlights cases of significant PTV dose deviations.

3.3.1 DVH

The data analysis was done in MS Excel. For each combination of variables S, A, and d, the relative D_{mean} deviation, Δ_{mean}, was calculated for the three different treatment plans using:

$$\Delta_{mean} = \frac{|D_{m,air} - D_{m,bone}|}{D_{m,air}} \cdot 100\,\% \tag{3.1}$$

where $D_{m,air}$ and $D_{m,bone}$ are the measured values of D_{mean} for the air and bone cases, respectively. The instrument precision in the TPS, σ_p, of the D_{mean} measurement is 0.0005 Gy. Through propagation of uncertainty [38], the uncertainty of the indirect measurement of Δ_{mean} was approximated using:

$$\sigma_\Delta \approx \left|\frac{D_{m,air} - D_{m,bone}}{D_{m,air}}\right| \left(\frac{2\sigma_p}{|D_{m,air} - D_{m,bone}|} + \frac{\sigma_p}{D_{m,air}}\right) \frac{100\,\%}{D_{m,air}} \tag{3.2}$$

As $D_{m,air}$ and $D_{m,bone}$ vary with Δ_{mean}, σ_Δ varies as well. However, this variation is very small. The uncertainty of the Δ_{mean} measurement is approximately 0.025 % for low values of Δ_{mean} and 0.026 % for high values of Δ_{mean}.

As shown in Figure 3.4, Δ_{mean} was plotted for A. This was done for every combination of S and d. As a linear relationship between Δ_{mean} and A was expected [12], the data points were fitted to a linear function ($y = ax + b$) using MS Excel's linear regression model. Here $y = \Delta_{mean}$ and $x = A$. The purpose of this was to find the point $I = (d, A)$ where the linear function intersects the Δ_{mean} limit of 2 %. This was simply done by using the *SLOPE* and *INTERSECT* functions in MS Excel to find a and b, respectively, before solving $y = 0.02 = ax + b$ for x. The fit of the data to the model, measured by R^2, was obtained (Appendix A.8.1). Values of R^2 close to 1 indicate good fits.

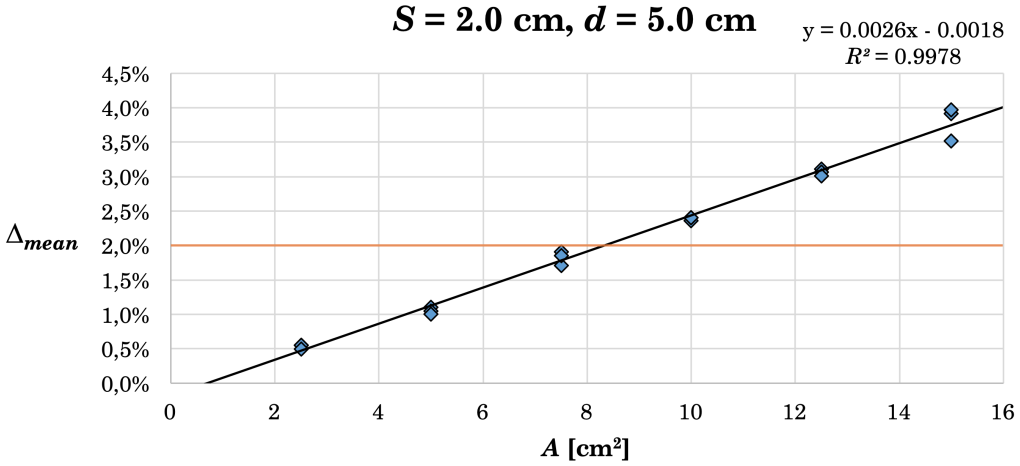


Figure 3.4: Δ_{mean} plotted as a function of A for the case where $S = 2.0\,\text{cm}$ and $d = 5.0\,\text{cm}$. The red line at $\Delta_{mean} = 2\,\%$ shows the border for acceptable mean dose deviation. The black line, resulting from linear regression of the 18 data points, is given by the linear equation in the top right corner ($y = ax + b$). R^2 indicates the fit of the equation to the measurements of Δ_{mean}.

However, as regression is imperfect (Section 2.8.1), there is uncertainty associated with these lines. These uncertainties, σ_L, were calculated through the equations shown in Section 2.8.1, ending with equation (2.17), where $\sigma_e = \bar{\sigma}_\Delta$. Here, $\bar{\sigma}_\Delta$ is the mean of all σ_Δ relevant for the regression line. The uncertainties, σ_L, provide the uncertainties of the measurements of I, σ_I. These uncertainties were computed by adjusting equation (2.18) to find the two intersections, $I(d, A_{left})$ and $I(d, A_{right})$, with the 2 % line by solving the following equation for $x = A$:

$$0.02 = ax + b \pm \sqrt{\sigma_m^2 + \bar{\sigma}_\Delta^2} \tag{3.3}$$

The uncertainties in I could now be calculated:

$$\sigma_I = \frac{A_{right} - A_{left}}{2} \tag{3.4}$$

By applying the relevant values of $t_{\alpha,\nu}$, the 95 % confidence bands of the regression lines (equation (2.18)) were also approximated (Section 2.8.1), as shown in Figure 3.5. Similarly, the 95 % confidence intervals of I were calculated, illustrated by the left-most and right-most intersections of the confidence band with the 2 % line in Figure 3.5.

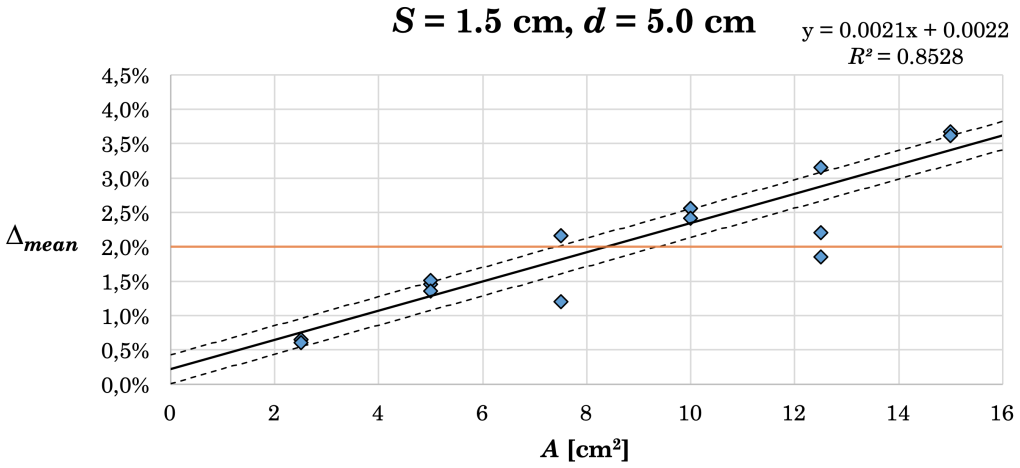


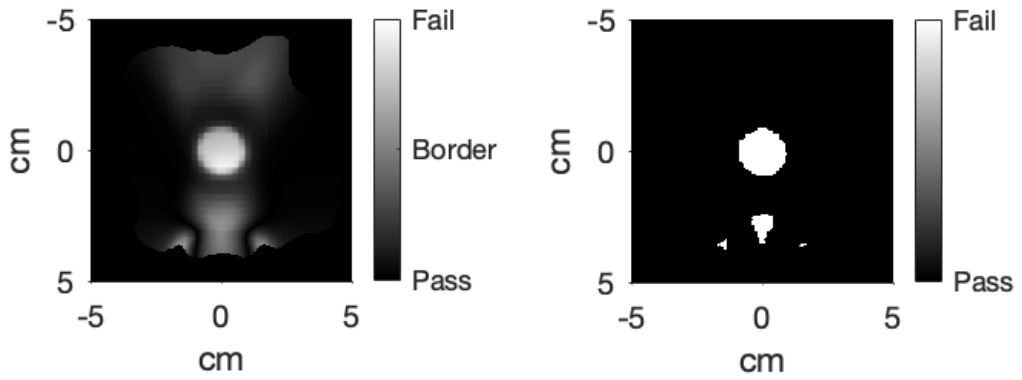
Figure 3.5: The 95 % confidence band (dotted lines) surrounding the regression line at $S = 1.5\,\text{cm}$ and $d = 5.0\,\text{cm}$. The relatively bad fit of the data ($R^2 = 0.8528$) creates a relatively wide confidence band.

For each S, the points $I(d, A)$ were plotted. Once again, linear regression was performed, resulting in new regression lines (Section 4.1). The same procedure as described above was performed again in order to create approximate 95 % confidence bands. In this case, the mean uncertainty connected to $I(d, A)$, $\bar{\sigma}_I$, replaced σ_e in equation (2.18).

To investigate the importance of the variable S, all three final regression lines, plotted in Figure 4.4, were compared, two at a time, using the Chow test (Section 2.8.2). All three resulting test statistics were then fed into the MS Excel function *F.DIST.RT* in order to calculate the p-values (Section 4.1).

3.3.2 Gamma evaluation

All gamma analysis was done using MATLAB® version R2018b (9.5.0.944444). The code for performing 2D gamma evaluation was used with permission from its author [40]. This code provides a matrix of gamma values, describing the coherence of dose distribution in each point. An example is given in Figure 3.6 along with a matrix adjusted to show whether a pixel has passed or failed. All code that was created and used is given in Appendix A.9.

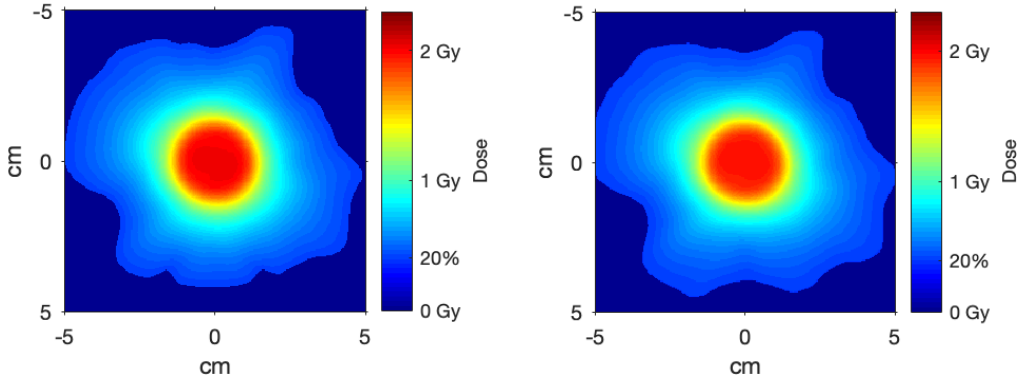


(a) Gamma value matrix. Border at $\gamma = 1$.

(b) Binary gamma matrix.

Figure 3.6: Gamma matrices found by performing global gamma evaluation on the dose matrices in the case of $A = 12.5\,\text{cm}^2$ and $d = 5.0\,\text{cm}$. Gamma criteria: 2 %/2 mm. Dose cutoff: 20 %.

The code given in Appendix A.9 used the above-mentioned gamma evaluation code to find the gamma pass rates given a global gamma evaluation and with a dose cutoff value of 20 %, illustrated in Figure 3.7. All pixels that were cut did not contribute to the final pass rate. Global gamma evaluation and this cutoff were chosen to match what is used in the radiotherapy clinic at UNN, Tromsø.



(a) Edited dose matrix, air volume case.

(b) Edited dose matrix, bone volume case.

Figure 3.7: Dose matrices where $S = 2.0\,\text{cm}$, $A = 12.5\,\text{cm}^2$, and $d = 5.0\,\text{cm}$, edited to illustrate the 20 % cutoff by setting all dose values below 20 % of maximal dose to 0 Gy.

After documentation of the gamma pass rates (Appendix A.2), the results were analysed in a similar fashion to that of the DVH experiment, as described

in Section 3.3.1. Here, $\sigma_p = 0$ and the line $\Delta_{mean} = 2\,\%$ was replaced with the gamma pass rate at 95 %. However, as it is not possible to obtain gamma pass rates above 100 %, a linear fit is difficult to achieve when including all data points, as illustrated in Figure 3.8.

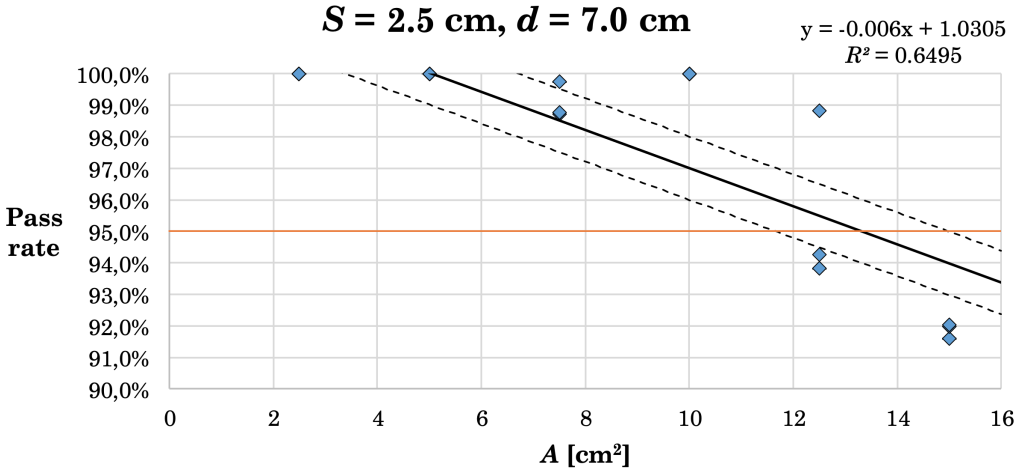


Figure 3.8: A bad linear fit of the data is worsened by multiple data points at maximum value (100 %). The 95 % confidence band surrounds the regression line.

Therefore, for all A where every single gamma pass rate measurement was 100 %, the data points were excluded, as shown in Figure 3.9.

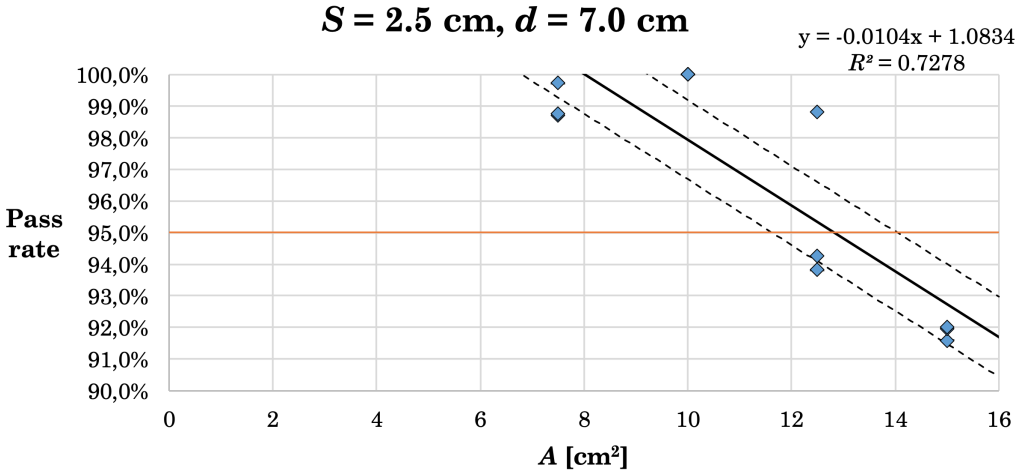


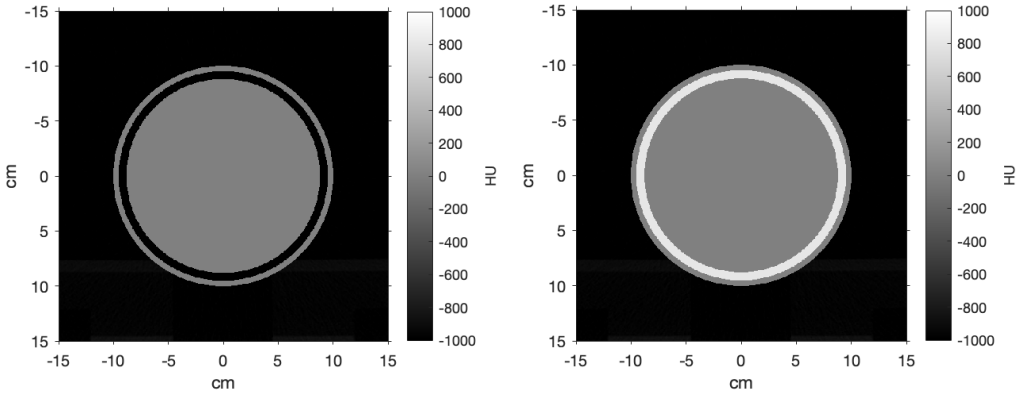
Figure 3.9: A slightly improved version of the fit to the linear data in Figure 3.8. For all A where every data point was 100 %, the data points have been excluded. The 95 % confidence band surrounds the regression line.

The results of this process are presented in Section 4.2, where they are visually compared to the results of Section 3.3.1. The Chow test was once more utilised in order to test for significant difference between the final lines using the gamma evaluation data compared to the DVH data.

3.4 Skull experiment

Because of the differing technique in generating sCT images of the head compared to the pelvis (Section 2.5), with the head technique focusing more on preventing misclassification of bone, interest arose in the consequences of misclassification of an entire skull. Why is more effort put into avoiding misclassification of bone tissue in the head? In order to investigate the radiotherapeutic consequences of bone misclassification of the skull, the experiment described below was set up.

The initial 20 cm diameter water sphere planning CT image (Section 3.1) was edited again to create an experiment simulating a bone misclassification error where an entire skull volume is replaced by air, shown in Figure 3.10. The water sphere is approximately the size of an average human head when measured from back to front [41]. The average human skull thickness is 6.5 mm for men and 7.1 mm for women [42]. Therefore, a 7.0 mm thickness was chosen for the misclassification volume. The misclassification volume was placed inside the water sphere, leaving a 5.0 mm thick layer of water representing the layer of soft tissue on the outside of the skull. The two-dimensional area of bone that has been misclassified is given by $A = \pi(R^2 - r^2)$, where $R = 9.5\,\text{cm}$ is the outer radius of the spherical shell and $r = 8.8\,\text{cm}$ is the inner radius. This gives $A = 40.2\,\text{cm}^2$.



(a) Centre axial slice of water sphere with air skull. Representing the bone misclassification case.

(b) Centre axial slice of water sphere with bone skull. Representing the real case.

Figure 3.10: A set of otherwise equal CT images with equal sized air/bone volumes in the shape of spherical shells of thickness 7.0 mm. The outside water shell is 5.0 mm thick.

In this experiment, the size of the PTV was kept constant ($S = 2.0\,\text{cm}$) and the variable tested was the distance, x, between the inner edge of the misclassification volume and the centre of the PTV, as is illustrated in Figure 3.11. Six different distances were tested, ranging from $x = 2.3\,\text{cm}$ to $x = 8.8\,\text{cm}$, where the latter is in the origin of the water sphere. The distance between all neighbouring PTV centres was 1.3 cm. The method for creating dose matrices was identical to that described in Section 3.2, with one exception; the two plans for the PTVs closest to the skull were optimised with two half rotations instead of the one full rotation, as recommended by the TPS. The MU values are given in Appendix A.6.

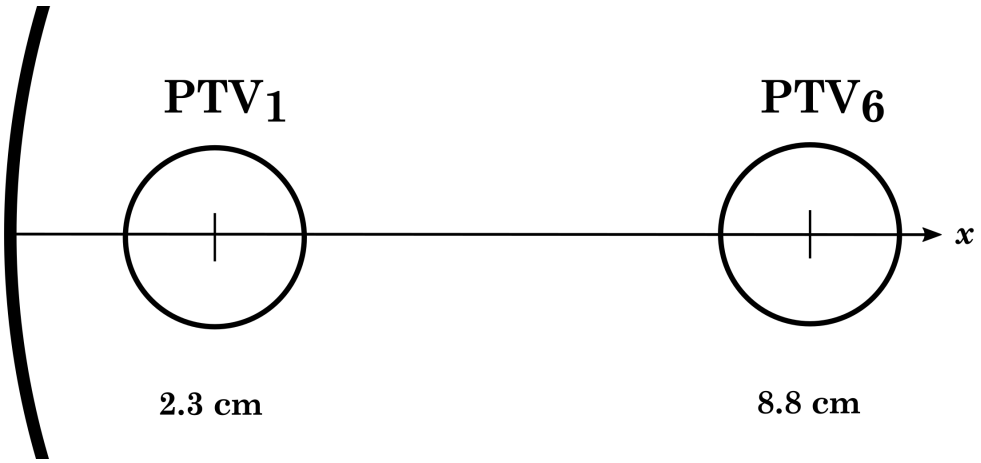


Figure 3.11: Two-dimensional geometric illustration of experimental set-up, showing the PTV closest to, and farthest from, the inner edge of the spherical shell misclassification volume. PTV_1 has its centre at $x = 2.3\,\text{cm}$ from the skull, while PTV_6 ($x = 8.8\,\text{cm}$) is in the origin of the water sphere.

The gamma pass rate and Δ_{mean} (equation (3.1)) were recorded (Appendix A.3) and averaged for each x (Section 4.3). Because of clarity of results, no statistical analysis was performed.

3.5 Synthetic CT scenarios

To test whether the results of the experiments above were realistic, real sCT images were created, featuring this author. This author was at the time a 25 year old, healthy male of approximately average height and weight (178 cm and 75 kg). A *Siemens Biograph mMR* 3T PET/MRI scanner was used to scan the author's pelvis and head, using the techniques described in Section 2.5, before the sCT images were created in *MM RT Image Suite*. The resultant sCT of the pelvis was a 224 x 224 pixel image of 0.2 cm resolution. The slice distance was also 0.2 cm. The head sCT image was 199 x 296 pixels with a resolution and slice distance of 0.1 cm. Both sCT images simulated the CT values at 120 kV maximum energy (Section 2.4). Using *MM RT Image Suite*, some of the bones in the images were replaced with air before all images were exported into the TPS. The contouring tool in the TPS was used to calculate the mean CT values of the bone in the sCT images. Both the pelvis and head case showed a mean value of bone (cortical and cancellous bone combined) of approximately 400 HU. This discrepancy in CT value from the other experiments performed in this paper is further discussed in Section 5 as it, naturally, influences the results of the experiments described in this section.

In the TPS, spherical PTVs of 2.0 cm diameter were placed at certain distances from bone misclassification volumes. $S = 2.0\,\text{cm}$ was chosen, as the results in Section 4.1 showed least uncertainty for this PTV size. Only one value of S was tested, as no significant statistical difference was found for the PTV size (Section 4.1). Only two PTVs at different distances, d or x, were needed to observe the dependency of the distance between PTV and misclassification volume. More PTVs could have been created, but because of the difference in bone CT value between the primary and skull experiments and the sCT experiments, the only valuable results that could be observed here was that Δ_{mean} decreased for increased d or x, while simultaneously being below 2 % if, but not only if, A and

d indicated that this magnitude of error was acceptable (Table 4.4). Because of the time constraints mentioned in Section 3.2, the method described there was repeated three times for each PTV ($E = \{1,2,3\}$). The resultant Δ_{mean} (equation (3.1)) and pass rates were recorded and averaged (Section 4.4.1 and Section 4.4.2). These results were compared to that of the main experiment and skull experiment (Section 3.3.1 and Section 3.4). The MU values were also documented (Appendix A.7).

3.5.1 Pelvis sCT

Prostate cancer is the most common type of cancer in men and is often treated with radiotherapy. To simulate cases of prostate cancer, the prostate in the sCT was approximately located by comparing with images in an online source showing the prostate in CT [43]. Once the prostate had been located, one of the bones close by, the right-sided *pubis* [44], was replaced by air, as shown in Figure 3.12. The average two-dimensional size of this bone, spanning the slices covering the PTV and the margins, was $A \approx 7.5\,\text{cm}^2$. Two PTVs were placed at distances $d \approx 4.0\,\text{cm}$ and $d \approx 7.0\,\text{cm}$ between the PTV centres and the misclassification volume centre. The dark grey voxels were avoided when placing the PTVs as the CT value in these pixels are that of fat (-75 HU).

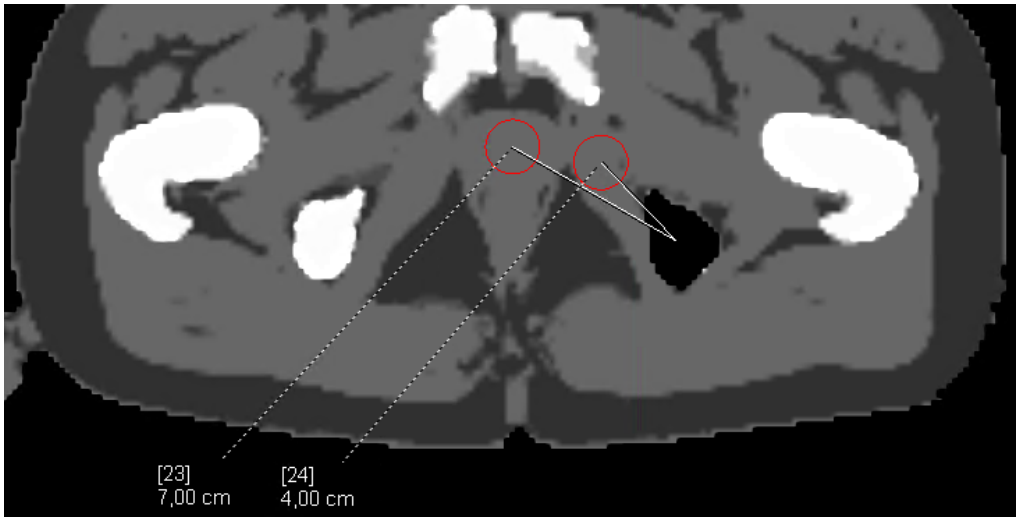


Figure 3.12: The pelvis sCT image showing two PTVs (red circles) in the prostate area, the centres of which are at approximate distances 4.0 cm and 7.0 cm from the approximated centre of a misclassified pubic bone. Screenshot from *syngo*.via.

To the left in Figure 3.12, an arm is visible. This is because the arms were lying down the side of the body during the scan. During radiotherapy of the pelvis, patients lie with their arms in the opposite direction in order to avoid giving radiation to the arms and improving the beams coming in from the sides. To avoid the arms being a part of the dose calculation, the arms were removed from the body volume that automatically generates in the contouring tool in the TPS.

3.5.2 Head sCT

The sCT image of the head was edited twice. First; replacing a segment of the skull of two-dimensional size $A \approx 4.4\,\mathrm{cm}^2$ with air. This is interesting because of the shape of the misclassification volume. In the pelvic sCT, the misclassification volume had a cylindrical-like shape, like in the primary experiment. In this experiment, the misclassification volume had a much flatter shape, meaning the misclassification volume occluded the PTV to a greater extent in relation to its size. Here, the length of the photon path through the misclassification volume was shortened. In this way, comparisons with the pelvic sCT experiments could be made in order to observe the effect of the misclassification volume's shape. Upon the second iteration of editing, in order to emulate the experiment performed in Section 3.4, the entire skull was replaced with air. The experiment in Section 3.4 showed no clinically acceptable results for the DVH measurements (Section 4.3). However, as the CT values of bone are significantly lower in the sCT images, a new experiment with a fully misclassified skull was performed. In both the partial and full skull misclassification, the automatic body segmenting of the edited images in the TPS failed to accurately delineate the body, as shown by the green line in Figure 3.13. This was easy to correct by copying the body volume from the sCT image with no induced misclassification. However, it is important to note that this would not be as easy with no such error-free image. For bone misclassifications close to the skin, this presents a reoccurring problem in the TPS.

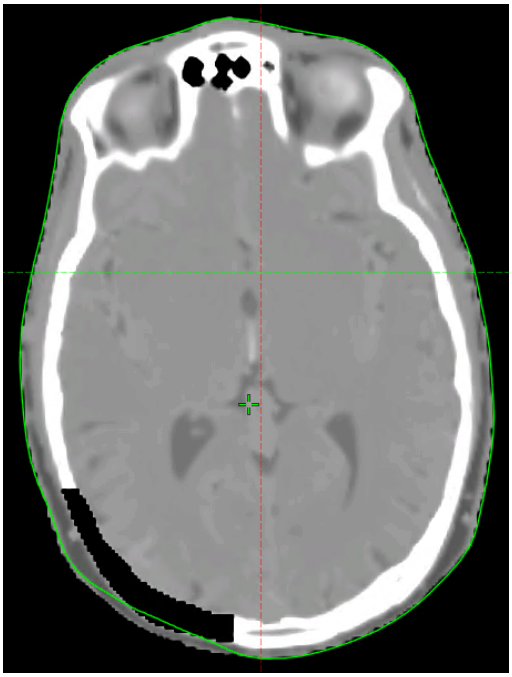


Figure 3.13: Contour of the body (green line) error at misclassified volume close to the surface of the skin. Screenshot from TPS.

For the first simulated misclassification, two PTVs were placed, with their centres at distances approximately 3.0 cm and 4.0 cm away from the approximated centre of the bone misclassification volume, as shown in Figure 3.14.

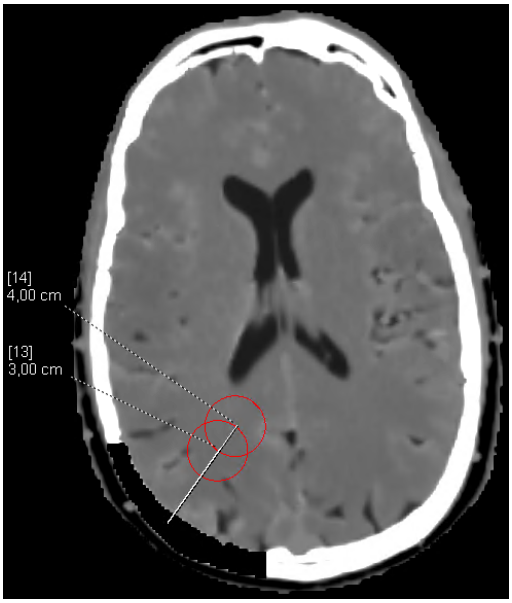


Figure 3.14: The head sCT image showing two PTVs (red circles) in the brain, the centres of which are at approximate distances 3.0 cm and 4.0 cm from the approximated centre of the misclassified piece of the skull. Screenshot from *syngo*.via.

To emulate the experiment in Section 3.4, using the sCT image with the entire skull replaced by air, two PTVs were placed in the skull at distances 2.3 cm and 8.8 cm from the inner edge of the back of the skull, as shown in Figure 3.15.

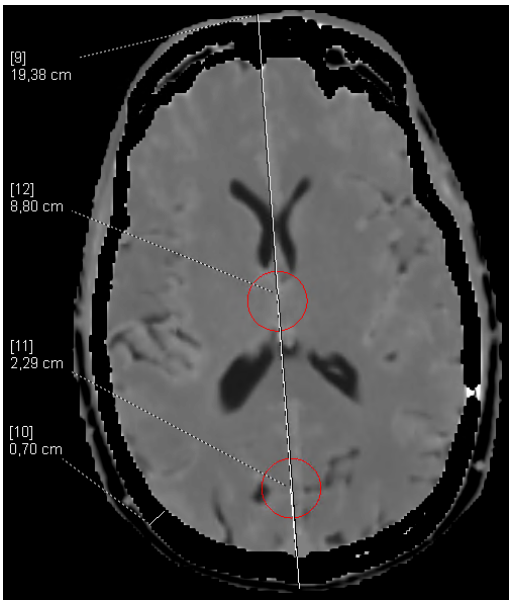


Figure 3.15: The head sCT image showing two PTVs (red circles) in the brain, the centres of which are at approximate distances 2.3 cm and 8.8 cm from the inner edge of the back of the misclassified skull. The back-to-front length of the skull and the length of 0.7 cm is also shown. Screenshot from *syngo*.via.

From Figure 3.15, it is apparent that the approximations of skull size and thickness made in Section 3.4 were fitting. However, the CT values making up

the brain tissue are slightly higher than that of water. Therefore, the PTVs and surrounding area in these tests have CT values of approximately 30 HU, with some exceptions. The cerebrospinal fluid (large dark grey areas in Figure 3.15) has CT values close to 0 HU.

4 Results

Here, the results of the experimental method described in Section 3 are given, featuring the variables S, A, and d, illustrated in Figure 3.2. First and foremost, this embroils the *lines of acceptance* created by statistical analysis of the DVH data. One such line is presented for each PTV size, S, that was tested, along with an approximate 95 % confidence band. Similar results for the gamma evaluation data are presented later. Thereafter, the results of the smaller experiments conducted are given.

4.1 DVH - lines of acceptance

In this section, the results of the data analysis, described in Section 3.3.1, on the DVH data (Appendix A.2) are presented.

Figure 4.1 shows the final line of acceptance; the line for which points above represent an 800 HU bone misclassification error too great to be acceptable ($\Delta_{mean} > 2\,\%$, red zone) and the line's approximate 95 % confidence band, for the case where $S = 1.5\,\text{cm}$. The error bars represent the 95 % confidence interval of the points $I(d, A)$ ($\alpha = 0.0025$). Figure 4.2 and Figure 4.3 presents the same data, though for $S = 2.0\,\text{cm}$ and $S = 2.5\,\text{cm}$, respectively.

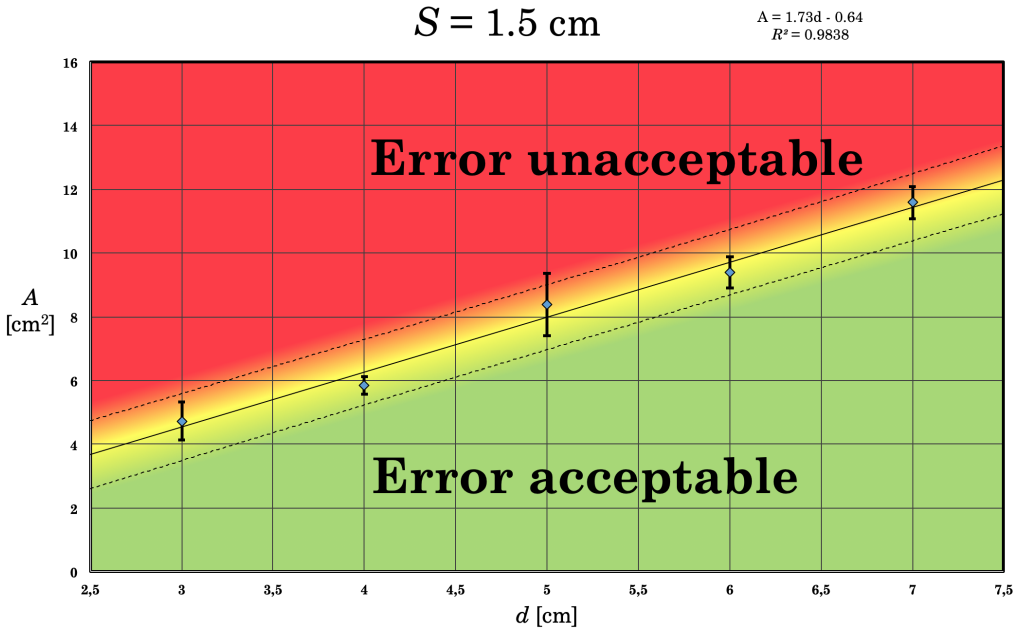


Figure 4.1: The line resulting from linear regression of the points $I(d, A)$ and surrounding 95 % confidence band (dotted lines) for $S = 1.5\,\text{cm}$. The error bars connected to $I(d, A)$ represent the 95 % confidence intervals of $I(d, A)$. The red zone represents combinations of A and d that results in a Δ_{mean} above 2 % and the green zone represents the combinations that give values of Δ_{mean} below 2 %.

Table 4.1 shows the numerical data in Figure 4.1.

Table 4.1: The values of A and d making up the points $I(d, A)$ in Figure 4.1, where $S = 1.5\,\text{cm}$. The 95 % confidence intervals of A are also given.

d [cm]	$A \pm t_{\alpha,\nu}\sigma_I$ [cm^2]
3.0	4.73 ± 0.60
4.0	5.85 ± 0.26
5.0	8.38 ± 0.97
6.0	9.39 ± 0.48
7.0	11.59 ± 0.50

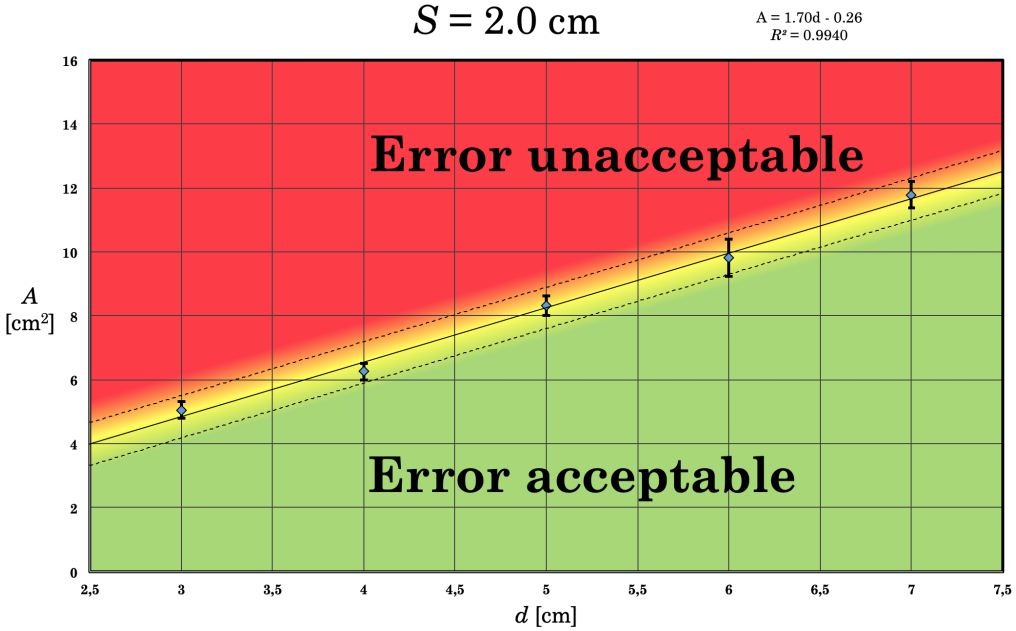


Figure 4.2: Same as Figure 4.1, except for $S = 2.0\,\text{cm}$.

Table 4.2 shows the numerical data in Figure 4.2.

Table 4.2: The values of A and d making up the points $I(d, A)$ in Figure 4.2, where $S = 2.0\,\text{cm}$. The 95 % confidence intervals of A are also given.

d [cm]	$A \pm t_{\alpha,\nu}\sigma_I$ [cm^2]
3.0	5.06 ± 0.27
4.0	6.25 ± 0.26
5.0	8.32 ± 0.30
6.0	9.81 ± 0.58
7.0	11.79 ± 0.42

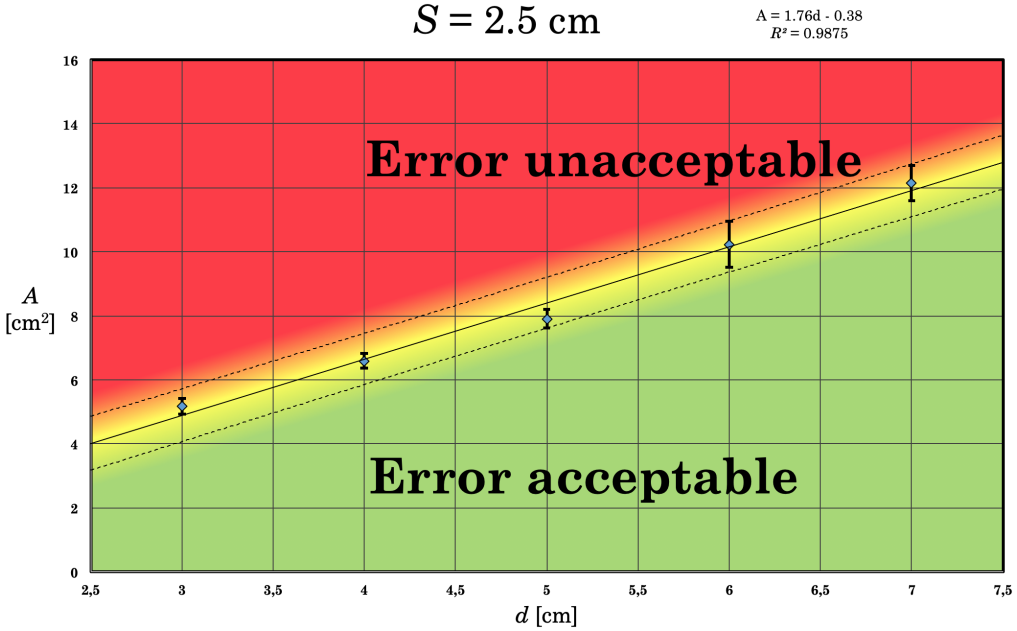


Figure 4.3: Same as Figure 4.1, except for $S = 2.5\,\text{cm}$.

Table 4.3 shows the numerical data in Figure 4.3.

Table 4.3: The values of A and d making up the points $I(d, A)$ in Figure 4.3, where $S = 2.5\,\text{cm}$. The 95 % confidence intervals of A are also given.

$\boldsymbol{d}$ [cm]	$\boldsymbol{A \pm t_{\alpha,\nu}\sigma_I}$[cm^2]
3.0	5.17 ± 0.25
4.0	6.59 ± 0.22
5.0	7.90 ± 0.28
6.0	10.23 ± 0.73
7.0	12.14 ± 0.55

The resultant equations and values of R^2 for each S is given in Table 4.4.

Table 4.4: The equations describing the linear relationship of A and d for each S. R^2, indicating the fit of the linear model to the data, is also given.

$\boldsymbol{S}$ [cm]	Equation	$\boldsymbol{R^2}$
1.5	$A = 1.73d - 0.64$	0.9838
2.0	$A = 1.70d - 0.26$	0.9940
2.5	$A = 1.76d - 0.38$	0.9875

With an average R^2 of approximately 0.9884 all lines of acceptance in this section indicate a good fit of the points $I(d, A)$ to the linear model.

Figure 4.4 shows all of the acceptance lines above in the same diagram. The background colour matches that of Figure 4.2.

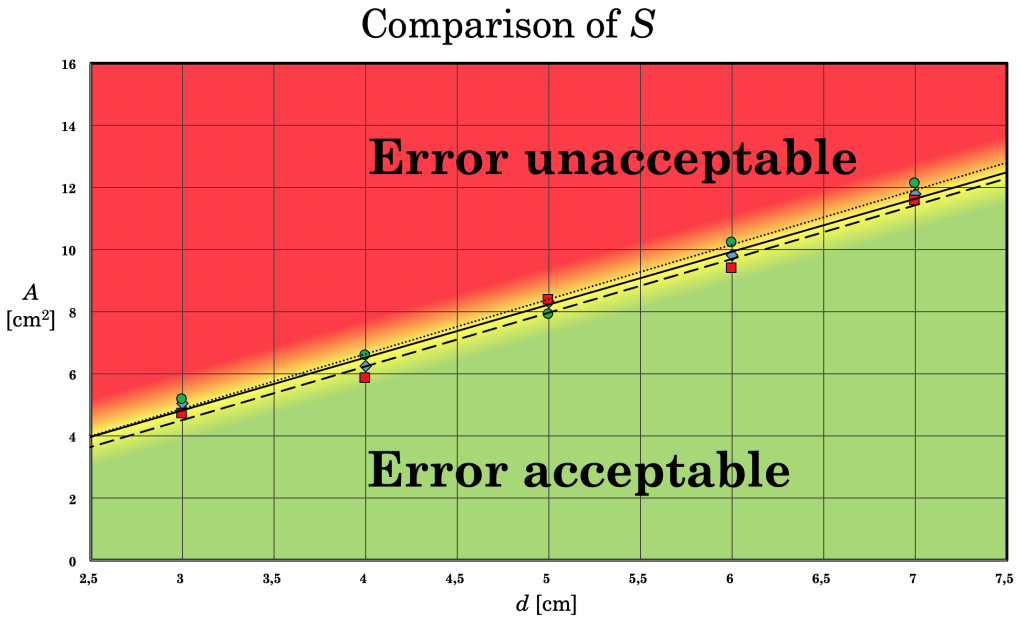


Figure 4.4: Comparison of regression lines where $S = 1.5\,\text{cm}$ (red squares, staggered line), $S = 2.0\,\text{cm}$ (blue squares, unbroken line), and $S = 2.5\,\text{cm}$ (green circles, dotted line). The coloured background is identical to that in Figure 4.2. The red zone represents combinations of A and d that results in a Δ_{mean} above $2\,\%$ and the green zone represents the combinations that give values of Δ_{mean} below $2\,\%$, for the case where $S = 2.0\,\text{cm}$.

As described in Section 3.3.1, the Chow test was performed in order to test for statistically significance of the change in the regression lines for differing S. Table 4.5 shows the result of these tests. None of the regression lines significantly ($p < 0.05$) differ from one another. The null hypothesis that all lines are equal cannot be rejected, meaning this thesis has found no significance of the volume changes, with the spherical PTVs' diameters ranging from 1.5 cm to 2.5 cm.

Table 4.5: p-values for the different combinations of S. There is no statistically significant difference between the lines.

Combination	p
$\boldsymbol{S = 1.5\ \text{cm} + S = 2.5\ \text{cm}}$	0.29
$\boldsymbol{S = 1.5\ \text{cm} + S = 2.0\ \text{cm}}$	0.50
$\boldsymbol{S = 2.0\ \text{cm} + S = 2.5\ \text{cm}}$	0.67

4.2 Gamma evaluation - lines of acceptance

Similarly to Section 4.1, the results of the data analysis, described in Section 3.3.2, on the gamma pass rate data (Appendix A.2) are presented. The difference in this section is that the coloured backgrounds of the plots do not correspond to the data in the plot, but to the results in Section 4.1. This is for the purpose of visually comparing the DVH and gamma evaluation results.

Figure 4.5 shows the final line of acceptance; the line for which points above represent a 800 HU bone misclassification error too great to be acceptable (gamma pass rate $< 95\,\%$) and the line's $95\,\%$ confidence band, for the case where $S = 1.5\,\text{cm}$. The error bars represent the $95\,\%$ confidence interval of the points $I(d, A)$ ($\alpha = 0.0025$). Figure 4.6 and Figure 4.7 presents the same data, though for $S = 2.0\,\text{cm}$ and $S = 2.5\,\text{cm}$, respectively.

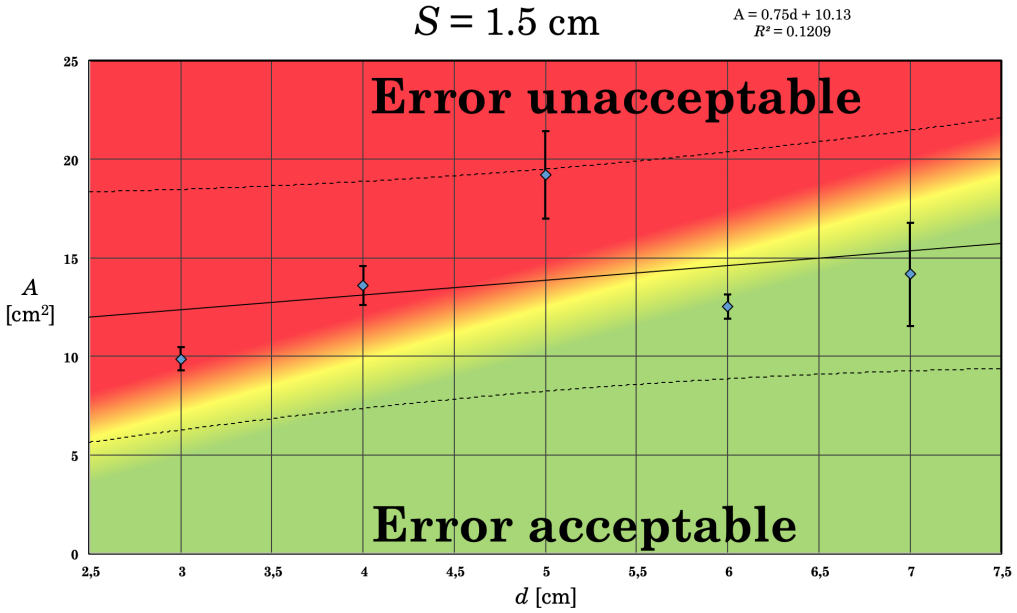


Figure 4.5: The line resulting from linear regression of the points $I(d, A)$ and surrounding $95\,\%$ confidence band (dotted lines) for $S = 1.5\,\text{cm}$. The error bars connected to $I(d, A)$ represent the $95\,\%$ confidence intervals of $I(d, A)$.

Table 4.6 shows the numerical data in Figure 4.5.

Table 4.6: The values of A and d making up the points $I(d, A)$ in Figure 4.5, where $S = 1.5\,\text{cm}$. The 95 % confidence intervals of A are also given.

d [cm]	$A \pm t_{\alpha,\nu}\sigma_I$ [cm^2]
3.0	9.88 ± 0.57
4.0	13.59 ± 1.00
5.0	19.20 ± 2.22
6.0	12.52 ± 0.62
7.0	14.16 ± 2.63

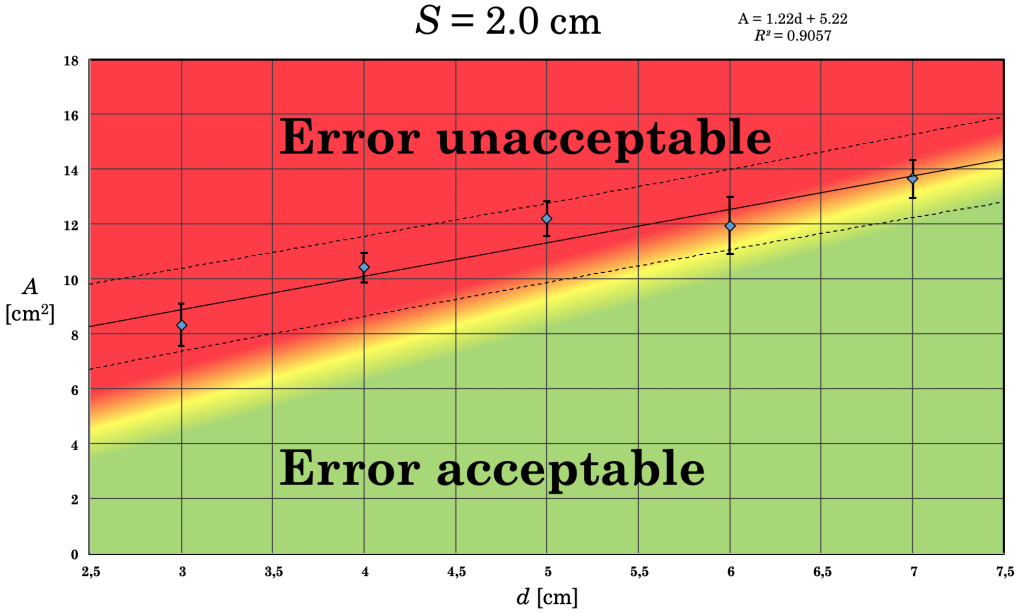


Figure 4.6: Same as Figure 4.5, except for $S = 2.0\,\text{cm}$.

Table 4.7 shows the numerical data in Figure 4.6.

Table 4.7: The values of A and d making up the points $I(d, A)$ in Figure 4.6, where $S = 2.0\,\text{cm}$. The 95 % confidence intervals of A are also given.

d [cm]	$A \pm t_{\alpha,\nu}\sigma_I$ [cm^2]
3.0	8.34 ± 0.77
4.0	10.43 ± 0.54
5.0	12.21 ± 0.65
6.0	11.95 ± 1.03
7.0	13.67 ± 0.69

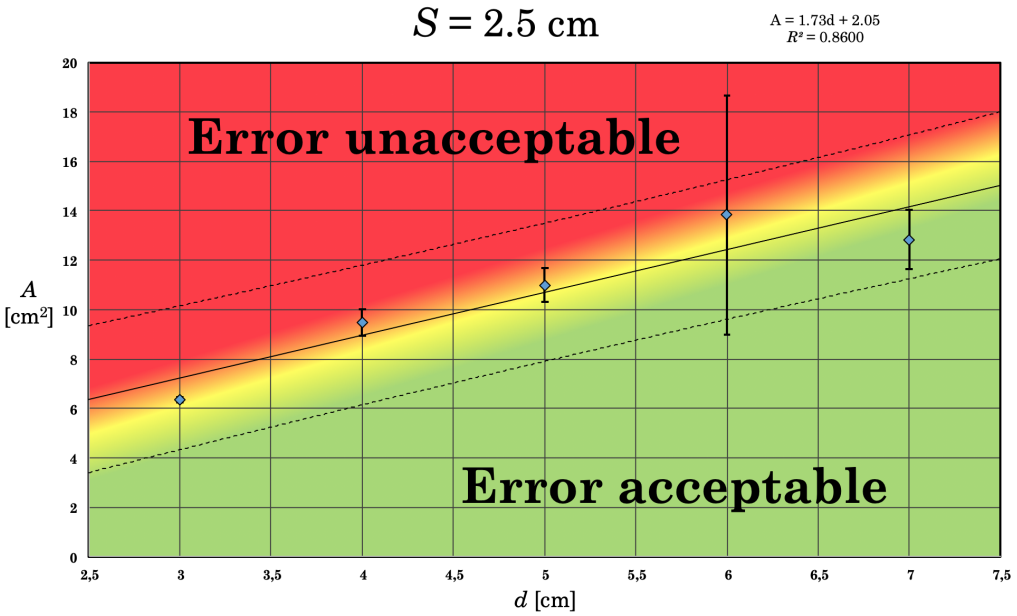


Figure 4.7: Same as Figure 4.5, except for $S = 2.5\,\text{cm}$.

Table 4.8 shows the numerical data in Figure 4.7.

Table 4.8: The values of A and d making up the points $I(d, A)$ in Figure 4.7, where $S = 2.5\,\text{cm}$. The 95 % confidence intervals of A are also given.

$\boldsymbol{d}$ [cm]	$\boldsymbol{A \pm t_{\alpha,\nu}\sigma_I}$[cm^2]
3.0	6.36 ± 0.10
4.0	9.48 ± 0.54
5.0	10.99 ± 0.69
6.0	13.83 ± 4.84
7.0	12.83 ± 1.21

The resultant equations and values of R^2 for each S is given in Table 4.9.

Table 4.9: The equations describing the linear relationship of A and d for each S. R^2, indicating the fit of the linear model to the data, is also given.

$\boldsymbol{S}$ [cm]	Equation	$\boldsymbol{R^2}$
1.5	$A = 0.75d + 10.13$	0.1209
2.0	$A = 1.22d + 5.22$	0.9057
2.5	$A = 1.73d + 2.05$	0.8600

For $S = 1.5\,\text{cm}$, it is apparent that this data does not fit well with the linear model, nor with the corresponding DVH data case (Figure 4.1). For $S = 2.0\,\text{cm}$, the data fits better with the linear model than for $S = 1.5\,\text{cm}$. However, for $d < 7.5\,\text{cm}$, the line of acceptance is higher (less strict) than that of the corresponding DVH data case (Figure 4.2). For $S = 2.5\,\text{cm}$, the line of acceptance fits better with the corresponding DVH data case (Figure 4.3).

Figure 4.8 shows all of the lines above in the same diagram. The background matches that of Figure 4.6.

Figure 4.8: Comparison of regression lines where $S = 1.5\,\text{cm}$ (red squares, staggered line), $S = 2.0\,\text{cm}$ (blue squares, unbroken line), and $S = 2.5\,\text{cm}$ (green circles, dotted line). The coloured background is identical to that in Figure 4.6.

The Chow test was performed to test for statistically significant difference of the regression lines. Table 4.10 shows the result of the Chow tests (Section 2.8.2). None of the regression lines significantly ($p < 0.05$) differ from one another. The null hypothesis that all lines are equal cannot be rejected, meaning this thesis has found no significance of the volume changes in the gamma evaluation data.

Table 4.10: p-values for the different combinations of S. There is no statistically significant difference between the lines.

Combination	p
$S = 1.5$ cm + $S = 2.5$ cm	0.22
$S = 1.5$ cm + $S = 2.0$ cm	0.36
$S = 2.0$ cm + $S = 2.5$ cm	0.40

Using the Chow test, the lines of acceptance in the gamma evaluation case were compared to the corresponding lines in the DVH case. Table 4.11 shows the resultant p-values. For all three S, statistically significant differences between lines were observed.

Table 4.11: p-values from the Chow test, testing for difference in the acceptance lines between the gamma evaluation and DVH data for all S. Statistically significant differences were observed for all lines.

S [cm]	$\boldsymbol{p}$
1.5	0.03
2.0	0.0003
2.5	0.02

For spherical PTVs with diameters smaller or equal to 2.5 cm, the gamma evaluation method, with gamma criteria of 2 %/2 mm and a dose cutoff of 20 %, does not perform well enough in signalling mean PTV dose deviations larger than 2 %.

4.3 Skull experiment

In this section, the results of the skull experiment (Section 3.4) are presented. Table 4.12 shows the average Δ_{mean} and gamma pass rates for the six different PTV distances, x. Because of the instrument precision in the TPS (Section 3.3.1) the measurements of Δ_{mean} have an uncertainty of approximately 0.03 %.

Table 4.12: Mean measured values of Δ_{mean} and gamma pass rates for each of the six PTVs at distance x.

$\boldsymbol{x}$ **[cm]**	**Mean** Δ_{mean} **[%]**	**Mean pass rate [%]**
2.3	5.18	83.53
3.6	5.06	88.08
4.9	4.97	94.81
6.2	4.84	94.97
7.5	4.76	95.03
8.8	4.72	94.97

None of the values in Table 4.12 are clinically acceptable, with the exception of the gamma pass rate at $x = 7.5$ cm. However, a larger average Δ_{mean} is observed for PTVs closer to the inner edge of the simulated skull, indicating a worsening of Δ_{mean} for decreased distance to areas of misclassification. This

behaviour fits well with previous results (Section 4.1). Nonetheless, misclassification of an entire skull with a CT value of 800 HU is not clinically acceptable.

4.4 Synthetic CT scenarios

In this section, the results of the experiments described in Section 3.5 are given. The values given here were found through averaging the measured values listed in Appendix A.4. Here, two 2.0 cm diameter PTVs were placed in a pelvic sCT image of the author to simulate prostate cancer treatment with bone misclassification. Similar experiments were conducted to simulate brain tumour treatment with partial and entire skull misclassification.

4.4.1 Pelvis sCT

The table in this subsection provides the results of the experiment in Section 3.5.1 for the two distances, d. The two-dimensional size of the misclassified bone was $A \approx 7.5\,\mathrm{cm}^2$.

Table 4.13: Mean measured values of Δ_{mean} and gamma pass rates for the two distances, d. Here, $A \approx 7.5\,\mathrm{cm}^2$.

$\boldsymbol{d}$ [cm]	Mean Δ_{mean} [%]	Mean pass rate [%]
4.0	1.64	99.63
7.0	0.72	100

For $d = 4.0\,\mathrm{cm}$, one expects a mean Δ_{mean} above 2.0 %, using the equation in Table 4.4 for $S = 2.0\,\mathrm{cm}$ and assuming bone CT value of 800 HU. As the mean CT value is lower than 800 HU in this experiment, a smaller Δ_{mean} is expected and observed. However, an increase in Δ_{mean} is still observed for a decrease in d, as one would expect.

4.4.2 Head sCT

The two tables in this subsection provide the results of the two experiments in Section 3.5.2. Table 4.14 shows the result of the experiment with the partial skull bone misclassification ($A \approx 4.4\,\mathrm{cm}^2$), while Table 4.15 shows the results of the entire skull misclassification experiment.

Table 4.14: Mean measured values of Δ_{mean} and gamma pass rates for the two distances, d, with partial skull bone misclassification. Here, $A \approx 4.4\,\mathrm{cm}^2$.

d [cm]	Mean Δ_{mean} [%]	Mean pass rate [%]
3.0	1.52	97.22
4.0	1.21	96.81

Once more, an increase in Δ_{mean} is observed for a decrease in d, as one would expect. However, an increase in gamma pass rate for a decrease in d is observed as well.

As was the case in the results of the first skull experiment (Table 4.12), an increase in mean Δ_{mean} was observed in Table 4.15 for a decrease in x. As the mean CT value of the misclassified bone is lower than 800 HU in this experiment, lower values of Δ_{mean} compared to that of the experiment described in Section 3.4 were expected.

Table 4.15: Mean measured values of Δ_{mean} and gamma pass rates for the two distances, x, with entire skull bone misclassification.

x [cm]	Mean Δ_{mean} [%]	Mean pass rate [%]
2.3	4.10	87.39
8.8	3.54	95.82

5 Discussion

As expected from the results of the project thesis [12], the DVH measurements show a lower D_{mean} when bone misclassifications occur (Appendix A.2, Appendix A.3, Appendix A.4). The drastically lowered electron density of the misclassification volumes in the planning CT/sCT images causes less attenuation to be planned for in these areas. Thus, during radiotherapy, the radiation beams have lower intensities after traversing the misclassification volumes than what is expected from the planning process, resulting in the observed decrease in D_{mean}. Naturally, a decrease in dose to the PTV is very undesirable, as it might affect chances of patient survival. This effect will be observed if and only if the treatment plan consists of beams that traverse a misclassification volume. The VMAT method is certain to be affected, assuming the misclassification volume is in the same sCT slice as the PTV. However, other methods might be affected differently by such errors dependent on the relative amounts of the treatment beams that traverse the misclassification volumes. Therefore, the results in this thesis only hold for the specific VMAT method described in Section 3.2.

The primary goal of this thesis was to find equations describing the limit of acceptable misclassification errors dependent on PTV size S, bone area A, and distance d (Figure 3.2), using the PTV dose deviation criterion of 2 %. Three such equations (for the three different S) are listed and plotted along with their approximate 95 % confidence bands, in Section 4.1, on the form: $A = ad + b$. These equations provide an easy-to-use tool for radiographers, radiologists, and/or medical physicists in evaluating whether a misclassification error of approximate two-dimensional size, A, and which centre is at a distance, d, to the centre of the PTV, is acceptable. For all $A < ad+b$, the misclassification error is acceptable, assuming the described VMAT plan, bone-to-air misclassification, and average CT value of the misclassified bone $\leq$ 800 HU. This thesis does not take bone-atlas misplacement into consideration [10]. As mentioned in Section 4.4.2, misclassifications of bone close to skin might create problems with the automatic body contouring in the TPS, meaning a new scan might be the more efficient choice in these cases.

The project thesis [12] showed slightly less strict results for $S = 2.0\,\text{cm}$ and $d = 5.0\,\text{cm}$ when approximating Δ_{mean} in two-dimensional dose matrices. In the project thesis, the 2 % PTV dose deviation line was crossed at $A \approx 8.9\,\text{cm}^2$. The corresponding result in this thesis is $A \approx 8.2\,\text{cm}^2$ (Table 4.4). However, in addition to not utilising real DVH points, the project thesis had some key differences in the experimental set-up. Most importantly; it did not use separately *optimised* treatment plans for each A. The same treatment plan was used, with the same MU value, for all experiments. This thesis, more realistically, went through the optimisation process for each combination of S, A, and d (Section 3.2). This might explain the $0.7\,\text{cm}^2$ difference in results.

The values of R^2 for the lines in Section 4.1 show good fits to linear models. The same is indicated for the lines created to find the data points, $I(d, A)$, for the final lines. The average value of R^2 shown in Table A.13 (Appendix A.8.1) is 0.9650. However, the best fitted lines are found for $S = 2.0\,\text{cm}$, where the average R^2 is 0.9824. This has resulted in a lower $\bar{\sigma}_I$ than the other PTV sizes. Thus, the thinnest 95 % confidence band is observed for $S = 2.0\,\text{cm}$. Considering the widths of the confidence bands and the observed proximity of the lines of acceptance in Figure 4.4, it is no surprise that the Chow test found no significant difference between the lines for PTV size variables $S = 1.5\,\text{cm}$ through

$S = 2.5\,\text{cm}$. However, an even larger difference in S and/or more data creating smaller uncertainties might have resulted in a statistically significant difference. It is logical to assume that a larger PTV, which would have a smaller percentage of volume that is occluded by the misclassification volume, would be less affected by the misclassification. In theory (Section 2.8.2), as there is no statistically significant difference between the regression lines, one combined regression line found from linear regression of all values in Table 4.1, Table 4.2, and Table 4.3 ($A = 1.73d - 0.43$) could have been given as the main result in this thesis. However, because of the known structural difference in the data (varying S) and assumed significance of S at larger differences, this author found it fitting to provide the results for each PTV size. In hindsight, values of S larger than 2.5 cm and smaller than 1.5 cm should have been tested. A maximum of $S = 2.5\,\text{cm}$ was chosen because of the number of obtainable data points; with the experimental set-up shown in Figure 3.2, an even larger S would have resulted in fewer possible data points because of the PTV overlapping the closest misclassification volumes (Section 3.1). This problem is seen when comparing Table A.1 and Table A.3 (or from Table A.8), where it is apparent that the statistical analysis of the $S = 1.5\,\text{cm}$ case utilised a total of 87 Δ_{mean} data points, while the analysis of the $S = 2.5\,\text{cm}$ case utilised a total of 81 Δ_{mean} data points.

The shape of the misclassification volume, logically, influences Δ_{mean}. The length, L, of traversed matter with attenuation coefficient, μ, for a beam of photons is important for the intensity of the beam, as seen in equation (2.7). Thus, in the skull experiment (Section 3.4), where $L = 0.7\,\text{cm}$ and $A = 40.2\,\text{cm}^2$, one would expect a higher beam intensity on the inner side of the skull than on the inner side of a cylindrical bone misclassification of the same size, as L would have a maximum of 7.15 cm. This might explain why the skull experiment *only* showed values of Δ_{mean} close to 5 % despite a very large A and full occlusion of the PTV (Section 4.3). The choice of a cylindrical misclassification volume in the primary experiment therefore represents a *worst case scenario*, as flatter bone misclassifications have reduced effects on the beam intensity with respect to their sizes, A. The pelvic sCT experiment and partial skull sCT experiment contained two differently shaped misclassification volumes. The pelvic sCT experiment had a cylinder-like misclassification volume, while the partial skull sCT experiment had a curved plate-like misclassification volume. Using the results of the primary experiments for $S = 2.0\,\text{cm}$ (Table 4.4), one expects a slightly larger Δ_{mean} at $d = 4.0\,\text{cm}$ in the sCT pelvis experiment, illustrated in Figure 5.1, than for $d = 3.0\,\text{cm}$ in the sCT partial skull experiment, illustrated in Figure 5.2. One also expects a larger difference in Δ_{mean} between the two points in Figure 5.1 than between the two points in Figure 5.2. The observed results in Table 4.13 and Table 4.14 meet both of these expectations, indicating that the shape of the misclassification volume has not had a substantial effect. However, this comparison of four averaged measurements is not satisfactory to fully explore the effects of misclassification volume shape. This should be explored in a separate study. It is also important to note that the misclassification volume is much closer to the external surrounding air in the sCT partial skull experiment than in the primary experiment, meaning there is less dose build-up and attenuation before the beam reaches the misclassification volume. The effects of this in relation to the results of this thesis are unknown.

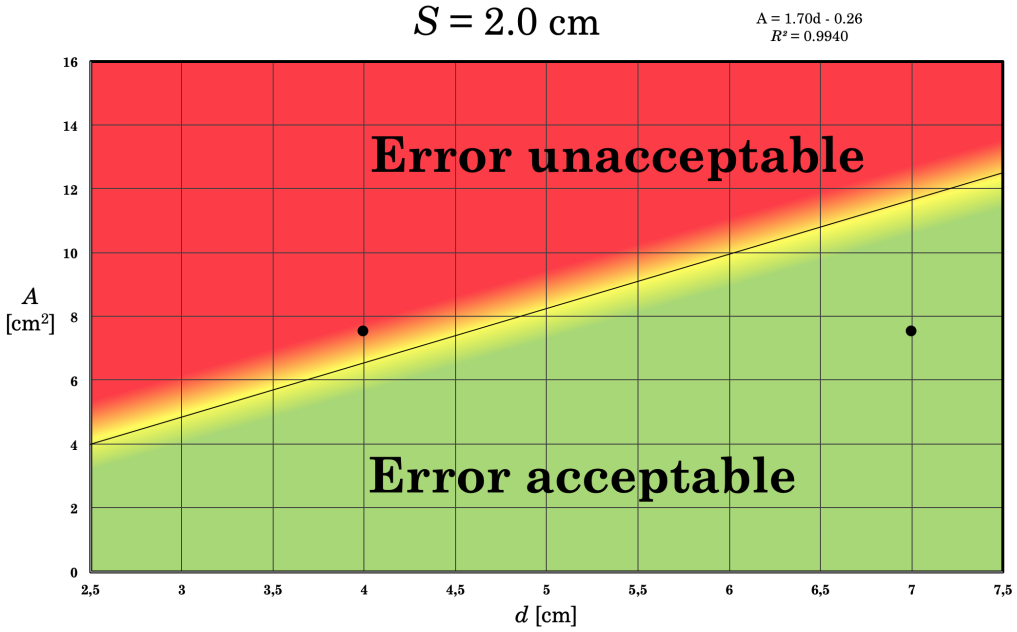


Figure 5.1: The points (4.0, 7.5) and (7.0, 7.5) in the sCT pelvis experiment (Section 3.5.1) plotted with the line in Figure 4.2.

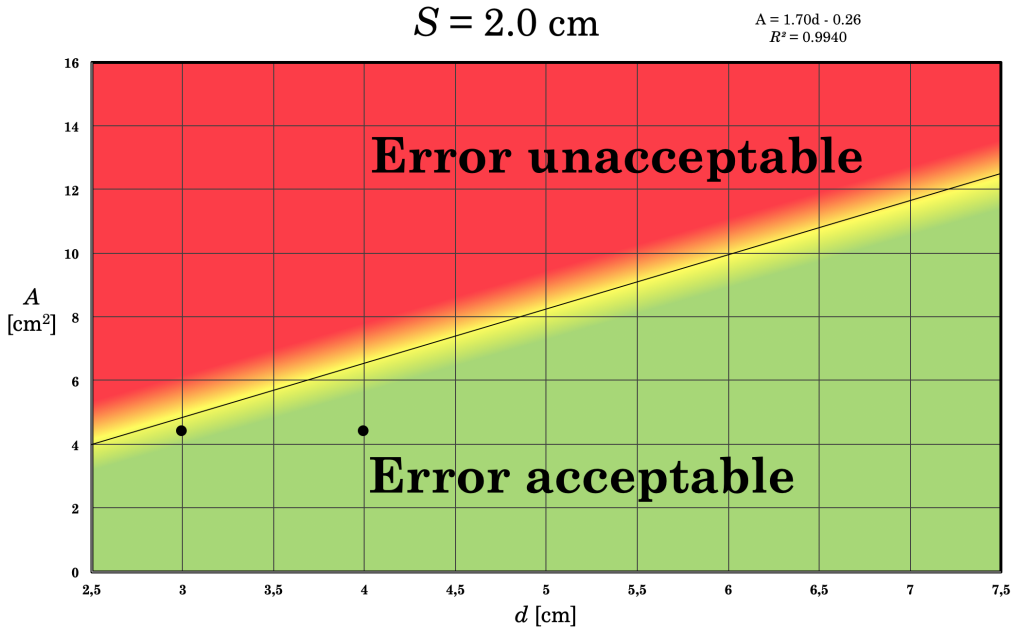


Figure 5.2: The points (3.0, 4.4) and (4.0, 4.4) in the sCT partial skull experiment (Section 3.5.2) plotted with the line in Figure 4.2.

Another important choice made in these experiments, as observed in Section 4.4, is the CT value (HU) of the misclassified bone. From the line of acceptance in Figure 5.1 and Figure 5.2, one expects larger values of Δ_{mean} in Table 4.13 and Table 4.14, respectively. The point (4.0, 7.5) in Figure 5.1 should have an average Δ_{mean} above 2 % in Table 4.13. The same behaviour of mean Δ_{mean} is observed when comparing the clinically unacceptable results of the skull experiment (Table 4.12) to the corresponding sCT experiment (Table 4.15). However, this can be attributed to the relatively low CT value of

bone in the sCT images ($\sim$ 400 HU). Chronologically, the primary experiment was performed long before the sCT images were created; thus the average CT values of bone in the sCT images were not known when choosing 800 HU for the bone tissue. Due to the coronavirus pandemic, the sCT images were not created until the beginning of May 2020, at which point there was too little time to make changes to the CT values of bone in the other experiments. However, the average CT values of real cases of sCT bone misclassification is not known. Choosing 400 HU for the experiments could have been proven to be too low if bone misclassifications generally consist of larger parts cortical bone and smaller parts cancellous bone. If only cortical bone is subject to misclassification in real sCT images, 800 HU is too low instead of too high, meaning this thesis has not taken the worst case scenario into consideration when choosing the CT value. More data on the expected CT values in real misclassification cases should be gathered before these experiments are repeated for another CT value of bone. It is also not known whether real misclassifications always put air in place of bone. Cortical and cancellous bone might be misclassified as water instead (0 HU), meaning larger and closer misclassifications would be acceptable. Again, these experiments have considered the *worst case scenario* by using air (-1000 HU) instead of water (0 HU, Section 2.4). Using the worst case scenarios in these experiments is natural, as one must avoid mistakenly accepting sCT images featuring too large errors. One logical conclusion that can be drawn from the comparisons of the results of the primary experiment and sCT experiments is that the equations in Table 4.4 should provide the limits for acceptable bone misclassification for misclassified bones of CT value $\leq$ 800 HU.

One point of interest, in regards to the CT values in the images, is that the original CT scan in the primary experiment was performed at 100 kV maximum energy (Section 3.1). For high density tissue, such as bone, the CT value varies dependent on the energy and the scanner [45]. The CT value of bone tissue of known electron density relative to water, RED, will be higher in a 100 kV CT image compared to in a 120 kV image, as seen in Figure 2.9. Using this figure, one would estimate $RED \approx 1.4$ for 800 HU at 100 kV, and $RED \approx 1.5$ at 120 kV. This is an approximate 6.67 % RED error. Energy-dependent HU-to-RED conversion is required to accurately obtain electron density from a CT image. According to Constantinou et al. [46], a 5 % error in RED in a voxel leads to approximately 1 % error in dose. However, in order to avoid these hassles at UNN, all CT scans for the purpose of radiotherapy planning are performed at 120 kV, and only this curve is utilised in the TPS. Therefore, even though the original scan in the primary experiment and skull experiment was performed at 100 kV, the TPS has treated the CT values in the images as though they were taken at 120 kV. Thus, the difference in CT energy between the primary experiment (100 kV) and the sCT experiments (120 kV) was inconsequential for the dose calculation in this thesis. However, for the purpose of reproducibility, the original CT scan should be performed at 120 kV when using a TPS that utilises different conversion curves for different energies.

The changes in D_{mean} indicate the bone misclassification consequences for the PTVs. In order to quantify the consequences in the entire irradiation volume, including possible OARs, the gamma evaluation is utilised. However, this thesis observed less strict limits of acceptance using the gamma evaluation data (Section 4.2), as was predicted from the results of the project thesis [12] (Section 1). The gamma evaluation and criteria used in this thesis are not sufficient to highlight unacceptable PTV dose deviations. The gamma evaluation data also

have much worse fits to the linear models, even with the steps taken in order to improve the fits in Section 3.3.2. The average value of R^2 in Table A.14 is 0.7730, with one value as low as 0.1470 (Appendix A.8.2). As a larger PTV will cover a greater number of pixels in the dose matrices, a change in the gamma evaluation data dependent on S was expected. However, no statistically significant difference was observed (Section 4.2). This might be due to the large uncertainties in the gamma evaluation regression lines. The gamma evaluation data might have been worsened by one particular issue that can be seen in Figure 3.7b. For some dose matrices, the exported 10 x 10 cm plane matrices were too small to include the entirety of the pixels that were above the 20 % cutoff. This affects the number of pixels that should have been tested and, by extension, the gamma pass rates (Section 4.2).

To ensure accurate error detection by the DVH measurements and gamma evaluation, it is important that the dose matrices are correctly calculated. At UNN, calculation is done using an analytical anisotropic algorithm (AAA). This algorithm is considered better than pencil beam models, but inferior to Monte Carlo simulations [47]. Monte Carlo simulations, however, have long computation times, making them impractical for clinical use. Using Monte Carlo simulation models to calculate the dose matrices would have given a better representation of the physical consequences of misclassification, but it would not have given an accurate representation of the errors that would have been detected by the gamma evaluation in a clinic that uses an AAA.

It is important to note that there is no misregistration in this experiment; the *anatomy* of the water sphere in the sets of CT images lines up perfectly, ensuring that the differences that are detected in the experiment are solely due to the misclassification error of interest. This is not clinically realistic, as both external and internal movement will occur when the patient is moved from the MRI scanner to the LINAC, despite many efforts to keep everything in the same position. When investigating the misregistration of the prostate, which is an area often in need of superior soft tissue contrast (MRI) for delineation, Roberson et al. [48] concluded that the minimum axial MR to axial CT registration error is approximately 2 mm. Such an error would worsen the Δ_{mean} and gamma pass rates, meaning a smaller magnitude of misclassification error would be acceptable. To quantify the effects of such a misregistration error, one could perform a similar experiment with an induced misregistration in addition to a bone misclassification. Such an experiment would provide more clinically relevant data for acceptable misclassification. MRI is also prone to geometric distortions, providing another possible source of error. These errors are considered beyond the scope of of this study.

As mentioned above, additional future work should encompass a number of things, either before or during the repetition of a similar experiment. Some of these are: 1) Obtaining the real sCT values that are prone to misclassification, before repeating this experiment for these Hounsfield units; or slightly higher to ensure a worst case scenario. 2) Increasing the range of PTV sizes, S, in order to fully understand the effects of PTV size in bone misclassification. 3) Making changes in the optimised VMAT treatment plan (Section 3.2), for example by changing the upper and lower DVH goals or treatment beam energy. 4) For partial misclassification of the skull specifically, repeating the experiment with a different, flatter shape of the misclassification volume, closer to the surrounding external air, and a CT value of 30 HU for the brain and PTV. Alternatively, real sCT images of the head, with realistic bone misclassification errors, could be

utilised directly. 5) Accounting for common misregistration errors by inducing a slight shift between the air and bone CT/sCT images. 6) Exploring the added effects of other possible MRI artefacts, like geometric distortion.

6 Conclusion

Using the deviation, Δ_{mean}, in the DVH point measurements, D_{mean}, and a 2 % acceptance criterion, this study found three linear equations describing the maximum acceptable sizes of misclassified bone given by the two-dimensional size, A, dependent on the distance from the centre of the spherical PTV to the centre of the misclassification volume, d. The three equations each describe the line of acceptance for a differently sized PTV. For a PTV of diameter $S = 1.5\,\text{cm}$, the equation is given by: $A = 1.73d - 0.64$. For a PTV of diameter $S = 2.0\,\text{cm}$, the equation is given by: $A = 1.70d - 0.26$. And for a PTV of diameter $S = 2.5\,\text{cm}$, the equation is given by: $A = 1.76d - 0.38$. This assumes a specific, clinically relevant VMAT treatment plan, perfect image registration, a CT value of bone of 800 HU, a PTV CT value of 0 HU, and a cylindrical misclassification error orthogonal to the treatment beam that replaces the bone with air (-1000 HU). These equations might provide a useful tool for radiographers, radiologists, and/or medical physicists in evaluating whether a misclassification error in sCT is too great to be acceptable for radiotherapy treatment planning. If the error is too great, inclusion of bone in the sCT must be pursued, e.g. by bone-atlas techniques [6] and/or UTE MR sequences [10], if possible. The gamma evaluation method utilised at UNN did not adequately identify cases of too large PTV dose deviation. Both simulated misclassifications of an entire skull volume showed clinically unacceptable results for all PTV distances, x, from the inner edge of the skull. As expected for bone of CT value $<$ 800 HU, the sCT experiments of partial bone misclassifications in the pelvis and the skull showed increase of Δ_{mean} for a decrease in d, and clinically acceptable results.

References

[1] Hall, Eric J. and Amato J. Giaccia. *Radiobiology for the Radiologist.* 7th ed. Philadelphia, Pa: Lippincott Williams & Wilkins, (2011).

[2] Mayles, P., A. Nahum, and J.C. Rosenwald. *Handbook of Radiotherapy Physics.* 1st ed. CRC Press, (2007).

[3] Khan, Faiz M., John P. Gibbons, and Paul W. Sperduto. *Khan's Treatment Planning in Radiation Oncology.* 4th ed. Philadelphia, Pa: Wolters Kluwer, (2016).

[4] Westbrook, Cathrine. *MRI in Practice.* 4th ed. Oxford: Wiley-Blackwell, (2011).

[5] Young, Lorna et al. "IMV Radiation Therapy Market Summary Report", (2018).

[6] Uh, Jinsoo, Thomas E. Merchant, Yimei Li, Xingyu Li, and Chiaho Hua. "MRI-based Treatment Planning with Pseudo CT Generated through Atlas Registration." *Medical Physics* 41, no. 5 (2014): N/a.

[7] "MR-only RT planning for the brain and pelvis with Synthetic CT." Siemens Healthcare GmbH. (2019). (accessed 18.10.2019). `https://www.siemens-healthineers.com/magnetic-resonance-imaging/clinical-specialities/synthetic-ct`.

[8] Palmér, Emilia, Emilia Persson, Petra Ambolt, Christian Gustafsson, Adalsteinn Gunnlaugsson, and Lars E. Olsson. "Cone Beam CT for QA of Synthetic CT in MRI Only for Prostate Patients." *Journal of Applied Clinical Medical Physics* 19, no. 6 (2018): 44-52.

[9] Persson, Emilia, Christian Gustafsson, Fredrik Nordstrom, Maja Sohlin, Adalsteinn Gunnlaugsson, Karin Petruson, Niina Rintela, et al. "MR-OPERA: A Multicenter/Multivendor Validation of Magnetic Resonance Imaginge-Only Prostate Treatment Planning Using Synthetic Computed Tomography Images." *International Journal Of Radiation Oncology Biology Physics* 99, no. 3 (2017): 692-700.

[10] Hsu, Shu-Hui, Yue Cao, Theodore S Lawrence, Christina Tsien, Mary Feng, David M Grodzki, and James M Balter. "Quantitative Characterizations of Ultrashort Echo (UTE) Images for Supporting Air–bone Separation in the Head." *Physics in Medicine and Biology* 60, no. 7 (2015): 2869-880.

[11] Broder, Joshua. "Imaging of Nontraumatic Abdominal Conditions." In *Diagnostic Imaging for the Emergency Physician.* W.B. Saunders. (2011). Pages 445-577.

[12] Jensen, Leif-André Bjerkeli. "Radiotherapeutic consequences of bone misclassification in sCT synthesised from MRI." Project thesis conducted at *Norwegian University of Science and Technology* and *University Hospital of North Norway*, submitted fall 2019. (2019)

[13] Low, Daniel A., William B. Harms, Sasa Mutic, and James A. Purdy. "A Technique for the Quantitative Evaluation of Dose Distributions." *Medical Physics* 25, no. 5 (1998): 656-61.

[14] Noufal, Manthala Padannayil P., Puzhakkal K. Niyas, Kallikuzhiyil Kochunny A Abdullah, and Pallimanhayil Abdul Raheem Subha. "Study of Impacts of Different Evaluation Criteria on Gamma Pass Rates in VMAT QA Using MatriXX and EPID." *Polish Journal of Medical Physics and Engineering* 23, no. 4 (2017): 99-107.

[15] Sjölin, Maria, and Jens Morgenthaler Edmund. "Incorrect Dosimetric Leaf Separation in IMRT and VMAT Treatment Planning: Clinical Impact and Correlation with Pretreatment Quality Assurance." *Physica Medica* 32, no. 7 (2016): 918-25.

[16] Korsholm, Marie, Line Waring, and Jens Edmund. "A Criterion for the Reliable Use of MRI-only Radiotherapy." *Radiation Oncology* 9, no. 1 (2014): 16.

[17] AD Elster, ELSTER LLC. "Spin Echo." *MRIquestions.com*. `http://mriquestions.com/spin-echo1.html`. (accessed 14.10.2019).

[18] Bloch, F. "Nuclear Induction." *Physical Review* 70, no. 7-8 (1946): 460-74.

[19] Wurnig, Moritz C., Maurizio Calcagni, David Kenkel, Magdalena Vich, Markus Weiger, Gustav Andreisek, Felix W. Wehrli, and Andreas Boss. "Characterization of Trabecular Bone Density with Ultra-short Echo-time MRI at 1.5, 3.0 and 7.0 T – Comparison with Micro-computed Tomography." *NMR in Biomedicine* 27, no. 10 (2014): 1159-166.

[20] Pixabay contributor "toubibe", *pixabay.com*. `https://pixabay.com/illustrations/mri-magnetic-x-ray-skull-head-782457/`. (accessed 03.06.2020).

[21] AD Elster, ELSTER LLC. "Chemical Shift Artefact." *MRIquestions.com*. `http://mriquestions.com/chemical-shift-artifact.html`. (accessed 14.10.2019).

[22] AD Elster, ELSTER LLC. "Dixon Method." *MRIquestions.com*. `http://mriquestions.com/dixon-method.html`. (accessed 14.10.2019).

[23] IMAIOS SAS. "Chemical shift artefacts in MRI." *IMAIOS.com*. `https://www.imaios.com/en/e-Courses/e-MRI/Image-quality-and-artifacts/chemical-shift`. (accessed 14.10.2019).

[24] Grodzki, David M., Peter M. Jakob, and Bjoern Heismann. "Ultrashort Echo Time Imaging Using Pointwise Encoding Time Reduction with Radial Acquisition (PETRA)." *Magnetic Resonance in Medicine* 67, no. 2 (2012): 510-18.

[25] AD Elster, ELSTER LLC. "Radial sampling." *MRIquestions.com*. `http://mriquestions.com/radial-sampling.html`. (accessed 05.05.2020).

[26] Eiknes, Live. "Spins in Motion: Flow". Lecture in FY8408, NTNU Trondheim, autumn 2019 (2019).

[27] AD Elster, ELSTER LLC. "Flow and MRA." *MRIquestions.com*. `http://mri-q.com/time-of-flight-effects.html`. (accessed 05.05.2020).

[28] Haase, A., J. Frahm, D. Matthaei, W. Hanicke, and K.-D. Merboldt. "FLASH Imaging: Rapid NMR Imaging Using Low Flip-angle Pulses." *Journal of Magnetic Resonance* 213, no. 2 (2011): 533.

[29] Wikimedia Commons contributors, "File:Mra-mip.jpg." *Wikimedia Commons, the free media repository.* `https://commons.wikimedia.org/wiki/File:Mra-mip.jpg`. (accessed 05.05.2020).

[30] Wikimedia Commons contributors, "File:MLCShape.svg." *Wikimedia Commons, the free media repository.* `https://commons.wikimedia.org/w/index.php?title=File:MLCShape.svg&oldid=259907810`. (accessed 31.10.2019).

[31] Scarbrough, Todd J. "File:Cumulative dose-volume histogram.jpg." *Wikipedia, the free encyclopedia.* `https://en.wikipedia.org/wiki/File:Cumulative_dose-volume_histogram.jpg#file`. (accessed 05.05.2020).

[32] Lilley, J.S. *Nuclear Physics : Principles and Applications.* The Manchester Physics Series. Chichester: Wiley, (2001).

[33] "SPACE" Siemens Healthcare GmbH. (2020). `https://www.siemens-healthineers.com/magnetic-resonance-imaging/options-and-upgrades/clinical-applications/syngo-space`. (accessed 05.05.2020).

[34] Depuydt, Tom, Ann Van Esch, and Dominique Pierre Huyskens. "A Quantitative Evaluation of IMRT Dose Distributions: Refinement and Clinical Assessment of the Gamma Evaluation." *Radiotherapy and Oncology* 62, no. 3 (2002): 309-19.

[35] Low, Daniel A., and James F. Dempsey. "Evaluation of the Gamma Dose Distribution Comparison Method." *Medical Physics* 30, no. 9 (2003): 2455-464.

[36] Johnstone, Emily, Jonathan J Wyatt, Ann M Henry, Susan C Short, David Sebag-Montefiore, Louise Murray, Charles G Kelly, Hazel M Mccallum, and Richard Speight. "Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging–Only Radiation Therapy." *International Journal of Radiation Oncology, Biology, Physics* 100, no. 1 (2018): 199-217.

[37] Jeter, Sheldon M. "Applications of error propagation analysis to the uncertainties of regression models in experimental thermal and fluids engineering." *Georgia Institute of Technology*, Session 1566 (2002)

[38] Palmer, M. "Propagation of Uncertainty Through Mathematical Operations" `http://web.mit.edu/fluids-modules/www/exper_techniques/2.Propagation_of_Uncertaint.pdf` (accessed 07.05.2020)

[39] Chow, Gregory. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* (pre-1986) 28, no. 3 (1960): 591-605.

[40] Geurts, Mark. "CalcGamma: 1D, 2D, or 3D gamma computation in MATLAB." *GitHub repository*, (2015). `https://github.com/mwgeurts/gamma`.

[41] Wikipedia contributors, "Human head," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=Human_head&oldid=951058670 (accessed 26.04.2020)

[42] Li, Haiyan, Jesse Ruan, Zhonghua Xie, Hao Wang, and Wengling Liu. "Investigation of the critical geometric characteristics of living human skulls utilising medical image analysis techniques." *International Journal of Vehicle Safety (IJVS)* 2, no. 4 (2007)

[43] Fishman, Elliot K. MD, FACR. "CT Case Studies." *CTisus.com* (2012) https://www.ctisus.com/redesign/teachingfiles/genitourinary/291964. (accessed 20.05.2020).

[44] Fidoe, Sophie. "The Hip Bone." *TeachMeAnatomy.info* (2019). https://teachmeanatomy.info/pelvis/bones/hip-bone/. (accessed 20.05.2020).

[45] Das, Indra, Chee-Wai Cheng, Minsong Cao, and Peter Johnstone. "Computed Tomography Imaging Parameters for Inhomogeneity Correction in Radiation Treatment Planning." *Journal of Medical Physics* 41, no. 1 (2016): 3-11.

[46] Constantinou, Chris, James C. Harrington, and Larry A. DeWerd. "An Electron Density Calibration Phantom for CT-based Treatment Planning Computers." *Medical Physics* 19, no. 2 (1992): 325-27.

[47] Hasenbalg, F., H. Neuenschwander, R. Mini, and E. J. Born. "Collapsed Cone Convolution and Analytical Anisotropic Algorithm Dose Calculations Compared to VMC Monte Carlo Simulations in Clinical Cases." *Physics in Medicine and Biology* 52, no. 13 (2007): 3679-691.

[48] Roberson, Peter L., P. William McLaughlin, Vrinda Narayana, Sara Troyer, George V. Hixson, and Marc L. Kessler. "Use and Uncertainties of Mutual Information for Computed Tomography/magnetic Resonance (CT/MR) Registration Post Permanent Implant of the Prostate." *Medical Physics* 32, no. 2 (2005): 473-82.

A Appendix

A.1 Guide to appendix

This appendix contains all measured and recorded values that are relevant to this thesis. These are the measurements of D_{mean} and gamma pass rates in the experiments (Appendix A.2, Appendix A.3, Appendix A.4) and the recorded MU values for each experiment (Appendix A.5, Appendix A.6, Appendix A.7). The values of R^2 from the statistical analysis of the data in Appendix A.2 and Appendix A.3 are given in Appendix A.8.1 and Appendix A.8.2, respectively. The appendix also contains the MATLAB code that was created in order to obtain the gamma pass rates (Appendix A.9).

A.2 Measurements

In this section, all measurements used for data analysis in this paper are listed. The tables show the measurements of $D_{m,air}$, $D_{m,bone}$, and the gamma pass rates (PR) for each time the experiment was performed, $E = \{1,2,3\}$. The difference between tables is the value of the PTV size, S.

Table A.1: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each combination of A and d. Here, $S = 1.5\,\text{cm}$.

E	1	1	1	2	2	2	3	3	3
d/A	$D_{m,air}$	$D_{m,bone}$	PR	$D_{m,air}$	$D_{m,bone}$	PR	$D_{m,air}$	$D_{m,bone}$	PR
[cm]/[cm^2]	[Gy]	[Gy]	[%]	[Gy]	[Gy]	[%]	[Gy]	[Gy]	[%]
3.0/2.5	1.993	1.974	99.95	1.992	1.969	100.00	1.992	1.971	99.99
3.0/5.0	1.990	1.947	98.34	1.992	1.948	98.47	1.992	1.949	98.57
3.0/7.5	1.995	1.932	97.11	1.993	1.925	95.19	1.993	1.927	94.94
3.0/10.0	1.990	1.934	95.36	1.991	1.912	94.56	1.991	1.912	95.37
3.0/12.5	1.991	1.899	93.60	1.990	1.900	93.59	1.991	1.900	93.56
4.0/2.5	1.991	1.974	99.98	1.993	1.976	100.00	1.994	1.976	99.92
4.0/5.0	1.991	1.960	100.00	1.992	1.957	99.78	1.995	1.962	99.98
4.0/7.5	1.992	1.941	98.03	1.992	1.944	96.17	1.991	1.938	96.97
4.0/10.0	1.990	1.918	96.03	1.991	1.919	95.23	1.992	1.921	95.15
4.0/12.5	1.991	1.910	95.40	1.991	1.909	95.72	1.991	1.908	95.40
4.0/15.0	1.993	1.893	95.26	1.991	1.895	94.63	1.992	1.893	95.40
5.0/2.5	1.992	1.979	100.00	1.992	1.979	100.00	1.991	1.979	100.00
5.0/5.0	1.992	1.963	100.00	1.990	1.960	100.00	1.994	1.967	100.00
5.0/7.5	1.990	1.947	99.97	1.991	1.948	98.07	1.993	1.969	100.00
5.0/10.0	1.992	1.941	97.35	1.991	1.940	97.15	1.990	1.942	97.28
5.0/12.5	1.991	1.954	99.75	1.991	1.947	97.94	1.996	1.933	96.29
5.0/15.0	1.990	1.918	96.25	1.990	1.917	96.25	1.992	1.920	96.28
6.0/2.5	1.991	1.983	100.00	1.991	1.980	100.00	1.991	1.981	100.00
6.0/5.0	1.991	1.970	100.00	1.990	1.970	100.00	1.990	1.967	100.00
6.0/7.5	1.990	1.957	100.00	1.990	1.957	99.99	1.994	1.963	100.00
6.0/10.0	1.993	1.958	99.25	1.994	1.961	99.53	1.993	1.954	99.55
6.0/12.5	1.991	1.936	95.64	1.992	1.937	93.39	1.990	1.936	95.53
6.0/15.0	1.992	1.921	91.28	1.991	1.924	91.78	1.990	1.924	92.01
7.0/2.5	1.994	1.984	100.00	1.994	1.985	100.00	1.994	1.984	100.00
7.0/5.0	1.994	1.980	94.92	1.994	1.978	100.00	1.992	1.977	100.00
7.0/7.5	1.993	1.967	96.03	1.995	1.970	100.00	1.991	1.966	100.00
7.0/10.0	1.993	1.962	99.96	1.994	1.962	99.97	1.994	1.961	100.00
7.0/12.5	1.992	1.944	93.03	1.994	1.946	95.90	1.992	1.945	94.40
7.0/15.0	1.995	1.944	93.18	1.995	1.944	93.27	1.994	1.946	95.88

Table A.2: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each combination of A and d. Here, $S = 2.0\,\text{cm}$.

E	1	1	1	2	2	2	3	3	3
d/A	$D_{m,air}$	$D_{m,bone}$	PR	$D_{m,air}$	$D_{m,bone}$	PR	$D_{m,air}$	$D_{m,bone}$	PR
[cm]/[cm^2]	[Gy]	[Gy]	[%]	[Gy]	[Gy]	[%]	[Gy]	[Gy]	[%]
3.0/2.5	1.993	1.971	100.00	1.993	1.970	100.00	1.995	1.972	100.00
3.0/5.0	1.990	1.955	99.32	1.991	1.955	99.50	1.991	1.950	97.45
3.0/7.5	1.995	1.936	94.65	1.996	1.938	94.79	1.995	1.941	95.37
3.0/10.0	1.990	1.912	93.71	1.990	1.913	93.81	1.991	1.913	93.85
4.0/2.5	1.995	1.980	100.00	1.996	1.981	100.00	1.994	1.981	100.00
4.0/5.0	1.992	1.957	99.43	1.992	1.958	99.45	1.992	1.957	99.32
4.0/7.5	1.993	1.944	96.76	1.994	1.944	96.81	1.991	1.941	96.24
4.0/10.0	1.991	1.930	94.70	1.990	1.931	95.07	1.992	1.931	93.99
4.0/12.5	1.991	1.911	93.45	1.989	1.910	93.33	1.991	1.913	93.44
4.0/15.0	1.991	1.896	91.98	1.991	1.899	92.62	1.992	1.900	92.77
5.0/2.5	1.991	1.980	100.00	1.991	1.980	100.00	1.995	1.985	100.00
5.0/5.0	1.993	1.971	99.99	1.992	1.971	100.00	1.990	1.970	100.00
5.0/7.5	1.994	1.960	99.89	1.990	1.952	99.25	1.992	1.955	99.34
5.0/10.0	1.994	1.947	95.86	1.995	1.947	95.71	1.993	1.945	95.74
5.0/12.5	1.991	1.929	94.40	1.989	1.928	94.65	1.990	1.930	94.62
5.0/15.0	1.990	1.920	94.68	1.990	1.912	92.66	1.991	1.912	92.51
6.0/2.5	1.993	1.982	100.00	1.992	1.982	100.00	1.993	1.983	100.00
6.0/5.0	1.994	1.973	100.00	1.994	1.972	100.00	1.992	1.970	100.00
6.0/7.5	1.991	1.971	100.00	1.991	1.963	100.00	1.991	1.966	100.00
6.0/10.0	1.990	1.948	95.96	1.991	1.962	100.00	1.991	1.950	96.45
6.0/12.5	1.990	1.938	94.34	1.991	1.937	93.25	1.991	1.938	93.28
6.0/15.0	1.990	1.925	92.35	1.990	1.925	92.26	1.991	1.927	91.86
7.0/2.5	1.991	1.982	100.00	1.991	1.983	100.00	1.991	1.983	100.00
7.0/5.0	1.990	1.973	100.00	1.992	1.972	100.00	1.990	1.975	100.00
7.0/7.5	1.997	1.974	100.00	1.995	1.973	100.00	1.997	1.974	100.00
7.0/10.0	1.991	1.960	99.89	1.991	1.959	99.93	1.991	1.960	99.92
7.0/12.5	1.992	1.947	96.32	1.990	1.948	98.27	1.991	1.947	94.23
7.0/15.0	1.992	1.941	93.62	1.991	1.938	92.71	1.993	1.941	93.81

Table A.3: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each combination of A and d. Here, $S = 2.5\,\mathrm{cm}$.

E	1	1	1	2	2	2	3	3	3
d/A	$D_{m,air}$	$D_{m,bone}$	PR	$D_{m,air}$	$D_{m,bone}$	PR	$D_{m,air}$	$D_{m,bone}$	PR
[cm]/[cm^2]	[Gy]	[Gy]	[%]	[Gy]	[Gy]	[%]	[Gy]	[Gy]	[%]
3.0/2.5	1.992	1.972	99.99	1.993	1.973	99.97	1.992	1.973	99.97
3.0/5.0	1.991	1.953	97.12	1.990	1.950	96.91	1.993	1.951	96.66
3.0/7.5	1.991	1.936	93.58	1.992	1.935	93.41	1.990	1.934	93.34
4.0/2.5	1.990	1.976	100.00	1.990	1.977	100.00	1.992	1.977	99.99
4.0/5.0	1.992	1.961	99.57	1.991	1.960	99.64	1.991	1.962	99.71
4.0/7.5	1.990	1.946	96.20	1.990	1.945	95.54	1.989	1.944	95.53
4.0/10.0	1.989	1.931	94.08	1.995	1.937	93.84	1.989	1.928	93.85
4.0/12.5	1.991	1.911	92.80	1.991	1.913	92.88	1.992	1.911	91.88
4.0/15.0	1.990	1.894	91.13	1.990	1.894	92.03	1.991	1.895	91.19
5.0/2.5	1.993	1.979	100.00	1.990	1.978	99.99	1.993	1.981	100.00
5.0/5.0	1.991	1.969	99.85	1.990	1.965	99.95	1.989	1.965	99.75
5.0/7.5	1.994	1.955	97.49	1.992	1.953	96.95	1.992	1.953	97.22
5.0/10.0	1.992	1.946	95.96	1.992	1.942	94.98	1.991	1.940	94.91
5.0/12.5	1.992	1.927	92.70	1.993	1.927	92.81	1.991	1.925	92.91
5.0/15.0	1.989	1.914	93.63	1.990	1.916	93.38	1.990	1.914	93.29
6.0/2.5	1.991	1.980	100.00	1.992	1.981	100.00	1.991	1.979	100.00
6.0/5.0	1.993	1.976	100.00	1.991	1.970	100.00	1.989	1.971	100.00
6.0/7.5	1.992	1.961	99.03	1.992	1.959	98.14	1.994	1.962	98.92
6.0/10.0	1.992	1.950	94.67	1.993	1.946	92.87	1.992	1.945	92.20
6.0/12.5	1.989	1.936	93.56	1.995	1.954	96.81	1.994	1.956	98.76
6.0/15.0	1.992	1.935	95.10	1.991	1.935	95.00	1.991	1.936	94.86
7.0/2.5	1.992	1.986	100.00	1.991	1.985	100.00	1.991	1.985	100.00
7.0/5.0	1.992	1.979	100.00	1.990	1.977	100.00	1.992	1.980	100.00
7.0/7.5	1.992	1.968	99.74	1.993	1.965	98.71	1.993	1.966	98.77
7.0/10.0	1.994	1.961	100.00	1.992	1.963	100.00	1.991	1.963	100.00
7.0/12.5	1.994	1.950	93.82	1.994	1.951	94.25	1.992	1.960	98.83
7.0/15.0	1.992	1.940	91.59	1.992	1.940	91.97	1.992	1.941	92.02

A.3 Measurements - skull

In this section, all measurements obtained from the experiments described in Section 3.4 are listed, for each time the experiment was performed, $E = \{1,2,3\}$.

Table A.4: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each combination of A and d.

E	1	1	1	2	2	2	3	3	3
x [cm]	$D_{m,air}$ [Gy]	$D_{m,bone}$ [Gy]	PR [%]	$D_{m,air}$ [Gy]	$D_{m,bone}$ [Gy]	PR [%]	$D_{m,air}$ [Gy]	$D_{m,bone}$ [Gy]	PR [%]
2.3	1.996	1.891	83.58	1.994	1.894	83.73	1.996	1.891	83.28
3.6	1.995	1.894	88.93	1.995	1.894	87.22	1.998	1.897	88.09
4.9	1.991	1.891	94.90	1.991	1.893	94.75	1.991	1.892	94.79
6.2	1.991	1.895	94.86	1.991	1.894	95.00	1.990	1.894	95.06
7.5	1.994	1.899	95.02	1.994	1.899	95.02	1.995	1.900	95.05
8.8	1.992	1.898	94.89	1.994	1.900	95.03	1.994	1.900	94.99

A.4 Measurements - sCT

In this section, all measurements obtained from the experiments described in Section 3.5 are listed, for each time the experiment was performed, $E = \{1,2,3\}$. Table A.5 shows the measurements of the pelvic sCT experiment in Section 3.5.1. Table A.6 and Table A.7 show the measurements of the first and second experiments described in Section 3.5.2, respectively.

Table A.5: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each d in the experiment described in Section 3.5.1.

E	**1**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**3**
d	$D_{m,air}$	$D_{m,bone}$	**PR**	$D_{m,air}$	$D_{m,bone}$	**PR**	$D_{m,air}$	$D_{m,bone}$	**PR**
[cm]	**[Gy]**	**[Gy]**	**[%]**	**[Gy]**	**[Gy]**	**[%]**	**[Gy]**	**[Gy]**	**[%]**
4.0	1.993	1.960	99.77	1.995	1.962	99.27	1.995	1.963	99.84
7.0	1.993	1.980	100.00	1.996	1.981	100.00	1.994	1.979	100.00

Table A.6: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each d in the first experiment described in Section 3.5.2.

E	**1**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**3**
d	$D_{m,air}$	$D_{m,bone}$	**PR**	$D_{m,air}$	$D_{m,bone}$	**PR**	$D_{m,air}$	$D_{m,bone}$	**PR**
[cm]	**[Gy]**	**[Gy]**	**[%]**	**[Gy]**	**[Gy]**	**[%]**	**[Gy]**	**[Gy]**	**[%]**
3.0	1.991	1.960	98.08	1.992	1.960	99.27	1.993	1.965	99.84
4.0	1.992	1.968	96.79	1.992	1.968	96.69	1.991	1.967	96.81

Table A.7: The measured values of $D_{m,air}$, $D_{m,bone}$, and gamma pass rates (PR), in each E, for each d in the second experiment described in Section 3.5.2.

E	**1**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**3**
d	$D_{m,air}$	$D_{m,bone}$	**PR**	$D_{m,air}$	$D_{m,bone}$	**PR**	$D_{m,air}$	$D_{m,bone}$	**PR**
[cm]	**[Gy]**	**[Gy]**	**[%]**	**[Gy]**	**[Gy]**	**[%]**	**[Gy]**	**[Gy]**	**[%]**
2.3	1.992	1.910	86.58	1.991	1.909	88.70	1.991	1.910	86.91
8.8	1.990	1.920	95.82	1.989	1.919	95.79	1.989	1.918	95.85

A.5 MU

In this section, all MU values that were found in the optimisation process and used in the specific plans for the experiment described in Section 3.2 are listed. The values 1, 2, and 3, in the top of each table, each represent one of the experiments, $E = \{1,2,3\}$, that were performed for each S.

Table A.8: The MU values resultant of optimisation, in each E, for each combination of S, A, and d.

A/d	S = 1.5 cm			S = 2.0 cm			S = 2.5 cm		
[cm^2]/[cm]	1	2	3	1	2	3	1	2	3
2.5/3.0	495.824	491.436	486.964	459.168	465.032	466.639	486.092	493.200	480.679
2.5/4.0	501.285	509.980	509.137	448.846	447.205	440.685	458.925	459.05	459.895
2.5/5.0	467.428	466.830	466.765	447.869	448.620	448.563	471.192	417.466	463.323
2.5/6.0	490.541	504.735	483.689	430.226	419.292	429.445	459.661	464.017	458.546
2.5/7.0	517.111	460.405	522.685	418.392	390.495	411.372	438.070	458.521	431.152
5.0/3.0	503.980	513.032	508.477	405.918	402.825	402.337	456.179	455.333	461.372
5.0/4.0	496.301	500.513	512.353	433.521	435.552	432.478	454.527	466.727	447.751
5.0/5.0	522.481	518.969	520.410	427.769	426.633	401.481	452.613	461.138	454.891
5.0/6.0	469.063	476.604	465.508	452.672	455.443	460.700	458.026	444.241	451.171
5.0/7.0	470.189	508.835	495.379	396.544	430.555	394.542	475.189	467.374	466.201
7.5/3.0	488.122	497.228	488.480	464.123	474.713	434.096	450.322	451.785	452.911
7.5/4.0	461.360	457.100	463.087	461.048	458.816	452.595	405.018	403.639	402.661
7.5/5.0	474.090	473.797	478.236	467.035	460.187	459.283	466.136	449.946	461.917
7.5/6.0	462.709	464.751	509.731	405.174	402.334	398.815	459.932	465.936	461.058
7.5/7.0	497.975	492.382	497.911	495.629	503.960	544.064	437.671	437.718	438.809
10.0/3.0	484.260	491.524	485.659	406.898	402.457	403.181	-	-	-
10.0/4.0	505.155	505.083	510.752	396.582	397.900	395.187	449.067	463.372	448.562
10.0/5.0	473.400	472.995	469.109	412.237	413.618	413.026	434.690	440.017	442.268
10.0/6.0	505.757	501.187	489.319	407.162	397.003	403.046	443.841	458.507	468.850
10.0/7.0	513.007	508.305	519.097	399.546	410.121	396.453	451.343	463.717	435.183
12.5/3.0	509.417	498.267	507.150	-	-	-	-	-	-
12.5/4.0	467.750	462.464	478.198	390.488	394.734	404.072	459.111	454.990	452.704
12.5/5.0	492.598	492.454	498.741	387.608	392.528	391.679	444.584	447.419	446.686
12.5/6.0	441.498	439.895	440.317	387.991	401.962	406.629	449.272	476.355	468.533
12.5/7.0	498.717	510.948	500.853	438.189	448.754	396.099	469.533	465.802	449.935
15.0/4.0	473.467	466.734	467.131	390.354	385.150	387.126	459.678	451.855	455.807
15.0/5.0	466.030	475.934	474.444	396.582	391.879	392.525	438.225	441.751	441.499
15.0/6.0	469.798	447.948	447.397	402.915	394.205	403.009	442.895	446.252	442.801
15.0/7.0	508.187	506.388	497.140	401.019	406.548	403.639	438.387	456.797	458.974

A.6 MU - skull experiment

In this section, the resultant MU values of the optimisation process in the experiment descried in Section 3.4 are listed, for each time the experiment was performed, $E = \{1,2,3\}$. Table A.9 shows the MU values for the plans created at each PTV distance, x.

Table A.9: The MU values resultant of optimisation, for each distance, x, and E in the skull experiment, described in Section 3.4. Two fields were applied for $x = 2.3\,\text{cm}$ and $x = 3.6\,\text{cm}$.

$\boldsymbol{x}$ [cm]	1	2	3
2.3	209.699 + 206.619	207.960 + 206.728	203.873 + 213.027
3.6	199.501 + 214.084	193.060 + 216.816	206.951 + 215.954
4.9	366.679	380.507	370.621
6.2	395.514	387.804	395.636
7.5	417.318	413.696	423.498
8.8	430.294	423.590	420.952

A.7 MU - sCT experiments

In this section, the resultant MU values of the optimisation process in the experiments described in Section 3.5 are listed. For the experiment described in Section 3.5.1, Table A.10 shows the MU values for the plans created for the PTVs at distance d for each time the experiment was performed, $E = \{1,2,3\}$. Similarly, Table A.11 and Table A.12 show the same data for the two experiments described in Section 3.5.2.

Table A.10: The MU values resultant of optimisation, for both distances, d, and E in the experiment described in Section 3.5.1.

***d* [cm]**	**1**	**2**	**3**
4.0	520.039	516.543	521.071
7.0	531.570	534.760	533.767

Table A.11: The MU values resultant of optimisation, for both distances, d, and E in the first experiment described in Section 3.5.2.

***d* [cm]**	**1**	**2**	**3**
3.0	385.172	367.110	386.232
4.0	410.333	408.064	395.483

Table A.12: The MU values resultant of optimisation, for both distances, x, and E in the second experiment described in Section 3.5.2.

***x* [cm]**	**1**	**2**	**3**
2.3	349.113	335.355	347.723
8.8	380.216	383.736	382.960

A.8 R^2

A.8.1 DVH experiment

In this section, all R^2 values from the linear regression of the DVH data is listed, with the exceptions of those that are given in Section 4.1.

Table A.13: The values of R^2 for all S and d.

S [cm]	1.5	2.0	2.5
d [cm]	R^2	R^2	R^2
3.0	0.9211	0.9873	0.9901
4.0	0.9919	0.9930	0.9956
5.0	0.8528	0.9978	0.9930
6.0	0.9671	0.9500	0.9212
7.0	0.9710	0.9837	0.9592

A.8.2 Gamma evaluation experiment

In this section, all values of R^2 from the linear regression of the gamma pass rate data are listed, with the exceptions of those that are given in Section 4.2.

Table A.14: The values of R^2 for all S and d.

S [cm]	1.5	2.0	2.5
d [cm]	R^2	R^2	R^2
3.0	0.9214	0.8515	0.9968
4.0	0.8362	0.9392	0.9455
5.0	0.4621	0.9154	0.9147
6.0	0.9087	0.7590	0.1470
7.0	0.3952	0.8750	0.7278

A.9 Gamma pass rate code

The code that was created in order to obtain the gamma pass rates from comparing the dose matrices is given below. The function *CalcGamma* was used with permission from its author [40]. MATLAB® version R2018b (9.5.0.944444) was used in this thesis.

```
clear; clc;

% Script for running 2D gamma evaluation
% Using the function CalcGamma


% Gamma evaluation criteria
percent = 2;
dta = 2; % res*3 = min dta. (0.3906 res. -> 1.1718 min dta)
local = 0; % Local gamma (1), global gamma (0)

% Cut-off value (low-dose gamma evaluation not useful)
cutoff = 0.20; % 20% cut-off

% Choosing reference and target dose matrices.
% Bone case = ref/measured = actual CT case (ground truth)
% Air case = target/calculated = synthetic CT (test)

refdicomfile = 'RD.7,5Bone5.dcm';
targetdicomfile = 'RD.7,5Air5.dcm';

% Retrieving dicom files and adding data to fit the input of CalcGamma
% width = vector containing pixelwidth and height
% start = vector describing start of evaluation (negative half of FOV in
% both directions used here (start lower left corner)).

ref = dicominfo(refdicomfile);
ref.data = double(dicomread(refdicomfile)); % relative matrix values
ref.data(ref.data > 10^9) = ref.data(1,2);
% replacing burn points incorners
ref.width = [ref.PixelSpacing(1,1) ref.PixelSpacing(2,1)];
ref.start = [-(ref.width(1)*double(ref.Rows))/2 -(ref.width(2)*...
    double(ref.Columns))/2];
ref.data = ref.DoseGridScaling*ref.data; % matrix values in Gy


target = dicominfo(targetdicomfile);
```

```
38 target.data = double(dicomread(targetdicomfile)); %
       relative matrix values
39 target.data(target.data > 10^9) = target.data(1,2);
40 % replacing burn points in corners
41 target.width = [target.PixelSpacing(1,1) target.
       PixelSpacing(2,1)];
42 target.start = [-(target.width(1)*double(target.Rows))/2
       -...
43      (target.width(2)*double(target.Columns))/2];
44 target.data = target.DoseGridScaling*target.data; %
       matrix values in Gy
45
46 % Performing Gamma evaluation
47 gamma = CalcGamma(ref, target, percent, dta, 'local',
       local); %Function
48
49
50 % Setting to either fail (1) or pass (0)
51 gammaPass = gamma > 1;
52
53 % Finding pixels to cut because of dose threshold
54 targetMask = (target.data < max(max(target.data))*cutoff)
       ;
55 refMask = (ref.data < max(max(ref.data))*cutoff);
56 mask = targetMask | refMask;
57
58 % Number of pixels cut
59 numCut = length(find(mask == 1));
60
61 % Setting cut values to pass
62 gammaPass(logical(mask)) = 0;
63 gamma(logical(mask)) = 0;
64
65
66 % Finding pass rate
67 [row, col] = size(gammaPass); % Using gammaPass to to
       find rows and columns
68 numTot = row*col - numCut; % Total number of pixels -
       number cut
69 numPass = length(find(gammaPass == 0)) - numCut; % Number
        of pixels that
70                                                      %
                                                           passed
                                                            -
                                                           number
                                                           cut
71
72 % Finding mean gamma value
```

```
meanGamma = sum(sum(gamma))/(numTot);

% Finding max gamma value
maxGamma = max(max(gamma));


passRate = numPass/numTot; % Pass rate, final result of
    script
disp(passRate);

% Setting figure axes
RI = imref2d(size(ref.data));
RI.XWorldLimits = [-col*ref.width(1,1)/20 col*ref.width
    (1,1)/20];
% Image width in cm
RI.YWorldLimits = [-row*ref.width(1,2)/20 row*ref.width
    (1,2)/20];
% Image height in cm


% Figure showing gamma evaluation image pass or fail
fig1 = figure(1);
hold on
im1 = imshow(gammaPass, RI, [0 1]);
%title('Global Gamma Pass Rate: 50 Degree Case');
xlabel('cm');
ylabel('cm');
c1 = colorbar;
c1.Ticks = [0 1];
c1.TickLabels = {'Pass', 'Fail'};
hold off

% Figure showing gamma evaluation image
fig2 = figure(2);
hold on
im2 = imshow(gamma, RI, [0 2]);
%title('Global Gamma Evaluation: 50 Degree Case');
xlabel('cm');
ylabel('cm');
c2 = colorbar;
c2.Ticks = [0 1 2];
c2.TickLabels = {'Pass','Border','Fail'};
hold off
```