UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

In-house Implementation and Validation of the Mid-Position CT approach for the Treatment Planning of Respiration-induced Moving Tumours in Radiotherapy for Lung and Upper abdomen cancer

Mariana Nunes Pereira

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Dr. Joep Stroom, Departamento Radioterapia, Fundação Champalimaud
Dr. Nuno Matela, Faculdade de Ciências da Universidade de Lisboa

2021

Acknowledgements

Firstly, I would like to extend my gratitude to all members of the Radiotherapy Department of Champalimaud Foundation for welcoming and offering me the opportunity to be part of the team.

I am especially thankful to Dr. Joep Stroom, my supervisor, for his guidance, support and for contributing to my personal and scientific growth. I also want to express my gratitude to Ariana Rocha and Dr. Sandra Vieira for being so kind and friendly towards me.

I would also like to thank Firass Ghareeb for sharing his knowledge and for accompanying me during this journey. His research work about the commissioning of Mid-position packages was an enormous contribution to my project. I am also grateful to Dr. Beatriz Nunes and Dr. Nuno Pimentel for spending their time and contributing to this work.

I want to thank Edward Jackson and Djamal Boukerroui for sharing their expertise with me and for all the useful suggestions. It was a pleasure to work with you.

Gostava também de agradecer ao professor Nuno Matela, não só pelos comentários construtivos à minha tese, mas também por todo o apoio ao longos dos anos de curso.

Muito obrigada aos meus amigos, em especial à Tânia por ter sido um apoio incondicional durante o meu percurso e por se tornar uma amiga para a vida.

Acima de tudo, quero agradecer à minha família – mãe, pai, mano, avó e Guilherme – obrigada pela vossa paciência, apoio e carinho. Sem vocês nada disto seria possível.

Resumo

A Radioterapia é uma das modalidades principais para tratamentos de foro oncológico que visa destruir a ação proliferativa das células cancerígenas e reduzir o volume tumoral. A sua ação terapêutica através do uso de radiação ionizante tem, subjacente, a máxima de irradiar o tumor com uma elevada dose, ao mesmo tempo que os órgãos de risco (OARs) adjacentes, são tanto quanto possível protegidos. Quando um tumor se localiza no pulmão ou abdómen superior, como no fígado ou pâncreas, o seu movimento devido à respiração pode alcançar até 4 cm, especialmente na direção crânio-caudal, aumentando as incertezas relativas à posição do tumor. No Centro Clínico Champalimaud (CCC), o planeamento convencional dos tratamentos de radioterapia faz uso de uma tomografia computadorizada (CT) que é adquirida aquando da respiração livre do doente e que, por isso, apresenta geralmente artefactos que podem ser uma fonte de erro durante o planeamento. Nos casos em que o movimento do tumor é considerável, é ainda adquirida uma tomografia computadorizada quadrimensional (4DCT) que consiste entre 8 e 10 CTs que representam fases do ciclo respiratório. Posteriormente, a 4DCT é utilizada para delinear o volume interno do alvo (ITV) que engloba toda a extensão do movimento do tumor. Apesar da estratégia do ITV garantir uma adequada cobertura do volume-alvo, os OARs ficam expostos a doses de radiação desnecessárias e a um maior risco de toxicidade. Este efeito é ainda mais preocupante em tratamentos hipofracionados, onde doses mais elevadas são administradas num número reduzido de frações. Nos últimos anos têm sido desenvolvidas estratégias que visam tornar os tratamentos de radioterapia mais eficazes. Uma delas é a reconstrução de uma CT que representa a posição média do doente ao longo do ciclo respiratório (Mid-P CT). Esta estratégia resulta em volumes de tratamento menores do que a estratégia do ITV, possibilitando o aumento da dose e maior controlo tumoral local. O primeiro passo para a reconstrução do Mid-P CT é o registo deformável de imagens (DIR) entre uma das fases da respiração (uma CT da 4DCT), definida como a fase de referência, e as restantes fases. Deste processo resultam campos vetoriais deformáveis (DVF) que contém informação do deslocamento dos tecidos. Os DVFs são subsequentemente utilizados para transformar cada uma das fases da respiração para a posição média. O método do Mid-P foi implementado com sucesso no Instituto do Cancro Holandês (NKI) em 2008. Apesar dos bons resultados clínicos, o número de centros de radioterapia que utiliza esta técnica é muito reduzido. Tal deve-se, por um lado, à inexistência de soluções comerciais com esta funcionalidade e, por outro, ao esforço necessário alocar para implementar e validar soluções desenvolvidas internamente.

O presente projeto teve como principal objetivo implementar a estratégia do Mid-P no CCC (Portugal). Para tal, foi otimizado um módulo – *RunMidP* – desenvolvido para o software 3D Slicer, que calcula o Mid-P CT e estima a amplitude do movimento do tumor e OARs com base nos DVFs. Considerando que a precisão do módulo e a qualidade de imagem do Mid-P CT devem atender os requisitos para o planeamento em radioterapia, foram realizados testes para validar o módulo. Sempre que possível, a sua performance foi comparada com outras aplicações desenvolvidas para a implementação da técnica do Mid-P, nomeadamente com um protótipo desenvolvido pela empresa Mirada Medical Ltd. (Reino Unido) – *Mirada* – e com o software desenvolvido no NKI (Holanda) – *Wimp*. Os testes foram divididos em três estudos diferentes, cada um com um conjunto de dados diferente. No primeiro estudo (estudo A), foram utilizadas 4DCT de 2 fantomas digitais, cuja função respiratória e cardíaca foi modelada de forma simplificada, e de 18 doentes com tumores localizados no pulmão (N = 8), no fígado (N = 6) e no pâncreas (N = 4). Neste estudo, foram comparados dois algoritmos DIR disponíveis no software 3D Slicer, o Plastimatch e o Elastix, em termos da precisão do registo e da qualidade de imagem do Mid-P CT reconstruído. Foi ainda avaliado a capacidade dos softwares *RunMidP* e *Mirada* representarem corretamente a posição média do doente e as diferenças das amplitudes do movimento do tumor estimadas pelos dois softwares. No estudo B, foram realizados testes de verificação semelhantes aos supre mencionados, em imagens sintéticas provenientes de 16 doentes, desta vez com a vantagem de se conhecer o “verdadeiro” Mid-P CT e as “verdadeiras” amplitudes do movimento do tumor. Estes foram comparados com os resultados obtidos com os softwares *RunMidP* e *Mirada*. Ainda, as unidades de Hounsfield (HU) no Mid-P CT reconstruído por *RunMidP* e *Mirada* foram comparadas com as HU na fase de referência, de modo a verificar se os Mid-P CTs produziriam diferenças dosimétricas relevantes. No último estudo (estudo C), a qualidade de imagem do Mid-P CT foi avaliada quantitativamente e qualitativamente. Durante a análise qualitativa, foi pedido a dois médicos especialistas que avaliassem a viabilidade dos Mid-P CTs, reconstruídos pelos três softwares (*RunMidP*, *Mirada* e *Wimp*), para o planeamento dos tratamentos. O tempo da reconstrução do Mid-P CT a partir da 4DCT foi de cerca de 1h. Ambos os algoritmos, Plastimach e Elastix, demonstraram ser adequados para DIR de imagens do pulmão e abdómen superior, com diferenças estatisticamente não significativas ($p > 0.05$) em termos da precisão do registo. Contudo, o Mid-P CT reconstruído com Elastix apresentou uma melhoria na qualidade de imagem, sendo assim o algoritmo DIR escolhido para ser implementado no *RunMidP*. Em termos de métricas aplicadas a contornos definidos manualmente, tais como a distância de Hausdorf (HD) e coeficiente de Dice (DSC), o erro do registo de imagem foi menor que 1 mm, dentro do contorno do tumor, e 2 mm no pulmão. Os Mid-P CTs reconstruídos com o *RunMidP* e *Mirada* apresentaram maiores diferenças, relativamente ao “verdadeiro” Mid-P CT, na região do diafragma e zonas de maior homogeneidade como, por exemplo, no ar presente no intestino. Contudo, para a maioria dos doentes do estudo B, o Mid-P CT reconstruído com o software *Mirada* apresentou maior índice de similaridade estrutural (SSIM) relativamente ao “verdadeiro” Mid-P CT. Estes resultados podem estar na origem do uso de diferentes algoritmos DIR, mas deveram-se principalmente a uma falha na aplicação das transformações deformáveis pelo módulo *RunMiP* que foi corrigida posteriormente. Ainda, as diferenças entre as amplitudes estimadas e previstas foram menores que 1 mm para 37 tumores (78,9%), que resultam em diferenças menores que 0.3mm quando convertidas em margens de planeamento. Para além disso, as diferenças nos valores de HU dos Mid-P CTs comparativamente à fase de referência foram, em média, de 1 HU no tumor e OARs. Foram também observadas melhorias na qualidade de imagem do Mid-P CT, nomeadamente um aumento da relação sinal-ruído (SNR) e diminuição dos artefactos. Estes resultados estão de acordo com a avaliação dos médicos que, em geral, consideraram que os Mid-P CTs reconstruídos pelos três softwares são adequados para o planeamento dos tratamentos. No entanto, os Mid-P CTs reconstruídos com dados 4DCT provenientes do CCC apresentaram classificações inferiores aos reconstruídos com dados 4DCT do NKI. Em suma, as modificações do algoritmo DIR Plastimach para Elastix e a correção do método para aplicar as transformações deformáveis, permitiram uma melhoria na qualidade de imagem do Mid-P CT e melhor performance do algoritmo, respetivamente. O módulo *RunMidP*, neste projeto otimizado e validado, apresenta um forte potencial para a reconstrução e implementação da estratégia do Mid-P na clínica, com performance comparável a outras aplicações existentes (*Mirada* e *Wimp*). Atenção especial deve ser dada aos dados 4DCT de *input* que parecem afetar a qualidade de imagem final do Mid-P CT. No futuro, valerá a pena otimizar os parâmetros de aquisição e reconstrução da 4DCT de modo a melhorar a qualidade de imagem e, ainda, o módulo *RunMidP* pode potencialmente ser otimizado no que respeita ao tempo de reconstrução do Mid-P CT e à precisão do DIR.

Palavras-chave: Radioterapia; Movimento do Tumor; Tomografia Computadorizada Quadrimensional; Registo Deformável de Imagens; Tomografia Computadorizada de Posição Média.

Abstract

Radiotherapy for tumours in the thorax and upper abdomen is challenging since they move notably with breathing. To cover the whole extent of tumour motion, relatively large margins are added to treatment volumes, posing a higher risk of toxicity for surrounding organs-at-risk (OARs). The Mid-Position (Mid-P) method accounts for breathing motion by using deformable image registration (DIR) to transform all phases of a 4DCT scan to a time-weighted average 3DCT scan (Mid-P CT). The Mid-P strategy results in smaller treatment volumes, potentially boosting the delivery of hypofractionated treatments. To bring the Mid-P approach to the Champalimaud Clinical Centre (CCC), an in-house Mid-position software module – *RunMidP* – was optimized. The module reconstructs the Mid-P CT and estimates breathing motion amplitudes of tumours and relevant OARs. In addition, this project presents a set of experiments to evaluate the performance of the Mid-P method and its feasibility for clinical implementation. The experiments were conducted throughout three different studies using 4DCT data from 18 phantoms and 23 patients. In Study A, the accuracy and image quality of two DIR algorithms (Plastimatch and Elastix) were assessed using quantitative metrics applied on either warped images or manually delineated contours. The reproduction of the patient's mean position by the Mid-P CT and the estimation of motion amplitudes were compared to a soon-to-be Mid-P commercial software developed by Mirada Medical Ltd. In Study B, similar experiments were performed, this time using a more rigorous reference – "true" Mid-P CT scans and "true" motion estimations. In Study C, the image quality of Mid-P CT scans was assessed quantitatively and qualitatively. Both Plastimatch and Elastix registration showed comparable registration accuracy, although Elastix showed superior image quality of reconstructed Mid-P CTs. Based on contour metrics, the registration error was less than 2 mm. In-house Mid-P CTs showed a slightly lower match to ground truth Mid-P CTs than the ones reconstructed by the Mirada prototype due to differences in DIR methods and small shifts to the original image geometry. Higher image differences were found in the diaphragm lung interface, where the patient's anatomy moves faster due to breathing, and in homogeneous regions such as the air regions in the bowel. On the other hand, differences (estimated-predicted) in motion amplitudes smaller than 1 mm were observed in 37 moving tumours (78.7%), showing a good performance of the Mid-P algorithm. Regarding the image quality, improvements in the signal-to-noise ratio and removal of image artefacts in Mid-P CTs are great advantages for using them as the planning CT. Clinicians also gave a good assessment of the suitability of Mid-P CT scans for treatment planning. No significant differences were found in the performance of the *RunMidP* compared to other Mid-Position packages, although worse scores were given to the CCC dataset than the dataset from another hospital. The in-house Mid-position algorithm shows promising results regarding the use of the software module in radiotherapy for lung and upper abdomen cancer. Further exploration must be given to improve the registration accuracy, image quality of the input data, and speed up the reconstruction of the Mid-P CT scan.

Keywords: Radiotherapy; Tumour Motion; 4DCT; Deformable Image Registration; Mid-Position CT.

Contents

List of Tables

List of Figures

List of Abbreviations

3D CRT	Three-Dimensional Conformal Radiotherapy
4DCT	Four-Dimensional Computed Tomography
ABC	Active Breathing Control
AIP	Average Intensity Projection
AP	Anterior-Posterior
CBCT	Cone-Beam Computed Tomography
CC	Cranio-Caudal or Cross-Correlation
CCC	Champalimaud Clinic Centre
CM	Centroid Motion
COM	Centre of Mass
CT	Computed Tomography
CTV	Clinical Target Volume
CTK	CLIP Tool Kit
DIBH	Deep-Inspiration Breath-Hold
DICOM	Digital Imaging and Communications in Medicine
DIR	Deformable Image Registration
DMLC	Dynamic Multi-Leaf Collimators
DNA	Deoxyribonucleic Acid
DOF	Degree-of-Freedom
DRR	Digitally Reconstructed Radiographs
DSC	Dice Similarity Coefficient
DVF	Deformation Vector Field
DVH	Dose-Volume Histogram
EBRT	External Beam Radiation Therapy
EORTC	European Organization for Research and Treatment of Cancer
EPIDs	Electronic Portal Imaging Devices

FF	Flattening Filter
FFF	Flattening Filter Free
FSB	Forced Shallow Breathing
GTV	Gross Target Volume
HU	Hounsfield Units
ICRU	International Commission on Radiation Units and Measurements
IMRT	Intensity-Modulated Radiotherapy
ITK	Insight Toolkit
ITV	Internal Target Volume
LINAC	Linear Accelerator
LQ	Linear-Quadratic
LR	Left-Right
MI	Mutual Information
Mid-P CT	Mid-Position CT scan
Mid-V CT	Mid-Ventilation CT scan
MIP	Maximum Intensity Projection
MLC	Multi-Leaf Collimators
MLD	Mean Lung Dose
MRI	Magnetic Resonance Imaging
MRML	Medical Reality Modeling Language
MSE	Mean Squared Error
MV	Megavoltage
NKI	Netherlands Cancer Institute
NSCLC	Non–Small Cell Lung Cancer
NTCP	Normal Tissue Complication Probability
OARs	Organs at Risk
pCT	Planning CT
PET	Positron Emission Tomography
PTV	Planning Target Volume

QA	Quality Assurance
RM	ROI Motion
RMSE	Root Mean Square Error
ROIs	Regions of Interest
RPM	Real-Time Position Management
RT	Radiotherapy
RTTT	Real-Time Tumour Tracking
SBRT	Stereotactic Body Radiation Therapy
SCLC	Small Cell Lung Cancer
SMLC	Segmented Multi-Leaf Collimators
SNR	Signal-to-Noise Ratio
SSD	Squared Intensity Differences
SSIM	Structural Similarity
TPS	Treatment Planning System
TRE	Target Registration Error
VMAT	Volumetric-Modulated Arc Therapy
VTK	Visualization ToolKit
XCAT	4D Extended Cardiac-Torso

1 Introduction

1.1 Motivation

Radiotherapy (RT) treatments, which have been shown to achieve good clinical outcomes for patients with cancer, are being increasingly adopted by healthcare providers globally as a key part of personalized medicine approaches. Given that radiation can cause harmful side effects on healthy tissues, radiotherapy for tumours in the lung and upper abdomen is challenging since the tumour and organs at risk (OARs) move notably with breathing, inducing large geometrical uncertainties in RT planning and treatment.

In the current radiotherapy practice, relatively large margins are added to tumour volumes to compensate for geometrical uncertainties, resulting in increased exposure of normal tissue to high doses. Even so, the current prognosis for lung cancer, the first and increasing cause of cancer-related mortality in both men and women, is poor with a 5-year survival rate of only 21% [1]. The quest for ever-increasing tumour doses to increase the tumour control [2], [3] combined with the goal of sparing healthy tissue of the irradiation field results in smaller margins around the tumour and a need to deal effectively with organ motion.

Several motion-management tools have been developed to account for organ motion and to increase the effectiveness of radiotherapy. One of them is the acquisition of a four-dimensional computed tomography (4DCT), which allows viewing of computed tomography (CT) images of respiratory-induced anatomical motion changing over the fourth dimension – time. Although 4DCT provides information on the breathing motion, the method to account for this motion varies between RT departments. Gating [4] and tracking [5] are two commonly used methods that require continuous real-time monitoring and LINAC response, causing a larger online workload in the clinic.

Two approaches are currently in use to account for 4DCT motion offline, the first is the internal target volume (ITV) approach, which encompasses the clinical target volume (CTV) position in all phases of the 4DCT. The ITV approach treats the respiratory motion as a systematic error, yielding too large treatment volumes and increasing the radiation toxicity to healthy tissues [6]. The other approach to account for breathing motion during radiotherapy is the Mid-Position strategy. Deformable image registration (DIR) methods are used to transform all breathing phases of the 4DCT scan to the time-weighted mean position over the respiratory cycle, yielding the Mid-Position CT scan (Mid-P CT) [7]. In addition, deformation vector fields (DVFs) acquired during DIR can be used to estimate motion amplitudes for relevant regions of interest (ROIs) [8].

The Mid-P strategy results in smaller treatment volumes than the more commonly used ITV strategy [9], which potentially allows dose escalation, and thus high probability of tumour control and overall survival for hypofractionated treatments (the delivery of high doses of radiation in a few number of fractions). Other advantages to this approach include improvements in image quality [8], leading to a higher signal-to-noise ratio (SNR), and removal of motion artefacts. Despite that, only 6% of the radiotherapy centres around the world use the Mid-P strategy [10], since each Mid-P method must be developed in-house and requires extensive validation.

1.2 Purpose of this dissertation

The current project was developed at Champalimaud Clinic Centre (CCC), in Portugal, where about 865 new patients including 1300 new plans are being treated per year, and about 40% are hypo-fractionated or single-shot treatments. The low number of fractions and the high dose per fraction require high-precision treatment planning and delivery to reduce treatment volumes and radiation toxicity.

Currently, due to the lack of commercial software to provide the Mid-P method, CCC defaults to the Internal Target Volume (ITV) approach for defining the extent of tumour motion. Using the ITV strategy, each patient receives a higher level of radiation than may be therapeutically necessary and may also encroach on organs at risk (OARs) which must not be exposed, leading to inefficient plans.

The main purpose of this work is to bring the Mid-Position (Mid-P) strategy to the CCC which has the potential to positively impact the treatment planning of a large number of patients with tumours in the thoracoabdominal region that move due to breathing. The work involves the optimization of a software module, developed for the medical image freeware 3D Slicer, that reconstructs the Mid-P CT and calculates breathing amplitudes for relevant anatomical structures. Taking into account that the image quality and the accuracy of the Mid-P method must attend to the requirements for radiotherapy planning, a set of experiments to validate the in-house Mid-P algorithm are outlined and benchmarked to two separate Mid-P packages: (1) a package in current clinical use at the Netherlands Cancer Institute (NKI) (NL); and (2) an automated science prototype developed by Mirada Medical Ltd[1] (UK).

This dissertation is divided into 6 chapters. Chapter 2 summarises the theoretical concepts behind the radiotherapy treatments, including the biological and physical aspects. Chapter 3 is dedicated to the description of state-of-the-art motion management techniques to account for breathing motion in clinics. In Chapter 4, the data and methods used for the Mid-P module implementation and validation are presented and discussed in Chapter 5. Finally, Chapter 6 concludes the work, and future improvements are considered.

---

[1] Mirada Medical, https://mirada-medical.com/ (Accessed on 03/07/2021).

2 Background Information

2.1 Upper abdomen cancer

Cancer is a major public health problem worldwide, being the second leading cause of death in the United States, with an estimate of 1,600 cancer deaths each day in 2021 [1]. Lung cancer is the first and increasing cause of cancer-related mortality in both men and women. According to the American Cancer Society, the current prognosis for non–small cell lung cancer (NSCLC) is poor, with a 5-year survival rate[2] of only 21%. In 2021[3], the disease is expected to cause approximately 131,880 deaths – more than colorectal, breast, and prostate cancers combined – and about 235,760 patients will be diagnosed with lung cancer [1]. Lung cancers are generally divided into two main categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC accounts for around 10-15% of the cases, while NSCLC concerns about 80-85% of the cases [11]. These two types of cancer differ in clinical and biological characteristics and therefore they have different treatment approaches.

Although the type of cancer that is most affected by breathing is lung cancer, other tumours located in the thorax, in the abdomen, and even in the pelvis are affected by respiratory motion, and some have lower survival rates than lung cancer: 10% for pancreas and 20% for liver cancer [1]. The treatment of thoracic and abdominal tumours will highly benefit from clinical advances in respiratory motion-management techniques.

2.1.1 Cancer treatment modalities

Radiotherapy (RT) is one of the main treatment modalities in oncology practice, being used in approximately 50% of all cancer treatments. RT (alone or combined with surgery or chemotherapy) is a curative treatment for 40% of the cancer patients receiving it [12]. Surgery has long been considered the treatment of choice for all patients with early-stage NSCLC (stage I and II) [11]. Unfortunately, at the time of diagnosis, most NSCLC patients have a poor physical condition or a disease that has advanced beyond the point where local treatment alone can provide a cure. Moreover, surgery is a complex operation and can have serious consequences such as lobectomy, pneumonectomy, and segmentectomy. Thus, most lung cancer patients are medically inoperable [6]. For that reason, radiotherapy has become the standard of care in inoperable early-stage NSCLC, aiming to deliver high doses to the tumour while sparing the surrounding healthy tissues, thereby increasing the probability of cure without serious complications.

2.1.2 Local tumour control and complications

Stereotactic Body Radiation Therapy (SBRT), a recent type of radiation delivery technique, consists in the delivery of single-dose and hypofractionated treatments, increasing the biological effectiveness and the local tumour control (Figure 2.1.a). SBRT has been shown to result in three-year local control rates of more than 90% for inoperable stage I NSCLC [6], [13]. Yet, surrounding healthy lung tissue,

---

[2] The five-year survival rate describes the percentage of patients that are alive five years after their disease is diagnosed.

[3] In Portugal, more than 4,000 new cases of lung cancer are diagnosed per year, which is the major cause of cancer death in men and the fourth largest in women [118].

such as the heart, blood vessels, the spinal cord, and the oesophagus are dose-limiting organs that must be spar from the radiation.

At CCC, about 865 new patients including 1300 new plans are being treated per year. Treatment outcomes for NSCLC using conventional radiation doses are poor, and the local recurrence rate is high [3]. The clinical service of CCC offers different radiotherapy modalities – single-dose session treatments and hypofractionated (> 5Gy per fraction) treatments. About 40% of patients receive hypo-fractionated or single-shot treatments at CCC. In these SBRT treatments, especially when tumours have a significant motion due to breathing, meticulous attention has to be paid toward ensuring optimal target coverage and guarantying minimum damage to healthy tissues.

In order to exclude sensitive surrounding tissues from the radiation and reduce the risk of complications, the treatment volumes must be smaller. However, due to the geometrical uncertainties that encompass inaccuracies of the radiation treatment (Section 2.5), large irradiated volumes are required to take into account these uncertainties and prevent geometrical misses. Radiation-induced respiratory toxicity ranges from impairment of lung function to fibrosis and pneumonitis – the most common complication of lung radiotherapy – which can develop a life-threatening complication [14]. Figure 2.1.b shows that the incidence of radiation pneumonitis increases as the mean lung dose (MLD[4]) increases. Thus, reducing the irradiated volume will reduce the risk of pneumonitis.

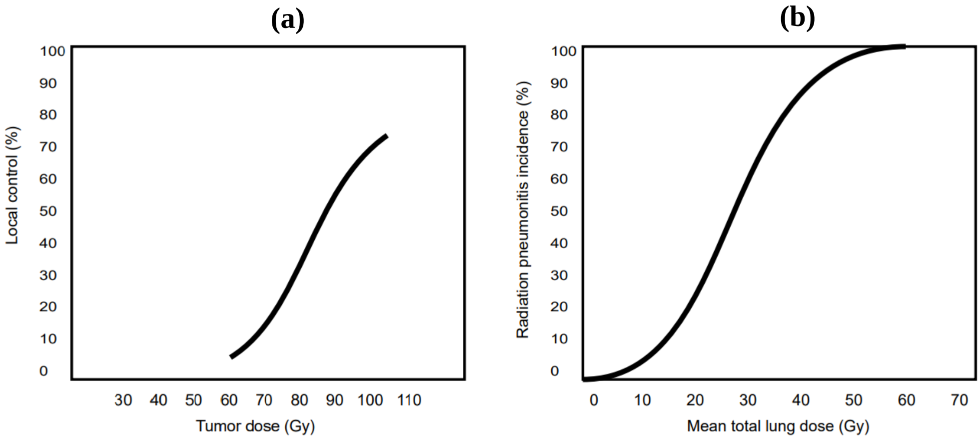


Figure 2.1: A schematic representation of (a) the dose-effect relation and (b) the relation between complications (radiation pneumonitis ≥ grade 2) and the mean lung dose (MLD). Retrieved from [15].

For upper abdomen cancer patients, technological improvements in radiation therapy delivery aim to widen the therapeutic window by accounting for geometrical uncertainties due to respiratory-induced tumour motion and prevent radiotherapy complications.

2.2 Mechanisms of action of Radiotherapy

Radiation therapy aims to control malignant cell growth by means of delivering a high dose of ionizing radiation to cancerous tissues whilst sparing healthy tissues as much as possible. The reason for ionizing radiation has been used to treat cancer is based on the fact that cells in the division phase (mitosis) or in the G2 phase (preparing to divide) are more sensitive to radiation. Therefore, rapidly

---

[4] Mean lung dose (MLD) is the integral radiation dose to the lung divided by the total lung volume.

proliferating cancer cells are more prone to DNA damage during radiotherapy than normal healthy cells [16]. When damages in the DNA structure of a cell occur, either directly or indirectly, its reproduction is interrupted, causing cancer cells to die and, subsequently, the tumour stops growing. Direct breakage in the DNA chain by ionizing radiation produces major cell damage, though the indirect effect, caused by the production of free radicals when the radiation interacts with cellular water, is more predominant in radiotherapy carried out by photon beams [16]. The presence of oxygen in cancer cells boosts the harmful effect of the free radicals which then will attack the DNA and cause irreversible cell damage.

For years, radiation biologists believed that delivering small daily fractions (usually 2 Gy per day for a total of 60-66 Gy over multiple weeks), also known as conventional radiation therapy, would preserve the integrity of normal tissue with acceptable tumour control [17]. In the last decades, hypofractionated treatments, where high doses are delivered in 1 to 5 fractions have gained more interest mainly influenced by good clinical results of SBRT in the treatment of tumours in the brain, lung, and liver. SBRT has indeed become the standard of care of NSCLC, increasing the patient's survival rates and local tumour control compared to conventional radiotherapy [13]. Hypofractionated schemes differ from conventional radiotherapy from a radiobiological point of view, and they have raised questions of whether the classic radiobiologic principles are still valid for high doses per fraction. The classic radiobiology concepts which try to explain the biological effect of radiation in the tumour and normal tissues are dependent on five factors, known as the five "R's": **R**epair, **R**epopulation, **R**edistribution, **R**eoxygenation, and **R**adiosensitivity [18].

The intrinsic molecular mechanism of DNA damage **repair**, allow cells to repair sublethal damage induced by radiation. This phenomenon occurs at a faster rate for normal cells than cancer cells, which will eventually lead to normal cells survival and cancer cells death over the treatment fractions. On the other hand, **repopulation** of tumour cells refers to the proliferation of surviving tumour cells during the course of a fractionated therapy [19]. Some studies demonstrated that lengthening the irradiation time, as it occurs in SBRT, allows the repair of some sublethal damage in rapidly proliferating tumour cells [19]. However, the repopulation biological effect after radiation has a low impact in tumours treated with SBRT, due to the reduction of the number of fractions [20]. The process in which cells go through the cell cycle after a dose of radiation, reaching the most sensitive mitosis phase, is called **redistribution** while **reoxygenation** brings hypoxic cells to a well-oxygenated state after irradiation. These two biological effects facilitate the success of radiation treatment since mitotic and oxygenated cells respond better to radiation. The effect of SBRT or single-dose treatments in these biological mechanisms is poorly understood yet but is probably less remarkable than in conventional radiotherapy. The last radiobiological effect – the **radiosensitivity** of tumour cells – is related to the susceptibility of cells to be damaged by ionizing radiation.

The linear-quadratic (LQ) model is commonly used to compare fractionation sensitivity and to predict tumour response to altered fractionation schemes. The model describes the cell survival (S) in terms of dose prescribed ($D = n \times d$), where d is the dose per fraction, n is the number of fractions, and α and β represent the lethal and sublethal damage, respectively, as follows [21]:

$$S = e^{-(\alpha D + \beta D^2)} \tag{2.1}$$

The cell's radiosensitivity is quantified by the ratio of α and β. The higher the sensitivity of tissues to an increase of dose per fraction, the lower the ratio. α/β is typically 3 Gy for late-responding tissues, such as the body and OARs, and 10 Gy for early-responding tissues, such as the tumour (Figure 2.2) [21].

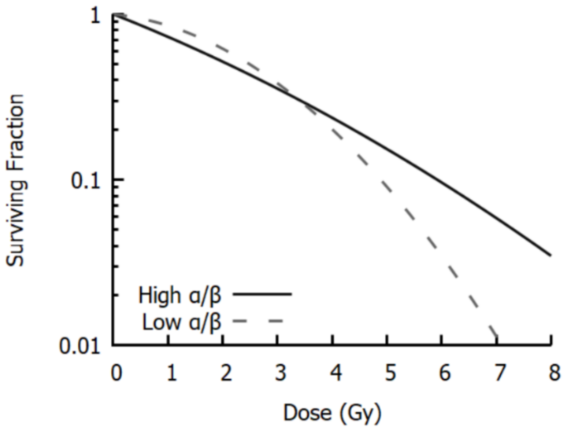


Figure 2.2: LQ curves for late-responding tissues with low *α/β* and early-responding tissues with high *α/β* [21].

Because the LQ model becomes progressively steeper with increasing radiation dose, it may overpredict tumour killing and late-responding normal tissue damage. Interestingly, clinical results have shown that the LQ model underestimates tumour control in SBRT. Therefore, it has been hypothesized that high-dose fraction treatments may cause significant vascular damage in tumours, leading to indirect cell death [22].

2.3 Physics of Radiation Therapy

External Beam Radiation Therapy (EBRT) is the most applied radiotherapy technique worldwide. It directs high-energy beams of ionizing radiation at a tumour from outside the body with a linear accelerator (LINAC). At CCC, megavoltage (MV) X-rays are the most frequent type of ionizing radiation to stop the multiplication potential of malignant cells, although protons and electrons can also be used for EBRT [23]. The other form of delivery radiation, called brachytherapy, is by using radioactive sources placed inside of or next to the tumour.

EBRT of tumours located within or near the thoracic cavity, especially in lungs, inevitably results in partial irradiation of surrounding normal tissues. Advances in systems integrated into medical LINACs and in radiation treatment delivery techniques have been allowed to widen the therapeutic window and consequently prevent tissue complications caused by radiation.

2.3.1 Medical LINACs

Modern medical LINACs (Figure 2.3) produce MV photons beams by accelerating electrons to high kinetic energies ranging from 4 to 25 MeV. Magnetrons and klystrons are special evacuated devices to generate a high-power radio-frequency electromagnetic field required for the acceleration of electrons produced by the electron gun. Magnets in the beam transport system (slalom bending system) guide the accelerated electrons until they collide with a metallic target, usually tungsten, producing X-rays photon beams by bremsstrahlung effect. The photon beam can be used for treatment purposes after (optional) flattening filtering, collimation and shaping by multi-leaf collimators (MLC) in order to obtain dose distributions that conform to the target volume while sparing neighbouring healthy tissue. X-rays photon beams can be delivered from any angle by rotating the gantry and couch [23].

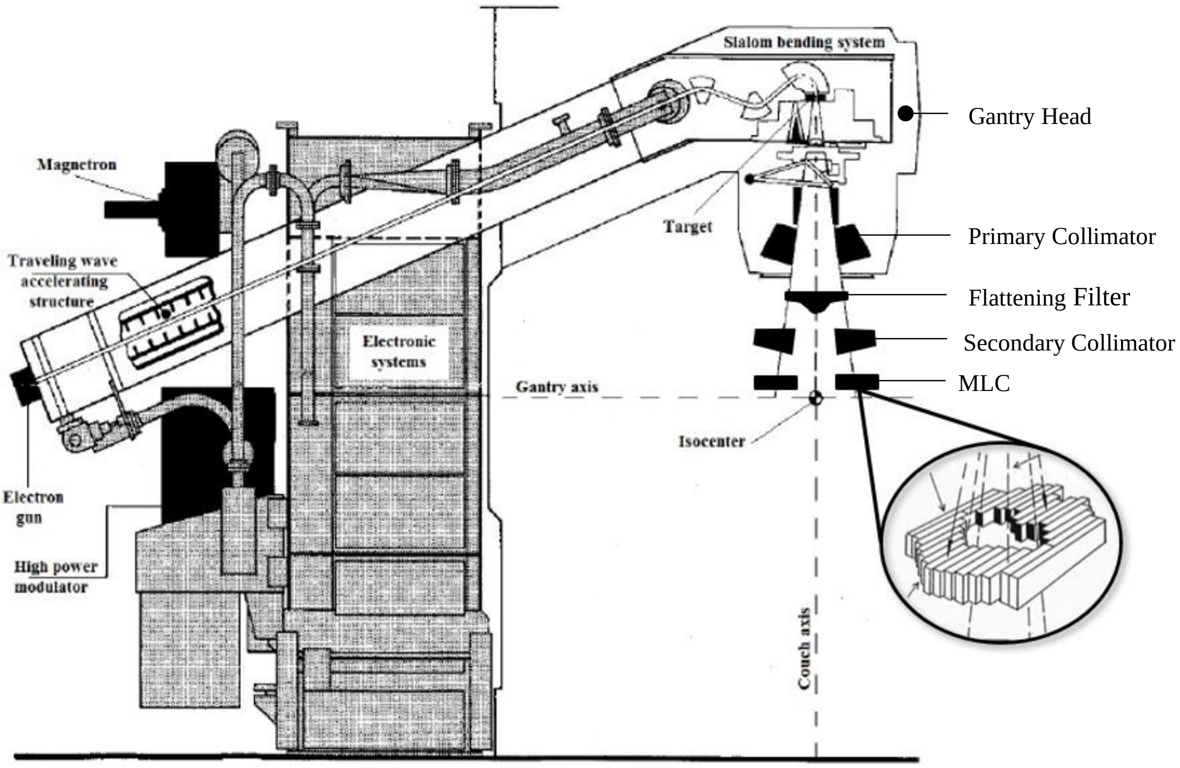


Figure 2.3: Main components of a medical LINAC. Electrons are injected by the electron gun and then accelerated through an accelerating tube by a high-power radiofrequency system (magnetron or klystron). The electron beam is shaped by the slalom bending system before it collides with a target (tungsten) creating high-energy photons. Collimation systems integrated into the gantry head are used to create the final photon beam distribution [24].

The first collimation system consists of a conical hole through a tungsten shielding block defining the largest available circular field size of the beam. It is designed to attenuate the X-ray beam intensity to less than 0.1% of the initial value [23]. After the primary collimator system, a flattening filter (FF) can be used to create uniform beam profiles. Despite that, advanced treatment techniques such as SBRT or IMRT have increased the interest in using a flattening filter-free (FFF) beam mode, where the X-ray beam has an irregular dose profile being more intense at the centre than in the periphery. The most significant benefit of using FFF for high-dose radiotherapy is the increased dose rate which provides faster treatments and improves clinical efficiency [25]. However, the non-uniform dose profile in FFF beams generates lower mean energy compared to FF beams of the same nominal energy, which raises the question of whether it can be compensated during planning [26]. In addition, more complex dose profiles of FFF beams increase plan complexity. The secondary beam collimators consist of two adjustable jaws that can move independently and can provide rectangular or square fields at the LINAC isocentre. They usually set the beam aperture to a maximum field size of 40×40 cm^2. Finally, the beam can be shaped by an MLC located inside the treatment head of a LINAC [23]. The MLC is the most recent collimator device integrated into modern LINACs, and it typically consists of 80 to 160 computer-controlled metallic leaves which can be positioned individually to modulate the intensity and to produce complex and irregular beam shapes. The introduction of MLC enabled to move from treatment field shape limited to squares and rectangles to more conformal shaped fields around the target. MLCs are therefore a key element in supplying intensity-modulated fields that conform the tumour and shield the OARs [23].

2.3.2 Radiation Treatment Delivery Modalities

The ultimate goal of radiotherapy is to achieve high local tumour control while minimizing the risk of toxicity of healthy tissues. To achieve this goal, three MLC-based external beam radiotherapy techniques have been routinely used in clinical practice: three-dimensional conformal radiotherapy (3D CRT), intensity-modulated radiotherapy (IMRT), and volumetric-modulated arc therapy (VMAT). At CCC, VMAT is the most currently used radiation delivery technique for the treatment of many types of cancers, apart from breast cancer which is mostly treated by IMRT.

2.3.2.1 Three-dimensional conformal radiotherapy (3D CRT)

Conventional three-dimensional conformal radiation therapy (3D CRT) uses beam collimator systems to achieve a good geometrical conformation of the treatment field to the shape of the tumour, allowing the delivery of high doses to the tumour volume. Still, the shaping of 3D CRT fields is a complex and time-consuming task, as it involves trying different combinations of beam shapes, weights, and gantry angles, conducted through forward treatment planning until a satisfactory dose distribution is achieved [27]. Moreover, 3D CRT is not able to change the intensity of the radiation across each beam, thus increasing the dose to healthy tissues, and neither to achieve optimal conformal fields for irregular-shaped tumours.

2.3.2.2 Intensity-modulated radiotherapy (IMRT)

Intensity-modulated radiotherapy (IMRT) is an advanced form of 3DCRT that modulates the intensity and shape of the radiation field while maintaining highly conformal dose distributions adapted to the tumour. The shapes and intensities of each field are optimized by computer algorithms to sculpt the high-dose volume around the tumour and thereby sparing adjacent OARs [28].

IMRT treatments can be delivered with the MLC operating in one of two modes: (1) the segmented MLC (SMLC) mode, also known as *step-and-shoot*; and (2) the dynamic MLC (DMLC) mode, also known as *sliding window*. In SMLC mode, each IMRT field is segmented into several shaped subfields (called segments), each one defined by a uniform intensity. The modulation of the beam fluence is achieved by altering the MLC configuration, the number of segments and the intensity of each segment individually (Figure 2.4). The *step-and-shoot* approach emphasizes that there is no MLC motion while the beam is on. The DMLC is a more complex technique, where the leaves of the MLC move during the irradiation of the patient. In the DMLC mode, the desired fluence profile is accomplished by varying the position and speed of each leaf of the MLC. These two modes (SMLC and DMLC) deliver radiation from a number of fixed gantry angles [23].

As opposed to 3D CRT, which uses forward planning, this method is impractical for IMRT, due to the large numbers of fields requiring several MLC arrangements. Therefore, IMRT planning utilizes computerized inverse planning to achieve the optimal dose distribution.

IMRT has been applied to treat many cancers, such as prostate, head and neck, and breast cancer [29]–[31]. However, several concerns have been raised regarding the use of IMRT for the treatment of lung tumours. This may in part be due to increased time during the treatment required to change the gantry angle. Longer treatment times can increase the patient’s discomfort and geometrical uncertainties during the treatment, such as intrafractional tumour motion and the patient’s anatomical changes.

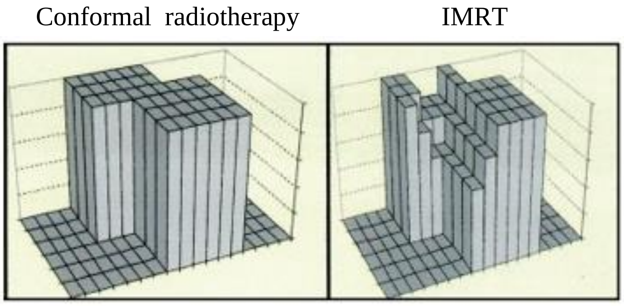


Figure 2.4: Comparison of intensity profiles of conformal radiotherapy and IMRT. Contrary to conformal radiotherapy, IMRT enables non-uniform intensity of radiation beams [32].

2.3.2.3 Volumetric-modulated arc therapy (VMAT)

Volumetric modulated arc therapy (VMAT) is the newest technique for delivering external photon beam radiotherapy. It consists of rotational IMRT, whereby the gantry of the LINAC rotates 360 degrees around the patient while the radiation beam is constantly on. The gantry rotation speed, dose rate and the shape of the irradiation field are parameters that can be modified to achieve highly conformal dose distributions around the tumour [33]. Another advantage of VMAT compared to 3D CRT and IMRT treatments is the reduction of the treatment time, decreasing the risk of intrafraction motion, and improving the patient's comfort during the treatment. The major challenge with designing VMAT treatment plans is that VMAT requires more time to plan optimization than IMRT, which increases the clinical workload.

2.4 Workflow of Radiation Therapy

Modern radiation therapy is a complex multi-step process as depicted in Figure 2.5. It starts with an evaluation of the patient's disease stage, followed by the treatment planning process, the treatment quality assurance (QA) and treatment delivery. The treatment planning involves a pre-treatment image acquisition, the target volume delineation and dose planning. The treatment quality assurance can occur before or during the treatment. The final step is the patient's position verification and treatment delivery.

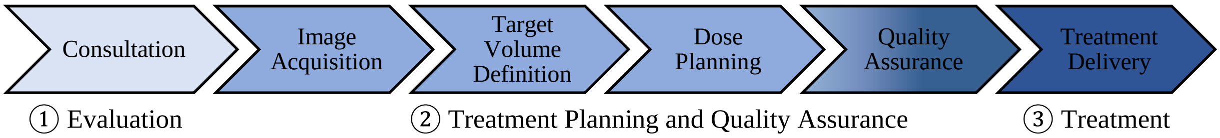


Figure 2.5: The pipeline of the radiotherapy workflow.

2.4.1 Image acquisition for Radiotherapy planning

The first step of the RT chain is the clinical evaluation of the patient's disease stage. The doctor selects the most appropriate treatment modality (radiotherapy, chemotherapy, or both) taking into account the characteristics of the tumour. If radiotherapy is the treatment chosen, the position and anatomy of the patient during treatment delivery must coincide exactly with the anatomical images acquired during the treatment planning, to spar the surrounding normal tissues from the ionizing radiation as much as possible. To ensure perfect alignment, several immobilization accessories are used during image acquisition and treatment delivery.

The current clinical practice utilizes X-ray CT as the imaging modality of choice for the assessment of tumour location and proximity to critical organs, planning the treatment, and calculating the dose delivery to the patient. This CT scan is commonly referred to as the planning CT (pCT). The gray-values derived from CT imaging represent the tissue attenuation, given by Hounsfield Units (HU), which are necessary for the calculation of doses in radiotherapy [23]. Even so, the delineation of target volumes in CT scans can be challenging due to the limited soft-tissue contrast. Magnetic resonance imaging (MRI), with its increased soft-tissue contrast, or positron emission tomography (PET), which provides cellular metabolism information, can be used as secondary imaging modalities to overcome difficulties associated with delineation of the tumour and OARs. Despite that, neither MRI nor PET provides information about tissue attenuation, and therefore they are not suitable for radiotherapy planning. Recently, image registration tools and machine-learning approaches have been exploited to generate synthetic computed tomography scans from MRI scans to overcome this limitation [34], [35].

2.4.2 Target Volume Definition

Target volume definition can be a challenging process in the case of upper abdominal cancers since tumours move notably with breathing. This treatment-planning step is therefore a prerequisite for minimising the effects of geometrical uncertainties, such as the tumour geometrical miss and tumour underdose during the treatment.

The previously acquired pCT is transferred to a treatment planning system (TPS) which allows the radiation oncologist to delineate the tumour and OARs. Eclipse$^{\text{TM}}$ from Varian is the preferred TPS used at CCC. To standardize the delineation process, the International Commission on Radiation Units and Measurements (ICRU) gives a detailed guideline to define critical structure volumes [36]. These volumes and margins are summarized in Figure 2.6. Delineation of the visible tumour volume on medical images results in the generation of a gross target volume (GTV). The GTV should be expanded to cover subclinical malignant disease, yielding a clinical target volume (CTV). For lung tumours, this empirical margin is generally taken to be 0 or 5 mm [37] although the final margin is at the discretion of the radiation oncologist based on clinic-pathological studies and clinical experience. The internal target volume (ITV) is a volume encompassing the CTV taking into account variations due to organ motion such as breathing and bladder or stomach contents. To deal with all relevant geometrical uncertainties (e.g. internal anatomical motion, patient's anatomical changes, and set-up uncertainties[5]) and to ensure that the prescribed dose will actually be delivered to the tumour, an additional margin is considered to the CTV, creating the planning target volume (PTV) [36].

Those margins aim to account for errors for specific patient populations and are derived from dose-coverage probabilities. The most widely used CTV-to-PTV margin is the margin recipe of van Herk *et al.* [38] given in the form of Equation 2.2, where $a = 2.5$, $b = 0.7$ and $\sum$ and σ represent standard deviations for systematic and random errors, respectively. This margin ensures that the CTV receives at least 95% of the prescribed dose for 90% of patients when all geometric uncertainties are included. Studies carried out at CCC [39] purposed different factors in the margin recipe for hypofractionated

---

[5] Set-up uncertainties are uncertainties that can occur during patient positioning in the treatment couch, such as the alignment of therapeutic beams during treatment planning and through all treatment sessions.

treatments, yielding $a = 2.1$ and $b = 0.8$. A more detailed description of the CTV-to-PTV margin is given in Section 2.5.1.

$$PTV_{margin} = a \sum + \ b \ \sigma \tag{2.2}$$

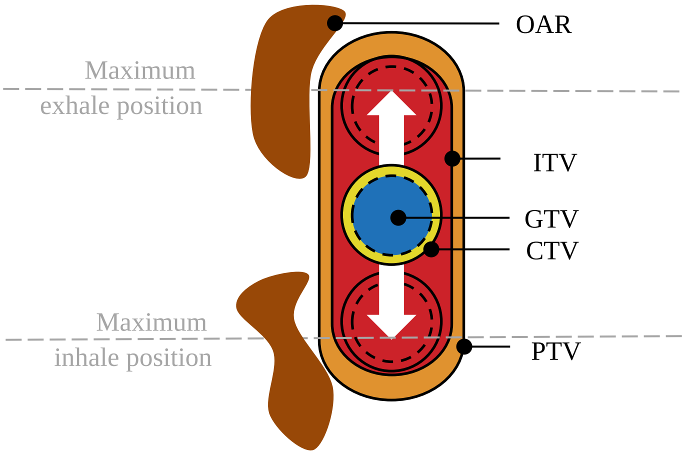


Figure 2.6: Schematic representation of the volumes of interest as defined by the ICRU 62 report. Adapted from [9].

2.4.3 Dose planning

Once target volumes are precisely delineated on the pCT, the goal of the treatment planning is to find the optimal dose distribution to deliver to the PTV while the dose to OARs does not exceed the predefined dose constraints. This is done by dosimetrists in order to maximise tumour control and minimise normal tissue complications due to radiation [23].

The treatment modalities used at CCC are the IMRT and VMAT which uses inverse treatment planning as the method to automatically determine the optimal dose distribution and corresponding LINAC settings for each patient treatment. With the pCT and the clinical goals[6] set – the main inputs of the inverse planning algorithms – an optimized patient-specific plan is created where all the input criteria are fulfilled. Dose calculations are performed using tissue attenuation information from the pCT and Monte Carlo algorithms integrated into the TPS [23].

The dose distribution can be then evaluated by a medical physicist using dose-volume histograms (DVHs) which is a valuable tool to verify if the target volume is correctly covered by the prescribed dose and the irradiation of the OARs does not exceed the international recommendations.

2.4.4 Quality Assurance

With the increasing use of high-dose fractionation radiotherapy, as in SBRT, treatment quality assurance has been gained more importance in clinical practice since any deviation of the planned dose

---

[6] Clinical goals include the dose prescription to the tumour and dose constraints to OARs. The most frequent dose prescription in SBRT for lung cancer is 50 Gy in 5 fractions [119]. The clinical objectives can be formulated in the form "*over 95% of the PTV volume should receive the prescription dose* ($D_{95\%} \geq 5000$ *cGy*)". Dose constraints can be formulated in the form ''*no more than X% of the OAR volume are allowed to receive more than Y Gy*''.

may cause severe damage to healthy tissues. Therefore, rigorous QA protocols are required to ensure the quality of the treatments and the safety of patients.

As aforementioned, the QA protocols can take place before treatment delivery (pre-treatment QA) or during the treatment delivery (*in vivo* QA). The pre-treatment QA is routinely performed with the use of electronic portal imaging devices (EPIDs) or, to a less extent, the ArcCheck phantom (SunNuclear Corporation), a heavy and cylindric-shape phantom with diode detectors, to verify LINAC-related settings such as beam flatness and stability, MLC position, and gantry rotation. In addition to machine-specific QA, patient-specific QA aims to find any discrepancy between the delivered and planned dose before the treatment session. Even more interesting would be the possibility to detect errors during the actual treatment delivery (*in vivo* QA) and stop the treatment before an error is committed. *In vivo* dosimetry has been successfully used in many clinics on a global scale [40].

If the physician approves the treatment plan, i.e., the differences between the planned and delivered doses are still clinically acceptable, the patient can be referred for the radiotherapy treatment. Otherwise, the treatment must be replanned.

2.4.5 Position verification and Treatment Delivery

Several strategies are used to reproduce as consistently as possible the patient's position of the pCT and deliver the treatment at the right position, thus avoiding anatomical inconsistencies before the treatment irradiation. One of them is the use of image-guided radiotherapy for adequate patient set-up and accurate location of the target. During the last few years, cone-beam CT (CBCT), mounted orthogonally to the treatment beam in the LINAC, has been the gold standard for daily in-room treatment imaging, although EPIDs can also be used to support patient positioning [41], [42]. Recently, there have been substantial developments in the industry of LINACs, which attempt to integrate MRI scans into LINACs with the advantage of having superior soft-tissue contrast than CBCT. Examples of commercial MRI-LINACs are the MRIndian system from ViewRay, in clinical use since 2014, and the Elekta MRI-LINAC Unity, showing clinical feasibility for high-precision radiotherapy [43], [44].

Image registration tools can also assist with patient set-up. Once the transformation of the patient's referential from the acquired daily images to the pCT is determined, the patient's position can be adjusted accordingly to the planned position [45]. Finally, if the position of anatomical structures matches in both images, the treatment is delivered by a LINAC.

2.5 Geometrical Uncertainties in Radiotherapy

A radiation treatment normally consists of one planning session and multiple irradiation sessions. Geometrical uncertainties may hamper the exact delivery of a treatment plan during all irradiation sessions. Beyond uncertainties in tumour delineation, the patient's movement, organ motions due to patient's physiology (e.g., respiration, cardiac motion) [46], anatomical shifts (e.g., different bladder or stomach filling) [47], and set-up uncertainties are common sources of geometrical uncertainties. The dosimetric consequences of this variety of uncertainties are almost an order of magnitude larger than those caused by respiration-induced motion [48]. Geometrical uncertainties can be modelled as systematic and random errors as will be discussed in Section 2.5.1.

2.5.1 Uncertainty margins

Several error sources can be defined as systematic errors ($\sum$) and random errors (σ). The first type of errors, the systematic errors, are identical for each treatment fraction of a specific patient [49]. Throughout all treatment sessions, systematic errors result in a shift in dose delivery, and consequently, CTV underdose and OARs overdose. A typical systematic error is the uncertainty due to the tumour delineation which is influenced by the image resolution and image artefacts during the entire course of the treatment. The random errors are deviations that can change per treatment fraction, resulting in a blurring of the cumulative dose distribution around the target. This displacement can subsequently lead to deviations between the intended and delivered dose distributions. A random error can be, as an example, a cycle-to-cycle variation in breathing pattern [50].

To account for geometrical uncertainties, the CTV is generally expanded by an additional margin, yielding the PTV. As mentioned before, systematic and random errors have different dosimetric impacts and should be treated differently in margin calculations. Margin recipes purposed by van Herk *et al.* [38] and Stroom *et al.* [49] show that the dosimetric consequences of systematic errors are three times more important than random errors (Section 2.4.2). Systematic and random errors are estimated based on a previously treated patient population and should be accurately quantified for each radiotherapy clinic, as they can depend on the image guidance procedures and treatment delivery techniques in use [15]. Since error sources are generally independent, the total uncertainty can be obtained by the root-sum-of squares of the existing uncertainties [39].

2.5.2 Uncertainty in tumour delineation

One of the first sources of error occurs just before the radiation treatment. At CCC, the planning CT (pCT) for thoracic and abdominal cancers is acquired while the patient breaths freely (free-breathing pCT), in which CT slices are arbitrary snapshots without any time information of the moving tumour. Consequently, image artefacts appear on the pCT.

Image artefacts manifest themselves as tumour and normal tissue delineation errors as well as dose-calculation errors since the patient's anatomy is not correctly represented [51]. An example of the image artefacts in the free-breathing pCT is shown in Figure 2.7.

To reduce uncertainties and artefacts induced by respiratory motion during pCT scanning, several respiratory-correlated acquisition techniques, such as the acquisition of a 4DCT, have been introduced in clinics and will be discussed in Section 3.

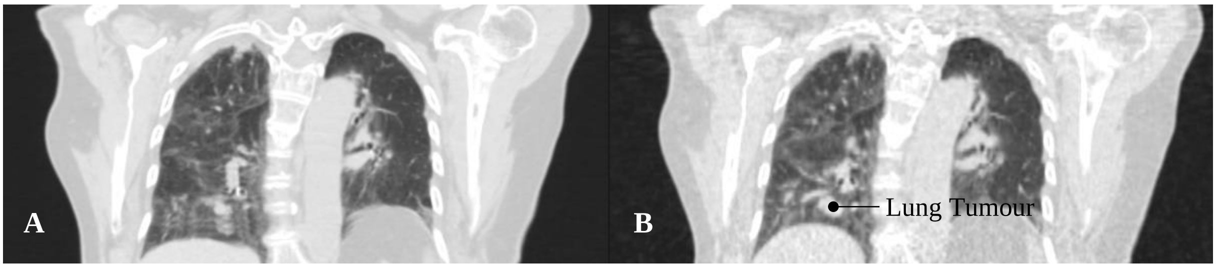


Figure 2.7: Coronal views of CT scans taken (A) during free breathing and (B) during inspiration at phase 30% (one of the breathing phases of a 4DCT). Image artefacts are seen in the free-breathing planning CT due to internal structures (e.g., tumour and diaphragm) motion.

2.5.3 Tumour motion due to respiration

The motion amplitude of lung tumours in most cases is smaller than 20 mm, although the tumour can move up to 40 mm due to breathing [52], [53]. As for liver tumours, the average displacement is 13 mm but it can reach up to 50 mm [54]. Tumour motion due to the respiratory cycle is complex, consisting of translations, rotations, deformations, and possible hysteresis[7] [52], [53]. Seppenwolde *et al*. [52] and Sonke *et al*. [53] demonstrated that tumour motion predominantly occurs in the cranio-caudal (CC) direction and it is usually bigger for lower lobe than upper lobe tumours, although substantial tumour motion can also be observed in the anterior-posterior (AP) direction and for tumours located in the upper lobe of the lung (Figure 2.9). The breathing motion is also patient-, fraction- and time-dependent, further increasing geometrical uncertainties during treatment delivery.

The difficulty of accurately reproduce the tumour's position through all the treatment fractions, according to the tumour's position during radiotherapy planning, requires large PTV margins to ensure total coverage of the target (Appendix I). This procedure is considered suboptimal because this increases the radiation field size and, consequently, the volume of healthy tissues exposed to high doses. The reduction of margins around the tumour demands strategies to deal effectively with tumour motion as will be mentioned in Section 3.

2.5.4 Baseline shift uncertainty

Apart from the variation in tumour position over the respiratory cycle (Section 2.5.3), the baseline shift is another challenging phenomenon when treating upper abdominal tumours. The baseline shift is defined as the variation in the mean tumour position (baseline) with the respect to the body anatomy over the treatment fractions. A schematic representation is shown in Figure 2.8.

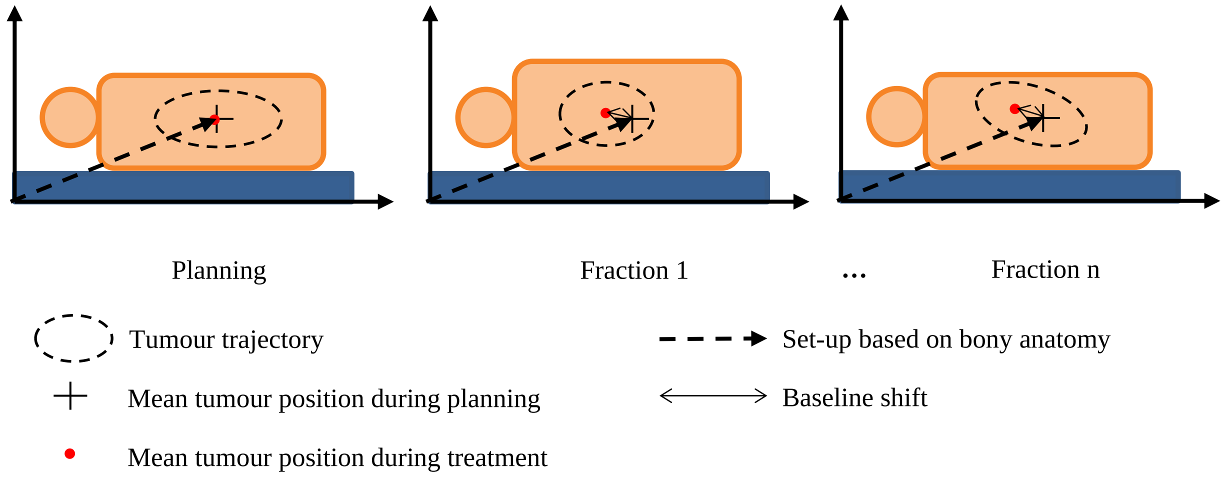


Figure 2.8: Schematic representation of the baseline shift uncertainty. The time-weighted mean position of the tumour is calculated during treatment planning, using e.g., a 4DCT scan. However, over the treatment fractions, the tumour's mean position and breathing motion amplitude can vary, which introduces geometrical uncertainties.

---

[7] Hysteresis: tumour trajectory from inhale to exhale is not equal to that of exhale to inhale, resulting in a three-dimensional elliptical tumour trajectory instead of a linear trajectory.

Sonke *et al*. [53] analysed the inter-fraction variability in the tumour mean position in terms of systematic and random errors. It was demonstrated that baseline variations can be in the same order of magnitude as bony set-up errors, with greater variability for lower-lobe tumours (Figure 2.9). A possible explanation for baseline shifts relies on some diaphragm position interference that may be caused by differences in stomach-filling, or even factors related to the patient's physiological processes, such as stress and cardiac status. Changes in tumour trajectory (breathing motion amplitude) can also play a role. Those baseline shifts need to be considered for the reduction of planning-treatment margins. For instance, the use of precise image guidance protocols allows for the correction of tumour position, which could impact the sparing of healthy tissues.

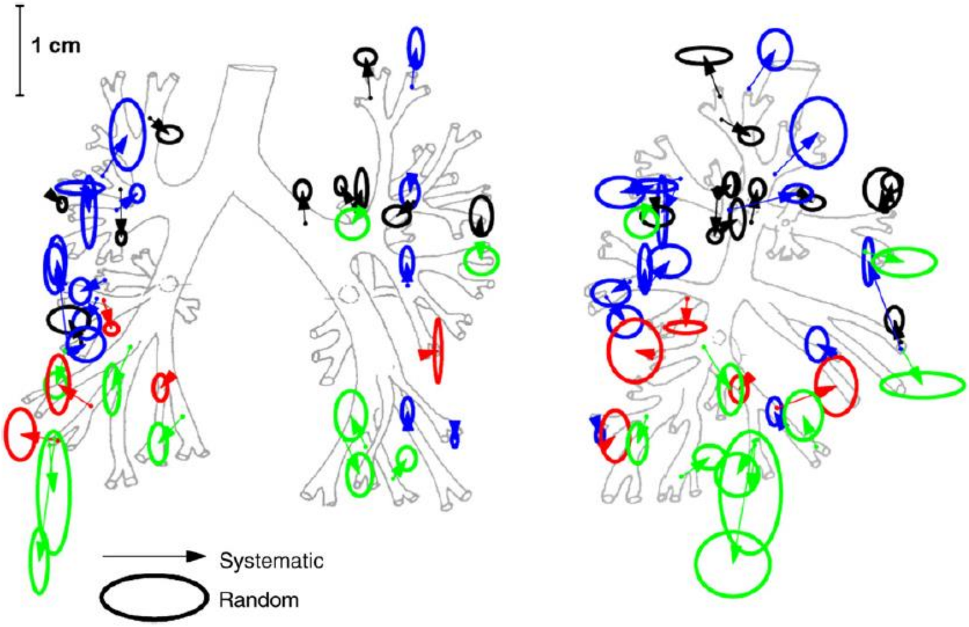


Figure 2.9: Graphical representation of systematic (arrows) and random (ellipses) baseline variations projected on coronal (left) and sagittal (right) planes of a schematic bronchial tree. The average amplitude of the bronchial tree and therefore the tumours are represented by different colours. Black: 0-5 mm; Blue: 5-10mm; Green: 10-15 mm; Red: ≥ 15mm [53].

2.5.5 Setup uncertainty

Set-up errors consist of the patient's position variation during the treatment course, regarding the patient's anatomical position in the planning CT (pCT). The positioning of the patient on the treatment couch on a daily basis is prone to patient set-up errors. The magnitude of these errors can be assessed by comparison of images acquired during the treatment using EPIDs or a CBCT. Besides, daily images can be registered to the pCT scan, correcting both set-up and baseline uncertainties [42].

2.6 Image Registration

Patient anatomy usually changes over the course of radiation therapy treatments. Organ deformations caused by patient motion, breathing, weight loss, tumours and OARs shrinkage, or organ filling must be correctly modulated to avoid tumour underdose and OARs overdose. Finding the spatial relationship between images (e.g., treatment CT and pCT) relies on the application of image registration methods.

Image registration can be categorized into two groups: rigid and non-rigid. Rigid registration is widely used for motion corrections in radiotherapy treatments, while deformable (non-rigid) image registration (DIR) is a powerful tool for motion compensation [8] and for automating tasks related to radiotherapy, such as recontouring of anatomical structures [55] and calculation of accumulated dose delivered during a re-irradiation treatment course [56]. Image registration applications will be mentioned in Section 2.6.2.

2.6.1 Image registration process

The image registration process involves estimating the geometric transformation $T(x')$ between two images, a reference (fixed) image $F(x)$ and a moving one $M(x')$, that minimises the difference between $F(x)$ and $M(T(x'))$. The result of the image registration is a deformation vector field (DVF, $u(x')$) that makes the moving image spatially aligned to the fixed image (Equation 2.3) and it is defined on the reference image coordinate system [57]. Figure 2.10 describes the image registration process where the estimated DVF is applied to the moving image (phase 0%) to fit the reference image (phase 50%).

$$F(x) = M(T(x')) = M(x' + u(x')) \tag{2.3}$$

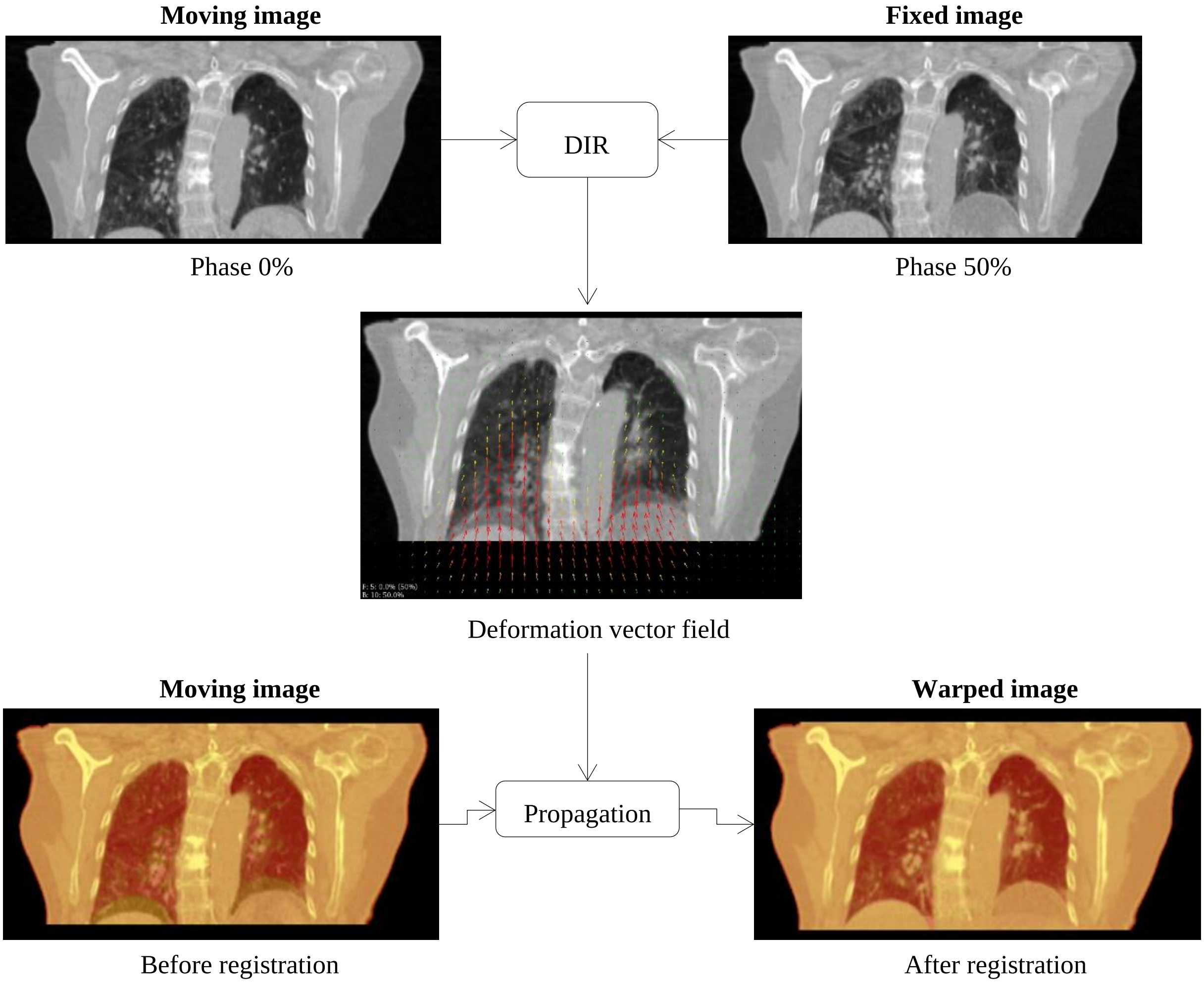


Figure 2.10: Flow chart of the image registration process for phase 0% and phase 50%. The output of the image registration between two images is a deformation vector field (DVF). After applying the DVF to the moving image, yielding a warped image, the two images (warped and fixed images) are completely aligned.

An image registration algorithm has four main components: (1) the input images; (2) the cost function as the mean to determine a good matching between the fixed and moving image; (3) the transformation model, which specifies how the moving image can be transformed to match the fixed image; and (4) the optimization algorithm that varies the parameters of the transformation model to maximize the matching criterion [57].

The image registration is formulated as an optimisation problem in which the cost function (C) is minimised with respect to the transformation model. The cost function consists of a trade-off (balanced by the factor γ) of a similarity measure (S), which defines the quality of the alignment, and a regularisation or penalty term (P) that constrains the transformation (Equation 2.4) [58].

There are two main approaches to solve the minimisation problem: parametric and non-parametric. In parametric methods, as is the case of the method that will be used in this work, the number of possible transformations is limited by introducing a parametrisation (model) of the transformation. The optimization problem thus becomes [58]:

$$C(T; I_F, I_M) = -S(T; I_F, I_M) + \gamma P(T) \tag{2.4}$$

$$\widehat{T_\mu} = \arg\min C(T_\mu; I_F, I_M) \tag{2.5}$$

where the vector μ contains the parameters that will be used in the transformation T. It implies that the transformation parameters depend on the transformation model used (rigid or non-rigid).

In other words, an image registration algorithm starts with the calculation of the similarity index between two input images. This initiates an iterative process where the optimization algorithm tries to maximize the similarity index by changing the DVF. The DVF is then used to deform the moving image according to the transformation model. The deformed moving image and the fixed image update the cost function and the optimization algorithm is done iteratively until the minimum of the cost function is achieved (Figure 2.11).

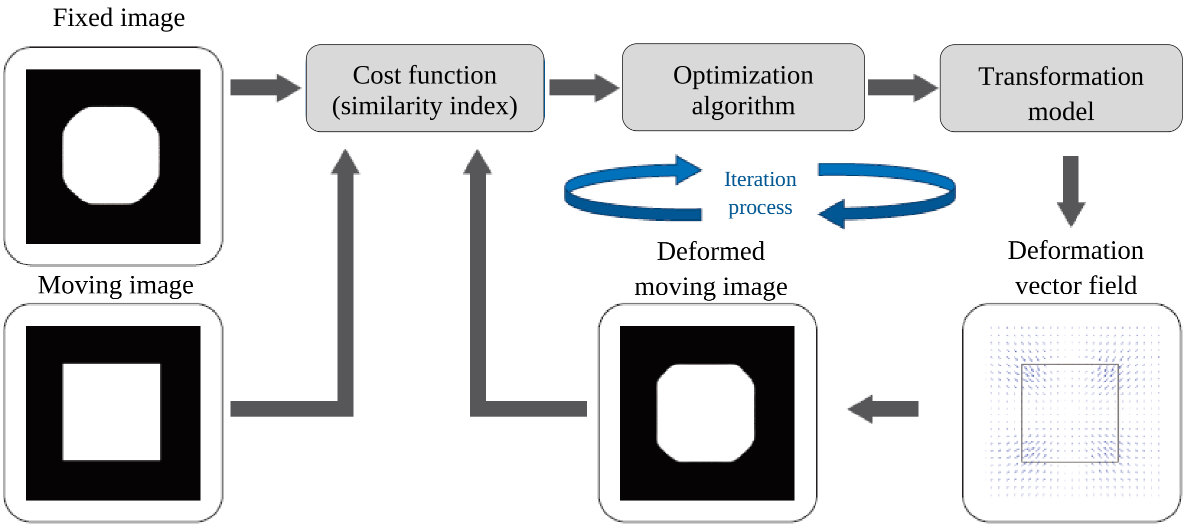


Figure 2.11: Flow chart of deformable image registration process [57].

2.6.1.1 Cost Function

The registration metric (or similarity index) quantifies the degree to which the reference and moving images are aligned. There are a large number of different similarity metrics. The selection of the proper measure to use depends on the imaging modality and the registration goal. For instance, features-based metrics make use of features extracted from the image data, such as anatomic or artificial landmarks and organ contours to align the two images. In turn, intensity-based metrics use the intensity information to find the spatial correspondence between images, assuming that the pixel values of the same anatomical region are similar among image sets [45].

This work is focused on intensity-based image registration algorithms. The more common intensity-based similarity measures in clinical use include: sum of squared intensity differences, cross-correlation, and mutual information.

The sum of squared intensity differences (SSD) is computed as the average squared intensity difference between the fixed (I_F) and moving image (I_M) and the optimal transformation is found by minimising Equation 2.6, where N is the number of voxels [45]. This metric is simple to compute and is effective for registering two imaging studies that have identical intensities for corresponding anatomies, such as 4DCT data.

$$SSD(I_F, I_M) = \frac{1}{N}\sum (I_F - I_M)^2 \tag{2.6}$$

Alternatively, the cross-correlation (CC) metric measures the similarity between two images assuming that they have a linear intensity relationship, therefore, this metric can handle differences in image contrast and brightness in the image (Equation 2.7).

$$CC(I_F, I_M) = \frac{\sum_x (I_F(x) - \overline{I_F})(T(I_M(x)) - \overline{I_M})}{\sqrt{\sum_x (I_F(x) - \overline{I_F})^2 \sum_x (T(I_M(x)) - \overline{I_M})^2}} \tag{2.7}$$

The SSD and CC similarity index are commonly used in mono-modal registration. For data from different modalities where the pixel intensities of the corresponding anatomy are typically different, it is not possible to simply minimise the error between intensities or calculate the products of intensities. Instead, the statistical mutual information (MI) is a very effective measure, which attempts to maximise the amount of information shared between the fixed (I_F) and moving image (I_M), based on their probability distributions of the voxel intensities [45]. In Equation 2.8, $p(I_F)$ and $p(I_M)$ are the probability distribution functions of the intensities I_F and I_M, respectively and $p(I_F, I_M)$ is the joint probability distribution function. This measure is often a good choice for multi-modal image registration, but also mono-modal image pairs.

$$MI(I_F, I_M) = \sum_M \sum_F p(I_F, I_M) log_2 \frac{p(I_F, I_M)}{p(I_F)p(I_M)} \tag{2.8}$$

2.6.1.2 Transformation Model

Since image registration is all about images, it is important to understand the fundamental underlying concepts of images from the Insight Toolkit (ITK). An ITK image is defined by a set of points on a grid occupying a physical region space, where the image's origin, spacing, size, and direction give a notion of the image's location in physical space. This significantly differs from many other image analysis libraries that treat an image as an array. Registration, for example, is performed in physical coordinates. Improperly defined spacing and origins will result in inconsistent results in such processes. Figure 2.12 illustrates the main geometrical concepts associated with the ITK image.

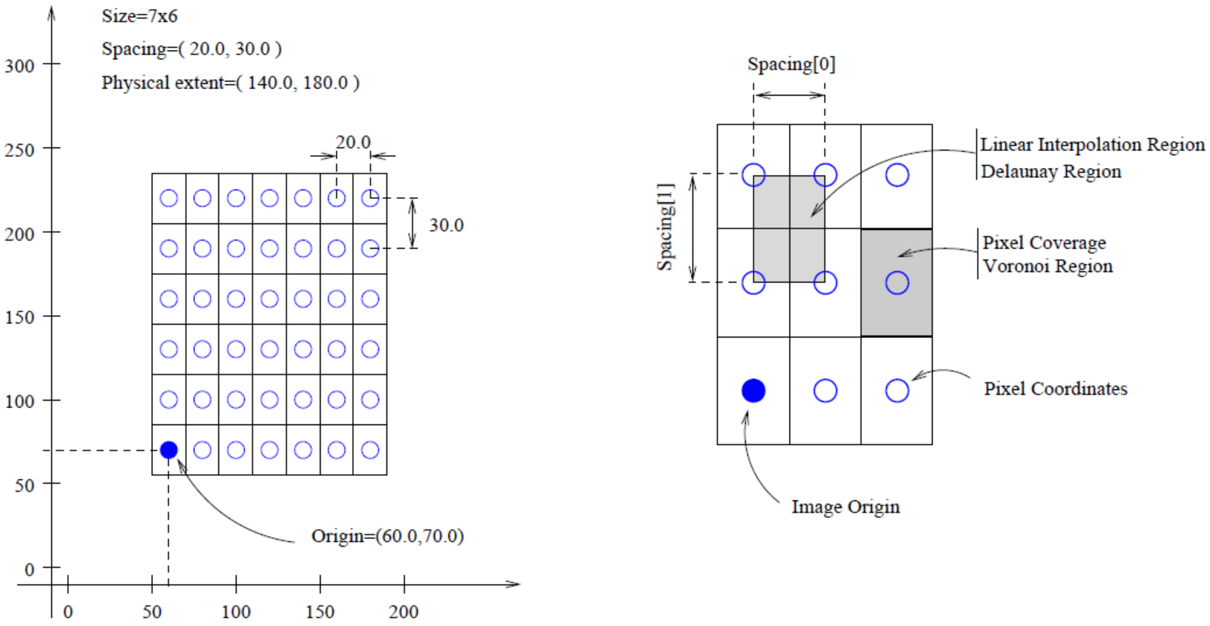


Figure 2.12: Geometrical concepts associated with the Insight Toolkit (ITK) image. An ITK image occupies a region in physical space which is defined by its meta-data (origin, size, spacing, and direction cosine matrix). Retrieved from [59].

As stated before, the transformation model determines the type of deformations between the fixed and moving image in order to improve the similarity. The transformation (T) can be rigid when the transformation preserves distances and angles between all points in the image (e.g., translation and rotation), and non-rigid (or deformable) when the estimated transformation does not include only rigid transformations but also deformations (e.g., shrinking or stretching) [57]. The most typical non-rigid parametric models are supported by a linear combination of basis-splines. B-spline image registration performs non-rigid alignment of two images driven by points on a (cubic) control grid and a spline function defines the spatial correspondence from these points [60].

The deformation vector (u) of a position vector x' in a moving image can be expressed as $u_k(x')$ after k iterations. Mathematically, the moving image with the size of $N_x \times N_y \times N_z$ is deformed by interpolating the displacement vector (ϕ) of control points on a rectangular voxel grid. The deformation vector at a pixel on the moving image can be expressed as follows:

$$u_k(x') = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} \beta_l(u)\,\beta_m(v)\beta_n(w)\phi(i+l, j+m, k+n) \qquad (2.9)$$

where u, v, w are the distances from x' to the nearest control points, (i, l, k), respectively. β represents the spline basis functions. Since the vector field can be defined in a parametric fashion (by B-spline coefficients), the number of control points determines the number of parameters [57]. B-spline transformations have a higher order of flexibility than rigid transformations. Therefore, the deformation vector fields are usually constrained by a regularization factor (P) to ensure that they represent an anatomical realistic transformation. The bigger the regularization term's (γ) is, the smoother will the transformation be (Equation 2.4) [59].

To speed up registration and make it more robust, a multiresolution approach is often used, which is accomplished by changing the grid spacing of control points.

2.6.1.3 Optimization

During optimization, the transformation is adjusted to find the minimum of the cost function, i.e., to improve the image similarity. There are several optimizers that can be chosen for this process (e.g., quasi-Newton, gradient descent and Robbins-Monro) [58]. Choosing a good optimizer depends on the registration problem which will influence the speed and accuracy of the registration.

In each iteration of the optimization, the value $M(T(x'))$ is evaluated at non-voxel positions, for which intensity interpolation is needed. Interpolators vary in quality and speed. The simplest interpolator is the nearest neighbour, in which the intensity voxel nearest in distance is returned. The linear interpolator returns the value, which is a weighted average of surrounding voxels, with the distance to each voxel taken as weight. Compared to the nearest neighbour, linear interpolators give higher quality registration. However, for high-quality registration algorithms, it is widely recommended to use a second-order (or higher) B-spline interpolator, although it requires more computation time [59]. The choice of the interpolator will also play a role in the image quality of the final image (warped image). Interpolation of image values is performed inside the Delaunay region (Figure 2.12).

2.6.2 Image registration applications

Image registration can play an important role in radiotherapy-related tasks, such as contouring anatomical structures [55], image guidance [45], and calculation of accumulated dose delivered during one or more radiotherapy treatments [56].

2.6.2.1 Contouring propagation

Image registration algorithms, especially DIR, can be used to perform automatic contouring. The idea behind this principle is to propagate already contoured structures to an image to be contoured, using the DVF estimated by sophisticated image registration algorithms. This has particular significance to reduce time and inter-physician variability associated with manual delineation of the GTV on each respiratory phase of a 4DCT to efficiently generate an ITV of lung tumours. Several studies have evaluated the accuracy of 4DCT contour propagation with results better than 2 mm [61]. Furthermore, contouring propagation can also be useful for inter-fraction delineation, e.g., to propagate contours from the planning CT to images acquired during treatment sessions [62].

2.6.2.2 Image guidance

Image registration is an integrated part of current treatment delivery systems to register and fuse the planning CT and images acquired at the time of the treatment (CBCT or EPID images) to aid in correcting patient positioning. After the calculation of the patient's offset between the planned and daily positions, the patient position can be adjusted based on the transformation estimated by the image registration algorithm [45].

2.6.2.3 *Dose accumulation*

DIR algorithms are used to perform dose accumulation between dose distributions in several situations: intra-fraction variations (breathing motion) and inter-fraction changes (session-to-session or re-irradiation). For instance, treatment planning systems only compute dose in static anatomies, although upper abdominal tumours move due to breathing. Therefore, the planning dose is not often the real delivered dose during the treatment. Ehrbar *et al*. [63] compared the accumulated tumour and lung doses for different respiratory motion-management techniques using 4D dose calculations and DIR. They found that the Mid-V, gating and couch tracking reduce the treatment volume and lung doses (see Section 3 for a detailed description of respiratory motion-management techniques). Moreover, they suggest the use of 4D dose calculations to evaluate the tumour dose coverage when these techniques are employed. Also, with the use of image registration methods, previous planning CT images can be mapped against current planning CT images to match the patient's anatomy along a re-irradiation treatment. The resulting transformation can hereafter be used to transform the previous dose distribution to the current treatment and then sum both doses. As a result, the previously delivered dose is considered in the treatment planning of the re-irradiation treatment [64].

2.6.3 Image registration validation and Quality assurance

Image registration errors can occur from image distortions, registration algorithm limitations, incorrect selection of registration algorithm parameters, among other factors. Image registration validation and verification are required especially if the intention is to use this tool integrated into a software for clinical purposes. The Therapy Physics Committee of the American Association of Physicists in Medicine (AAPM) commissioned Task Group 132 suggests that the overall image registration accuracy for most clinical applications, where an in-plane resolution is about 1 mm and slice thickness range approximately from 2 to 3 mm, is typically desired to be within 2 mm [45]. The methods for evaluation of the image registration accuracy can be classified broadly into two categories: qualitative and quantitative methods. Quantitative methods differ depending on the type of structure being evaluated.

2.6.3.1 *Qualitative methods*

Qualitative evaluation of the image registration often rely on the visualization of image fusion. Once the optimal transformation between a fixed and moving image is determined, the moving image can be mapped to the fixed image, resulting in a warped image. The overlay of the images (fixed and warped images) is a useful approach for the visualization of the success of the registration method [45].

2.6.3.2 *Quantitative methods*

Images: The determination of the mean square error (MSE) is a useful metric to evaluate the extent of mismatch between the original and deformably registered image. For a perfect image match between fixed and warped images, the MSE error should be zero [56].

Fiducials: The identification of anatomical points on the fixed and moving image can be a simple way to verify the accuracy of the registration method. The target registration error (TRE) is the average residual error between the two identified anatomical points mapped onto the warped image through image registration. The more accurate is the registration, the smaller is the TRE [45]. In practice, it may be difficult to accurately define the corresponding points in each image, meaning that there will always be some degree of uncertainty in the location of the identified fiducial points, resulting in a non-zero TRE, even after the highest quality image registration.

Volumes: The alignment of structures delineated on the fixed image overlaid onto the warped image can provide a qualitative assessment of the accuracy of registration. Under perfect registration (and perfect delineation), the contours should be perfectly matched, although in practice this is almost impossible to achieve. For instance, Persson *et al*. [65] quantified the interobserver delineation variation to be 2.6 mm for SBRT of peripherical lung tumours. However, the discrepancy between the contours should be within the variation of the observer contouring. Quantitative comparison between the contours can be performed using the Dice similarity coefficient [45] and the Hausdorff distance [66].

The Dice similarity coefficient (DSC) computes the number of pixels that overlap between the two volumes and normalizes it by half of the sum of the number of non-zero pixels in the two volumes, see Equation 2.10, where A is the segmentation in the fixed image and B is the segmentation mapped from the deformably registered image. As the contours approach agreement, the DSC value approaches 1; as the volumes diverge into two non-overlapping structures, the DSC value goes to 0.

The Hausdorff distance is defined as the maximum distance in pixel between two contours A and B (Equation 2.11).

$$DSC = \frac{|A \cap B|}{|A \cup B|/2} \tag{2.10}$$

$$H(A,B) = max\big(h(A,B), h(B,A)\big) \text{ with } h(A,B) = \underset{a \in A}{max} \ \underset{b \in B}{min} |a - b| \tag{2.11}$$

3 State-of-the-art

The Management of Respiratory Motion in Radiotherapy

This Chapter describes the respiratory-motion management techniques that are currently clinically implemented to deal with tumours affected by respiratory motion during radiotherapy, elucidating the advantages and disadvantages of each of them. According to the AAPM report no.91 [48], many technological solutions have been developed to account for respiration-induced organ motion and they can be separated into methods to deal with tumour motion *during treatment planning* and *during treatment delivery* (Figure 3.1). The technologies covered in this Chapter are four-dimensional imaging methods, respiratory-gated techniques, breath-hold techniques, forced shallow breathing methods, and respiration-synchronized techniques.

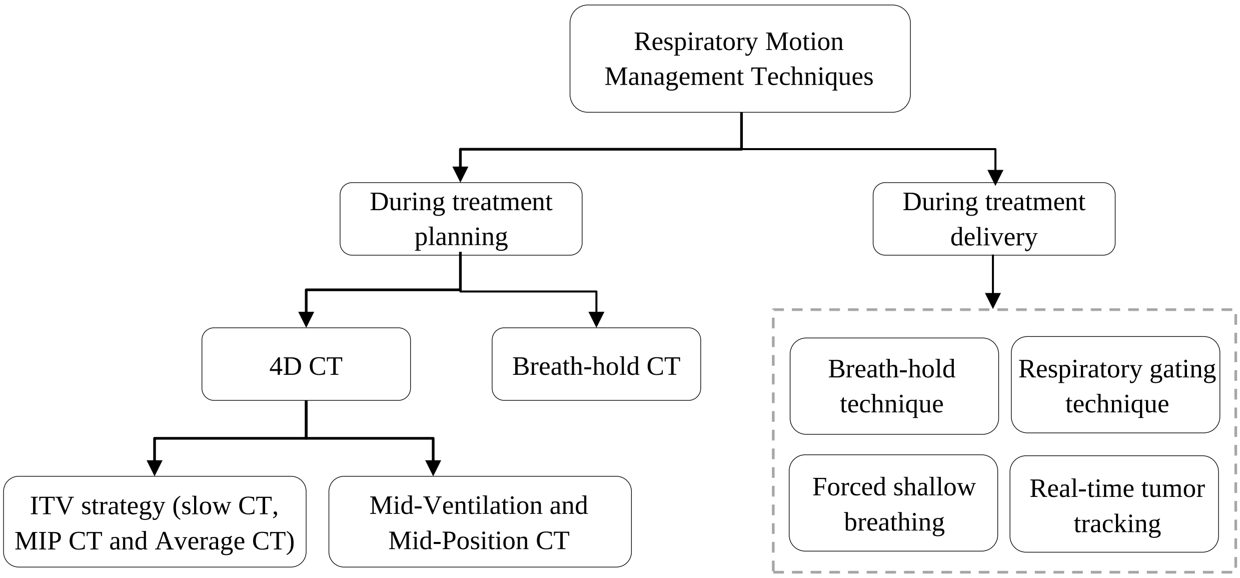


Figure 3.1: Overview of the respiratory motion-management techniques.

3.1 Measuring organ and tumour motion

The lungs, oesophagus, liver, pancreas, breast, and kidneys, among other organs, are known to move with breathing [48]. Together with 3DCT scanning [67]–[69], fluoroscopic imaging [52], [70], [71], MRI [72], and ultrasound [73] have been used to obtain organ motion information. In many cases, however, the target being tracked is the tumour, although fiducial markers implanted in or near the tumour [4], [52] or a surrogate organ (e.g. diaphragm) [7], [74] can also be used to correlate with tumour motion. On the other hand, both surface markers, spirometers, and thermocouples provide signals that try to predict the tumour motion, and they are crucial to getting information about the phase of the breathing motion used for reconstructing respiratory-correlated 3DCT (4DCT). Some CT scanners, such as the Brilliance CT Big Bore (Philips Healthcare, Cleveland, OH), are integrated with devices to measure the respiratory signal, such as the Philips Bellows, a pressure transducer, or the commercially available real-time position management (RPM^{TM}) system from Varian.

3.2 Managing tumour motion in treatment planning

3.2.1 Breath-hold CT

During the planning CT acquisition, tumours that move due to breathing can be incorrectly detected, resulting in a smaller planning volume or a distorted tumour shape (Figure 3.2) [68], [75]. On the other hand, image artefacts caused by respiratory motion can result in larger planning volumes (Figure 2.7 and Appendix I). Both effects may lead to erroneous delineation of the CTV. The reliability of the location and area size of the tumour in the planning CT depends on the magnitude of tumour motion during the examination. To obtain a more representative volume and shape of the tumour, a breath-hold CT is preferable. The tumour encompassing volume can be acquired in both inhalation and exhalation breath-hold CT scans, where the blurring and image artefacts caused by the motion during free-breathing is significantly reduced. This technique has the disadvantages of increasing the CT scanning time and relies on the ability of the patient to hold his breath, which is too difficult for many radiotherapy patients, especially for those with poor pulmonary function [48].

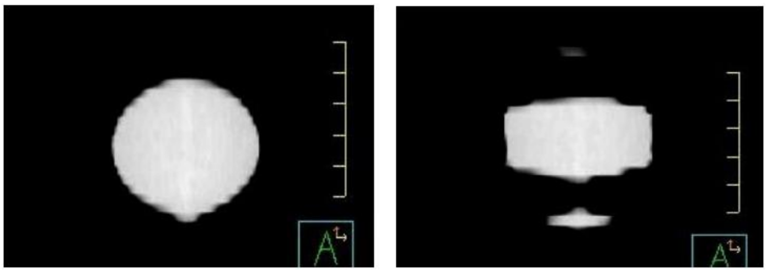


Figure 3.2: Illustration of image artefacts that are caused by superior-inferior (SI) motion during 3DCT imaging of a test sphere. Left: CT coronal section of a static sphere. Right: CT coronal section of the same sphere in oscillatory motion (range = 2 cm, period = 4 s) [75].

3.2.2 Four-dimensional (4D) CT

Four-dimensional computed tomography (4DCT) correlates the image acquisition with the patient's breathing, resulting in a sequence of 3D image volumes that represent the entire breathing cycle. It thus enables to visualize and to estimate the tumour and organs trajectory over the breathing cycle. Beyond that, a major contribution of the 4DCT is the reduction of imaging artefacts that appear due to breathing [15], rendering delineation of structures for radiotherapy more accurate than conventional free-breathing CT scanning. Therefore, the use of 4DCT for treatment planning is highly recommended by the European Organization for Research and Treatment of Cancer (EORTC) since 2010 [76].

The breathing signal of the patient can be recorded either by a thermocouple, air bellows, Varian RPM™ system, among other devices. The CT scanner usually can generate a signal when slices are acquired ("X-ray ON status"), enabling a correlation of each slice with its respiratory phase. Retrospectively, the acquired data is binned into a set of eight or ten 3DCT scans, each corresponding to a particular breathing phase. Sorting of imaging data into different bins can be performed with respect to the time (phase-based binning) or amplitude (amplitude-based binning) of the breathing signal [77]. During phase-based binning, each bin represents an equal percentage of the period of the respiratory waveform, while in the amplitude-based binning each bin represents a percentage of the inhale or exhale amplitude of the respiratory waveform. It is known that amplitude-based binning is less prone to imaging artefacts [78], [79], especially in the vicinity of the diaphragm. However, phase binning is often favoured for the use within radiotherapy because it provides relative velocity information, allowing to define a proper time-averaged position or perform dose accumulation without additional information [79]. Figure 3.3 depicts the sorting of imaging data into eight bins with respect to amplitude.

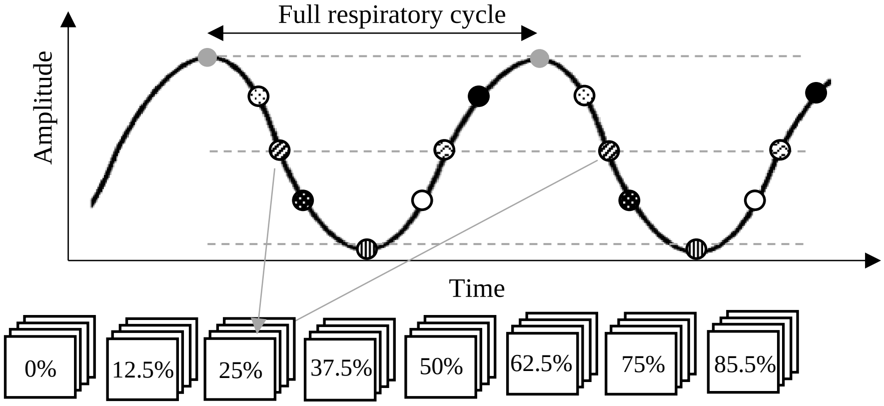


Figure 3.3: Amplitude-based retrospective sorting for the reconstruction of eight 4DCT frames. Adapted from [77].

4D imaging techniques do not differ from the conventional procedure for imaging acquisition or treatment delivery. In some oncology institutes, respiration-correlated CT scans are easy to be performed. Still, the implementation of 4D imaging in the treatment-planning stage is not yet standardized and some techniques to handle the 4DCT data set (Section 3.4) are not yet commercially available. It usually requires in-house developed software and multiple manual operations, resulting in high variability in outcomes and increase workload in clinical practice.

3.3 Managing tumour motion during treatment delivery

3.3.1 Breath-hold

Breath-hold techniques have been predominantly applied to lung cancer radiotherapy in order to reduce respiratory tumour motion. In this technique, the patient is instructed to hold his breath (usually during the maximum inhale phase) during beam delivery, which is turned on and off to treat only in a predefined position of the breathing cycle. There are two breath-hold techniques that can be employed: deep-inspiration breath-hold (DIBH) [80] and active breathing control (ABC) [81]. The DIBH technique involves coaching the patient to a reproducible deep inhalation breath-hold during planning treatment and treatment delivery. The applicability of DIBH is limited by patient compliance and most patients cannot perform the manoeuvre reproducibly [80]. ABC has the same principle as DIBH but it assists patients in holding their breath if they cannot do it on their own [81]. When the predefined respiratory phase is achieved, a balloon valve is actuated to block airflow to the patient, suspending the breathing. The result is a repeatable breath-hold that provides the same volumes each time.

Although PTV margins can be reduced with breath-hold techniques as there is very limited respiratory-induced motion, the treatment time is significantly increased compared to other strategies. In addition, the implementation of these techniques is extremely dependent on patient selection criteria (e.g., patient compliance and the existence of diaphragm motion during breath-hold) and it increases the patient's discomfort. For those reasons, free-breathing imaging techniques are usually preferred.

3.3.2 Forced shallow breathing

The forced shallow breathing (FSB) technique uses a stereotactic body frame with an attached plate that is pressed against the patient’s abdomen. The applied pressure to the abdomen reduces diaphragmatic motions, while still allowing limited normal breathing [82]. This technique has the disadvantage of increasing the patient's discomfort, which might limit the successful employment of this strategy for patients with comorbidities and poor lung function.

3.3.3 Respiratory gating methods

Respiratory gating techniques consist of the administration of radiation within a particular portion of the patient’s breathing cycle, commonly referred to as the “gate” [4]. To accomplish this technique, the position and width of the gate within a respiratory cycle are determined by monitoring the patient’s respiratory motion, using either an external respiration signal or internal fiducial markers. The RPM^{TM} system (Varian Medical Systems, Palo Alto, CA) is a current commercially available respiratory gating system that uses an external respiration signal (e.g., up and down movement of the diaphragm). The target motion is then correlated with the motion of the acquired respiration signal. The radiation beam is activated whenever the respiratory signal, and thus the tumour position, is within the predefined gate. Otherwise, the radiation beam is switched off. Figure 3.4 depicts a respiratory surrogate signal together with the beam-on and beam-off conditions during free-breathing respiratory-gated radiotherapy.

In this technique, the radiation to the healthy tissues is minimised and the geometrical uncertainties are reduced. Owing to its non-invasive nature, gating using an external respiration signal can be applied to more than 90% of patients [48]. However, due to the intermittent treatment delivery and increased set-up time, gated procedures are longer than conventional procedures. Besides, for internal and external tracking systems, a possible source of error is the fact that the surrogate for tumour motion tracked by the gating system does not accurately correspond to the time-dependent target position. This can cause the timing of the beam-on pulse to shift relative to the actual respiratory cycle of the target [48].

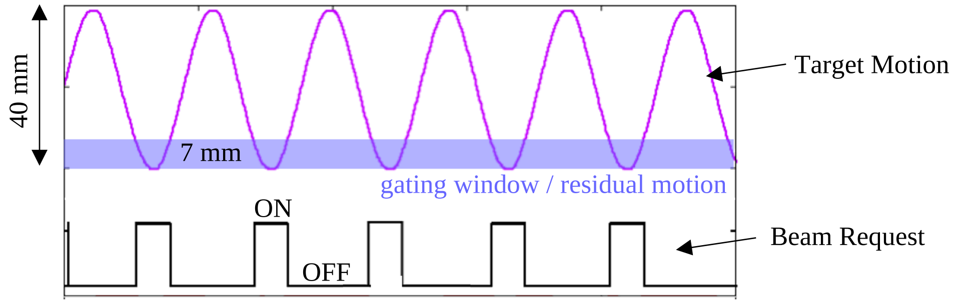


Figure 3.4: Respiratory surrogate signals of tumour motion and beam-on conditions during free-breathing respiratory-gated radiotherapy.

3.3.4 Real-time tumour tracking

Another means of accommodating respiratory motion is to dynamically align the radiation beam with the target during the delivery of radiation dose, referred to as real-time tumour tracking (RTTT). These RTTT techniques permit adequate dose coverage of moving targets whilst keeping the PTV relatively small if the radiation delivery is synchronized with the tumour movement and position [83]. Real-time alignment of the treatment beam and target lesion can be accomplished by two different approaches: moving the treatment beam by moving the entire LINAC or by moving the treatment couch or MLCs. The application of RTTT techniques is extremely hardware limited. The employment of RTTT techniques by moving the entire LINAC has been restricted to only two dedicated machines clinically in use: the CyberKnife Robotic Radiosurgery system (Accuray, Sunnyvale, CA, USA) and the Vero SBRT system (Mitsubishi Heavy Industries, Tokyo, Japan and BrainLab AG, Feldkirchen, Germany). The CyberKnife Robotic Radiosurgery system consists of a 6 MV linear accelerator mounted on a six-degree-of-freedom (DOF) industrial robot, which can track the tumour and deliver doses in different angles with pinpoint accuracy [84]. For the couch-based motion compensation method, a robotic couch moves in real-time in the opposite direction of tumour motion to cancel out the tumour motion [85]. For this method, there are concerns related to patient comfort and the latency between patient motion and internal organ motion. During MLC tracking, the MLC aperture is shifted according to the internal motion of the target lesion. Although this approach has been successfully employed in several radiotherapy applications, including lung cancers [86], it is not routinely used in the clinical mode (only in the research mode).

3.4 4D imaging techniques in the planning process

4D imaging techniques is a recent and promising technique in radiotherapy to deal with tumour motion due to breathing. However, 4DCT data cannot be handled as input for treatment-planning by the available commercial systems. Currently, several strategies have been developed to overcome this limitation and bring 4D data into the radiotherapy treatment-planning process, some of them still depend on in-house developed algorithms.

3.4.1 The ITV strategy

The internal target volume (ITV) strategy is the most widely used strategy in clinics to deal with moving targets during treatment planning [10], [87]. It aims to include all possible positions that the tumour can occupy during the respiratory cycle, i.e., in all phases of the 4DCT. The fluctuations in size and position of CTV because of respiratory motion are generally accounted for by adding an internal motion margin to the CTV, resulting in an ITV. The ITV method treats all the respiratory motion as a systematic error, therefore, patients with considerable tumour motion consequently have large planning margins, resulting in increased exposure of normal tissue to high doses and increased risks of radiation side effects. The strategies to determine the ITV from a 4DCT scan range from fully manual methods, where the radiation oncologist delineates the target in all breathing phases, to more automated methods such as the slow CT scan, the maximum intensity projection (MIP) scan, and average intensity projection (AIP) scan.

3.4.1.1 Slow CT

Slow CT scan consists of the acquisition of multiple CT scans in which the tube rotation time and pitch are considerably reduced. Then, these slices are averaged such that multiple respiration phases are recorded per slice. This results in a respiratory-blurred scan (which can be considered as the time average of the 4D data set) in which the image of the tumour should show the full extent of respiratory motion. The low image resolution due to motion blurring, which potentially leads to larger observer error during tumour/organs volumes definition, and the variation in the respiratory motion that may occur between imaging and treatment, lead to additional larger treatment-planning margins [48]. Another disadvantage is the increased dose from slow CT scanning compared to conventional CT scanning.

3.4.1.2 Maximum Intensity Projection (MIP) CT

The Maximum Intensity Projection (MIP) scans are created by selecting the maximum CT number (in Hounsfield units) at each image position in the entire 4DCT image set, thus representing the motion envelope of the tumour [87], [88]. Although this ITV technique is a valid strategy for appropriate dose coverage of the entire mobile target lesion, it results in larger target volumes than the planning volumes necessary for appropriate radiotherapy treatment [9] as mentioned before. This ends up increasing the exposure of normal tissues to high radiation doses and subsequently higher normal tissue complication probability (NTCP), limiting the possibility to perform dose escalation.

3.4.1.3 Average Intensity Projection (AIP) CT

The Average CT scan, or Average Intensity Projection (AIP) CT, is reconstructed taking the average voxel intensity values throughout the 4DCT data set. Although this is an easy and viable approach for defining the ITV without increasing the acquisition time, it results in a blurry scan and increased treatment planning margins.

3.4.2 The Mid-ventilation CT

An alternative approach to the ITV strategy is to perform the radiotherapy treatment planning on the Mid-Ventilation CT scan (Mid-V CT) since this approach aims to eliminate the systematic error caused by respiration and to obtain a more representative scan for the delineation of tumour and normal healthy tissues.

The Mid-V approach takes advantage of the cyclical nature of the respiratory motion, restricting the CTV to the area where the target is most often located during the breathing cycle and boosting the dose to this region [6]. Engelsman *et al*. [70] and Witt *et al*. [89] showed that if the tumour is irradiated at its average position during the respiration cycle, a good dose coverage can still be obtained even if the tumour is not fully within the high-dose region during a small fraction of the breathing cycle. Such an approach makes margin reduction possible because it is based on the criteria that the CTV receives at least 95% of the prescribed dose to 90% of the patient population and treats the motion amplitude as an additional random error during treatment. The ITV method expands the CTV to the total volume that the moving tumour covers, and the tumour motion amplitude is added linearly to create the ITV-based PTV margin, in which 100% of the dose is applied to that area. Also, the improvement in the visualization of tumour shape in the Mid-V CT will reduce the delineation uncertainties that are usually taken into account in the CTV-to-PTV margins.

The Mid-Ventilation technique relies on finding a time-weighted mean position[8] of the target lesion over the respiratory cycle. This information can be derived from 4DCT imaging data. Starting from the ten CT breathing phases, the tumour's position is defined in each breathing phase. An image registration procedure to a reference frame can then be used to access tumour motion and amplitudes along all three directions (cranio-caudal (CC), left-right (LR), anterior-posterior (AP)). Based on tumour motion curves, the time-weighted mean tumour position is calculated, and the Mid-Ventilation phase is defined as the breathing phase where the target is located closest to the mean position [15]. By using Mid-V CT scans for treatment planning, it is possible to reduce the treatment volume [6], [15], as depicted in Figure 3.5.

---

[8] The mean position of a moving tumour is the tumour-weighted mean position of the tumour's centre-of-gravity in all three directions (cranio-caudal (CC), left-right (LR), anterior-posterior (AP)).

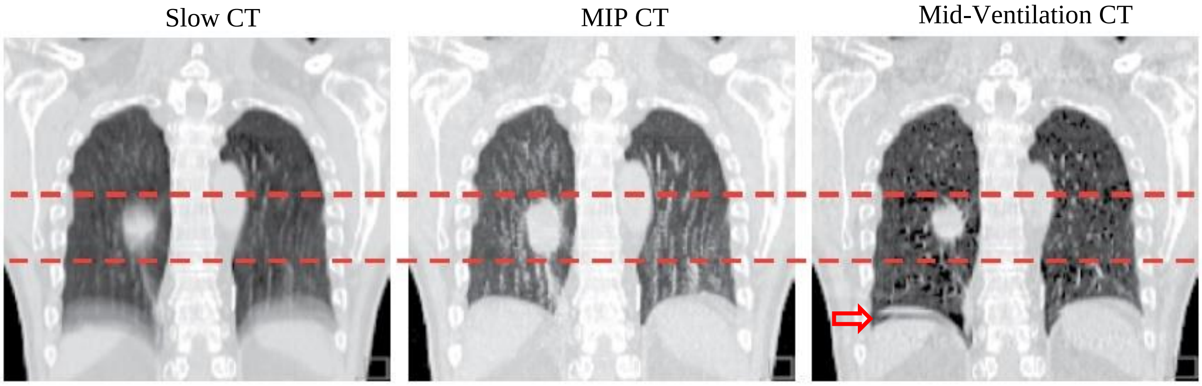


Figure 3.5: Comparison of a Slow CT, a Maximum Intensity Projection (MIP) CT and a Mid-Ventilation CT for the same patient [15]. The tumour appears larger on the MIP CT when compared to the Mid-V CT as indicated by the red lines. The boundary between the tumour and soft tissue can be more difficult to distinguish on the Slow CT than in the Mid-V CT. The clarity of the tumour boundaries is visually improved on the Mid-V CT but it has more noise and is more prone to motion artefacts as shown by the distortion of the dome at the lung–diaphragm interface (red arrow).

If hysteresis occurs, the time-averaged mean position is not however necessarily situated on the trajectory of the tumour [52]. Therefore, none of the phases of the 4DCT scan represents the mean position accurately, resulting in a remaining systematic geometrical error (Figure 3.6) [90]. Besides, the image quality in the Mid-V CT is poor since it only uses one-tenth of the 4DCT data, which itself might contain artefacts, and it is noisier than conventional CT scan since the radiation exposure per frame is lower. The loss of image quality related to the Mid-V CT scan inspired the development of the Mid-Position CT scan.

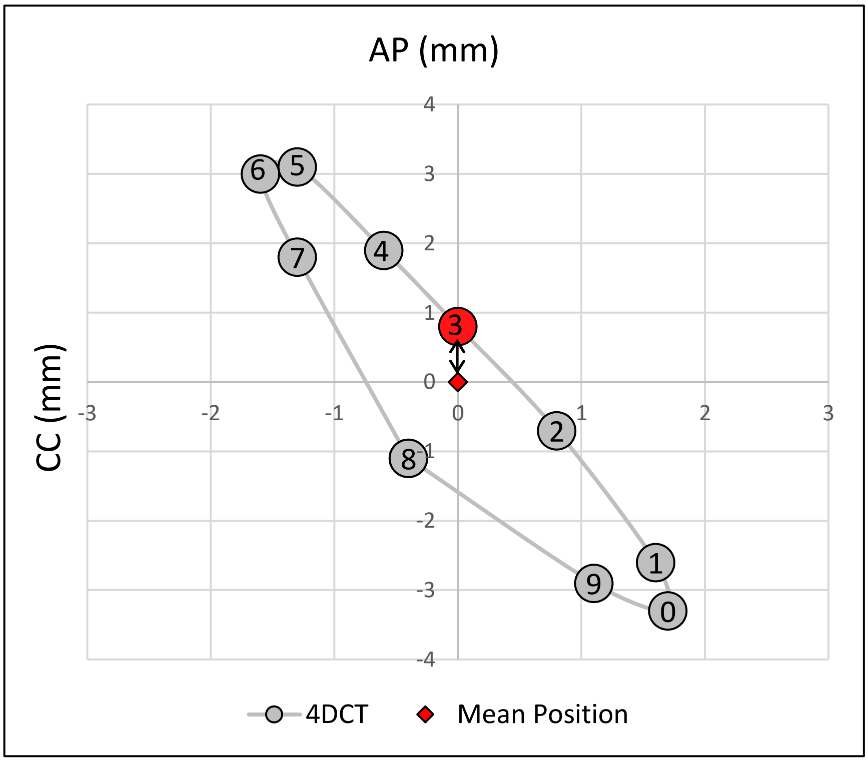


Figure 3.6: Schematic explanation of the reconstruction of the Mid-Ventilation (Mid-V) CT scan based on the tumour trajectory along the 4DCT. Hysteresis of the tumour motion due to respiration and a (small) systematic error between the Mid-V position and the time-averaged mean position is shown. The spheres show the positions of the tumour relative to the mean position. For simplicity, the third dimension (left-right direction) is not shown.

3.4.3 The Mid-Position CT

The Mid-Position CT scan (Mid-P CT) is considered a refinement of the Mid-V concept which represents all the internal structures, including the tumour, in their time-weighted mean position over the respiratory cycle, thus eliminating the (small) systematic error due to hysteresis, and improving image quality (Figure 3.7).

Wolthaus *et al*. [8] developed a pipeline to reconstruct the Mid-P CT. It is reconstructed from the entire 4DCT dataset by transforming each breathing phase to the time-averaged mean position using deformable image registration (DIR). Since the Mid-P CT uses the information of all frames (breathing phases) of the 4DCT, this approach is more efficient for dose calculations and it has higher image quality compared to one 4D phase and Mid-V CT scans, leading to a higher signal to noise ratio, and removal of motion artefacts [15].

With the Mid-P approach, systematic errors due to respiration are nearly reduced to zero and the absence of motion blurring will potentially improve the identification of tumour and OARs. Therefore, the use of the Mid-P CT results in smaller CTV-PTV margins in the treatment planning than the more commonly used ITV method, without significant reduction of tumour coverage [6], [91]. For the lung SBRT, liver SBRT, and pancreatic cancer RT, the Mid-P or Mid-V approach reduced the PTV volume (compared with the ITV approach) by 23% [63], 34% [92], and 13.9% [93], respectively. Besides, due to the elimination of the systematic error due to hysteresis, the Mid-P approach is more applicable to the outcome of Engelsman *et al*. [70] and Witt *et al*. [89] than the Mid-V approach. Nevertheless, variation in the average position of the tumour (baseline shift) gives a contribution that remains in 4D treatment-planning. This contribution can only be reduced when using offline or online 4D verification protocols [15].

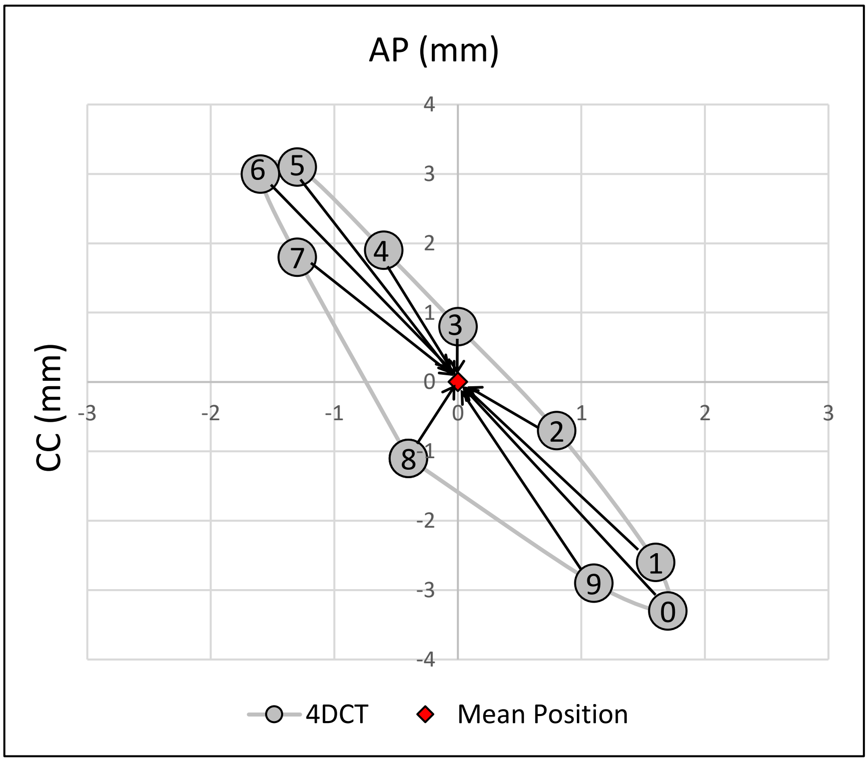


Figure 3.7: Schematic explanation of the reconstruction of the Mid-Position (Mid-P) CT based on the tumour trajectory along the 4DCT. Hysteresis of the tumour motion due to respiration is shown and there is no shift between the Mid-P and the time-averaged mean position. The spheres show the positions of the tumour relative to the mean position. For simplicity, the third dimension (left-right direction) is not shown.

The Mid-Position approach has been implemented at the Netherlands Cancer Institute (NKI) for over 10 years and shown to provide more accurate radiotherapy [6]. Despite the promising results, the current uptake of Mid-Position is still as low as 6% among fifty radiotherapy centres surveyed at the British Thoracic Oncology Conference in 2016 [10]. One of the reasons for the lack of adoption of the Mid-P method is the lack of commercial systems to calculate the Mid-P CT scan. It usually relies on in-house software solutions, and requires extensive validation prior to clinical deployment, increasing the workload at the institution.

4 Materials and Methods

This work aims to bring the Mid-Position approach to the clinic, namely at Champalimaud Clinic Centre (CCC). This Chapter is focused on the optimization of the Mid-Position module developed for the open-source 3D Slicer software as well as on a practical framework to validate its accuracy and feasibility to be clinically used. First, the 3D Slicer software and the main steps of the in-house module are presented. As a second goal, this Chapter details a set of experiments to validate the software which is divided into three studies as described in Table 4.1. The in-house developed Mid-P software was benchmarked to an automated science prototype developed by Mirada Medical Ltd (UK) and to a package in current clinical use at the Netherlands Cancer Institute (NKI) (NL) called Wimp. For simplicity, the three Mid-Position packages will be referred to as 3D Slicer (or *RunMidP*), Mirada and Wimp, respectively, throughout this work

Table 4.1: Dataset and objectives description of the three different studies conducted to validate the Mid-Position module.

Study	Dataset description	Objectives
A	4DCT data from 2 digital phantoms and 18 cancer patients	Image registration evaluation (comparison between the Plastimatch and Elastix software) and Image quality optimization Evaluation of the algorithm accuracy to reconstruct the time-weighted mean position and estimate motion amplitudes
B	4DCT data from 16 phantoms derived from real patients' data	Evaluation of the algorithm accuracy to reconstruct the time-weighted mean position and estimate motion amplitudes Evaluation of the effect of the differences in the motion amplitude estimation in the planning treatment margins Evaluation of the intensity voxel distribution in relevant anatomical structures Evaluation of choice of the reference phase (as an *input* to the algorithm) in the Mid-P CT reconstruction
C	4DCT data from 5 cancer patients	Image quality assessment of the Mid-Position CT scan both quantitatively (signal-to-noise ratio and image sharpness) and qualitatively (from the doctors' perspective).

4.1 3D Slicer

The Mid-Position software module aspires to be used in the 3D Slicer software. As a free, open-source and multi-platform software package, 3D Slicer[9] is widely used for biomedical research since it provides a powerful Graphical User Interface (GUI) for versatile image visualization and advanced analysis tools, such as image segmentation and registration. It also supports import and export data from a wide range of standard formats, including DICOM. The software uses several libraries which provide a high level of functionalities: the CLIP Tool Kit (CTK) and Qt packages for the GUI; the Insight ToolKit (ITK) for Inputs/Outputs and image processing; and the Visualization ToolKit (VTK) for

---

[9] 3D Slicer, https://www.slicer.org/ (Accessed on 13/07/2021).

visualization [94]. 3D Slicer is extended to enable users to download and install plugin modules through the *Slicer Extensions Manager*. Users can also develop their own modules, developed in either Python or C++, and integrate them into the 3D Slicer software through the *Extension Wizard*. All objects (volumetric images, surface models, transforms, etc.) in 3D Slicer are stored in a hierarchical structure of nodes encoded in the *Medical Reality Modeling Language* (MRML), enabling modules to have access to the MRML tree and additional characteristics of the data object [94]. For this work, the *SlicerRT*, the *Plastimatch* and the *SlicerElastix* modules were downloaded as plugin modules, and they are required to run the Mid-Position module.

4.1.1 Slicer RT extension

SlicerRT[10] is an extension of 3D Slicer comprising almost twenty modules for radiotherapy research, such as DICOM-RT import and export for structure sets, dose distributions and treatment plans visualization, dose calculation and dose comparison (with gamma analysis), dose-volume histogram calculation, among other functionalities [94].

4.1.2 Plastimatch software

Plastimatch[11] is open-source software focused on high-performance algorithms for automatic deformable image registration and other useful tools for radiotherapy research and clinical practice (e.g., CBCT reconstruction and generation of digitally reconstructed radiographs (DRRs)) [95].

4.1.3 Elastix software

Elastix[12] is a promising registration toolbox that includes several optimization methods, multiresolution schemes, interpolators, transformation models, and cost functions. Although the default settings work properly for a wide range of registration tasks, Elastix enables users to define their own registration settings for a specific application [58]. Elastix provides N-th order B-spline interpolation to deform the moving image, instead of using the nearest neighbour or the linear interpolation as the case of the Plastimatch software. Setting the B-spline interpolation order to 2 or higher is expected to improve the registration accuracy and the final image quality of the Mid-P CT.

4.2 In-house Mid-Position module implementation

Lacking commercial systems to reconstruct the Mid-Position CT scan (Mid-P CT), the first version of a Mid-Position software, written in Python programming language, was developed for the freeware 3D Slicer (v4.4) in 2015 by Kristjan Anderle, a Ph.D. student at that time [96]. Since 3D Slicer had great improvements in the latest nightly version and some previous modules were replaced, the module software was no longer in use.

To allow the Mid-P approach to be part of the radiotherapy clinical practice of moving tumours at CCC, the previous module was optimized for the 3D Slicer latest version (v4.13) and it was called the *RunMidP* module. Briefly, the software analyses 4DCT data of patients with moving tumours in the upper abdomen or lung and calculates the Mid-P CT by application of deformable image registration

---

[10] SlicerRT, http://slicerrt.github.io/ (Accessed on 13/07/2021).

[11] Plastimatch, http://plastimatch.org (Accessed on 13/07/2021).

[12] Elastix, https://elastix.lumc.nl/ (Accessed on 13/07/2021).

(DIR) between breathing phases. It subsequently can determine the breathing amplitudes of tumours and healthy organs, which should be used for the calculation of safety margins during planning.

As it will be mentioned throughout this work, the main improvements of the *RunMidP* module over its first version consist of: (1) generation of the Mid-P CT scan with minimal user interaction; (2) use of a recent and powerful DIR toolbox; (3) application of transforms without changing image geometry; (4) new approach to calculate motion amplitudes; and (5) option to visualize the original 4D data as a movie. The changes stated in points 2 and 3 have resulted in improvements regarding the image quality (sharpness) and precision of the Mid-P method, respectively.

To provide a clear description of the module, the *RunMidP* is divided into four main sections: (1) Inputs section; (2) 4DCT section; (3) Log text box, and (4) *Run* button as depicted in Figure 4.1. For comparison, the GUI of the first version of the module is shown in Appendix II.

- **Section 1**: In the "*Inputs*" section, the user can select the patient ID and the corresponding planning CT in which relevant anatomic structures, such as the tumour and OARs, were delineated. These data are later used to calculate the motion amplitudes for relevant ROIs.
- **Section 2**: The "*4DCT*" section is where the user can optionally create the Average CT scan and visualize the 4DCT data by pressing the *Create Sequence of 4DCT* button and then the *Play* button in the toolbar, which animates the time series at a chosen speed.
- **Section 3**: Text box containing all logs generated during module execution.
- **Section 4**: By hitting the *Run* button, the registration process and the calculation of the Mid-P CT start. Alternatively, the *Run* button can also be pressed if the user wants to calculate motion amplitudes. This only occurs if Section 1 is fulfilled with the patient's information.

The *RunMidP* module is an end-user application intending to be inserted into the clinical practice, without increasing the workload. Figure 4.2 shows the software pipeline to reconstruct the Mid-P CT and to calculate the motion amplitudes.

Firstly, the user imports the patient's data (e.g., 4DCT dataset, DICOM structures and the planning CT) into Slicer's DICOM database. After starting the *RunMidP* module, the user can immediately press the *Run* button without any additional input information. This starts the 4DCT registration required for the Mid-P CT calculation, thereby reconstructing the Mid-P CT scan for all patients in Slicer's DICOM database.

After the generation of the Mid-P CT, it is exported as DICOM to the patient's directory and imported automatically into the Slicer's DICOM database. After the task nr.6 finishes (Figure 4.2), the module checks (at every 10 s) if the user introduced any information at the *Inputs* section (Section 1). If not, the *RunMidP* module performs the Mid-Position workflow for the next patient in Slicer's database or it continues waiting until a new patient is added into the Slicer's database. This method enables the calculation of the Mid-P CT automatically whenever 4DCT and planning data from a new patient is imported into the database.

As soon as the Mid-P CT is reconstructed, doctors can use it for visualization and contouring of anatomical structures. At that time, the planning data can be further loaded into the 3D Slicer software and set at Section 1 to calculate the breathing motions. Finally, the *RunMidP* module enables the user to calculate the Average CT scan and to visualize the 4DCT data changing in time.

In the next sections, the reconstruction of the Mid-P CT scan and the calculation of motion amplitudes will be explained in more detail.

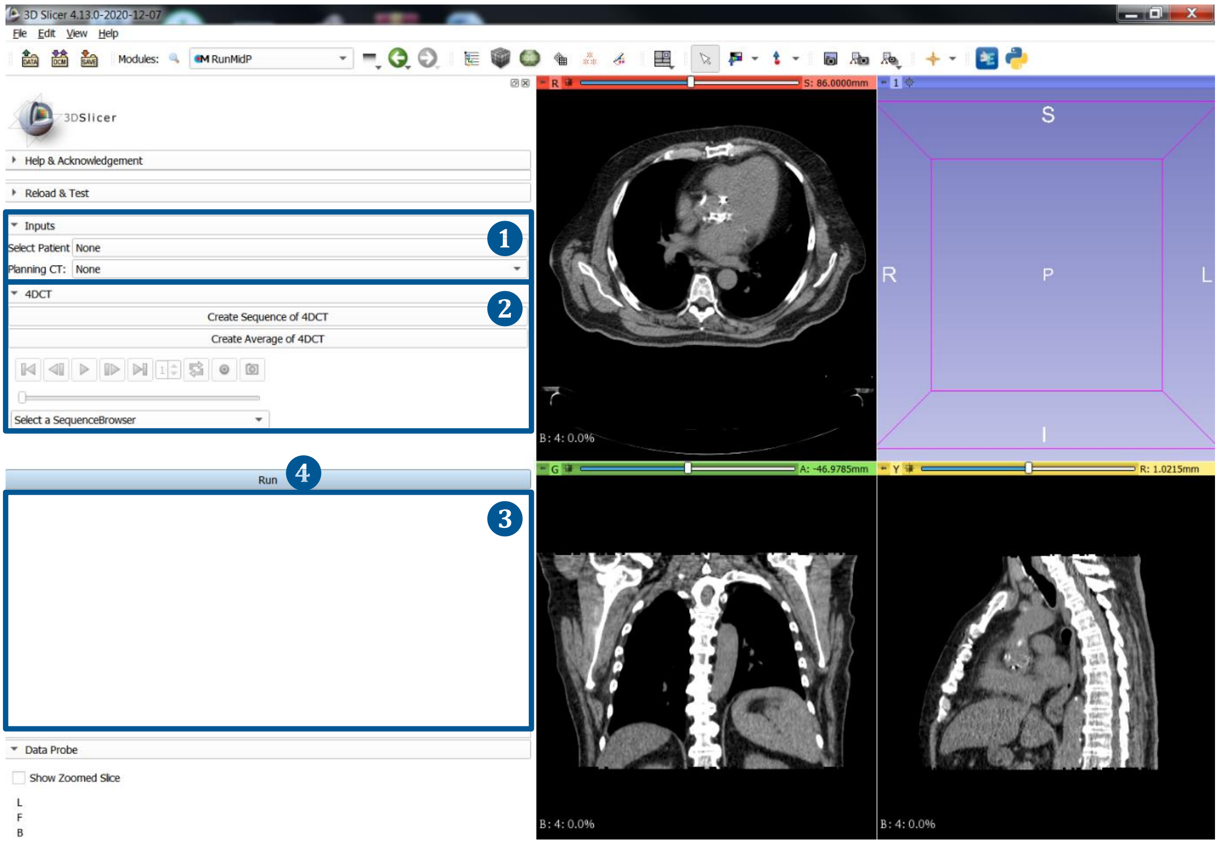


Figure 4.1: The left side shows the graphical user interface (GUI) of the *RunMidP* module. Four different parts of the module are highlighted: (1) Inputs section, (2) 4DCT section, (3) Log text box and (4) *Run* button. The axial, sagittal and coronal views show the first breathing phase of the 4DCT data (Phase 0%) but can also be used to visualize the Mid-P CT after the module is finished.

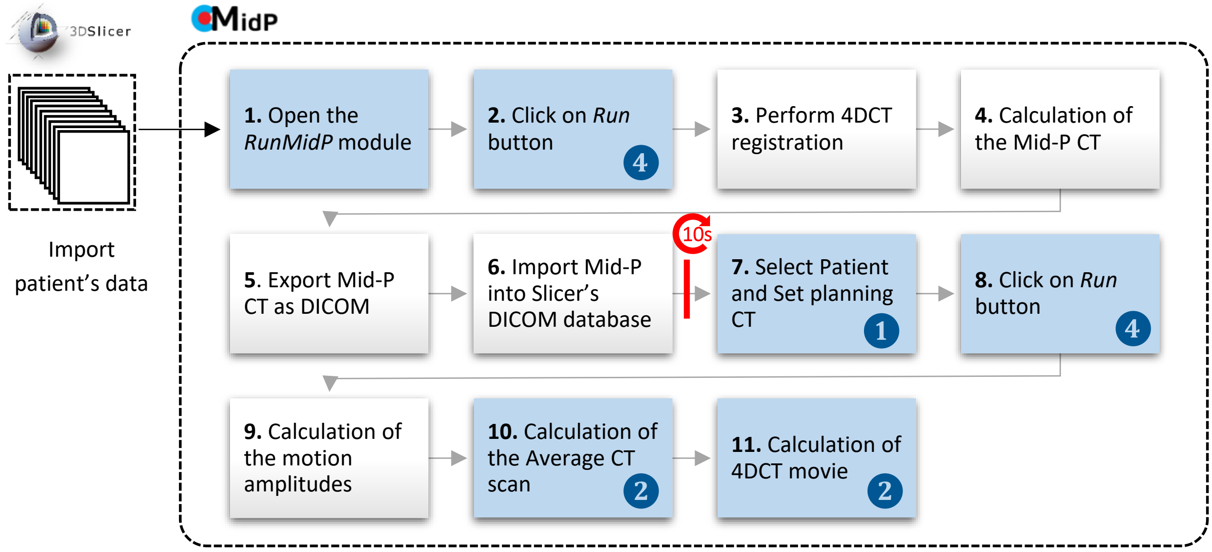


Figure 4.2: Pipeline for the *RunMidP* module. The highlighted tasks (nr. 1,2,7,8, 10 and 11) require user interaction at the specified module sections, while the other tasks run in the background. After task nr.6 is finished, if the user sets the Patient ID and the planning CT (at Section 1), the *RunMidP* module continues for task nr.9, otherwise it checks at every 10 seconds if

there is any new patient at the Slicer's DICOM database to calculate the Mid-P CT. The last two steps (tasks nr. 10 and 11) are optional, and they can be performed before or after the Mid-P CT calculation.

4.2.1 Reconstruction of Mid-Position CT

The reconstruction of the Mid-Position CT (Mid-P CT) followed a similar underlying algorithm detailed in Wolthaus *et al*. [8] and it is presented in Figure 4.3. From the ten phases of the 4DCT scan, one reference phase[13] is chosen, and all phases are deformably matched to the reference scan resulting in nine deformation vector fields (DVF). Since the same reference phase was used to compute each DVF, the mean motion of all voxels can be computed by averaging the nine DVFs, yielding the mean DVF. Each DVF is subsequently used to warp all breathing phases (expect the reference scan) to the reference phase. The mean DVF is then applied to the 10 scans (9 warped scans + 1 reference), transforming each 4DCT frame from reference to the time-weighted mean position. Finally, the ten transformed CT scans are averaged, taking the median of pixel intensities at each point, to obtain the Mid-P CT scan.

Differences between the method described in this work and detailed in [8] stem from differences in deformable image registration (DIR) methods, which are expected to affect their performance.

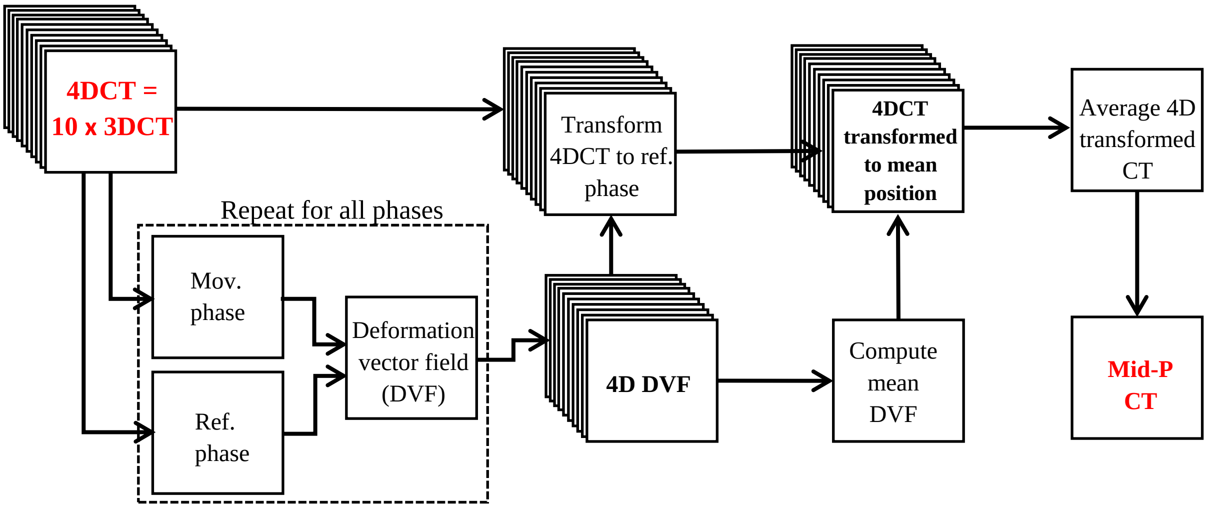


Figure 4.3: Pipeline to calculate the Mid-P CT scan. From the 4DCT scan, a reference frame and a moving phase are selected, which are registered to each other with deformable image registration. This process is repeated for all frames of the 4DCT scan, registering each of them to the same reference frame. The resulting deformation vector field (DVF) is recomputed to deform all CT frames to the time-weighted mean position. The deformed frames are averaged to obtain a high-quality Mid-Position CT scan. Adapted from [8].

Image registration is the process through which DVFs and warped scans are computed. When the objective is to register a pair of pulmonary CT images, DIR is required due to the elastic nature of the lung tissue. As depicted in Table 4.1, one of the objectives for Study A is to compare two DIR algorithms – Plastimatch and Elastix – which are described below.

---

[13] For study A, the reference phase was set as the breathing phase 30%. For study B and C, the reference phase was set as the phase 0%. Theoretically, the Mid-P CT should be independent of the choice of the reference phase, but small differences can be found between Mid-P CT scans reconstructed using different reference phases since the noise and image artefacts differ from each breathing phase of the 4DCT dataset (see Section 5.2.5 for the robustness study).

For the two image registration software, a B-spline-based registration using a pipelined, multi-stage registration framework was used to improve speed, accuracy and robustness of the algorithms. The parameter files which specify the settings to be used in each stage are presented in Appendix III (Plastimatch registration) and Appendix IV (Elastix registration).

For the Plastimatch software, the MSE was used as the similarity metric along two processing stages, where the image resolution increases and the B-spline grid spacing decreases at each stage. Specifically, the first stage was performed on a $4 \times 4 \times 2$ subsampled image using a grid size of 50 mm. The second stage was performed on a $2 \times 2 \times 1$ subsampled image with a grid size of 25 mm. Decreasing the image subsampling rate and the grid spacing allows performing a finer registration. For the two stages, a regularization term of 0.005 was used to control the smoothness of the vector field (see Appendix III). These parameters were set by the Ph.D. student when he started working on the Mid-Position module [96].

During the registration process and to generate the final (warped) image, interpolation is needed to calculate intensity values at non-voxel positions. However, the Plastimatch registration only provides the nearest neighbour and linear interpolation to achieve this goal. Therefore, the Elastix registration toolbox was implemented in the *RunMidP* module to register the breathing phases of the 4DCT with the advantage of using a higher order of interpolation (cubic), which usually produces high-quality images after registration.

Elastix provides a collection of successful parameter files that can be found at [97]. Most of them consist of multiple transformations, in which a pre-alignment of images is performed by rigid registration and the final alignment is completed by a B-spline deformable transformation. In 4DCT data, the patient's anatomy is nearly the same over the breathing phases, so there is no need to perform the pre-alignment stage. In this regard, a customized parameter file was added to the Elastix database in order to achieve a high amount of control over the registration which is specific for 4DCT data.

As described in Appendix IV, a multi-resolution non-rigid registration was performed along 4 stages, using the MI (Equation 2.8) as the registration's similarity metric and a grid spacing of 14 mm. The control point spacing of the B-spline transformation is one of the most important parameters to be set since it controls the accuracy of the registration. Low control point spacing may improve the accuracy of the registration but may also increase the computation time and can cause unrealistic deformations, especially on homogenous areas where no image features, such as edges, can guide the registration. Thus, a grid spacing of 14 mm was found to be the best trade-off between the accuracy, speed, and realism of the transformation for the majority of the cases presented in this work. To reduce the data complexity, the input images were smoothed (and not downsampled) using a gaussian filter and a smoothing schedule of 16 16 8, 8 8 4, 4 4 2, 2 2 1, 1 1 1, which means that at resolution level 0 images are blurred with $\sigma = 16/2$ mm in the x- and y-direction and with $\sigma = 8/2$ mm in the z-direction. In the last resolution level, the original images were used. Note that σ is half the pyramid schedule value and less smoothing were performed in the z-direction.

As a second multiresolution pipeline, the B-spline transformation was carried out by a grid spacing schedule of 8, 4, 2, 1. These factors, in combination with the grid spacing of 14 mm, determine the final grid spacing used for all resolution levels. For instance, a grid spacing of $8 \times 14 = 112$ mm in resolution level 0, via 56 mm and 28 mm, to 14 mm in the final resolution level was used. Starting with a coarse grid, larger anatomical structures are matched first. In later resolution stages, a finer grid is used to align smaller structures, up to the final precision. The maximum number of iterations was set to 500 at each stage. This iterative coarse-to-fine image scale approach with multiple registration steps improves the registration accuracy as well as the registration speed.

4.2.2 Calculation of motion amplitudes

The motion of tumours and OARs is hard for clinicians to assess visually. The resulting DVFs from the image registration process can be used to estimate the breathing motion of any planning volume or fiducial marker. In the *RunMidP* module, two different methods to estimate motion amplitudes were implemented:

Centroid Motion (CM) method: It computes the peak-to-peak motion amplitude of a volume of interest from the relative motion of the centre of mass (COM) of each contour. Firstly, the contours from the planning CT are warped to the reference phase and the COM's coordinates are taken as the reference origins. DVFs are then used to propagate the contours to the 9 breathing phases and the COM of each contour is calculated. The peak-to-peak amplitudes are estimated by subtracting the minimum and maximum relative motion in the three directions: LR, AP, CC.

ROI Motion (RM) method: The peak-to-peak motion amplitude is calculated by taking the average amplitude of the DVFs within ROIs. The contours from the planning CT are warped to the reference phase, and then to the other remaining breathing phases using the DVFs transforming each 4DCT phase to the Mid-Position, as in the CM method. Using the warped contours, the magnitude of the displacement of every voxel within ROIs can be determined for the three directions (LR, AP, and CC). Finally, motion amplitudes are estimated by taking the average of peak-to-peak amplitude of every voxel within the tumour contour.

The advantage of the RM method over the CM method is the fact that taking the standard deviation of motion amplitude estimation it is possible to have an idea of the uncertainty of the method. This is not possible using the CM method since only one point (the tumour's COM) is considered.

4.3 Mid-Position algorithm validation

4.3.1 Study A: Image Registration Evaluation and Image Quality Optimization

The Mid-Position algorithm validation was carried out over three different studies. The first study (Study A) aimed to compare the Plastimatch and the Elastix toolbox registration methods in terms of registration accuracy and image quality. The differences between the tumour position in the Mid-P CT and the actual mean position of the tumour were compared to evaluate if the Mid-P CT gives a good representation of the patient's anatomy in the time-weighted mean position. The peak-to-peak motion amplitudes calculated from the *RunMidP* module were also compared against ground truth estimates of tumour motion.

4.3.1.1 Data Sets

In the Department of Radiotherapy of the Champalimaud Foundation, the patients who have tumours that move with normal respiration undergo 4D studies as part of their radiation therapy planning according to routine clinical protocols. Study A includes anonymized 4DCT datasets from 2 digital phantoms and 18 cancer patients (8 lung, 6 liver and 4 pancreas). Tumour location and volume for each phantom and patient are summarized in Table 4.2. For liver- and pancreatic lesions, implanted gold markers close to the CTV were evaluated instead since the tumour itself is hardly visible in the absence of contrast agents on the 4DCT data.

For each patient, the 4DCT data were acquired on a combined PET-CT scanner (Vereos Digital PET/CT from Philips Healthcare, Cleveland, OH) in helical mode, with the patient in a head-first supine position on a couch using knee and neck support and arms raised above the head. Parameters for image acquisition were rotation time of 0.44s, matrix of 512 × 512 voxels and tube potential of 120 kV, and a dose range of 200-250 mAs. The pitch is set by the CT operator according to the patient breathing pattern and the dose is adapted for each patient using the iDose reconstruction algorithm (Philips Healthcare, Cleveland, OH). During CT acquisition, a Pneumo Chest Bellows system (Philips Healthcare, Cleveland, OH) was employed to record the respiratory signal while the patient breaths freely. The tension of the belt provides an external signal for synchronization to the phase of the breathing cycle. Lower signal amplitude corresponds to the exhalation and higher amplitude corresponds to the inhalation phase of the breathing cycle. Abdominal comprehension was used to reduce respiratory-induced motion.

The CT scanner is integrated with a dedicated software in which 10 respiratory frames were retrospectively reconstructed using phase sorting. Each frame represents 1/10th of the breathing period and contains about 150 slices with a slice spacing of about 1 mm in the LR and AP directions and 2 mm in the CC direction. Since the radiotherapy planning session is not yet fully relied on 4D imaging, a free-breathing planning CT is also acquired with or without the administration of a contrast agent. The radiation oncologist then uses the free-breathing planning CT for contouring the tumour and OARs using the Average CT scan to assess the extent of organs motion due to respiration.

Validation with real patient data is a complex process, requiring considerable additional effort in tumour delineation and in determining the exact tumour location. Therefore, two digital anthropomorphic phantoms based on the noise-free 4D extended cardiac-torso (XCAT) phantom [98] were also used in this study to validate the Mid-Position algorithm. XCAT phantoms provide simplified (but still realistic) models of the human anatomy upon which physiological functions can be modelled through user-defined parameters. Therefore, parameterized models for the cardiac and respiratory motions were incorporated into the XCAT phantoms based on the existing gated 4DCT datasets obtained from normal subjects. The phantom's CT and physiological models were provided by the XCAT phantom program (dxcat2[14]).

The respiratory mechanics of the XCAT phantoms are controlled by two curves: (1) the diaphragm curve, which controls the motion of the diaphragm and surrounding organs such as the liver, stomach, spleen, and kidneys; and (2) the chest curve which controls the expansion and contraction of the ribcage. The first XCAT phantom (LP1) consists of 10 frames with a voxel resolution of 1 mm with a cardiac and respiratory period of 1s and 5s, respectively. The maximum diaphragm motion and chest expansion were set 2 cm and 1.2 cm, respectively, according to normal breathing. A well-defined spherical lesion (2 cm in diameter) was placed in the right lobe near the diaphragm which moves according to the motion of the lungs. Figure 4.4 and Figure 4.5 show the 10 frames of the 4DCT data from LP1 and a lung cancer patient (L1) with a sphere/tumour following the respiratory motion, respectively. The second XCAT phantom (LP2) was designed similarly to LP1 but with a voxel resolution of 0.5 mm and a sphere of 1 cm in diameter placed in the left lobe of the lung.

---

[14] A trial version of the software was provided by Dr. Paul Segars.
Digital Phantoms, https://www.hopkinsmedicine.org/radiology/research/divisions/radiological-physics/research/projects/imaging-simulation-computer-phantoms.html (Accessed on 13/07/2021).

Table 4.2: Patient and tumour characteristics of Study A.

Patient ID	**CT size**	**CT resolution (mm)**	**Structure delineated**	**Structure location**	**Structure volume (cm^3)**
LP1	256 × 256 × 301	1 × 1 × 1	Sphere	RLL	1.5
LP2	500 × 500 × 501	0.5 × 0.5 × 0.5	Sphere	LLL	0.6
L1	512 × 512 × 89	1.172 × 1.172 × 2	CTV	RLL	2.8
L2	512 × 512 × 141	0.977 × 0.977 × 2	CTV	RUL	7.0
L3	512 × 512 × 120	0.879 × 0.879 × 2	CTV	RML	2.0
L4	512 × 512 × 73	0.977 × 0.977 × 2	CTV	LLL	1.3
L5	512 × 512 × 109	1.172 × 1.172 × 2	CTV	LUL	11.2
L6	512 × 512 × 104	1.021 × 1.021 × 2	CTV	LUL	1.9
L7	512 × 512 × 280	0.863 × 0.863 × 1	CTV	RLL	1.6
L8	512 × 512 × 136	0.977 × 0.977 × 2	CTV	LLL	0.5
LV1	512 × 512 × 182	0.992 × 0.992 × 2	Marker	Liver SegIII	17.8*
LV2	512 × 512 × 142	1.082 × 1.082 × 2	Marker	Liver SegII Liver	0.9* 184.0*
LV3	512 × 512 × 128	1.133 × 1.133 × 2	Marker	Liver SegVIII	17.2*
LV4	512 × 512 × 79	0.949 × 0.949 × 2	Marker	Liver	53.4*
LV5	512 × 512 × 110	0.929 × 0.929 × 2	Marker	Liver SegVII	38.9*
LV6	512 × 512 × 103	0.836 × 0.836 × 2	Marker	Liver SegV	3.6*
P1	512 × 512 × 78	1.172 × 1.172 × 2	Marker	Pancreas	76.9**
P2	512 × 512 × 99	1.172 × 1.172 × 2	Marker	Pancreas	15.0**
P3	512 × 512 × 110	0.934 × 0.934 × 2	Marker	Pancreas	66.9**
P4	512 × 512 × 132	1.172 × 1.172 × 2	Marker	Pancreas	87.8**

Abbreviations:

LP: Lung Phantom

L: Lung

LV: Liver

P: Pancreas

CTV: Clinical Target Volume

RLL: Right Lower Lobe | LLL: Left Lower Lobe | RUL: Right Upper Lobe | RML: Right Median Lobe | LUL: Left Upper Lobe

*/** These volumes refer to volumes of liver (*) or pancreatic (**) lesions close to the implanted gold marker

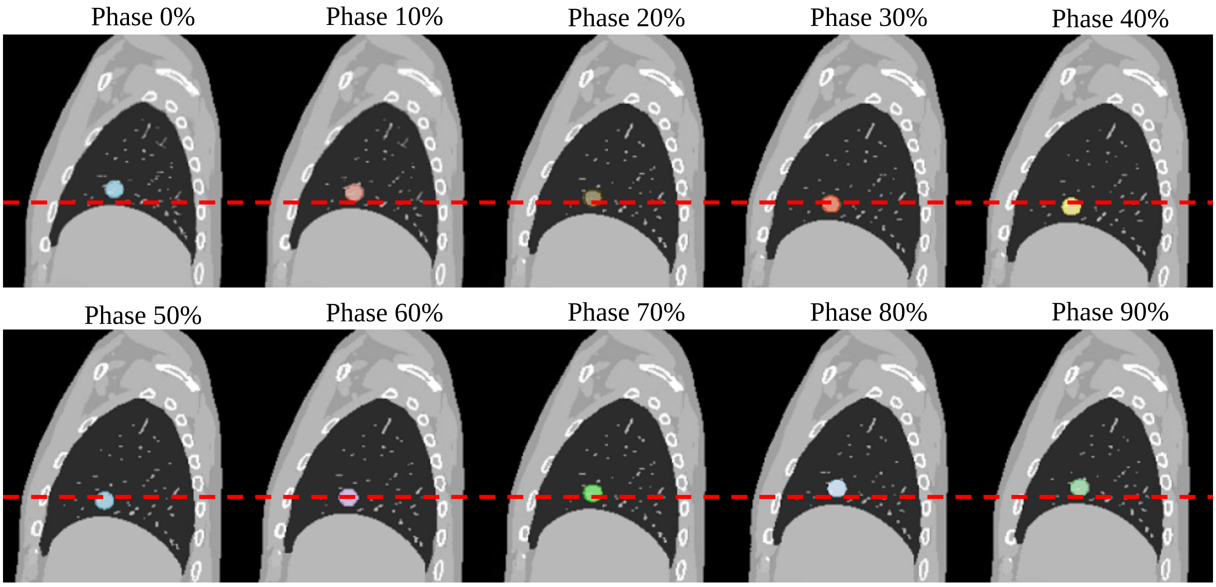

Phase 0% Phase 10% Phase 20% Phase 30% Phase 40%
Phase 50% Phase 60% Phase 70% Phase 80% Phase 90%

Figure 4.4: Simulated 10-phase 4DCT images based on the XCAT phantom (LP1) with a sphere of 2 cm for a regular breathing profile. Dashed lines were added to facilitate the visualization of respiratory motion.

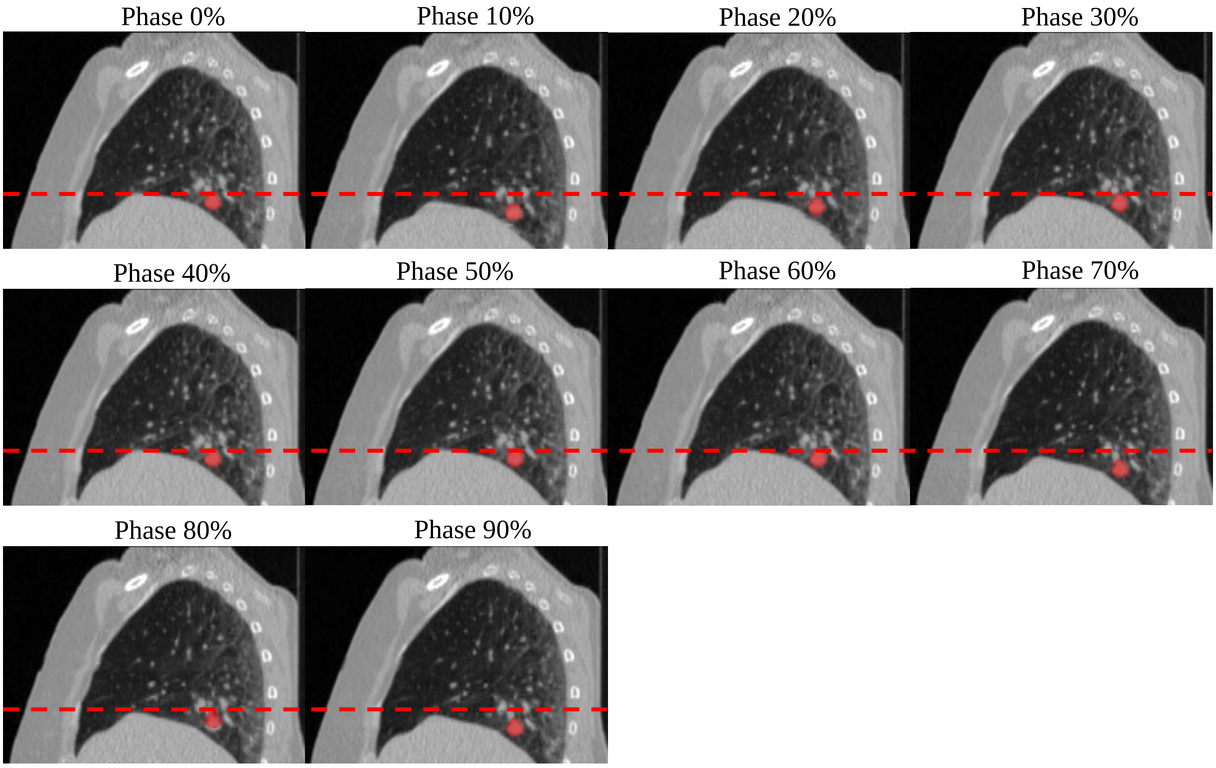

Phase 0% Phase 10% Phase 20% Phase 30%
Phase 40% Phase 50% Phase 60% Phase 70%
Phase 80% Phase 90%

Figure 4.5: 10 frames of the 4DCT from a lung cancer patient with a ROI encompassing the tumour. Dashed lines were added to facilitate the visualization of respiratory motion.

4.3.1.2 Image Registration Evaluation

Plastimacth and Elastix registration software were used to perform deformable registration using a cubic B-spline in order to match the original and deformed frame corresponding to two different breathing phases. The registration accuracy of both registration methods (Plastimatch and Elastix) was quantified using the Root Mean Square Error (RMSE). The warped image should, ideally, exactly match the original (reference) frame, yielding a small image difference. The larger the RMSE, the larger the mismatch between the original and deformed image. The RMSE was calculated between all breathing phases against the reference phase, which is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(I_i^{reference} - I_i^{warped})^2}{N}} \tag{4.1}$$

where $I_i^{reference}$ and I_i^{warped} represents the intensity of the ith voxel within the reference and warped image, respectively.

In addition to accuracy, the choice of the registration method depends on the image quality of the Mid-P CT. The image quality was evaluated by visual inspection of the Mid-P CT reconstructed by the two registration methods and by using the Structural Similarity (SSIM) index. The SSIM index extracts three features from an image: luminance (l), contrast (c), and structure (s). Given two images x and y, the SSIM index between both images is given by [99]:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{4.2}$$

where $\alpha > 0, \beta > 0$ and $\gamma > 0$ are parameters used to adjust the relative importance of the three components. For simplicity, α, β and γ were set equal to 1. The SSIM index was quantified (slice by slice) for the Mid-P CTs reconstructed with Plastimatch and Elastix software, using the Mid-P CT calculated by an automated research prototype (developed by Mirada Medical Ltd.) as reference.

For the Elastix registration method, the Hausdorff distance (HD) and the Dice similarity coefficient (DSC) between manually delineated tumours in the 4DCT with those obtained after warping the original RT structures, with the DVFs derived after DIR, were calculated for all frames of the 4DCT of the lung phantoms (LP1 and LP2) and lung patients (L1 to L8). HD and DSC measure the anatomical correspondence between contours. For a perfect registration (and perfect delineation), the two contours should match. In practice, they will not perfectly match due to contour variation. As an example, Persson *et al.* [65] quantified the interobserver delineation variation to be 2.6 mm for SBRT of peripheral lung tumours. Interobserver variation is typically larger than intraobserver variability [65], so when evaluating the registration accuracy, having the same observer contouring the structures on images, as is the case in this study, reduces the uncertainty in the metrics. Besides, using well-defined tumours in lung phantoms minimizes the contouring variability. The degree of mismatch was quantified using the HD and DSC and it indicates the quality of the registration from an anatomical perspective. Liver and pancreatic cancer patients were excluded from this experiment since image artefacts/distortions in the CT arise for the implementation of fiducial markers, preventing its precise delineation.

The last quantitative test to evaluate the accuracy of the Elastix registration algorithm used the publicly available POPI-model data set [100], containing a list of 41 landmarks identified manually by a medical expert in each frame of the 4DCT data. Each set of landmarks was warped to the reference phase using the DVFs. Figure 4.6 shows the projection of the landmarks identified in phase 0% (= reference phase, yellow) and in phase 60% (= moving phase, red) of the 4DCT.

Statistical significance between the two registration methods (Elastix and Plastimatch) was evaluated using the Mann–Whitney U test for two independent samples ($p < 0.05$).

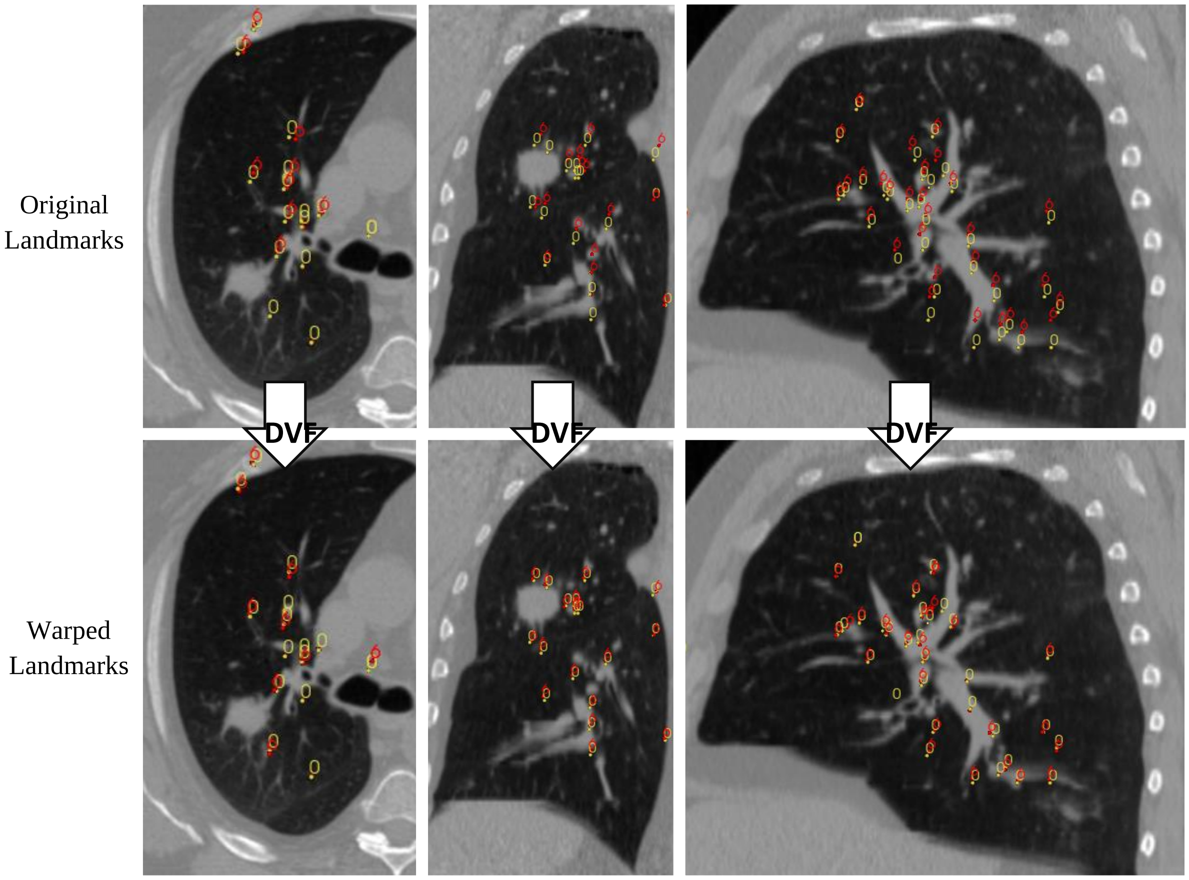


Figure 4.6: Upper row: Projection on the three body planes of the manually identified landmarks in phase 0% (yellow) and phase 60% (red); Bottom row: Projection on the three body planes of the warped landmarks (red) with respect to the original landmarks identified in the phase 0% (yellow). After applying the DVFs generated by image registration, the warped landmarks match better the original landmarks.

4.3.1.3 Mean position evaluation

Figure 4.4 and Figure 4.5 show examples of tumour contours manually delineated in each 4DCT frame for one lung phantom (LP1) and one lung cancer patient (L1). To give a ground truth estimate of the actual tumour mean position, the average position of the ten tumour COMs was calculated for the lung phantoms (LP1 and LP2) and lung cancer patients (L1 to L8). Unless intravenous contrast is synchronized with 4DCT acquisition, liver- and pancreatic tumours are not well visualized on the 4DCT data. Therefore, the position of the marker COM was measured manually for each patient’s breathing phase instead.

The tumour (or markers) mean position was compared to the COM of the target contour on the Mid-P CT scan in terms of target offset. This experiment evaluates whether the Mid-Position module correctly represents the tumour’s mean position over the respiratory cycle.

4.3.1.4 Motion amplitude evaluation

For each patient, the peak-to-peak motion amplitude of the target (tumour or fiducial marker) was calculated using one of the methods described in Section 4.2.2 – the CM method. Briefly, the displacement of the tumour's COM in each breathing phase relative to its coordinates in the reference phase was calculated in LR, AP and CC directions. The motion amplitude is then given by subtracting the maximum from the minimum for COM's coordinates in each direction.

These motion amplitudes calculated by the *RunMidP* module were compared to ground truth estimates and motion amplitudes calculated by Mirada Medical's Mid-P module for the same RT structures. Ground truth estimates of target motion were computed from the COM's coordinates of the manually delineated CTV contours (or markers) on the 4DCT scans.

4.3.2 Study B: Extensive Module Validation

As mentioned earlier, the Mid-Position algorithm validation was carried out over three different studies. The second study (Study B) aimed to compare Mid-P CT scans from two different Mid-P implementations (3D Slicer and Mirada Medical) with a ground truth Mid-P CT image. Besides, reference values of motion amplitudes within ROIs, that do not rely on manually delineated contours, were used and were compared to the motion amplitudes calculated with both software. Also, the effect of differences in motion amplitudes estimation on planning margins were evaluated as well as the intensity distributions within several organs on each Mid-P CT scan. Finally, a robustness study was performed to evaluate the effect of different 4DCT frames as the reference phase in the Mid-Position reconstruction algorithm.

4.3.2.1 Data Sets

Image registration evaluation is of utmost importance if the intention is to use this tool integrated into software for clinical purposes, as is the case of the *RunMidP* module. However, an appropriate reference, upon which the evaluation of the registration performance is based, is generally not available. The registration accuracy is often measured via manually marked contours or anatomical landmarks by doctors or specialists, which adds itself a bias to the registration algorithms [45].

In addition to evaluating the accuracy of the image registration, the set of experiments presented in Section 4.3.1.3 and 4.3.1.4 to validate the Mid-P module were subjected to human inaccuracies during the contouring of ROIs and COM calculation of the implanted fiducial markers.

To overcome the aforementioned issues, a post-doctoral researcher of our research group at CCC created an algorithm that was used for the commissioning of the developed *RunMidP* module. The algorithm creates a synthetic 4DCT dataset (derived from real patient's data) where the ground truth Mid-P CT can be known, and the motion amplitudes within the ROIs can be accurately estimated. This work was presented at ESTRO 2021 [101] and is being included in one paper under printing [121]. Details about the algorithm will be briefly summarized here[15]. The generated synthetic 4DCT data was

---

[15] ***Reconstruction of the ground truth Mid-P CT***

Starting with clinical 4DCT data, one of the 10 frames was chosen as the reference frame. This reference frame was then registered to the other 9 frames using the Demons DIR algorithm from the commercial software Matlab R2020a (The MathWorks, Inc., United States). From the image registration process, 9 DVFs were acquired and they were all re-applied to the reference phase to create new (synthetic) 9 frames closely mimicking the original 4DCT but with precisely known ground truth DVFs. Since these DVFs have the same starting point, they were averaged to a single mean DVF which was consequently used to warp the reference phase to the Mid-Position, thus creating the ground truth Mid-P CT scan [101].

used by the *RunMidP* module and Mirada Medical's Mid-P algorithm to calculate a Mid-P CT that should be equal (or at least very similar) to the ground truth Mid-P CT.

Study B includes synthetic 4DCT data based on 16 treatment cases (8 lungs, 5 livers, 2 thorax and 1 pancreas tumours) acquired using the 4DCT protocol described in Section 4.3.1.1. Tumour location and volume for each case are summarized in Table 4.3 and shown in Figure 4.7.

Table 4.3: Patient and tumour characteristics of Study B.

Patient ID	CT size	CT resolution (mm)	Structure delineated	Structure location	Structure volume (cm^3)
S1	256 × 256 × 167	0.887 × 0.887 × 2	CTV1 CTV2 GTV	RML RML RML	93.3 29.1 19.7
S2	500 × 500 × 103	0.836 × 0.836 × 2	CTV	Liver SegV	4.1
S3	512 × 512 × 111	1.172 × 1.172 × 2	GTV	Liver	72.0
S4	512 × 512 × 142	1.082 × 1.082 × 2	CTV	Liver	184.0
S5	512 × 512 × 120	0.879 × 0.879 × 2	CTV1 CTV2	LUL RML	2.6 1.7
S6	512 × 512 × 136	0.977 × 0.977 × 2	CTV1 CTV2 CTV3	LLL LLL LLL	0.6 1.0 0.4
S7	512 × 512 × 110	0.934 × 0.934 × 2	CTV	Pancreas	105.7
S8	512 × 512 × 128	1.133 × 1.133 × 2	CTV1 CTV2 CTV3	Liver SegIVa Liver SegVIII Liver SegVII	251.5 16.9 26.0
S9	512 × 512 × 109	1.172 × 1.172 × 2	CTV1 CTV2	LLL LUL	1.6 12.0
S10	512 × 512 × 79	0.949 × 0.949 × 2	CTV	Liver	53.2
S11	512 × 512 × 137	0.975 × 0.975 × 2	CTV1 CTV2	RUL RLL	1.4 1.0
S12	512 × 512 × 127	1.172 × 1.172 × 2	CTV1 CTV2	RRib LThorax	43.2 10.8
S13	512 × 512 × 145	0.896 × 0.896 × 2	CTV	RUL	0.4
S14	512 × 512 × 119	1.172 × 1.172 × 2	CTV1 CTV2	RLL RUL	69.6 1.2
S15	512 × 512 × 147	1.008 × 1.008 × 2	CTV	RLL	1.7
S16	512 × 512 × 142	0.976 × 0.976 × 2	CTV	VIII RRib	7.0

Abbreviations:
GTV: Gross Target Volume
CTV: Clinical Target Volume
RLL: Right Lower Lobe | LLL: Left Lower Lobe | RUL: Right Upper Lobe | RML: Right Median Lobe | LUL: Left Upper Lobe
RRib: Right Rib
LThorax: Left Thorax

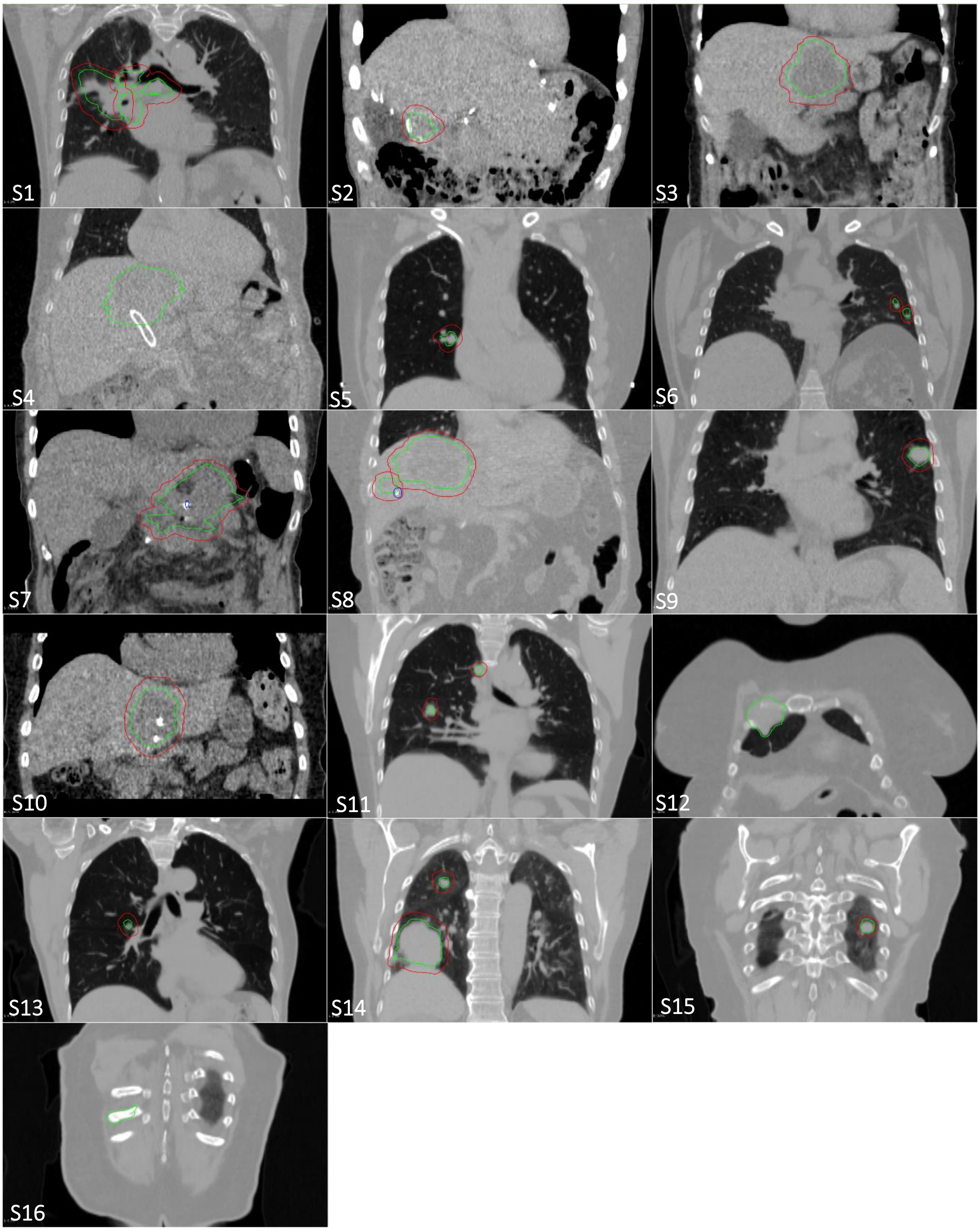


Figure 4.7: Coronal views of phase 0% of the synthetic 4DCT data derived from 16 patient cases (Study B). When available, planning contours of CTV (green), PTV (red) and implanted fiducial markers (blue) are shown.

4.3.2.2 Mid-Position CT evaluation

Calculation of the Mid-P CT scan was performed as described by Wolthaus *et al*. [8]. Details about the Mid-P CT scan calculation with the in-house developed Mid-P module (*RunMidP*) can be found in Section 4.2.1. Mid-P CT scans for the synthetic 4DCTs were calculated by the *RunMidP* module and by an automated research prototype (developed by Mirada Medical) and were compared to the ground truth Mid-P images (slice by slice) using the MSE and the SSIM index.

4.3.2.3 Motion amplitude evaluation

For each patient in Table 4.3, motion amplitudes of tumours were calculated by finding the peak-to-peak amplitudes of the displacement vectors within ROIs for each 4DCT frame and taking the average – the ROI Motion (RM) method. Tumour motion amplitudes were compared to ground truth estimates calculated with the Mid-P commissioning algorithm [101]. Using this approach, uncertainties that come from manual contour delineating are eliminated.

Besides, using the RM method to estimate motion amplitudes, motion amplitudes are evaluated in each voxel within the tumour, thus being a more precise method than the CM method which only uses the contour COM to calculate motion amplitudes.

4.3.2.4 Planning margins evaluation

Motion amplitudes can be converted into treatment margins following the margin recipe estimated at CCC for hypofractionated treatments (Equation 4.3). Systematic (Σ) and random (σ) errors used in this formula were estimated at the NKI where a frameless SBRT with 4D-CBCT guidance is in use for the treatment of lung cancer patients [102]. The accuracy of the tumour location and the intrafractional stability contributed to both Σ and σ. The respiratory motion and the delineation accuracy were the third factors considered to the total random and systematic errors, respectively. The geometrical error with respect to the respiratory motion is modelled as one third of the peak-to-peak amplitude (*A*), which is in agreement with the conclusions of van Herk *et al*. [46]. Tumour localization accuracy, intrafraction variability of the tumour and delineation accuracy in terms of Σ and σ errors for the LR, AP and CC directions are shown in Table 4.4.

$$PTV_{margin} = 2.1 \sum + \ 0.8\,\sigma \tag{4.3}$$

$$\Sigma \ = \ \sqrt{\Sigma^2_{localization\ accuracy} + \Sigma^2_{intrafraction\ stability} \ + \Sigma^2_{delineation}} \tag{4.4}$$

$$\sigma \ = \ \sqrt{\sigma^2_{localization\ accuracy} + \sigma^2_{intrafraction\ stability} \ + \left(\frac{A}{3}\right)^2} \tag{4.5}$$

Using the PTV margin (Equation 4.3) with the values of measurement uncertainty calculated at NKI, the treatment margins for the peak-to-peak motion amplitudes calculated by the two software (Mirada research prototype and *RunMidP*) were compared using the ground truth motion amplitudes as reference.

Table 4.4: Tumour localization accuracy, intrafraction variability of the tumour and delineation accuracy in terms of systematic (Σ) and random (σ) errors for the three directions [102]. LR: Left-Right; AP: Anterior-Posterior; CC: Cranio-Caudal.

	LR (mm)		**AP (mm)**		**CC (mm)**	
	Σ	σ	Σ	σ	Σ	σ
Tumour location accuracy	0.8	1.1	0.9	1.4	0.8	1.1
Intrafraction variability of tumour	1.2	1.3	1.8	1.8	1.2	1.5
Delineation accuracy	2.0	0.0	2.0	0.0	2.0	0.0
Total (without breathing)	**2.5**	**1.7**	**2.8**	**2.3**	**2.5**	**1.9**

4.3.2.5 Hounsfield Units distribution

To evaluate the feasibility of the two Mid-P CT scans (by Mirada and 3D Slicer) for dose calculation, the distribution of Hounsfield Units (HU) within the tumour and OARs were compared against one 4DCT frame by isotropically shrinking the original contours (manually delineated by doctors) in order to avoid regions of organ transition and consider only the bulk. A subtraction operation of 2 mm was used on the tumour and spinal cord, and 5 mm for the heart and liver. Nine phantoms of Study B were selected. The selection criteria consisted of: (1) tumours located in the lung, liver or pancreas; and (2) tumour volumes > 3 cm^3. Small tumours were excluded due to the sample of pixel intensities after shrinking would not be representative of the tumour HU distribution. Based on this criterion, patients S2, S3, S4, S7, S8, S9, S10, S12, and S14 were selected.

The Mann-Whitney U test for two independent samples was used to determine whether the voxels within the tumour contour and relevant OARs belong to the same distribution ($p < 0.05$).

4.3.2.6 Robustness study

Five patients of Study B were randomly selected to evaluate the effect of different 4DCT phases to be used as the reference phase in the Mid-P CT algorithm. For each patient, ten Mid-P CT scans were reconstructed using 10 different 4DCT frames as the reference phase. For simplicity, the reconstructed Mid-P CT scans will be referred to as Mid-P_i CTs (where $i = 0, 1, \ldots,$ and 9) where the phase 0%, 10%, …, 90% was used as the reference phase, respectively.

The Structural Similarity (SSIM) index between the reconstructed Mid-P_i CTs and the ground truth Mid-P CT was calculated slice by slice.

4.3.3 Study C: Image Quality Assessment and Mid-P approach through doctors' eyes

High image quality is essential for diagnostic purposes and accurate tumour and OARs contouring in radiotherapy treatment planning. As a common practice in radiotherapy planning, margins surrounding GTV and CTV are added to create the PTV. These margins are introduced to account for various factors, including structure delineation uncertainties that increase if the image quality is worse. Naturally, it is desirable to keep treatment margins small to reduce radiation dose to OARs.

The Mid-P approach incorporates the tumour breathing motion in the PTV margin, yielding significantly smaller PTV than the ITV method [9]. In addition, it is known to lead to a higher signal-to-noise ratio and removal of motion artefacts [8]. However, the reconstruction of the Mid-P CT scan involves averaging registered 4DCT scans, which inevitably lead to image blurring.

The purpose of Study C is to assess the image quality of the Mid-Position CT scan (quantitatively and qualitatively) to evaluate if it has acceptable image quality for radiotherapy treatment planning.

4.3.3.1 Data Sets

Study C includes 4DCT data of 5 cancer patients with tumours in the lung, bronchus, stomach, sternum and liver. 4DCT data was acquired using the 4D acquisition protocol described in Section 4.3.1.1. Tumour location and volume for each patient are summarized in Table 4.5.

Table 4.5: Patient and tumour characteristics of Study C.

Patient ID	CT size	CT resolution (mm)	Structure delineated	Structure location	Structure volume (cm^3)
1	512 × 512 × 133	0.910 × 0.910 × 2	CTV1	LUL	0.7
			CTV2	Lymph Node 4L	11.0
2	512 × 512 × 149	0.973 × 0.973 × 2	CTV1	LPB	34.4
			CTV2	Mediastinum	16.9
3	512 × 512 × 106	0.898 × 0.898 × 2	CTV	Stomach	183.3
4	512 × 512 × 130	1.172 × 1.172 × 2	CTV	Sternum	25.5
5	512 × 512 × 122	0.780 × 0.780 × 2	CTV1	Liver SegVIII	4.0
			CTV2	Liver SegVII	11.8

Abbreviations:
CTV: Clinical Target Volume
LUL: Left Upper Lobe
LPB: Left Pulmonary Bronchus

4.3.3.2 Image Quality Assessment

Quantitative Assessment

Signal-to-noise ratio (SNR) and image sharpness are related to anatomical structures visualization and delineating accuracy. These two image quality-based metrics were calculated to evaluate the feasibility of the Mid-P CT scan for radiotherapy planning purposes. SNR was determined by manually placing five circular ROIs (20 mm in diameter) in five neighbour axial slices in homogenous regions of the liver in the reference 4DCT phase. These ROIs were then warped to the other nine 4DCT phases, free-breathing planning CT (pCT) scan and Mid-P CT, using the DVFs of the registration process. This method establishes an anatomical correspondence between scans, thus enabling the evaluation of the same ROI in different scans[16]. The SNR in the liver was calculated by taking the average of the signal (mean CT number, HU) within ROIs divided by the noise (standard deviation of the signal, σ) within the same ROIs (Equation 4.6). SNR was compared on all 4DCT phases, pCT and Mid-P CT scan reconstructed by the *RunMidP* module of the five patients in Study C.

$$SNR = \frac{\overline{HU_{Liver}}}{\sigma_{Liver}} \tag{4.6}$$

To evaluate the image blurring, a line at the same position perpendicular to the transition between the diaphragm and lung was delineated on different CT scans. The line was drawn by finding the peak of the diaphragm of each lung with a standardized length of 20 mm. The sharpness of the transition between the diaphragm and lung was measured using the method described in [103], where it was quantified by fitting a cumulative Gaussian distribution to the intensity along the line. Sharpness was represented by the standard deviation of the fitted distribution and the Mid-P CT sharpness was compared against one of the 4DCT phases, a sharper image (the pCT) and a worst-case scenario (the Average CT scan).

Qualitative Assessment

The final decision on whether the image quality is acceptable for radiotherapy planning lies with the radiation oncologists. To further evaluate the image quality of the Mid-P CT scans, 2 reviewers from CCC (with 6 and 20 years of clinical experience in lung radiotherapy, respectively) blinded to the type of the planning CT used, were asked to independently analyse and score 81 lung scans from 18-randomly selected patients[17] (9 from CCC and 9 from NKI). This study includes 18 Average CT scans, 9 breath-hold CT scans, and 54 (18×3) Mid-P CT scans reconstructed by three different software applications (*RunMidP* module, Mirada Medical software, and Wimp[18]). At CCC, breath-hold CT scans are not routinely used for radiotherapy planning, so only breath-hold CT scans acquired in NKI were included in this study.

The scans were viewed using a 1920 × 1080- pixel and 27-inch monitor (Full HD LCD Monitor 273V7QDSB/00, *Philips*). For each scan, doctors were allowed to change the window level to visualize different structures as well as browse through each slice and zoom in and out.

The sharp reproduction of critical anatomical structures such as the heart and great vessels, oesophagus, lung parenchyma, tumour, carina and lymph node area, main and segmental bronchi, pleuromediastinal border, border between pleura and thoracic wall, and diaphragm, was qualitatively evaluated using a "YES/NO" criteria. The critical anatomical structures were determined according to the European Guidelines on quality criteria for CT [104] and doctors' suggestions. The image sharpness, image contrast, artefacts, and contouring acceptability were qualitatively assessed using a 5-point grading scale (Table 4.6 and Appendix VI to see the planning CT assessment form).

---

[16] It is important to note that image registration is not perfect and there are always some anatomical inaccuracies.

[17] This study includes more patients than those specified in Table 4.5 because it was performed in collaboration with another clinical institution - The Netherlands Cancer Institute (NKI).

[18] Wimp is a clinically used software to reconstruct Mid-P CT scans, which was developed by the NKI.

Table 4.6: 5-score grading system for the planning CT assessment form.

Classification	**1**	**2**	**3**	**4**	**5**
Sharpness	Blurry	Sub-optimal	Acceptable	Good	Sharpest
Image contrast	Very poor	Sub-optimal	Acceptable	Good	Excellent
Artefacts	Unacceptable artefacts	Many/large artefacts	Some artefacts but acceptable	Minimal artefacts	No artefacts
Contouring acceptability	Unacceptable	Suboptimal	Acceptable	Good	Excellent

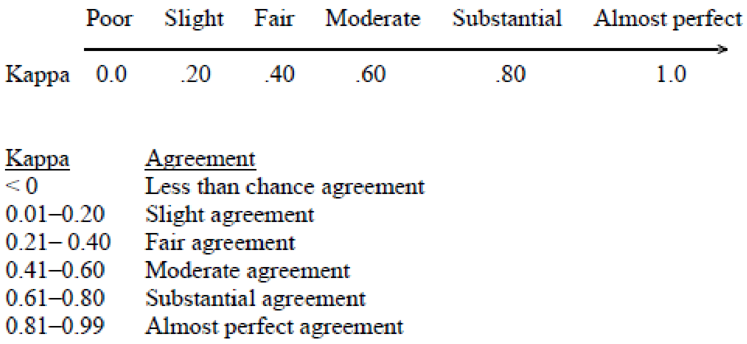
Grades of image sharpness, image contrast, image artefacts, contouring acceptability and visibility of anatomical structures were compared for each type of CT scan and averaged for both readers.

A Wilcoxon signed-rank test was performed to investigate statistically significant differences in averaged qualitative scores among alternative planning CTs (Average CT, Mid-P CT, and breath-hold CT). P-value < 0.05 indicated a statistically significant difference.

Interobserver agreement was measured using the Cohen's Kappa (κ). The calculation of κ is based on the difference between how much agreement is actually present (p_o, "observed" agreement) compared to how much agreement would be expected to be present by chance alone (p_e, "expected" agreement) and given by [105]:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{4.7}$$

κ ranges from −1 to +1, where 0 or lower represents the amount of agreement that can be expected from random chance, and 1 represents perfect agreement between raters [105]. A more detailed interpretation of Cohen's Kappa is shown in Figure 4.8.



Kappa	Agreement
< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21– 0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

Figure 4.8: Interpretation of Cohen's Kappa [105].

5 Results and Discussion

In addition to the implementation of the in-house-developed Mid-Position module (*RunMidP*) into the 3D Slicer platform with a GUI displayed in Figure 4.1, a set of experiments were carried out to evaluate the module performance. The results are divided as in the "*Materials and Methods*" section, i.e., into three studies conducted with three different datasets. Besides, three separate Mid-Position packages will be referred throughout this section: 3D Slicer (or *RunMidP*), Mirada, and Wimp software.

5.1 Study A: Image Registration Evaluation and Image Quality Optimization

5.1.1 Image Registration Evaluation

The evaluation of the registration accuracy of the Plastimatch and Elastix software was based on the RMSE of the difference between warped and reference images in the 4DCT dataset. The results are shown in Figure 5.1 and expressed as the mean RMSE over the 9 warped images and the standard deviation (error bars).

For the majority of patients, there is no statistically significant difference between the two registration methods which accurately estimated the deformations between the reference and the warped frames. The mean RMSE considering all patients in Study A was 42.1 ± 16.5 HU and 41.1 ± 15.9 HU for the Elastix and Plastimatch software, respectively ("*All_Pats*" column).

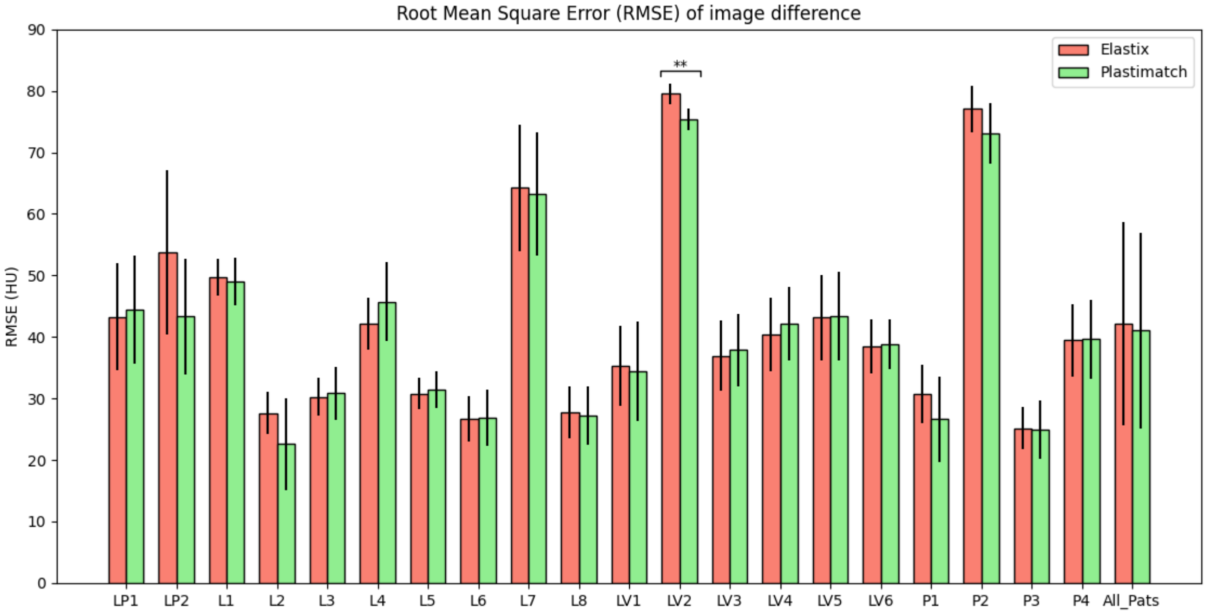


Figure 5.1: Root mean square error (RMSE) between the original and 9 warped images (after registration) for each patient. Two image registration methods – Elastix and Plastimatch – were evaluated. The right-most columns show the mean and standard deviation of the RMSE for all patients, emphasising that there is no significant difference between the two methods. ** = $p < 0.01$.

For a perfect registration, the RMSE should be zero, however, some issues may complicate the registration and lead to a larger RMSE. Artefacts in the 4DCT dataset are a common source of errors in the registration since they introduce anatomical inconsistencies between images. Large movements in tissues, particularly in areas that lack contrast, such as in the liver, can be especially prone to errors. Finally, the discontinuous motion between the lung tissue and the motion of the chest wall during the breathing may also lead to registration errors. Because of these issues, small differences between the original (reference) and the warped frames are expected. Even so, the RMSE values obtained with both registration methods agree with those found in the literature [106].

The maximum RMSE was found for patient LV2 resulting in an RMSE of 79.5 ± 1.7 HU and 75.3 ± 1.7 HU for the Elastix and Plastimatch software, respectively. This was caused by the presence of noise in the image. In Figure 5.2, the 4DCT of patient LV2, as well as the subtraction images of the original and warped frames, were compared to the phase 30% of the 4DCT (used as the reference phase), showing a strong reduction of the image difference after DIR. Residual differences between the original and warped images are due to the noise in the 4DCT, increasing the RMSE.

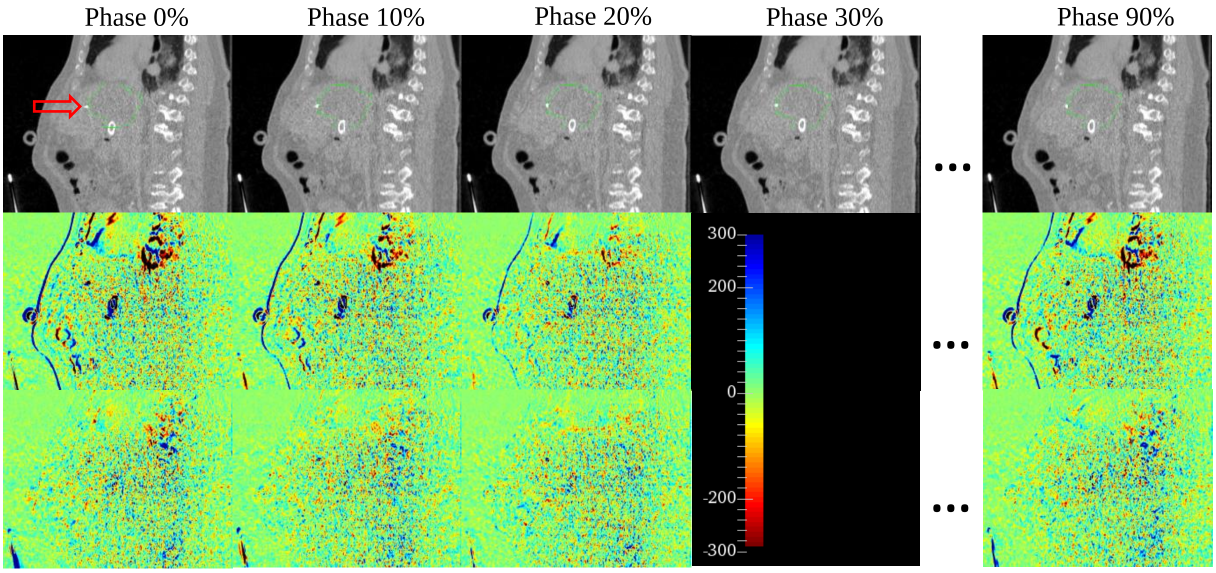


Figure 5.2: Upper row: sagittal slices of the 4DCT scan of patient LV2; Middle row: Image subtraction between original CT frames and the reference phase (phase 30%) with a colour scale range between -300 and 300 HU; and Bottom row; Image subtraction between warped CT frames and the reference phase (phase 30%) with the same colour scale range. After registration, clearly shown is a reduction of the image difference, where there are no edges visible in images. The red arrow shows the implanted gold marker close to the CTV (green).

Since the registration accuracy of both methods was comparable, the image quality of the Mid-P CT reconstructed with the two registration software dictated the final choice of the registration method to be used in the *RunMidP* module. Figure 5.3 shows a comparison between three Mid-P CTs of the same lung cancer patient (L5): two reconstructed by 3D Slicer, with Plastimatch and Elastix software, and the third reconstructed by a soon to be commercial Mid-P CT algorithm developed by Mirada Medical. By visual inspection, the image quality of the Mid-P CT from 3D Slicer with Elastix registration is better than with the Plastimatch software, since the blur is reduced and the pulmonary fissure is well visualized, being similar to the Mid-P CT reconstructed by Mirada software. A possible explanation for the difference between the two DIR methods is the higher B-spline order interpolation (cubic) that Elastix software provides during the resample process, while Plastimatch software only allows the nearest neighbour or linear interpolation.

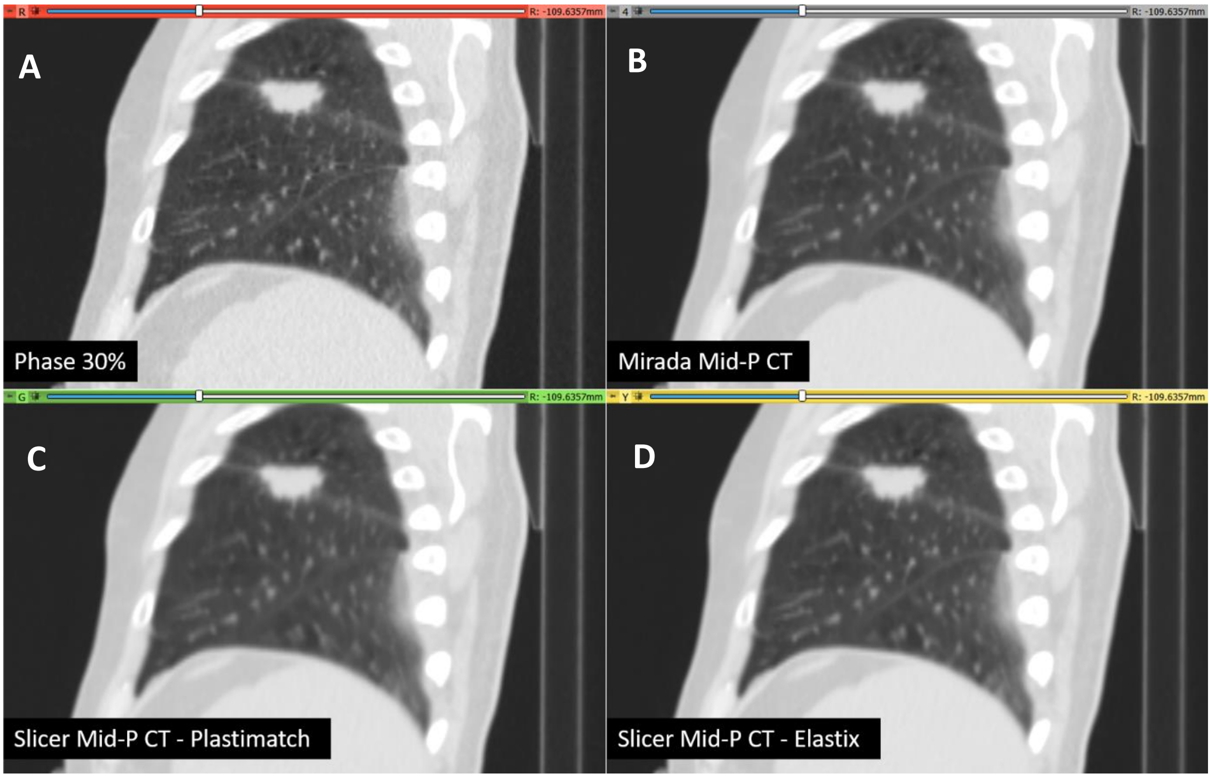


Figure 5.3: Comparison between (A) one frame of the 4DCT (phase 30%); (B) the Mid-P CT reconstructed by Mirada Medical software; (C) the Mid-P CT reconstructed by 3D Slicer with Plastimatch registration software; and (D) the Mid-P CT reconstructed by Elastix registration software.

In addition to the qualitative analysis, image quality was also assessed using the SSIM index, comparing the Mid-P CT generated by the two registration methods against the Mid-P CT generated by Mirada software for each patient. Overall, the average SSIM index between the Mid-P CT from 3D Slicer with Elastix software and Mirada software is significantly higher than with Plastimatch, indicating a comparable image quality between the two Mid-P CT scans (Figure 5.5). The minimum SSIM was 0.997 for the lung phantom LP2 using the Plastimatch registration, possibly because this method is more prone to errors when images lack texture (pixel intensity variations). The high motion amplitude of the sphere (about 14 and 18 mm in the AP and CC direction, respectively) and the fact that the organs and tissues in the XCAT phantom are modelled homogenously, as opposed to more intensity gradients in clinical CT images, may complicate the registration and lead to inaccuracies. Figure 5.4 shows that these registration inaccuracies with Plastimatch software resulted in a blurred Mid-P CT compared to Elastix and Mirada Mid-P CTs. It is important to note that this effect was more noticeable in LP2 than in LP1 possibly due to the lower CT resolution.

It was demonstrated that the accuracy of the Elastix registration is sufficient to register lung and upper abdomen CT scans with promising results in terms of image quality. For these reasons, the Elastix registration was implemented in the *RunMidP* module. All the results presented in the next sections were obtained using the Elastix registration software.

Figure 5.4: Axial views of the Mid-P CT reconstructed by 3D Slicer with Plastimatch registration software (left), with Elastix registration software (middle), and with Mirada Medical software (right) for LP2. Image blur is shown in the Mid-P CT reconstructed by 3D Slicer with Plastimatch software due to inaccuracies in image registration.

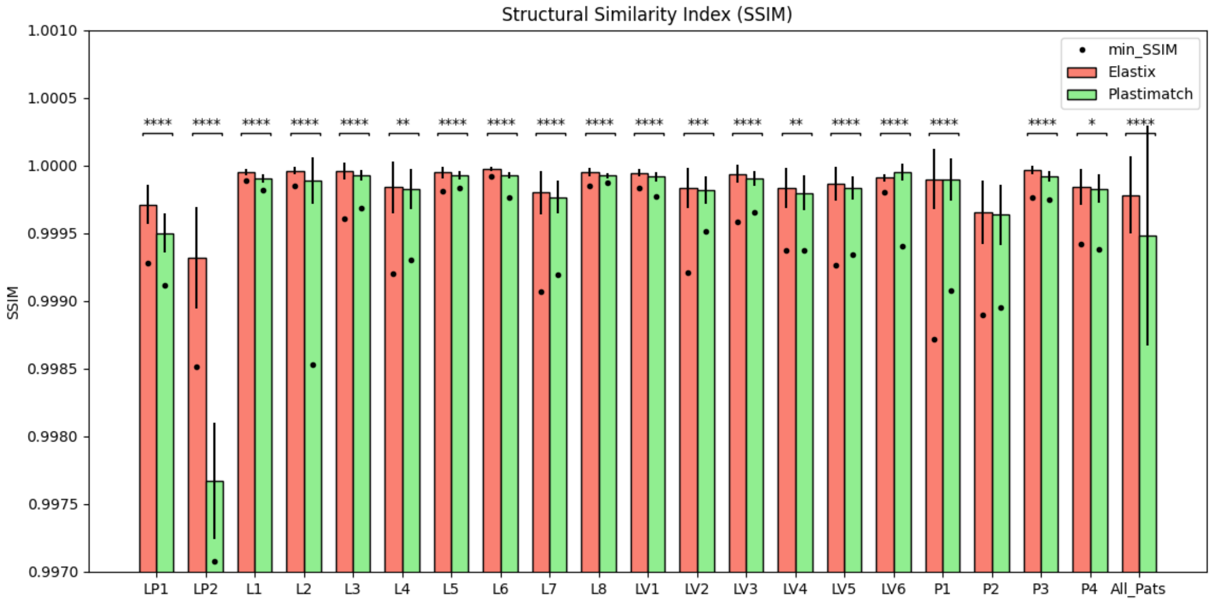


Figure 5.5: Red bars: SSIM index between the Mid-P CT reconstructed by 3D Slicer with Elastix registration and Mirada Medical software; Green bars: SSIM index between the Mid-P CT reconstructed by 3D Slicer with Plastismatch registration and Mirada Medical software. The black dots correspond to the minimum SSIM found along all slices of the Mid-P CT scans for each patient. The right-most columns show the mean and standard deviation of the SSIM for all patients, emphasising that there is a significant difference between the two methods in terms of image quality. * = $p < 0.05$ | ** = $p < 0.01$ | *** = $p < 0.001$ | **** = $p < 0.0001$.

The RMSE and SSIM index were used to evaluate the registration accuracy in the entire image voxel space, which may not be adequate for drawing a consistent conclusion at the tumour (or organs) level. Therefore, other metrics were used to investigate the accuracy within ROIs. After manually delineating the tumour in all frames of the 4DCT, DVFs were applied to these original RT structures to warp them to the reference phase (Figure 5.6). The warped contours should perfectly overlap with the original reference contour. The Dice similarity coefficient (DSC) and Hausdorf distance (HD) were used to quantify the matching between contours before and after registration. Figure 5.7 shows that the HD between the original and warped contour was less than 1 mm for the ten analysed patients, being on average 0.5 ± 0.3 mm, while the DSC ranged from 0.73 to 0.95. Both contour statistics show a good match between the two contours, implying that, at least for the tumour region, the Elastix DIR software is accurate enough for estimating and compensating tumour motion.

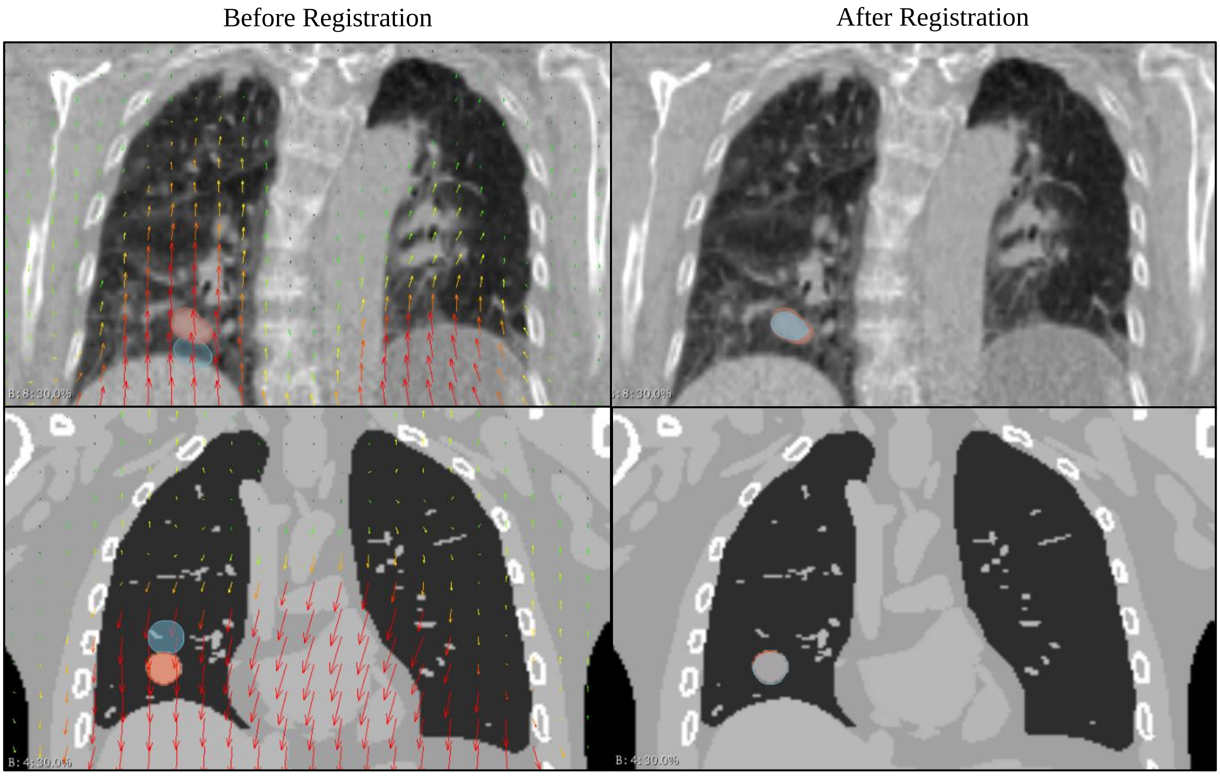


Figure 5.6: Left column: Coronal views of the patient L1 (upper row) and the phantom LP1 (bottom row) with original tumour contours in the reference phase (red) and moving phase (blue). Right column: Coronal views of the patient L1 (upper row) and the phantom LP1 (bottom row) with the warped contour (blue) which refers to the original contour after applying the DVF. The arrows show the magnitude and orientation of the DVF from the moving to the reference phase. The magnitude of the DVF is represented as follows: red colour for displacements above 10 mm; yellow for displacement equal to 5 mm; and green for displacement equal to 2 mm. It is shown that the warped contours match the reference ones after registration.

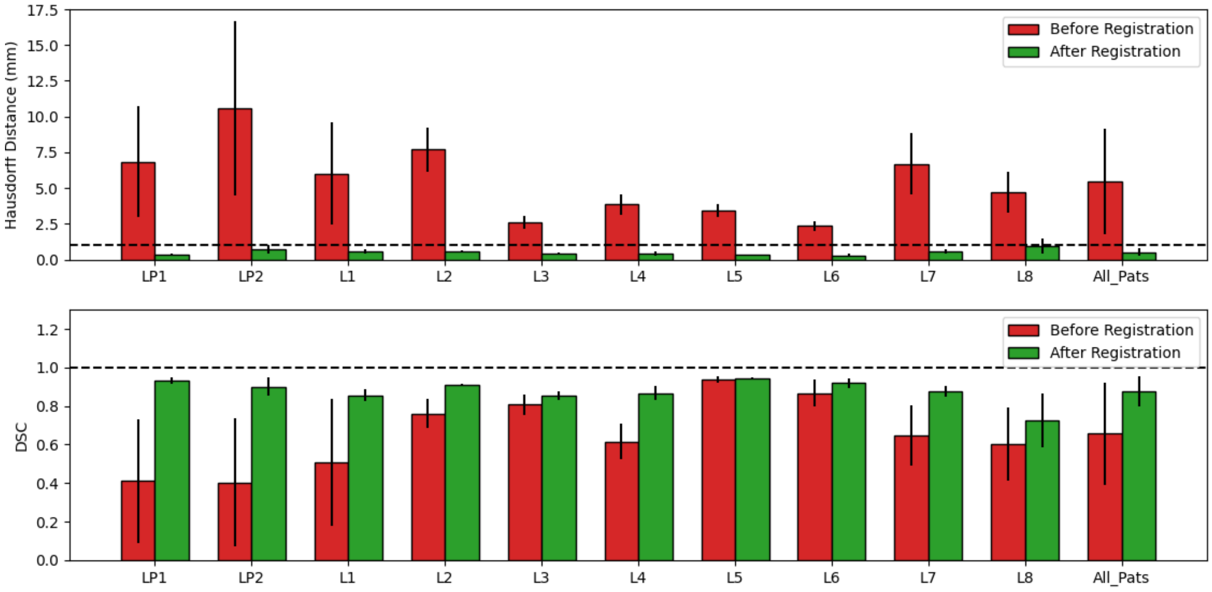


Figure 5.7: Hausdorff Distance (HD) and Dice similarity coefficient (DSC) between the original contours delineated on the reference and moving phase (red) and between the original contours delineated in the reference phase and the warped contours after applying the DVF (green). The black dashed line corresponds to HD = 1 mm and DSC = 1.

Finally, to assess registration accuracy outside of the tumour, fiducial distances of 41 lung landmarks from the POPI-model data were compared before and after applying DVFs to the manually identified landmarks by medical experts, using phase 0% as the reference phase. The average fiducial distance between the original fiducials (in phase 0%) and the 41 transformed fiducials to the reference phase was less than 2 mm through all 9 breathing phases (Figure 5.8). This difference is clinically acceptable as it may not only be due to the quality of the registration method but also due to the manual landmark identification which can induce an error of up to 2 mm [100].

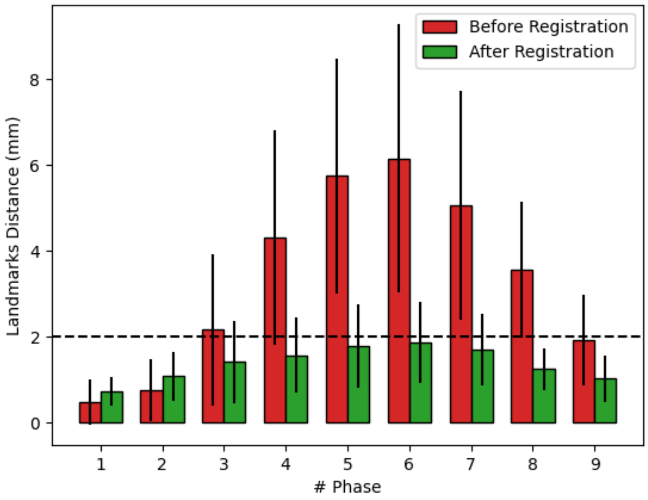


Figure 5.8: Mean distance between original manually identified 41 landmarks in 9 4DCT frames (from phase 10% to phase 90%) against the same points in the reference phase (phase 0%). The set of 41 landmarks in the 9 frames were then warped to the reference phase and the distance between the warped landmarks and the same points in the reference frame were calculated. The black dashed line corresponds to a fiducial distance equal to 2 mm.

Data from the EMPIRE10 challenge [107], where more than 40 image registration algorithms with a database of 30 pairs of thoracic CT images were compared, showed that the first 10 methods depicted a target registration error (TRE) lower than 0.9 mm. Also, for the Demons registration algorithm, the mean TRE for the 41 landmarks identified on the POPI-model data (the same data used in this study) was 1.2 ± 0.4 mm [100]. In practice, the TRE is equivalent to the landmark distance mentioned above (Figure 5.8), and the Elastix software showed a mean TRE of 1.4 ± 0.8 mm. Moreover, Piper *et al*. [108] reported average registration errors of 3.7 mm, and the MIDRAS study [109] reported errors between 1.8 mm and 8.2 mm. Thus, the Elastix software integrated into the 3D Slicer platform proves to be adequate for the registration of CT scans of the lung and upper abdomen with comparable results with other registration methods in terms of accuracy.

One drawback of the registration method is its computation time. A typical mono-modal registration takes about 5-20 min with default settings [110]. Yet, the total computation time of Elastix registration is about 40 min on a 3.40 GHz processor, while the Plastimatch software takes about 30 min. This ultimately leads to a total of about 1h to reconstruct the Mid-P CT using the Elastix software and about 45 min using the Plastimatch software. As future work, the parameters of the registration must be tailored to speed up the overall registration time.

5.1.2 Mean position evaluation

To evaluate whether the Mid-P CT scan can accurately reconstruct the patient's anatomy in its time-weighted mean position, the mean and standard deviation of the difference between the ground truth estimation of the target Mid-Position (based on manual delineations[19] of the target in 4DCT phases) and the target's position on Mid-P CT images generated by 3D Slicer's and Mirada Medical's Mid-P module were calculated. For the 3D Slicer's Mid-P module (*RunMidP*), the differences were 0.00 ± 0.26 mm (LR), 0.13 ± 0.30 mm (AP), and -0.06 ± 0.27 mm (CC) over all patients. For the Mirada Medical's Mid-P module, the differences were 0.04 ± 0.19 mm (LR), -0.03 ± 0.22 mm (AP), and 0.06 ± 0.28 mm (CC) for the same RT structures, as can be noticed from Figure 5.9.

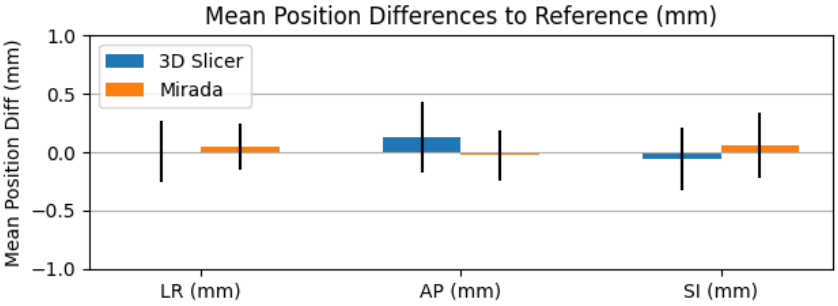


Figure 5.9: Target Mean position differences comparison for the 3D Slicer's and Mirada's Mid-P algorithms.

Both Mid-P algorithms have comparable results with differences in the target position less than 0.5 mm, demonstrating that they are adequate for estimating at least the tumour mean position. Outside of the tumour region, the patient's anatomy was validated using the POPI-model data [100]. The average position of the 4DCT landmarks was set as the 'true' Mid-Position. Each landmark was then warped to the Mid-Position (using the DVFs generated by the *RunMidP* module) and the average distance between the warped landmarks and the true Mid-Position was calculated. The differences were even smaller, with maximum differences of 0.12 mm, 0.16 mm and 0.02 mm in LR, AP and CC directions.

5.1.3 Mid-Position CT evaluation

Figure 5.10 shows a comparison between different CT scans: phase 30% of the 4DCT scan, the Average CT scan, the free-breathing planning CT (pCT) scan, and the Mid-P CT scan. Comparing the Mid-P CT to one single frame of the 4DCT and to the pCT, both appeared to have higher spatial resolution than the Mid-P CT, thus being more useful for planning volumes definition. However, it is important to note that separate 4DCT frames, and especially the free-breathing pCT, may have image artefacts which consequently can introduce serious target delineation errors. These inaccuracies are considerably reduced in the Mid-P CT scan, where the tumour and internal structures are better represented than in the free-breathing pCT. Despite that, averaging all frames of the 4DCT transformed to the mean position, in the final step of the Mid-P CT reconstruction process, results in blurring of

---

[19] The most recent version of the commissing Mid-P software [101] also includes a ground truth estimate of the tumour mean position which not rely on manual delineated contours. It uses the pre-defined DVFs to propagate the tumour mask to the exact Mid-Position. The results will be included in the paper under printing [121].

artefacts present in the 4DCT phases, thus reducing the visual sharpness of anatomical structures compared to the pCT. Even so, the blur is less pronounced than in the Average CT scan (Figure 5.10).

Currently, at CCC the free-breathing pCT is the scan used for planning. Besides the artefacts (Figure 5.10), using the pCT scan runs the risk of planning on a scan that is in an extreme breathing position adding a systematic error or overestimation of the PTV (when ITV is used) to the treatment. The consistent representation of the patient's mean anatomy in the Mid-P CT, combined with the possibility to reduce treatment planning margins, makes the Mid-P CT the best candidate to be used for treatment planning of patients with tumours in areas that move a lot due to breathing.

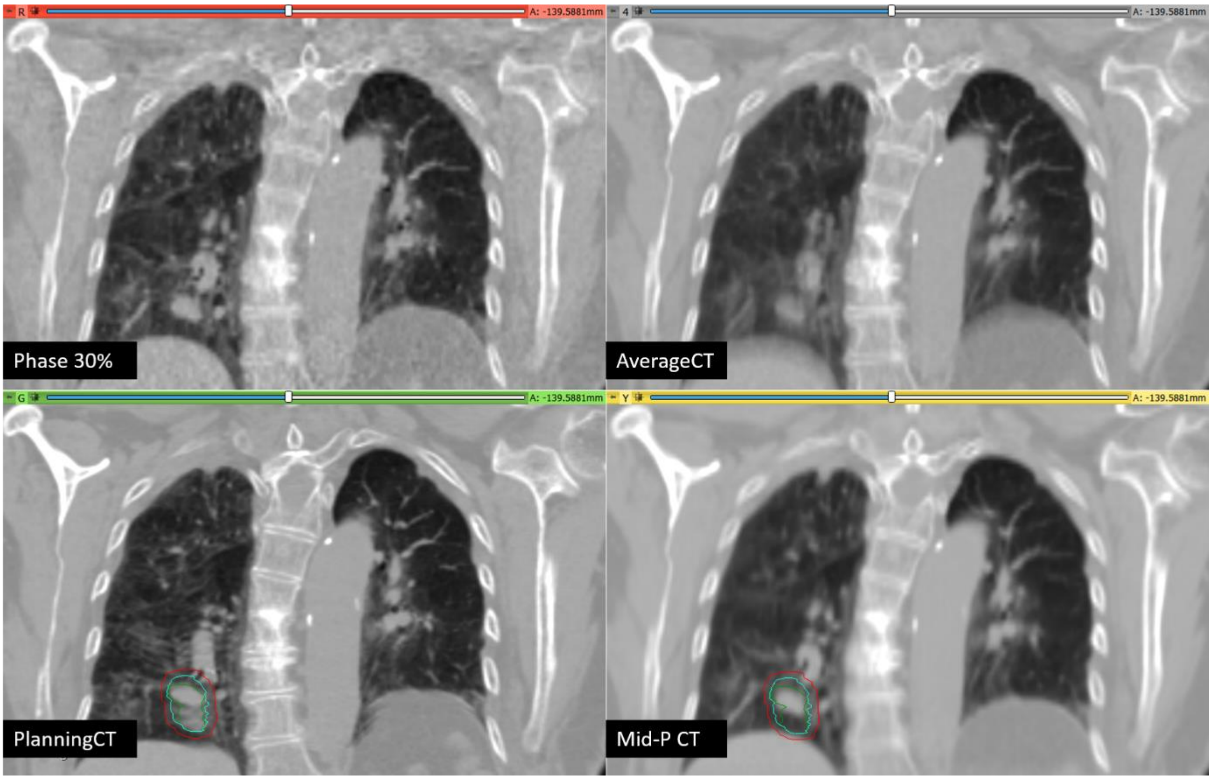


Figure 5.10: Comparison between phase 30% of the 4DCT scan, Average CT scan, free-breathing planning CT and Mid-P CT scan. Image artefacts in the planning CT due to breathing during CT acquisition led to an inaccurate delineation of CTV (green). The ITV (blue) covers all the tumour positions over the respiratory target and an additional margin is applied to the ITV, composing a larger PTV (red). The Mid-P CT reduces image artefacts, yielding a good representation of tumour and internal structures, which potentially allows the reduction of planning-treatment margins.

5.1.4 Motion amplitude evaluation

Ground truth motion amplitudes compared against the motion amplitudes calculated by the *RunMidP* module and Mirada Medical's algorithm for the group of 20 patients in Study A are shown in Figure 5.11. The mean and standard deviation of the range of motion was 1.4 ± 0.9 mm in LR, 3.4 ± 3.0 mm in AP, and 6.7 ± 4.8 mm in CC direction. The largest range observed in each direction was 4.0 mm, 14.4 mm, and 18.1 mm respectively, where the CC direction saw the largest motion in 15 patients (75%).

Motion differences compared to ground truth estimations were higher for the CC direction but still are within 1 mm difference (Figure 5.12), thus showing an accurate motion amplitudes calculation for both software. These small differences can be due to inaccuracies during the target delineation and registration process.

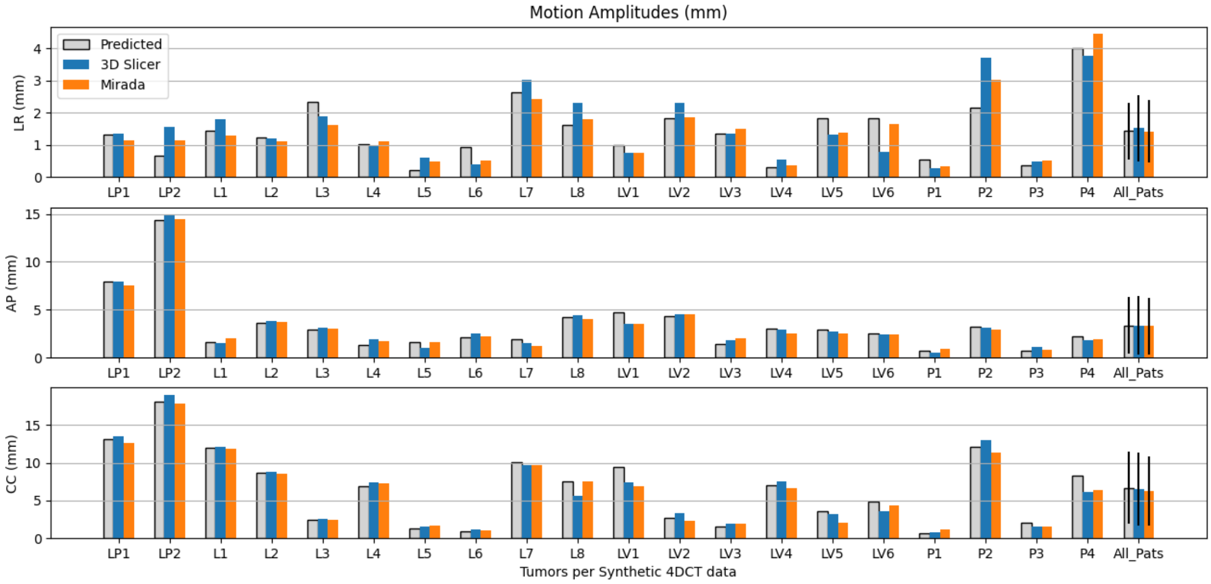


Figure 5.11: Comparison of tumour and marker motion amplitudes. Ground truth motion amplitudes (gray); motion amplitudes calculated with the *RunMidP* module integrated into 3D Slicer (blue) and motion amplitudes calculated with the Mirada Medical's algorithm (orange) in the three directions: left-right (LR), anterior-posterior (AP), and cranio-caudal (CC).

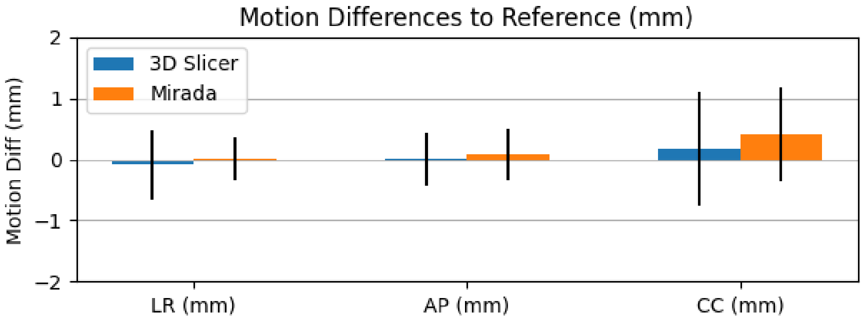


Figure 5.12: Motion amplitude differences between motion amplitudes estimated with two software – the *RunMidP* module integrated into 3D Slicer (blue) and Mirada Medical (orange) – and the ground truth motion estimation.

As mentioned in Section 4.2.2, the first step to calculate the motion amplitudes with the *RunMidP* module was to propagate the planning contours, delineated in the free-breathing scan, to the reference phase of the 4DCT. Image artefacts in the free-breathing scan can cause inconsistencies in the patient's anatomy which can lead to errors during registration. Ideally, when the Mid-P CT scan is used clinically for treatment-planning and contouring, these differences will be reduced, since the Mid-P CT is known to contain fewer image artefacts and noise. In addition, motion amplitude estimations from the contour's COM (using the CM method) do not provide information on the range of motion in all voxels within the target. It is based on only one point – the centroid of the contour – which may lead to larger differences compared to the true motion amplitude. Thus, the method to calculate motion amplitudes with the *RunMidP* module was optimized (to the RM method) and will be referred in Section 5.2.2.

5.2 Study B: Extensive Module Validation

5.2.1 Mid-Position CT evaluation

Mid-P CTs reconstructed by the *RunMidP* module and Mirada's Mid-P algorithm were compared against "true" Mid-P images created by a commissioning software developed in our institution according to [101], see Section 4.3.2.1. Figure 5.13 depicts a comparison (slice by slice) between both Mid-P CTs for patient S6 in terms of Mean Squared Error (MSE) and Structural Similarity (SSIM) and considering the ground truth Mid-P CT as reference. As expected, higher image differences were found in the diaphragm-lung interface, where the patient's anatomy moves faster due to breathing. Air regions in the bowel also produced higher HU differences compared to the predicted Mid-P scan for both Mid-P algorithms. This may be caused due to the limited accuracy of the image registration method to register homogeneous regions. Although in some anatomical regions the Mid-P CT difference was higher, the average image difference was only about 1 HU for all the analysed cases, suggesting that these HU differences will not produce meaningful differences in dose calculation.

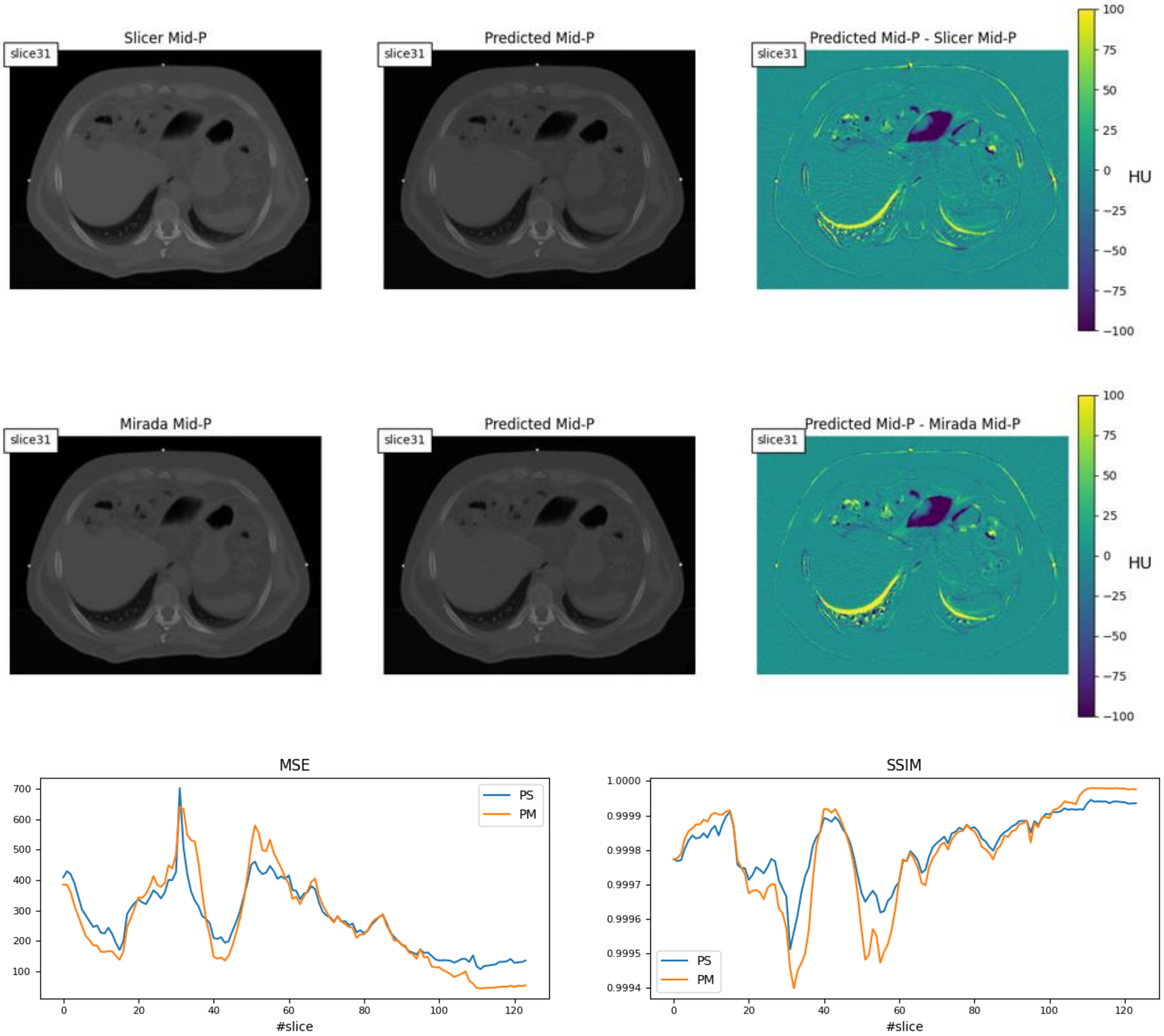


Figure 5.13: Upper row: Mid-P CT reconstructed by the 3D Slicer software and image difference between 3D Slicer Mid-P CT and ground truth Mid-P CT at slice nr. 31 for patient S6; Middle row: Mid-P CT reconstructed by Mirada Medical software and image difference between Mirada Mid-P CT and ground truth Mid-P CT at the same slice and for the same patient. Bottom row: Mean Squared Error (MSE) (left) and Structural Similarity (SSIM) index (right) between both Mid-P CTs. PS: Predicted Mid-P CT vs 3D Slicer Mid-P CT | PM: Predicted Mid-P CT vs Mirada Medical Mid-P CT.

Considering all patients of Study B, the mean SSIM between the ground truth Mid-P image and Mid-P CT reconstructed by 3D Slicer and Mirada Medical was 0.99985 ± 0.00009 and 0.99987 ± 0.00010, respectively. The minimum SSIM index was 0.99903 and 0.99889, respectively. Regarding MSE values, the mean MSE was 263 ± 104 HU for 3D Slicer and 208 ± 105 HU for Mirada Medical. The maximum MSE for 3D Slicer and Mirada was 1455 HU and 960 HU, respectively. According to these results, both Mid-P images are equally different from the ground truth Mid-P image. Even for liver and pancreas patients, in which the absence of contrast in the CT could complicate the registration, there were similar differences in the Mid-P images, suggesting that both Mid-P algorithms provide reliable tools to calculate the Mid-P CT scan for lung and upper abdomen cancer patients. Nonetheless, Mid-P images from Mirada's research prototype are slightly more similar to the ground truth Mid-P CT than Mid-P images from 3D Slicer's module as can be noticed from Table 5.1.

Table 5.1: Mean and standard deviation of the Structural Similarity (SSIM) index and Mean Squared Error (MSE) between the predicted (ground truth) Mid-P CT and the Mid-P CT reconstructed by 3D **S**licer and Mirada Medical software. Higher values of the SSIM index (and lower values of the MSE) between the ground truth Mid-P image and the Mid-P for both methods are highlighted in blue and orange for 3D Slicer and Mirada Medical software, respectively, indicating each of the Mid-P images are more similar to the ground truth Mid-P CT according to each metric.

	SSIM		**SSIM**		**MSE (HU)**		**MSE (HU)**	
	Predicted vs Slicer		**Predicted vs Mirada**		**Predicted vs Slicer**		**Predicted vs Mirada**	
Patient ID	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**
S1	#75	##11	#80	##11	411	117	299	129
S2	#83	##09	#86	##09	282	141	228	138
S3	#81	##15	#79	##24	310	150	282	206
S4	#77	##07	#83	##06	575	159	413	122
S5	#90	##05	#88	##07	194	54	235	89
S6	#82	##09	#80	##14	269	110	258	150
S7	#84	##13	#84	##17	239	114	213	134
S8	#90	##06	#91	##06	203	76	161	77
S9	#91	##07	#93	##03	203	68	137	51
S10	#82	##06	#86	##07	297	68	211	72
S11	#90	##12	#92	##11	167	146	124	138
S12	#84	##16	#89	##08	325	223	209	104
S13	#76	##13	#77	##19	351	115	321	160
S14	#92	##03	#95	##03	174	52	95	42
S15	#96	##04	#97	##03	69	31	40	26
S16	#92	##03	#94	##05	141	35	102	45

Abbreviations:
= 0.999
= 0.000

The remaining Mid-P differences can be caused by the use of different image registration methods and different resampling techniques to generate the warped scans during the Mid-P CT calculation. The Mid-P CTs generated by the Mirada Medical software were reconstructed using an optimized derivative of Lucas-Kanade optical flow registration algorithm [111]. The transformation model is free-form but it is regularized to produce continuous, biologically plausible deformation fields. Taking into account the results presented in this work, Mirada's CT DIR algorithm seems to be more accurate than the publicly available Elastix registration. It would be interesting to analyse if changing the registration parameters – for instance, decreasing the grid spacing or using another optimizer – would result in a more accurate registration. Despite that, Elastix registration was found to be suitable for the reconstruction of the Mid-P CT with comparable results to the soon to be commercial Mirada's Mid-Position algorithm.

Regarding the calculation of the resampled scans, it was found that the method in the *RunMidP* module to harden[20] non-linear transforms was slightly changing the image geometry (origin and size). For the majority of the 4D cases, the origin shift was within the pixel resolution, meaning that the shift can be up to 2 mm in the z-direction and skewing the results. To overcome this issue, another 3D Slicer's built-in module (*Resample Scalar/Vector/DWI Volume* module) was used instead, enforcing the resampled scan to have the same image geometry as the reference 4DCT frame. The results obtained with the optimized *RunMidP* module are not presented in this section but are presented in Section 5.2.5 and will be included in two papers in progress [120], [121]. The results showed better agreement with ground truth Mid-P CTs and results from other Mid-Position packages, including Mirada prototype and Wimp.

As mentioned in Section 4.2.1, the Mid-P CT was reconstructed taking the median of the pixel intensities of resampled scans warped to the Mid-Position, since it is known to result in a sharper image than using the standard arithmetic mean [15]. The same conclusion was drawn comparing the Mid-P$_{\text{mean}}$ and Mid-P$_{\text{median}}$ reconstructions for the 16 treatment cases of Study B (Figure 5.14).

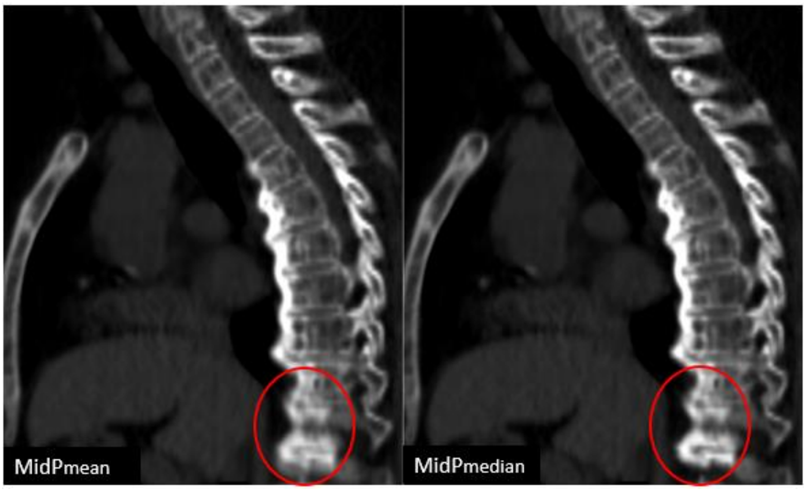


Figure 5.14: Sagittal views of the lung patient S12. The circle in the images focuses on a vertebra that was not registered correctly in phase 50%. Arithmetic averaging results in blurring of those structures, however, the median average is more robust against these misregistration artefacts, resulting in a visual sharper image.

---

[20] "Hardening a transform means permanently modify the node (e.g., 3D scan, segmentation, etc.), according to the currently applied transform. After hardening the transform, the coordinate values of model points are updated with transformed coordinates, or a warped image voxels are resampled." Source: 3D Slicer documentation, Transforms. https://slicer.readthedocs.io/en/latest/user_guide/modules/transforms.html (Accessed on 04-11-2021)

As depicted in Figure 5.14, arithmetic averaging of the transformed 4DCT scans results in blurring of the misregistration artefacts, while median averaging suppresses outlier values and preserves higher frequencies in the signal (edges and small similar artefacts of the 4D dataset), resulting in a visual sharper image.

5.2.2 Motion amplitude evaluation

Figure 5.15 shows ground truth tumour motion amplitudes of each patient in Study B compared against the motion amplitudes estimated by the *RunMidP* module integrated into 3D Slicer and Mirada Medical's software. The mean and standard deviation of the range of the ground truth motion was 1.0 ± 0.1 mm in LR, 1.9 ± 0.1 mm in AP, and 4.0 ± 0.1 mm in CC. The largest range observed in each direction was 2.3 mm, 3.6 mm, and 8.9 mm respectively, where the CC direction saw the largest motion in 24 tumours (89%).

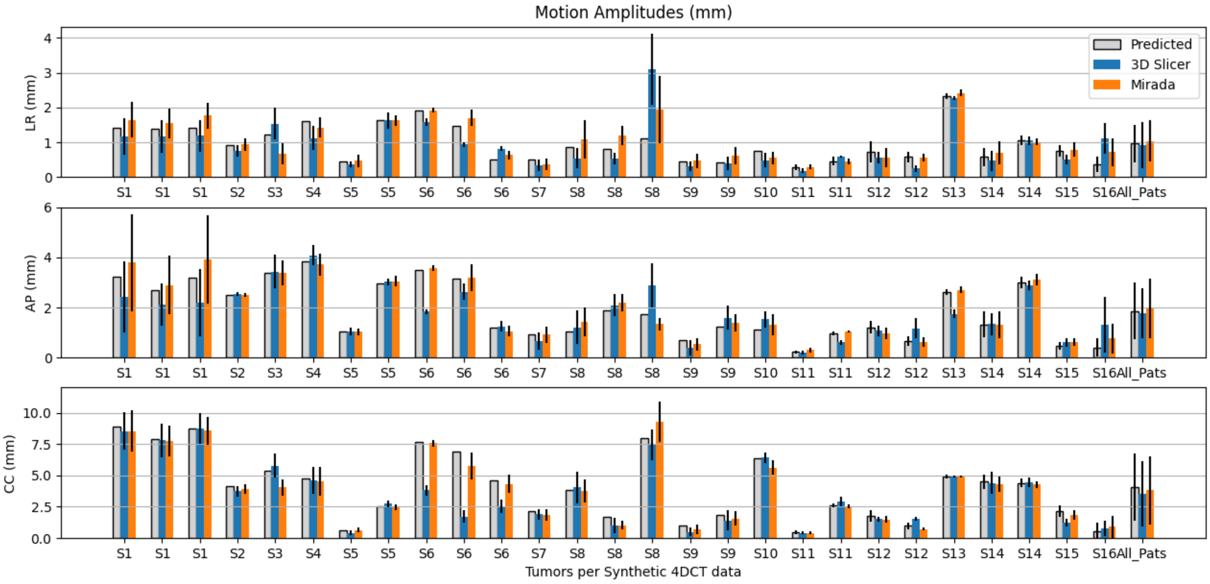


Figure 5.15: Comparison of tumour motion amplitudes. Ground truth motion amplitudes (gray); motion amplitudes calculated with the *RunMidP* module integrated into the 3D Slicer (blue) and motion amplitudes calculated with the Mirada Medical's algorithm (orange) in the three directions: left-right (LR), anterior-posterior (AP), and cranio-caudal (CC). Standard deviations for ground truth motion amplitudes are not available for the first 10 cases.

It is important to note that the motion amplitudes from the *RunMidP* module were calculated using the RM method described above (Section 4.2.2), although they look similar to the output of the centroid method (the CM method). The main advantage of the RM method over the CM method is that it is possible to have an estimation of the motion amplitude in all voxels within relevant ROIs, such as the tumour. Using the CM method, only one point is used to calculate the motion estimation.

Figure 5.16 shows the differences between motion amplitudes of the two Mid-Position methods compared against the ground truth estimation for each tumour of the synthetic 4DCT data of Study B. Motion amplitude differences were smaller than 1 mm for all tumours considering both Mid-Position methods, except for tumours of patient S6 (CTV1$_{S6}$, CTV2$_{S6}$, and CTV3$_{S6}$) and patient S8 (CTV3$_{S8}$) calculated with the *RunMidP* module. Table 5.2 summarizes motion differences for patients S6 and S8 in the three directions and estimated with both Elastix and Plastimatch algorithms. Using the Elastix registration, differences were 1.65 mm and 3.88 mm for CTV1$_{S6}$ in the AP and CC direction, respectively; 5.16 mm for CTV2$_{S6}$ in the CC direction; 2.05 mm for CTV3$_{S6}$ in the CC direction; and 1.98 mm and 1.16 mm for the CTV3$_{S8}$ in the LR and AP direction, respectively. The synthetic 4DCT data of patient S6 has small lung tumours (0.6, 1.0, and 0.4 cm^3) located in the lower lung lobe, where anatomical tissues move notably with breathing, and consequently, the Elastix registration may fail to register. This can be pointed out as one of the limitations of the *RunMidP* module compared to Mirada's Mid-Position algorithm. Patient S8 is based on clinical data of a liver patient. The homogeneous region of the liver may also difficult the registration for this case. Similar results were found using the Plastimatch image registration (Table 5.2).

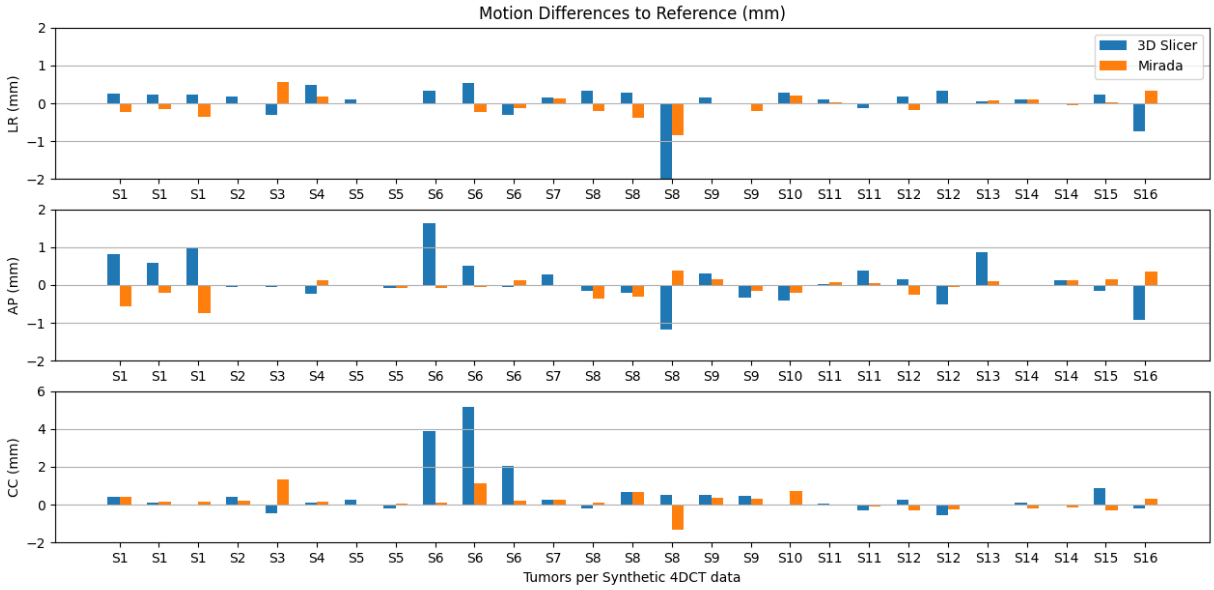


Figure 5.16: Motion amplitude differences between motion amplitudes calculated with both software – *RunMidP* module integrated into 3D Slicer (blue) and Mirada Medical (orange) – and ground truth motion estimations for each tumour of the synthetic 4DCT data of Study B.

Table 5.2: Motion differences with respect to ground truth motion amplitudes for patients S6 and S8 in the three directions: left-right (LR), anterior-posterior (AP), and cranio-caudal (CC), and estimated with Elastix and Plastimatch software. Motion differences higher than 1 mm are highlighted in red.

		Motion Differences to Reference (mm)								
		CTV1			**CTV2**			**CTV3**		
Patient ID	**DIR software**	**LR**	**AP**	**CC**	**LR**	**AP**	**CC**	**LR**	**AP**	**CC**
S6	Elastix	0.33	**1.65**	**3.88**	0.53	0.50	**5.16**	-0.31	-0.06	**2.05**
	Plastimatch	0.94	-0.21	**3.94**	0.72	0.08	**4.64**	-0.19	0.26	**1.78**
S8	Elastix	0.33	-0.16	-0.19	0.27	-0.21	0.69	**1.98**	**-1.16**	0.50
	Plastimatch	-0.47	-0.47	-0.38	-0.59	0.29	0.81	**-1.67**	**-1.41**	**1.14**

One of the limitations of this study is the limited respiratory motion in our data. The average tumour motion amplitude was significantly smaller than measured in other 4D-CT studies (4.0 ± 0.1 mm in CC direction versus 12.3 ± 7.1 mm in [15] and 12.4 ± 6.9 mm in [90]. An explanation for this is that our patients undergo abdominal compression during 4DCT scan acquisition, reducing the breathing motion.

5.2.3 Planning margins evaluation

Treatment margins in radiotherapy account for uncertainties that occur during treatment planning and treatment delivery (Section 4.3.2.4). Improvements in radiotherapy of lung and upper abdominal cancer patients are often focused on the reduction of the PTV to enable dose escalation. It is known from the literature that the Mid-P approach can significantly reduce the PTV volume compared to the ITV approach by using motion information from 4DCT scans in the treatment-planning process [6], [15].

This section focuses on the contribution of the tumour motion amplitude to the CTV-to-PTV margins. In particular, motion amplitudes differences of about 1 mm were found between both Mid-Position algorithms (3D Slicer and Mirada) against ground truth motion amplitudes. Figure 5.17 shows the impact of these motion amplitudes differences on treatment planning margins. As expected, the difference between the two Mid-Position methods is very small when motion estimations are converted into treatment-planning margins. This was expected since the effect of random uncertainties is smaller than the systematic uncertainties in the CTV-to-PTV margins (Equation 4.3) and the geometrical errors with respect to the respiratory motion are in turn modelled as about one third of the peak-to-peak amplitude. For the majority of the analysed cases, the margins are almost identical compared to ground truth estimations with differences within 0.3 mm in the three directions. Maximum differences were 0.7 mm, 0.8 mm and 0.3 mm for $CTV1_{S6}$, $CTV2_{S6}$, and $CTV3_{S6}$ in CC direction, respectively. This result was expected since the difference in motion amplitude estimation from the *RunMidP* module was also higher for these tumours.

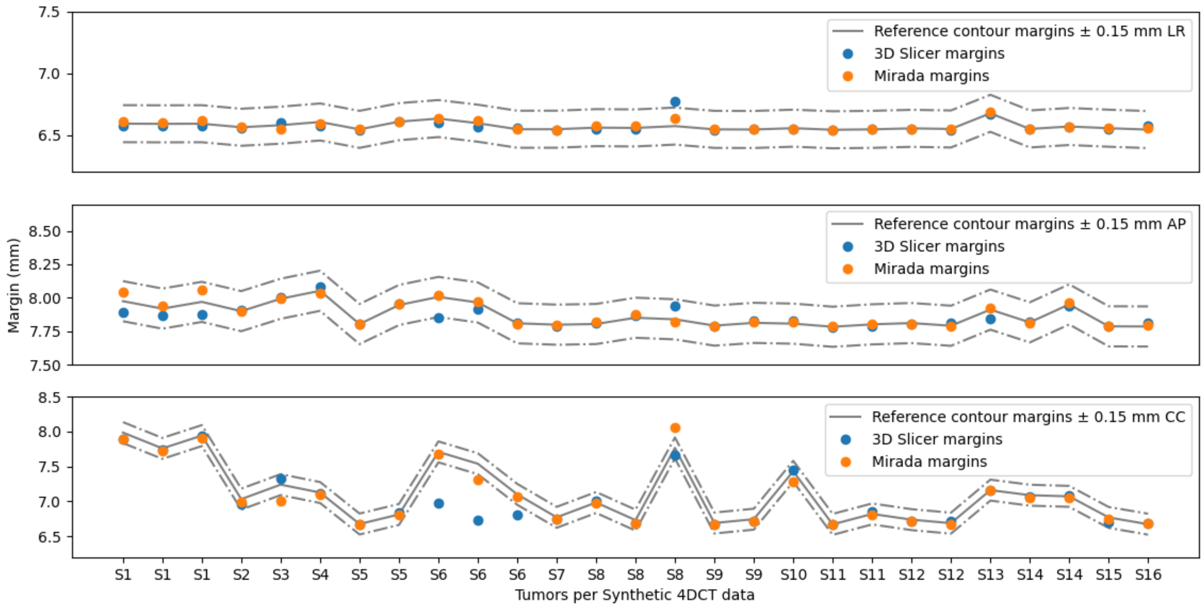


Figure 5.17: Treatment-planning margins for motion amplitude estimations with both software – *RunMidP* module integrated into 3D Slicer (blue) and Mirada Medical (orange) – compared against ground truth margins.

5.2.4 Hounsfield Units distribution

Hounsfield Units (HU) within ROIs on Mid-P CT scans should match the HU on the original 4DCT data. To evaluate the intensity distribution within contours, 10 patients of Study B were selected, yielding in total 13 tumour contours and 21 OARs contours, including the aorta, the heart, the spinal cord, and the liver.

According to the Mann-Whitney test, intensity distributions within the tumour matched the 4DCT scan in only 33% and 50% of the cases for 3D Slicer and Mirada software, respectively. Regarding OARs, Mid-P scans matched the 4DCT scan in only 33% and 38% of the cases, respectively. Figure 5.18 shows the HU intensity difference within tumour contours on the two Mid-P CT scans (reconstructed by 3D Slicer and Mirada) against tumour contours on one selected 4DCT phase. Less than 2% and 5% of the tumour voxels yielded HU intensity differences higher than 100 HU and 50 HU, respectively, compared to the 4DCT scan. Similar HU differences ($HU_{Mid\text{-}PCTs} - HU_{4DCT}$) were found for relevant OARs (see Appendix V).

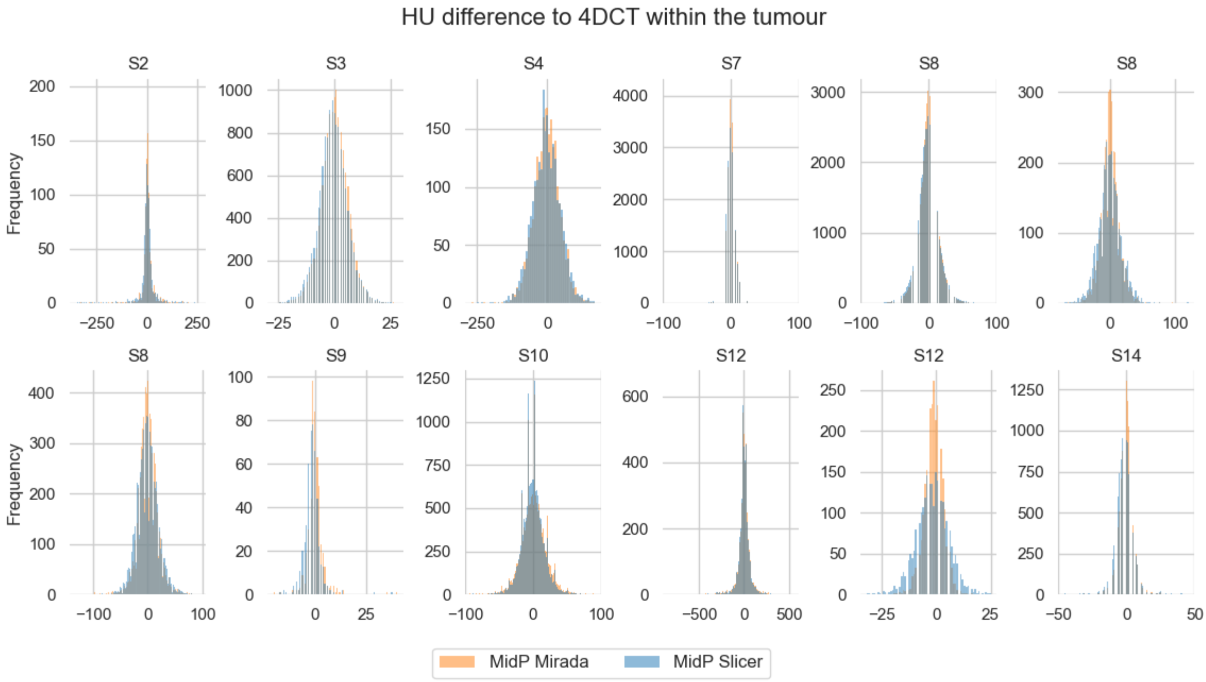


Figure 5.18: Histograms showing the distributions of the intensity difference within tumour contours between the two Mid-Position CT scans (reconstructed by 3D Slicer and Mirada software) and one 4DCT scan.

Table 5.3 summarizes the average and standard deviation of HU intensity differences within the tumour and other relevant ROIs, including the aorta, the heart, the spinal cord, and the liver. As shown in Table 5.3, the average HU difference to 4DCT is minimal. The remaining HU differences can be explained due to the fact that the patient’s anatomy in the chosen 4DCT phase does not perfectly match with the patient’s anatomy in its mean position over the respiratory cycle.

Table 5.3: Hounsfield Units (HU) difference on Mid-P CTs compared to 4DCT within selected ROIs for 10 patients of Study B.

		HU difference to 4DCT within ROIs			
PatientID	**Scan**	**ROI #1**	**ROI #2**	**ROI #3**	**ROI #4**
S2	MidP Slicer	0.0 ± 45.7	-0.6 ± 12.4	0.1 ± 7.7	-0.7 ± 11.1
	MidP Mirada	2.5 ± 37.1	0.1 ± 6.0	0.6 ± 4.5	0.3 ± 11.2
S3	MidP Slicer	-0.5 ± 6.3	-1.0 ± 19.3	-0.3 ± 8.8	1.8 ± 33.1
	MidP Mirada	0.3 ± 5.9	0.0 ± 18.7	0.3 ± 9.1	1.9 ± 34.6
S4	MidP Slicer	-3.0 ± 48.6	-0.1 ± 27.2	0.4 ± 30.9	-0.8 ± 42.8
	MidP Mirada	-1.4 ± 45.4	0.8 ± 24.7	0.0 ± 23.0	0.2 ± 39.9
S7	MidP Slicer	-0.6 ± 16.2	0.1 ± 7.6	-0.5 ± 4.6	-0.7 ± 8.2
	MidP Mirada	-0.1 ± 12.5	1.0 ± 7.0	0.3 ± 3.2	0.3 ± 7.9
S8	MidP Slicer	-0.8 ± 17.2	-0.5 ± 16.9	-0.5 ± 19.4	-0.7 ± 18.9
	MidP Mirada	0.1 ± 15.8	0.4 ± 14.1	0.2 ± 18.0	0.3 ± 16.5
S9	MidP Slicer	-1.4 ± 4.0	0.3 ± 12.8	-0.5 ± 7.3	-0.7 ± 12.4
	MidP Mirada	0.2 ± 4.2	1.0 ± 12.5	0.2 ± 5.4	0.2 ± 12.0
S10	MidP Slicer	-0.3 ± 43.4	-0.2 ± 10.3	-0.1 ± 11.5	-0.7 ± 15.4
	MidP Mirada	0.8 ± 59.6	0.5 ± 9.2	0.2 ± 4.1	0.3 ± 17.8
S12	MidP Slicer	0.1 ± 69.6	-1.7 ± 7.4	2.1 ± 40.0	-2.3 ± 26.7
	MidP Mirada	1.4 ± 70.3	-1.1 ± 4.2	2.6 ± 40.3	-1.4 ± 26.3
S14	MidP Slicer	-1.3 ± 8.1	-0.4 ± 7.2	0.5 ± 12.5	-0.7 ± 7.7
	MidP Mirada	-0.1 ± 6.2	0.3 ± 6.2	0.3 ± 4.0	0.2 ± 6.5

According to [112] and [113], variations of ±50, ±30, and ±100 HU in the CT number are required to produce 2% dose errors at 6 MV photon for lung, soft tissue, and bone respectively. Based on the results presented in this section, it is reasonable to believe that both Mid-P images can be used for relevant radiotherapy tasks, namely for dose calculation. Furthermore, the similarity between intensity distributions within tumour contours and OARs on the two Mid-P CT scans suggest that applying the same dose plan to both scans and observing the difference in the received doses will produce negligible dosimetric differences.

5.2.5 Robustness study

This study aims to evaluate the effect of choosing one particular breathing phase of the 4DCT as the reference phase (an input parameter of the Mid-P CT algorithm) in the reconstruction of the Mid-P CT. Figure 5.19 shows the SSIM calculated (slice by slice) between ten reconstructed Mid-P_i CTs (where i = 0, 1, … , and 9 corresponds to the reference phase) and the ground truth Mid-P CT for patient S8. Average SSIM results over all slices for five patients of Study B are shown in Table 5.4.

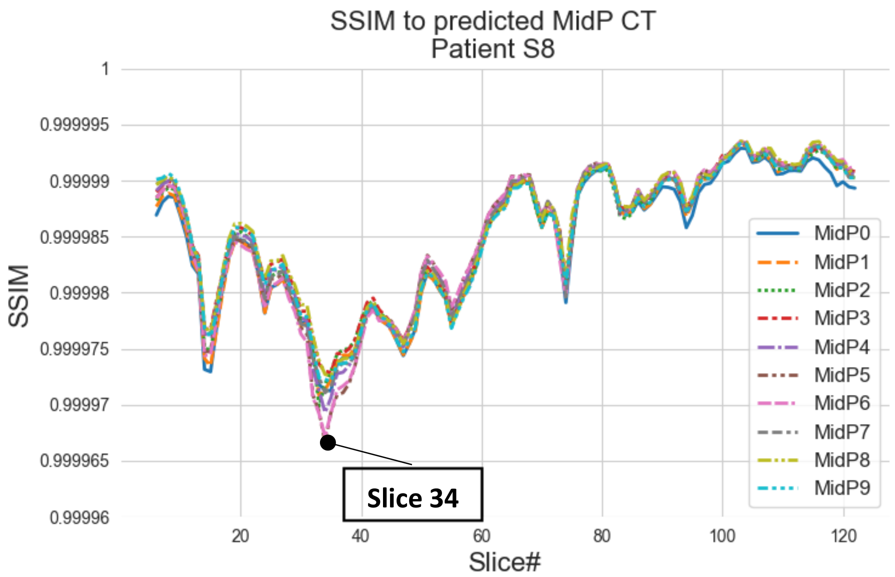


Figure 5.19: Structural Similarity (SSIM) index between ten reconstructed Mid-P CTs and the ground truth (predicted) Mid-P CT. Mid-P$_i$ stands for the Mid-P CT reconstructed using phase *i* of the 4DCT as the reference phase.

Table 5.4: SSIM results between 10 reconstructed Mid-P CTs and the ground truth (predicted) Mid-P CT averaged over all slices. The minimum average SSIM for each patient is highlighted in red, while the maximum is in green.

	Structural Similarity (SSIM) index									
Patient ID	**S1**		**S2**		**S5**		**S8**		**S9**	
Scans	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**
MidP$_0$ vs MidP$_{pred}$	#707	##94	#838	##71	#892	##41	#846	##60	#886	##57
MidP$_1$ vs MidP$_{pred}$	#718	##94	#842	##64	#894	##33	#848	##62	#893	##45
MidP$_2$ vs MidP$_{pred}$	#726	##90	#848	##68	#895	##32	#851	##59	#895	##43
MidP$_3$ vs MidP$_{pred}$	#726	##90	#850	##64	#894	##34	#856	##58	#897	##42
MidP$_4$ vs MidP$_{pred}$	#723	##89	#845	##71	#893	##34	#853	##62	#895	##41
MidP$_5$ vs MidP$_{pred}$	#719	##91	#845	##70	#891	##35	#852	##66	#893	##43
MidP$_6$ vs MidP$_{pred}$	#722	##91	#845	##71	#890	##37	#851	##67	#893	##41
MidP$_7$ vs MidP$_{pred}$	#729	##88	#847	##69	#895	##33	#853	##61	#894	##42
MidP$_8$ vs MidP$_{pred}$	#728	##89	#847	##69	#896	##31	#856	##60	#896	##41
MidP$_9$ vs MidP$_{pred}$	#717	##95	#846	##69	#896	##31	#852	##60	#894	##44

Abbreviations:
= 0.9999
= 0.0000

For each patient, the Mid-P_i that showed the minimum and maximum average SSIM (Table 5.4) was compared against the Mid-P_{pred} at the slice with minimum SSIM. As an example, for patient S8 the Mid-P_0 resulted in slightly lower SSIM compared to the other Mid-P_i CTs, while the Mid-P_8 showed the maximum average SSIM. Therefore, Mid-P_0 and Mid-P_8 were compared against the Mid-P_{pred} at slice nr. 34. The results are shown in Figure 5.20.

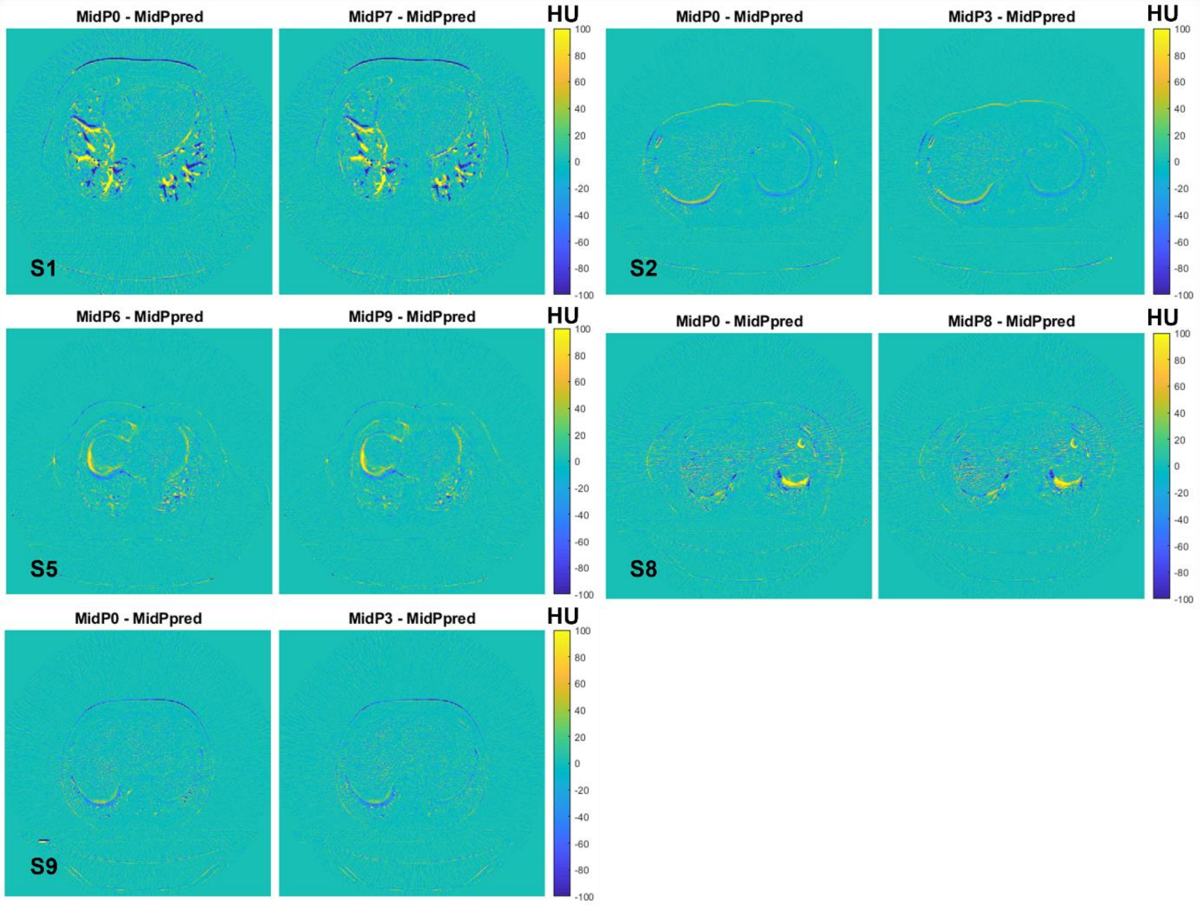


Figure 5.20: Axial views of the image difference between Mid-P CTs reconstructed using different 4DCT frames as the reference phase (Mid-P_i) and the ground truth Mid-P CT (Mid-P_{pred}). Colour scale range: -100 to 100 HU.

As shown in Figure 5.19 and Table 5.4, all Mid-P_i CTs are very similar to the ground truth Mid-P CT (SSIM ~ 1). Small shifts (in the order of magnitude of the CT resolution) were found consistently across all Mid-P_i CTs, with main incidence on the diaphragm's position and airways (Figure 5.20). Although theoretically, the Mid-P CT should be independent of the choice of the reference phase in the Mid-Position algorithm, the noise and image artefacts differ in each breathing phase of the 4DCT dataset, playing a role in the output of the image registration.

According to Kruis *et al*. [79], the maximum exhale phase contains fewer artefacts due to the reduced respiratory velocity and should be used as the reference phase. In addition, Mirada Medical software is optimized to select the max-exhale as the reference phase to reconstruct the Mid-P CT. Despite that, the results in this study suggest that the Mid-P algorithm itself is relatively robust to changes in the reference phase. Besides, there was not an agreement on the respiratory phase that would reconstruct the finest Mid-P CT, albeit it seems that the max-inhale (phase 0%) performs worse. Since the inspiration period is shorter than the expiration period, the maximum inspiration phase is expected to have more image artefacts than the maximum exhale phase which can difficult the registration.

By comparing the results presented in this section to the results in Section 5.2.1, there is a significant improvement of SSIM results for the same patients and using the same ground truth Mid-P CT as reference. As an example, for patients S2 and S5 the average SSIM was 0.99983 ± 0.00009 and 0.99990 ± 0.00005 before the module optimization (Table 5.1), while the most up to date results showed an average SSIM of 0.99998 ± 0.00007 and 0.99999 ± 0.00004 (Table 5.4), respectively. The reason for this is the fact that the module to hardening transforms was slightly changing the image geometry. After the software optimization with the *Resample Scalar/Vector/DWI Volume* module, a better agreement of reconstructed Mid-P CT scans with ground truth Mid-P CT was found.

5.3 Study C: Image Quality Assessment and Mid-P approach through doctors' eyes

5.3.1 Image Quality Assessment

Quantitative Assessment

SNR was compared in homogeneous regions in the liver in all 4DCT phases, Mid-P CT scans, and in the free-breathing planning CT (pCT).

The results shown in Figure 5.21 illustrate that the SNR in the Mid-P CT scan is higher than in the 4DCT frames and pCT. This was expected owing to the fact that the Mid-P CT scan is reconstructed averaging all ten 4DCT frames, reducing the noise (σ) in the image. The noise in the liver in the Mid-P CT scan was on average 5.1 HU (ranging from 3.1 to 8.0 HU), while in the pCT was 14.0 HU (ranging from 9.2 to 18.3 HU) and in the original 4DCT frames the noise was 18.5 HU (ranging from 10.2 to 34.2 HU). The reduction of the noise by a factor of the square root of the independent number of frames ($\sim\sqrt{10}$) is in good agreement with [114].

The increase in SNR in the liver by a factor of about three regarding each frame of the 4D CT scan was observed in the five patients analysed, suggesting an improvement in image quality using the Mid-P CT scan. Similar results were obtained in [15], where the noise in the air region as well as in the soft tissue region (mediastinum) was evaluated. The findings of Wang *et al.* [115] also suggest that Mid-P CT has superior SNR over the original 4DCT, due to a decrease in noise.

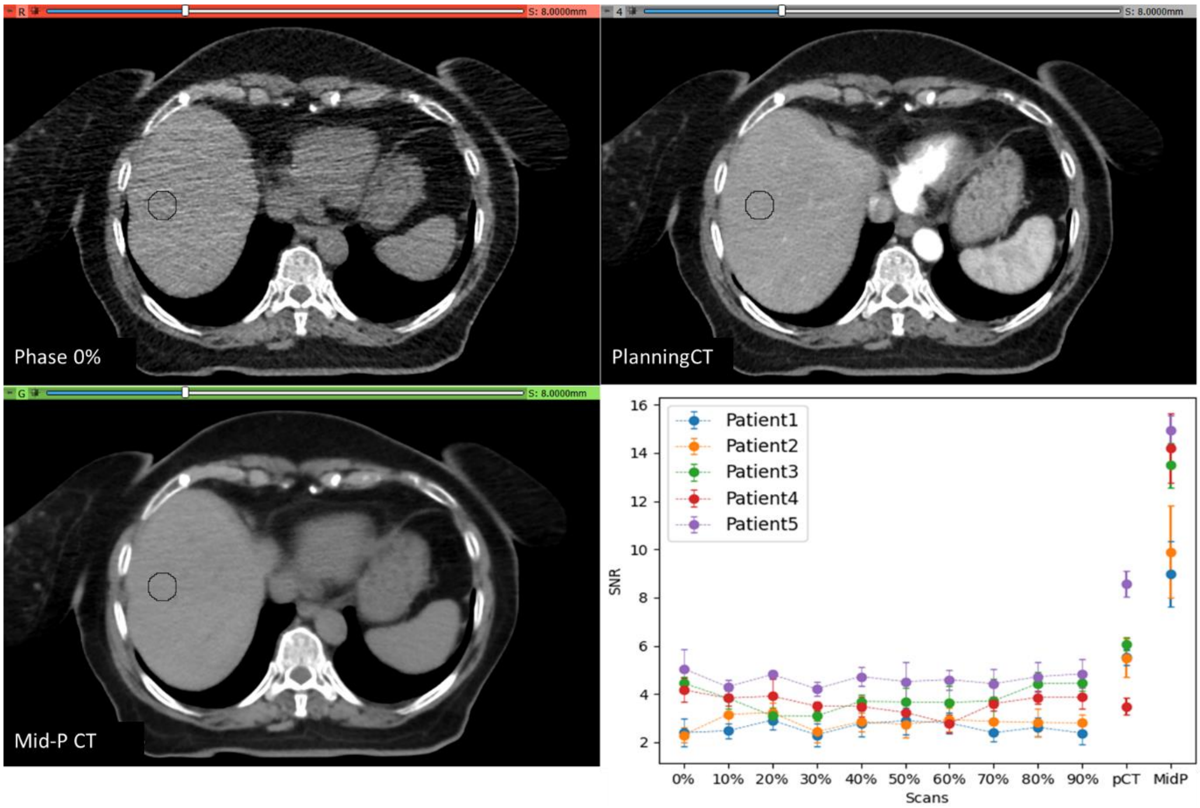


Figure 5.21: Axial slices from three different scans of Patient1: (A) one of the 4DCT frames (phase 0%); (B) the free-breathing planning CT; and (C) the Mid-P CT scan reconstructed by the *RunMidP* module. The SNR was calculated in each CT scan within circular ROIs in five adjacent slices in the liver, although only one ROI is shown in the image. (D) Mean and standard deviation of the SNR in homogeneous regions of the liver in ten 4DCT frames, in the free-breathing planning CT and in the Mid-P CT scan for the five patients in Study C. A reduction of the noise in the Mid-P CT scan compared to the planning CT and each frame of the 4DCT scan is clearly shown in the liver as well as in other organs.

Image sharpness is also important when images are intended to be used for radiotherapy planning. For instance, edges contain high variations in pixel intensities that create impacts on the perceived sharpness by the human vision system, facilitating organs differentiation and delineation. The sharpness of the Mid-P CT scan was quantified by the evaluation of intensity profiles along a user-defined line perpendicular to the diaphragm-lung interface. Lines were drawn on the diaphragm of five patients. A cumulative Gaussian distribution was then fitted to the intensity profile, and image sharpness was represented by the standard deviation. Results are shown in Figure 5.22.

The Average CT scan is visually blurrier than the Mid-P CT, while the pCT is sharper. Therefore, the image sharpness of the Mid-P CT scan was compared against one 4DCT frame, Average CT scan (as the worst-case scenario) and the pCT.

As depicted in Figure 5.23, the average width of the diaphragm on different CT scans was 2.0 mm, 1.8 mm, 1.6 mm and 4.3 mm for the Mid-P CT scan, pCT, one 4DCT frame and Average CT, respectively. This was to be expected since the Average CT has the lowest variations in pixel intensities in the dome of the diaphragm as well as in the lung parenchyma compared to the other CT scans. On the other hand, the image sharpness of the pCT is visually the highest although the width of the diaphragm was found to be slightly higher than in one of the 4DCT frames. A possible explanation for this might be the presence of image artefacts caused by respiratory motion that can be influencing the results (Figure 5.22).

Although the sharpness in the Mid-P CT scan is slightly lower compared to the pCT and one of the 4DCT frames, it has the main advantage of eliminating motion-induced artefacts due to patient breathing that are seen in the pCT (Figure 5.22). These image artefacts do not correspond to the patient's anatomy and may introduce serious errors in organ delineation. Therefore, a large improvement in image fidelity is possible when using Mid-P CTs for radiotherapy planning, with still comparable results in terms of image sharpness and increased SNR compared to the free-breathing pCT, which is currently the CT modality of choice for planning at the CCC.

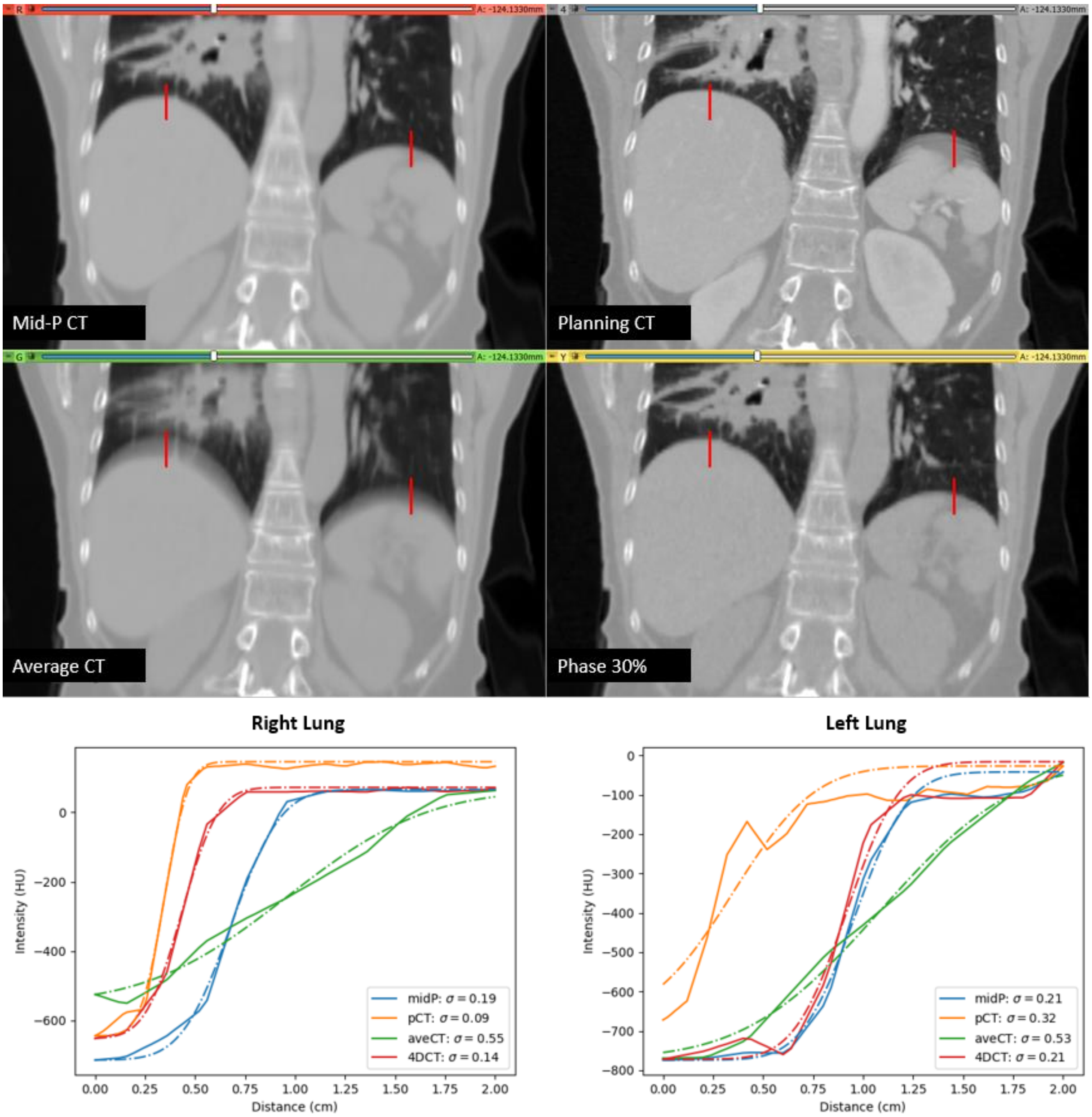


Figure 5.22: Image sharpness comparison between the Mid-P CT scan reconstructed by the *RunMidP* module, Average CT, free-breathing planning CT and one 4DCT frame (phase 30%) for Patient5. Intensity values along the red lines were plotted for the right and left diaphragm-lung interface. Standard deviations of the fitted cumulative Gaussian distribution were taken to evaluate the image sharpness of each scan (bottom row).

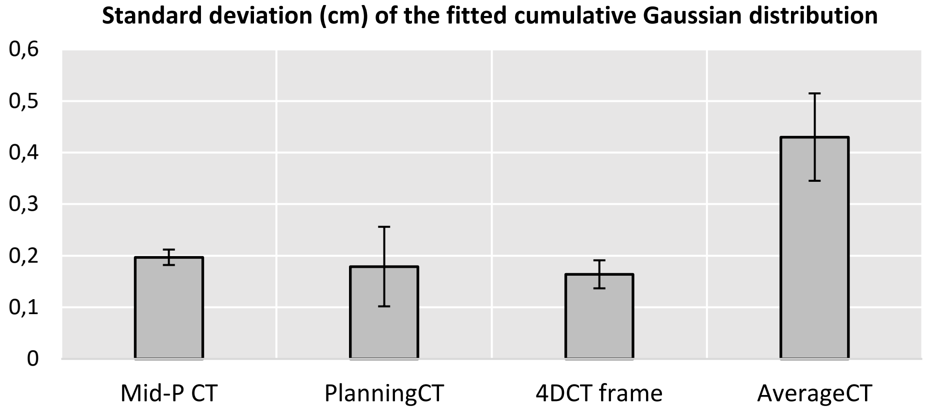


Figure 5.23: Mean and standard deviation of sigma values derived from the fitted cumulative Gaussian distribution along the lines drawn in the diaphragm-lung interface for five patients. The Mid-P CT has much higher image sharpness compared to the Average CT and just below that of the free-breathing planning CT and one 4DCT frame.

Qualitative Assessment

The purpose of this study was to evaluate the image quality of Mid-P CT scans and their feasibility for radiotherapy planning purposes. The qualitative assessment with regard to image sharpness, image contrast, image artefacts and contouring acceptability is shown in Figure 5.24 and the anatomical structures visualization assessment is shown in Figure 5.25. The statistical results (p-values) are shown in Appendix VII.

Regarding the data from NKI, image sharpness of Mid-P CT scans reconstructed by the three Mid-position software (3D Slicer, Wimp and Mirada) was significantly better than Average CT scans and not inferior to BH scans. On average, radiation oncologists scored Mid-P CT scans with *good* image sharpness, with 22% classified as *sharpest*. This comes in agreement with the increased (similar) sharpness of the diaphragm boundary on Mid-P CTs compared to Average CTs (pCTs) as measured earlier. The improved image quality of Mid-P CT scans over the Average CT scans and their comparable image quality to BH CT scans was confirmed by the results on image contrast, artefacts and contouring acceptability. Overall, Mid-P CT scans were considered *good* for contouring anatomical structures, in which 20% achieved the maximum score (*excellent* contouring acceptability). For the visually sharp reproduction of selected anatomical structures, the qualitative assessment revealed no statistically significant differences in the heart and great vessels, oesophagus, lung parenchyma, tumour, border between pleura and thoracic wall and diaphragm in the different types of scans analysed, whereas the carina and lymph node area and main and segmental bronchi are better visualized in Mid-P CT scans than in Average CT scans. Hence Mid-P CT scans produce a sharper reproduction of some anatomical structures compared to Average CT scans and do not differ from BH CT scans.

Regarding the data from CCC, Mid-P CT scans showed higher image sharpness, image contrast and contouring acceptability than Average CT scans, with *acceptable* image quality. At CCC, breath-hold CT scans are not routinely used for radiotherapy planning, so the comparison between Mid-P CT scans and breath-hold CT scans was not possible. Regarding image artefacts, there were no significant differences between these two planning CT scans. In addition, a downgrade of contouring acceptability was found for Mid-P CTs, only 22% of Mid-P CTs were considered *good* for contouring but never *excellent*. This is directly linked to the lower scores found in image sharpness and image contrast of Mid-P CT scans.

Figure 5.24 and Figure 5.25 show that the three Mid-Position packages performed similarly in image quality. In addition, it was found a significant decrease in the average scores of Mid-P CTs reconstructed with the CCC dataset when compared to NKI dataset. As an example, this resulted in a clearer appearance of the pleuromediastinal border and the diaphragm in the NKI dataset than in the CCC dataset. This strongly suggests that the input data itself can have an effect on the quality of the output (Mid-P) images. Different scanners, scanning parameters and 4D imaging acquisition protocols used in the two institutions may give rise to these differences. Therefore, the improvement of the image quality of Mid-P CTs and consequently, their clinical benefit for radiotherapy tasks, should start in the image acquisition stage.

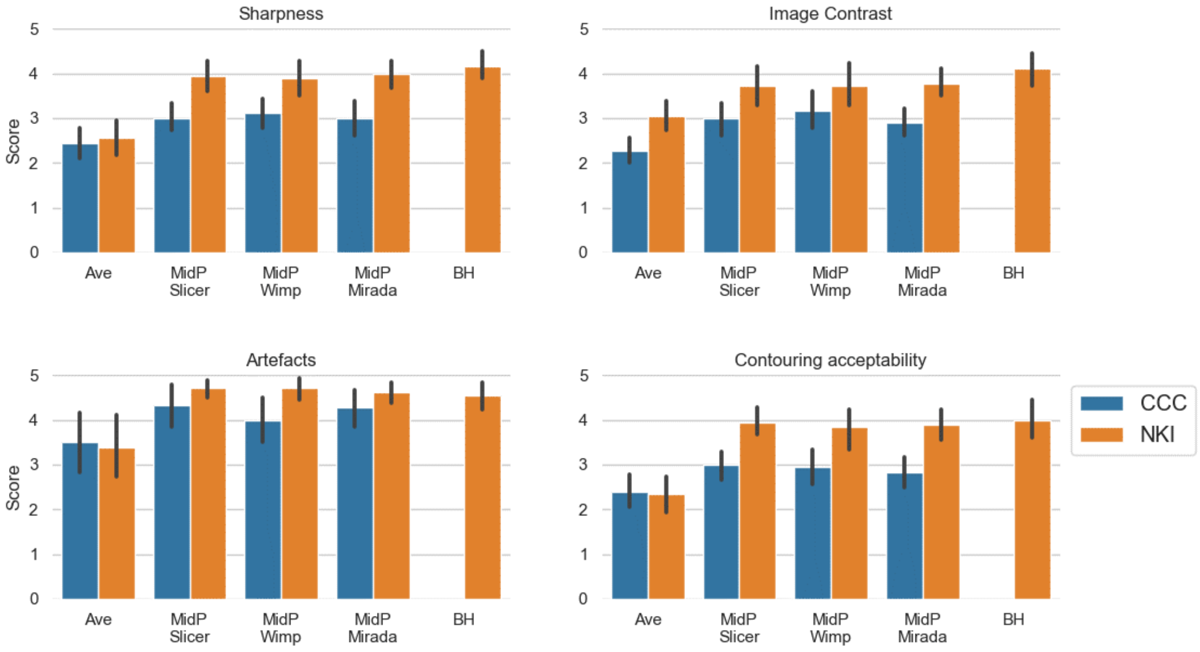


Figure 5.24: Average and standard deviation of image quality assessment scores (1-5) in three different types of planning CT scans (Average CT, Mid-P CT and BH CT). Mid-P CT scans were reconstructed by three different software (3D Slicer, Wimp and Mirada Medical Results are organized by institution (blue: CCC; orange: NKI).

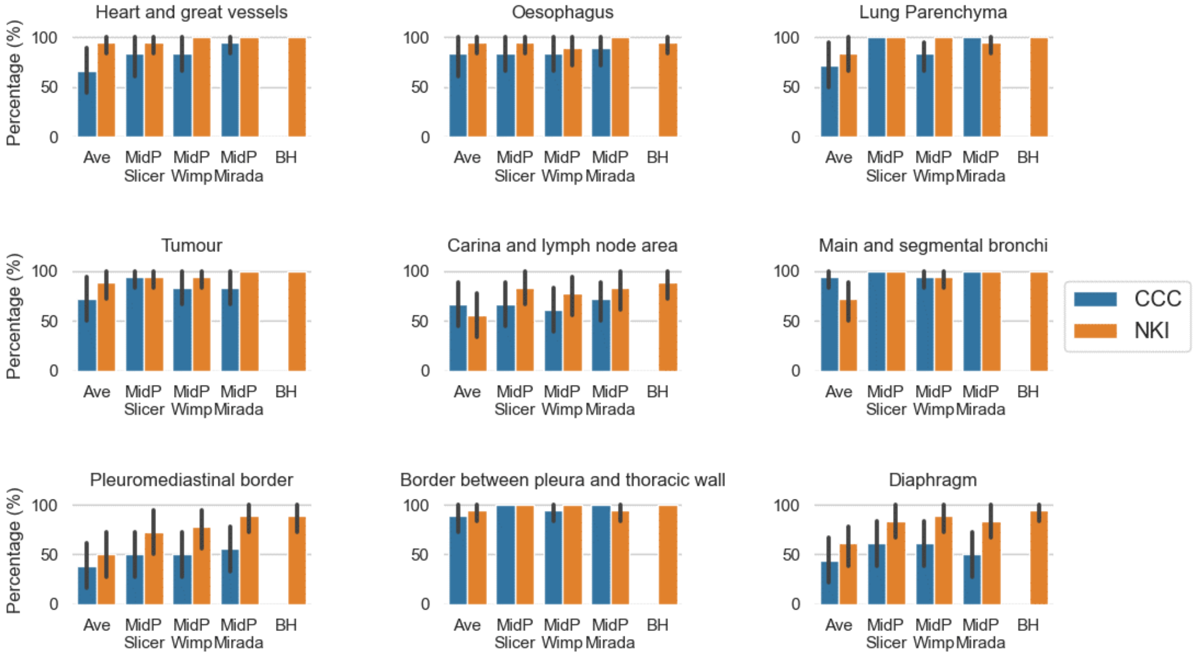


Figure 5.25: Average and standard deviations of visually sharp reproduction assessment scores in three different types of planning CT scans (Average CT, Mid-P CT and BH CT). Mid-P CT scans were reconstructed by three different software (3D Slicer, Wimp and Mirada Medical). Results are organized by institution (blue: CCC; orange: NKI).

It is noteworthy that there was a fair agreement for image sharpness (κ = 0.35), image contrast (κ = 0.21) and contouring acceptability (κ = 0.27) between the two reviewers, but only a slight agreement (κ = 0.08) for image artefacts. This means that both radiation oncologists perfectly agreed in the scoring of 40%, 27%, 43% and 32% of the scans, regarding image sharpness, image contrast, artefacts and contouring acceptability, respectively. However, their scores frequently show small differences (e.g., *acceptable* image sharpness *vs good* image sharpness). This slightly mismatching classification induced lower degrees of agreement between the two readers. This might have occurred due to their difference in clinical experience and the use of a 5-point grading scale with blurred boundaries between categories. Disagreement across major categories was only found for 16% of the scans regarding the image artefacts – only in Average CT scans. Reader 1 frequently scored the Average CT scans with *no* or *minimal artefacts*, whereas Reader 2 scored them with *unacceptable* or *many/large artefacts*.

On the basis of these results, both radiation oncologists from CCC would prefer to use Mid-P CT scans for the treatment planning in radiotherapy for lung, compared to the routinely used Average CT scans. However, the differences between the scores given to the two datasets by the doctors suggest that there is still room for image quality improvement before its full implementation in clinical practice. In addition, variations in the scores given by the two doctors show that visual assessment of any Mid-P method should be performed by the clinicians who will be using the output images at that institution, using input images generated by that institution.

6 Conclusion and Future Work

With the introduction of more demanding fractionation schemes, where fewer irradiation fractions with higher doses per fraction are delivered, motion management techniques have been developed to account for organ motion due to breathing and to spare healthy tissue from the irradiation field. Among them, the Mid-Position approach has gained interest since it was introduced in 2008 at the Netherlands Cancer Institute (NKI). It uses the complete 4D scan and deforms each separate frame to the time-weighted mean position. The resulting “mid-position” scan (Mid-P CT) allows the reduction of treatment volumes (compared to the ITV strategy), which consequently enables boosting the dose to the tumours and increases the overall survival for patients with tumours in the lung and upper abdominal region. In addition, displacements vector fields (DVFs) calculated during the image registration process can subsequently be used to estimate motion amplitudes of the tumour and OARs, which remains a hard radiotherapy task for clinicians to assess visually.

This work was motivated by the lack of commercial software to generate the Mid-P CT, which is hampering the introduction of such technique in clinics. In order to implement the Mid-Position approach at Champalimaud Clinical Centre (CCC), an in-house software module to reconstruct the Mid-P CT scan (the *RunMidP* module) was optimized and a set of experiments were performed to validate the software. Most of the results generated from the in-house software were benchmarked with the results from other Mid-Position packages, including a soon-to-be commercial software developed by Mirada Medical and the Wimp software that has been in regular clinical use at NKI for about a decade.

The experiments were conducted throughout three different studies. Study A evaluated the registration accuracy and image quality of two image registration algorithms (Plastimatch and Elastix) using quantitative metrics applied on either warped images or manually delineated contours. These contours and manually identified landmarks by medical experts (POPI-model data) were used to determine whether Mid-P CTs can represent the patient’s time-weighted mean position accurately. Study B continues to evaluate the feasibility of the in-house software to reconstruct the Mid-P CT scans and to estimate motion amplitudes, comparing the results against "true" estimations, thus excluding uncertainties that come from manual operations. In addition, Hounsfield Units (HU) within tumour and OARs contours on Mid-P CT scans were compared against 4DCT contours for a dosimetric evaluation. Finally, in Study C, the image quality of Mid-P CT scans was assessed quantitatively and qualitatively.

The results presented in this work are very promising regarding the use of the *RunMidP* module for the treatment planning of respiration-induced moving tumours in radiotherapy for lung and upper abdomen cancer, which ultimately can replace the currently used ITV strategy performed on standard methods (i.e., free-breathing planning CT and Average CT scan, not correlated to respiration).

Regarding the registration accuracy, both Plastimatch and Elastix registration showed good and comparable registration accuracy, but the superior image quality of the reconstructed Mid-P CTs by Elastix registration motivated the replacement of the registration method. Taking into account contour metrics – the Hausdorf Distance (HD) and Dice similarity coefficient (DSC) – the Elastix registration error was estimated to be less than 1 mm in the tumour contour and about 2 mm throughout the lung. The DSC ranged from 0.73 to 0.95, thus showing a good match between warped and reference tumour contours. In addition, differences in tumour mean position and motion amplitudes were in order of 0.5 and 1 mm in the three directions, respectively. Similar results were found for the Mirada Medical software. This implies that, at least in the tumour region, the *RunMidP* module is accurate enough to reconstruct the Mid-P CT scan and for estimating tumour motion amplitudes.

When a more rigorous validation test was used, comparing reconstructed Mid-P CTs with predicted Mid-P CTs (considered as “truth”), it appears that the *RunMidP* module performed slightly worse than Mirada Medical software, although both software presented higher image differences in the diaphragm lung interface, where the patient's anatomy moves faster due to breathing, and in homogeneous regions such as the air regions in the bowel. Differences in image registration algorithms might be partly responsible for the lower performance found for the *RunMidP* module. Still, a better match between reconstructed and predicted Mid-P CTs was accomplished after changing the method to harden transforms in the *RunMidP* module, enforcing the preservation of the original image geometry.

Regarding motion estimations, the developed Mid-P algorithm can accurately estimate motion amplitudes of anatomical structures from the 4DCT dataset. Motion amplitude differences (estimated-predicted) smaller than 1 mm were observed in 37 moving tumours (78.7 %) which result in differences in the planning treatment margins of less than 0.3 mm. Exceptions were found for patients with small lung tumours located in the lower lung lobe, where the motion amplitude is higher due to breathing, and in homogeneous regions in the liver. Intensity-based image registration methods, such as the Elastix registration, require the image to have some texture. The lack of contrast in the liver or other organs in the pelvic region will complicate the registration. These results led to the conclusion that the *RunMidP* module is suitable to reconstruct the time-weighted patient’s mean position and estimate motion amplitudes, but future work should be done to improve the registration accuracy and occasional registration failures.

To evaluate the use of the Mid-P CT scan for treatment planning, further analysis should be done to confirm whether the observed HU differences (<100 HU within 98% of tumour voxels and <50 HU within 95% of tumour voxels) will produce meaningful differences in dose calculations.

This work also showed that the reduced noise (higher signal-to-noise ratio) and absence of image artefacts in the reconstructed Mid-P scans are great advantages to using the Mid-P CT scan for radiotherapy treatment planning instead of using the Average CT and the free-breathing pCT, which are the currently used modalities for that purpose. The Average CT has low spatial resolution and the free-breathing pCT may not precisely represent patient anatomy due to the presence of artefacts. According to the radiation oncologists’ opinion, Mid-P CT scans have superior image quality than Average CT scans and do not differ significantly from breath-hold scans (the preferentially used planning CT scan at NKI). A downgrade of image quality was found for Mid-P CTs reconstructed with 4DCT data from CCC, compared to the ones reconstructed with 4DCT data from another hospital (NKI). This indicates that the input data itself can influence more the quality of the output images than the registration approach. Therefore, visual assessment of any Mid-P method should be performed by the clinicians who will be using the output images at that institution, using input images generated by that institution. More importantly, the qualitative assessment results show that there is room to improve the image quality at CCC by e.g. adjusting 4D imaging acquisition protocols.

Although this work constitutes one of the 4D studies with more data heterogeneity, since it is based on 4DCT data from patients with moving tumour in several regions (e.g. lung, liver, pancreas, stomach, sternum, bronchi), it has some limitations. One of the limitations is the fact that the extent of the motion amplitudes is reduced since 4DCT data is acquired with abdominal compression. Midway through the project, we started including data from NKI which show bigger motion amplitudes. Our papers will be ready for publication soon and the results show that the *RunMidP* module has a good performance on the reconstruction of the Mid-P CT and performs similarly to other Mid-Position packages, although motion amplitudes seem to be slightly underestimated. Another limitation of the module is the time required to reconstruct the Mid-P CT scan. The new versions of the software should provide a faster generation of the Mid-P CT, which will also facilitate the introduction of this approach in the clinical workflow (even though it might run in the background).

Finally, the Mid-P approach only involves the reconstruction of a new planning CT scan and leaves the other treatment planning and treatment delivery aspects unchanged, which makes the Mid-P approach easy to implement. This work was certainly a step forward on the implementation of the Mid-P approach in the clinic, however further exploration is required to ensure that it could be safely transferable to clinical practice. Possible improvements and future work will be discussed in the next section.

6.1 Future directions

6.1.1.1 Improvements in registration accuracy

In this work, an optimization-based algorithm called Elastix was employed. It uses a control point grid in combination with B-spline deformation functions. These functions regularize the control point grid preventing ill-conditioning and instability of solutions issues. Elastix registration showed to be appropriate for the intended use with similar performance to other image registration algorithms (e.g., optical flow algorithms, applied on both Mirada and Wimp software) that are known to allow more "fluid" (i.e., less constrained) deformations. Nonetheless, the optimisation solution in Elastix registration might be smoothing the vectors fields (DVFs), leading to registration errors and underestimation of motion amplitudes. In addition, limitations in the registration model can cause errors, such as the inability to handle tissue interfaces that are sliding (e.g., the lung sliding along the rib cage) and the large numbers of degrees of freedom in the registration may lead to ambiguity in DVFs, especially in areas with very low tissue contrast, thus leading to occasional registration failures as found in this work.

These issues can be addressed by adjusting the registration parameters, such as the similarity metric, transformation model, and optimizer. Elastix registration allows for altering those parameters, and they have not been all explored in this work. It might be recommended to use different parameters for different tumour localization and use-cases.

6.1.1.2 Fast-commissioning of the Mid-position software

The reconstruction of the Mid-position scan involves several steps that can be susceptible to errors and have a serious impact on the patient treatment. They can be divided into three broad areas: (1) 4D data transfer and usage by the *RunMidP* module; (2) image registration; and (3) actual reconstruction of the Mid-P CT scan.

The first step can be validated using the added feature of the *RunMidP* to create a movie of the 4DCT dataset that will be used to reconstruct the Mid-P CT scan. If any breathing phase is missing or if the breathing motion seems to be abnormal, a double-check of the input data should be performed.

Regarding the image registration, one essential question that remains unanswered is in which cases (patients) the registration will fail. *Are the cases with small tumours more prone to image registration errors? Or with big breathing motion amplitudes? What are the limits for a correct image registration?* As described in this work, several experiments can be performed for image registration validation. However, most of them require extensive time and effort thus being impractical in the clinic. A fast and easy validation test should be integrated into the Mid-position module. As an example, all warped phases should ideally be almost identical and match the reference phase. In addition, the patient's anatomy should not be shifted within 1-2 voxels regarding its original position [45]. Therefore, a 4D movie of all breathing phases warped to the reference phase should be provided and evaluated by the clinical personnel.

Finally, the patient’s mean position represented by the Mid-P CT scan should be commissioned, preferentially using a ground truth for comparison. The method developed at CCC to provide the “true” Mid-P CT scan would be able to verify that the reconstructed Mid-P CT was correctly calculated, and thus it should be added as an extension to the *RunMidP* module.

6.1.1.3 Fast generation of the Mid-P CT scan

The reconstruction of the Mid-P CT scan is currently taking about one hour per patient. Although it can run “offline” in the background, decreasing the computation time will facilitate its implementation in radiotherapy workstations. The number of samples (voxels) is an important registration setting that influences both registration speed and quality. If the sample number is too small, registration may fail because it is driven by image content that insufficiently represents the image. If the number is too large, registration can slow down significantly. In this work, the number of samples was set to 2048 that are randomly selected in the entire image domain. It can be more useful to use a small number of samples that are randomly selected within a user-defined ROI that fall inside the body contour. In addition, some optimisers available in Elastix registration can be faster than gradient descent optimisers. As an example, the Robbins-Monro optimization method is known to accelerate registration significantly, without compromising registration accuracy [59]. Finally, a maximum number of iterations of 200 (instead of 500) may suffice.

6.1.1.4 Improvements in CT imaging

In this work, superior image quality was found for the 4DCT dataset from NKI compared to the CCC dataset. At NKI, 4DCT scans were performed in a Somatom Open (Siemens, Forchheim, Germany) and 10 frames covering the breathing period were reconstructed using phase-based sorting. In addition to the different CT scanners, the breathing signal is acquired using a thermocouple inserted into the entry of a regular oxygen mask, which measures temperature changes in airflow during inhalation (cold) and exhalation (warm), instead of using an abdominal belt as at CCC.

Other 4D imaging parameters (e.g., pitch, detector rotation time, etc.) may have an impact on the image quality of the 4D data. Some image artefacts were moreover found in the CCC dataset. Image artefacts during 4DCT acquisition can be caused by the inappropriate setting of imaging parameters and can consequently cause registration errors and blurring of the final Mid-P CT scan. Some authors claim amplitude-based sorting results in fewer reconstruction artefacts than phase-based sorting [79], [116]. In the future, it might be worth fine-tuning such scanning parameters to avoid anatomy-based artefacts.

Audio-visual coaching during 4D data acquisition can also result in fewer artefacts in the 4D image reconstruction [117]. If image artefacts are still present on the 4DCT dataset, Wolthaus *et al.* [15] recommend excluding the CT frames with strong image artefacts to improve the image quality of Mid-P CT scans. This examination can be easily carried out using the 4D movie of the breathing phases.

References

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, “Cancer Statistics, 2021,” *CA. Cancer J. Clin.*, vol. 71, no. 1, pp. 7–33, Jan. 2021, doi: 10.3322/caac.21654.

[2] K. E. Rosenzweig *et al.*, “Results of a phase I dose-escalation study using three-dimensional conformal radiotherapy in the treatment of inoperable nonsmall cell lung carcinoma,” *Cancer*, vol. 103, no. 10, pp. 2118–2127, May 2005, doi: 10.1002/cncr.21007.

[3] X. Qiao, O. Tullgren, I. Lax, F. Sirzén, and R. Lewensohn, “The role of radiotherapy in treatment of stage I non-small cell lung cancer,” *Lung Cancer*, vol. 41, no. 1, pp. 1–11, Jul. 2003, doi: 10.1016/S0169-5002(03)00152-1.

[4] H. D. Kubo and B. C. Hill, “Respiration gated radiotherapy treatment: A technical study,” *Phys. Med. Biol.*, vol. 41, no. 1, pp. 83–91, Jan. 1996, doi: 10.1088/0031-9155/41/1/007.

[5] P. J. Keall, S. Joshi, S. Sastry Vedam, J. V. Siebers, V. R. Kini, and R. Mohan, “Four-dimensional radiotherapy planning for DMLC-based respiratory motion tracking,” *Med. Phys.*, vol. 32, no. 4, pp. 942–951, 2005, doi: 10.1118/1.1879152.

[6] H. Peulen, J. Belderbos, M. Rossi, and J. J. Sonke, “Mid-ventilation based PTV margins in Stereotactic Body Radiotherapy (SBRT): A clinical evaluation,” *Radiother. Oncol.*, vol. 110, no. 3, pp. 511–516, 2014, doi: 10.1016/j.radonc.2014.01.010.

[7] J. W. H. Wolthaus *et al.*, “Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 65, no. 5, pp. 1560–1571, Aug. 2006, doi: 10.1016/j.ijrobp.2006.04.031.

[8] J. W. H. Wolthaus, J. J. Sonke, M. Van Herk, and E. M. F. Damen, “Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registrationa,” *Med. Phys.*, vol. 35, no. 9, pp. 3998–4011, 2008, doi: 10.1118/1.2966347.

[9] J. W. H. Wolthaus *et al.*, “Comparison of Different Strategies to Use Four-Dimensional Computed Tomography in Treatment Planning for Lung Cancer Patients,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 70, no. 4, pp. 1229–1238, Mar. 2008, doi: 10.1016/j.ijrobp.2007.11.042.

[10] S. Mercieca *et al.*, “Interobserver variability in the delineation of the primary lung cancer and lymph nodes on different four-dimensional computed tomography reconstructions,” *Radiother. Oncol.*, vol. 126, no. 2, pp. 325–332, Feb. 2018, doi: 10.1016/j.radonc.2017.11.020.

[11] “Non-Small Cell Lung Cancer (NSCLC): Practice Essentials, Background, Pathophysiology.” Accessed: Jul. 03, 2021. [Online]. Available: https://emedicine.medscape.com/article/279960-overview.

[12] S. M. Bentzen *et al.*, “Towards evidence-based guidelines for radiotherapy infrastructure and staffing needs in Europe: The ESTRO QUARTS project,” *Radiother. Oncol.*, vol. 75, no. 3, pp. 355–365, 2005, doi: 10.1016/j.radonc.2004.12.007.

[13] P. Baumann *et al.*, “Outcome in a prospective phase II trial of medically inoperable stage I non-small-cell lung cancer patients treated with stereotactic body radiotherapy,” *J. Clin. Oncol.*, vol. 27, no. 20, pp. 3290–3296, Jul. 2009, doi: 10.1200/JCO.2008.21.5681.

[14] K. De Jaeger *et al.*, “Pulmonary function following high-dose radiotherapy of non-small-cell lung cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 55, no. 5, pp. 1331–1340, Apr. 2003, doi: 10.1016/S0360-3016(02)04389-4.

[15] J. W. H. Wolthaus, *Four-dimensional imaging in radiotherapy for lung cancer patients (Ph.D. dissertation)*. 2009.

[16] R. Baskar, J. Dai, N. Wenlong, R. Yeo, and K. W. Yeoh, “Biological response of cancer cells to radiation treatment,” *Front. Mol. Biosci.*, vol. 1, no. 24, Nov. 2014, doi: 10.3389/fmolb.2014.00024.

[17] C. F. Dunne-Daly, “Principles of radiotherapy and radiobiology.,” *Semin. Oncol. Nurs.*, vol. 15, no. 4, pp. 250–259, 1999, doi: 10.1016/S0749-2081(99)80054-0.

[18] J. M. Brown, D. J. Carlson, and D. J. Brenner, “The tumor radiobiology of SRS and SBRT: Are more than the 5 Rs involved?,” *International Journal of Radiation Oncology Biology Physics*, vol. 88, no. 2. Int J Radiat Oncol Biol Phys, pp. 254–262, Feb. 01, 2014, doi: 10.1016/j.ijrobp.2013.07.022.

[19] M. Macià i Garau, “Radiobiology of stereotactic body radiation therapy (SBRT),” *Reports of Practical Oncology and Radiotherapy*, vol. 22, no. 2. Urban and Partner, pp. 86–95, Mar. 01, 2017, doi: 10.1016/j.rpor.2017.02.010.

[20] S. S. Yom, “Accelerated Repopulation as a Cause of Radiation Treatment Failure in Non-Small Cell Lung Cancer: Review of Current Data and Future Clinical Strategies,” *Seminars in Radiation Oncology*, vol. 25, no. 2. W.B. Saunders, pp. 93–99, Apr. 01, 2015, doi: 10.1016/j.semradonc.2014.12.002.

[21] S. J. McMahon, “The linear quadratic model: Usage, interpretation and challenges,” *Phys. Med. Biol.*, vol. 64, no. 1, Jan. 2019, doi: 10.1088/1361-6560/aaf26a.

[22] J. P. Kirkpatrick, J. J. Meyer, and L. B. Marks, “The Linear-Quadratic Model Is Inappropriate to Model High Dose per Fraction Effects in Radiosurgery,” *Semin. Radiat. Oncol.*, vol. 18, no. 4, pp. 240–243, Oct. 2008, doi: 10.1016/j.semradonc.2008.04.005.

[23] I. Rosenberg, *Radiation Oncology Physics: A Handbook for Teachers and Students*, vol. 98, no. 5. Springer Nature, 2008.

[24] C. Hægeland, *Optimization and verification of dosimetric robustness of VMAT dose-plans (Ph.D. dissertation)*. 2016.

[25] B. M. Prendergast *et al.*, “Flattening filter-free linac improves treatment delivery efficiency in stereotactic body radiation therapy,” *J. Appl. Clin. Med. Phys.*, vol. 14, no. 3, pp. 64–71, 2013, doi: 10.1120/jacmp.v14i3.4126.

[26] T. M. Dang, M. J. Peters, B. Hickey, and A. Semciw, “Efficacy of flattening-filter-free beam in stereotactic body radiation therapy planning and treatment: A systematic review with meta-analysis,” *J. Med. Imaging Radiat. Oncol.*, vol. 61, no. 3, pp. 379–387, Jun. 2017, doi: 10.1111/1754-9485.12583.

[27] C. Chan, S. Lang, C. Rowbottom, M. Guckenberger, and C. Faivre-Finn, “Intensity-modulated radiotherapy for lung cancer: Current status and future developments,” *J. Thorac. Oncol.*, vol. 9, no. 11, pp. 1598–1608, Nov. 2014, doi: 10.1097/JTO.0000000000000346.

[28] A. Taylor and M. E. B. Powell, “Intensity-modulated radiotherapy - What is it?,” *Cancer Imaging*, vol. 4, no. 2, pp. 68–73, 2004, doi: 10.1102/1470-7330.2004.0003.

[29] N. K. Sharma, T. Li, D. Y. Chen, A. Pollack, E. M. Horwitz, and M. K. Buyyounouski, “Intensity-modulated radiotherapy reduces gastrointestinal toxicity in patients treated with androgen deprivation therapy for prostate cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 80, no. 2, pp. 437–444, Jun. 2011, doi: 10.1016/j.ijrobp.2010.02.040.

[30] M. Ashraf *et al.*, “Dosimetric comparison of 3DCRT versus IMRT in whole breast irradiation of early stage breast cancer,” *Int. J. Cancer Ther. Oncol.*, vol. 2, no. 3, p. 020318, Aug. 2014, doi: 10.14319/ijcto.0203.18.

[31] C. M. Nutting *et al.*, “Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial,” *Lancet Oncol.*, vol. 12, no. 2, pp. 127–136, Feb. 2011, doi: 10.1016/S1470-2045(10)70290-4.

[32] J. A. Bogers, R. W. M. Van der Maazen, and A. G. Visser, “Conformal photon-beam radiotherapy of prostate carcinoma,” *Eur. Urol.*, vol. 41, no. 5, pp. 515–522, May 2002, doi: 10.1016/S0302-2838(02)00017-9.

[33] M. Teoh, C. H. Clark, K. Wood, S. Whitaker, and A. Nisbet, “Volumetric modulated arc therapy: A review of current literature and clinical use in practice,” *British Journal of Radiology*, vol. 84, no. 1007. Br J Radiol, pp. 967–996, Nov. 2011, doi: 10.1259/bjr/22373346.

[34] T. Huynh *et al.*, “Estimating CT Image from MRI Data Using Structured Random Forest and Auto-Context Model,” *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 174–183, Jan. 2016, doi:

10.1109/TMI.2015.2461533.

[35] X. Han, “MR-based synthetic CT generation using a deep convolutional neural network method:,” *Med. Phys.*, vol. 44, no. 4, pp. 1408–1419, Apr. 2017, doi: 10.1002/mp.12155.

[36] T. Landberg *et al.*, “ICRU Reports,” *Reports Int. Comm. Radiat. Units Meas.*, vol. os-32, no. 1, pp. 48–51, Nov. 1999, doi: 10.1093/jicru_os32.1.48.

[37] S. Yuan *et al.*, “Determining optimal clinical target volume margins on the basis of microscopic extracapsular extension of metastatic nodes in patients with non-small-cell lung cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 67, no. 3, pp. 727–734, Mar. 2007, doi: 10.1016/j.ijrobp.2006.08.057.

[38] M. Van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque, “The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, no. 4, pp. 1121–1135, Jul. 2000, doi: 10.1016/S0360-3016(00)00518-6.

[39] J. Stroom *et al.*, “Combined recipe for clinical target volume and planning target volume margins,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 88, no. 3, pp. 708–714, Mar. 2014, doi: 10.1016/j.ijrobp.2013.08.028.

[40] W. van Elmpt, L. McDermott, S. Nijsten, M. Wendling, P. Lambin, and B. Mijnheer, “A literature review of electronic portal imaging for radiotherapy dosimetry,” *Radiother. Oncol.*, vol. 88, no. 3, pp. 289–309, Sep. 2008, doi: 10.1016/j.radonc.2008.07.008.

[41] D. Verellen, M. De Ridder, N. Linthout, K. Tournel, G. Soete, and G. Storme, “Innovations in image-guided radiotherapy,” *Nat. Rev. Cancer*, vol. 7, no. 12, pp. 949–960, Dec. 2007, doi: 10.1038/nrc2288.

[42] G. R. Borst, J. J. Sonke, A. Betgen, P. Remeijer, M. van Herk, and J. V. Lebesque, “Kilo-Voltage Cone-Beam Computed Tomography Setup Measurements for Lung Cancer Patients; First Clinical Results and Comparison With Electronic Portal-Imaging Device,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, no. 2, pp. 555–561, Jun. 2007, doi: 10.1016/j.ijrobp.2007.01.014.

[43] S. Mutic *et al.*, “TU-H-BRA-08: The Design and Characteristics of a Novel Compact Linac-Based MRI Guided Radiation Therapy (MR-IGRT) System,” *Med. Phys.*, vol. 43, no. 6, pp. 3770–3770, Jun. 2016, doi: 10.1118/1.4957630.

[44] B. W. Raaymakers *et al.*, “First patients treated with a 1.5 T MRI-Linac: Clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment,” *Phys. Med. Biol.*, vol. 62, no. 23, pp. L41–L50, Nov. 2017, doi: 10.1088/1361-6560/aa9517.

[45] K. K. Brock, S. Mutic, T. R. McNutt, H. Li, and M. L. Kessler, “Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132: Report,” *Med. Phys.*, vol. 44, no. 7, pp. e43–e76, Jul. 2017, doi: 10.1002/mp.12256.

[46] M. Van Herk, “Errors and Margins in Radiotherapy,” *Semin. Radiat. Oncol.*, vol. 14, no. 1, pp. 52–64, 2004, doi: 10.1053/j.semradonc.2003.10.003.

[47] Z. Chen, Z. Yang, J. Wang, and W. Hu, “Dosimetric impact of different bladder and rectum filling during prostate cancer radiotherapy,” *Radiat. Oncol.*, vol. 11, no. 1, Aug. 2016, doi: 10.1186/s13014-016-0681-z.

[48] P. J. Keall *et al.*, *The Management of Respiratory Motion in Radiation Oncology Report of AAPM Task Group 76*. 2006.

[49] J. C. Stroom and B. J. M. Heijmen, “Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report,” *Radiother. Oncol.*, vol. 64, no. 1, pp. 75–83, Jul. 2002, doi: 10.1016/S0167-8140(02)00140-8.

[50] J. Stroom, *Safety margins for geometrical uncertainties in radiotherapy (Ph.D. dissertation)*. 2000.

[51] J. M. Balter, R. K. Ten Haken, T. S. Lawrence, K. L. Lam, and J. M. Robertson, “Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 36, no. 1, pp. 167–174, 1996, doi: 10.1016/S0360-3016(96)00275-1.

[52] Y. Seppenwoolde *et al.*, “Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 53, no. 4, pp. 822–834, Jul. 2002, doi: 10.1016/S0360-3016(02)02803-1.

[53] J. J. Sonke, J. Lebesque, and M. van Herk, “Variability of Four-Dimensional Computed Tomography Patient Models,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 70, no. 2, pp. 590–598, Feb. 2008, doi: 10.1016/j.ijrobp.2007.08.067.

[54] E. D. Brandner *et al.*, “Abdominal organ motion measured using 4D CT,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 65, no. 2, pp. 554–560, Jun. 2006, doi: 10.1016/j.ijrobp.2005.12.042.

[55] W. Lu, G. H. Olivera, Q. Chen, M. L. Chen, and K. J. Ruchala, “Automatic re-contouring in 4D radiotherapy,” *Phys. Med. Biol.*, vol. 51, no. 5, pp. 1077–1099, Mar. 2006, doi: 10.1088/0031-9155/51/5/002.

[56] N. McVicar, S. Thomas, M. Liu, H. Carolan, and A. Bergman, “Re-irradiation volumetric modulated arc therapy optimization based on cumulative biologically effective dose objectives,” *J. Appl. Clin. Med. Phys.*, vol. 19, no. 6, pp. 341–345, Nov. 2018, doi: 10.1002/acm2.12481.

[57] S. Oh and S. Kim, “Deformable image registration in radiation therapy,” *Radiat. Oncol. J.*, vol. 35, no. 2, pp. 101–111, 2017, doi: 10.3857/roj.2017.00325.

[58] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, “Elastix: A toolbox for intensity-based medical image registration,” *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 196–205, Jan. 2010, doi: 10.1109/TMI.2009.2035616.

[59] S. Klein and M. Staring, *Elastix: the manual*. 2019.

[60] J. A. Shackleford, N. Kandasamy, and G. C. Sharp, “On developing B-spline registration algorithms for multi-core processors,” *Phys. Med. Biol.*, vol. 55, no. 21, pp. 6329–6351, Nov. 2010, doi: 10.1088/0031-9155/55/21/001.

[61] P. B. Johnson, K. R. Padgett, K. L. Chen, and N. Dogan, “Evaluation of the tool ‘Reg Refine’ for user-guided deformable image registration,” *J. Appl. Clin. Med. Phys.*, vol. 17, no. 3, pp. 158–170, 2016, doi: 10.1120/jacmp.v17i3.6025.

[62] D. Sarrut, T. Baudier, M. Ayadi, R. Tanguy, and S. Rit, “Deformable image registration applied to lung SBRT: Usefulness and limitations,” *Phys. Medica*, vol. 44, pp. 108–112, Dec. 2017, doi: 10.1016/j.ejmp.2017.09.121.

[63] S. Ehrbar *et al.*, “ITV, mid-ventilation, gating or couch tracking – A comparison of respiratory motion-management techniques based on 4D dose calculations,” *Radiother. Oncol.*, vol. 124, no. 1, pp. 80–88, Jul. 2017, doi: 10.1016/j.radonc.2017.05.016.

[64] E. Boman, M. Kapanen, L. Pickup, and S. L. Lahtela, “Importance of deformable image registration and biological dose summation in planning of radiotherapy retreatments,” *Med. Dosim.*, vol. 42, no. 4, pp. 296–303, Feb. 2017, doi: 10.1016/j.meddos.2017.06.006.

[65] G. F. Persson *et al.*, “Interobserver delineation variation in lung tumour stereotactic body radiotherapy,” *Br. J. Radiol.*, vol. 85, no. 1017, p. e654, Sep. 2012, doi: 10.1259/bjr/76424694.

[66] G. D. Hugo *et al.*, “A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer:,” *Med. Phys.*, vol. 44, no. 2, pp. 762–771, Feb. 2017, doi: 10.1002/mp.12059.

[67] J. Hanley *et al.*, “Deep inspiration breath-hold technique for lung tumors: The potential value of target immobilization and reduced lung density in dose escalation,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 45, no. 3, pp. 603–611, 1999, doi: 10.1016/S0360-3016(99)00154-6.

[68] S. Shimizu *et al.*, “Impact of respiratory movement on the computed tomographic images of small lung tumors in three-dimensional (3D) radiotherapy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 46, no. 5, pp. 1127–1133, Mar. 2000, doi: 10.1016/S0360-3016(99)00352-1.

[69] P. Giraud *et al.*, “Conformal radiotherapy (CRT) planning for lung cancer: Analysis of intrathoracic organ motion during extreme phases of breathing,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 51, no. 4, pp. 1081–1092, Nov. 2001, doi: 10.1016/S0360-3016(01)01766-7.

[70] M. Engelsman, E. M. F. Damen, K. De Jaeger, K. M. Van Ingen, and B. J. Mijnheer, “The effect of breathing and set-up errors on the cumulative dose to a lung tumor,” *Radiother. Oncol.*, vol. 60, no. 1, pp. 95–105, 2001, doi: 10.1016/S0167-8140(01)00349-8.

[71] S. Shimizu *et al.*, “Detection of lung tumor movement in real-time tumor-tracking radiotherapy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 51, no. 2, pp. 304–310, Oct. 2001, doi: 10.1016/S0360-3016(01)01641-8.

[72] C. Plathow *et al.*, “Analysis of intrathoracic tumor mobility during whole breathing cycle by dynamic MRI,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 59, no. 4, pp. 952–959, Jul. 2004, doi: 10.1016/j.ijrobp.2003.12.035.

[73] S. C. Davies, A. L. Hill, R. B. Holmes, M. Halliwell, and P. C. Jackson, “Ultrasound quantitation of respiratory organ motion in the upper abdomen,” *Br. J. Radiol.*, vol. 67, no. 803, pp. 1096–1102, 1994, doi: 10.1259/0007-1285-67-803-1096.

[74] S. S. Vedam, V. R. Kini, P. J. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan, “Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker,” *Med. Phys.*, vol. 30, no. 4, pp. 505–513, Apr. 2003, doi: 10.1118/1.1558675.

[75] S. S. Vedam, P. J. Keall, V. R. Kini, H. Mostafavi, H. P. Shukla, and R. Mohan, “Acquiring a four-dimensional computed tomography dataset using an external respiratory signal,” *Phys. Med. Biol.*, vol. 48, no. 1, pp. 45–62, Jan. 2003, doi: 10.1088/0031-9155/48/1/304.

[76] D. De Ruysscher *et al.*, “European organisation for research and treatment of cancer recommendations for planning and delivery of high-dose, high-precision radiotherapy for lung cancer,” *J. Clin. Oncol.*, vol. 28, no. 36, pp. 5301–5310, Dec. 2010, doi: 10.1200/JCO.2010.30.3271.

[77] Philips, *Respiratory motion management for CT*. 2013.

[78] H. Li *et al.*, “Clinical evaluations of an amplitude-based binning algorithm for 4DCT reconstruction in radiation therapy,” *Med. Phys.*, vol. 39, no. 2, pp. 922–932, 2012, doi: 10.1118/1.3679015.

[79] M. F. Kruis, J. B. Van De Kamer, J. S. A. Belderbos, J. J. Sonke, and M. Van Herk, “4D CT amplitude binning for the generation of a time-averaged 3D mid-position CT scan,” *Phys. Med. Biol.*, vol. 59, no. 18, pp. 5517–5529, Sep. 2014, doi: 10.1088/0031-9155/59/18/5517.

[80] K. E. Rosenzweig *et al.*, “The deep inspiration breath-hold technique in the treatment of inoperable non-small-cell lung cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 48, no. 1, pp. 81–87, Aug. 2000, doi: 10.1016/S0360-3016(00)00583-6.

[81] J. W. Wong *et al.*, “The use of active breathing control (ABC) to reduce margin for breathing motion,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 44, no. 4, pp. 911–919, Jul. 1999, doi: 10.1016/S0360-3016(99)00056-5.

[82] Y. Negoro *et al.*, “The effectiveness of an immobilization device in conformal radiotherapy for lung tumor: Reduction of respiratory tumor movement and evaluation of the daily setup accuracy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 50, no. 4, pp. 889–898, Jul. 2001, doi: 10.1016/S0360-3016(01)01516-4.

[83] E. Colvill *et al.*, “A dosimetric comparison of real-time adaptive and non-adaptive radiotherapy: A multi-institutional study encompassing robotic, gimbaled, multileaf collimator and couch tracking,” *Radiother. Oncol.*, vol. 119, no. 1, pp. 159–165, Apr. 2016, doi: 10.1016/j.radonc.2016.03.006.

[84] S. Sayeh, J. Wang, W. T. Main, W. Kilby, and C. R. Maurer, “Respiratory motion tracking for robotic radiosurgery,” in *Treating Tumors that Move with Respiration*, Springer Berlin Heidelberg, 2007, pp. 15–29.

[85] W. D. D’Souza, S. A. Naqvi, and C. X. Yu, “Real-time intra-fraction-motion tracking using the treatment couch: A feasibility study,” *Phys. Med. Biol.*, vol. 50, no. 17, pp. 4021–4033, Sep. 2005, doi: 10.1088/0031-9155/50/17/007.

[86] V. Caillet *et al.*, “MLC tracking for lung SABR reduces planning target volumes and dose to organs at risk,” *Radiother. Oncol.*, vol. 124, no. 1, pp. 18–24, Jul. 2017, doi: 10.1016/j.radonc.2017.06.016.

[87] E. Rietzel *et al.*, “Design of 4D treatment planning target volumes,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 66, no. 1, pp. 287–295, Sep. 2006, doi: 10.1016/j.ijrobp.2006.05.024.

[88] R. W. M. Underberg, F. J. Lagerwaard, B. J. Slotman, J. P. Cuijpers, and S. Senan, “Use of maximum intensity projections (MIP) for target volume generation in 4DCT scans for lung cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 63, no. 1, pp. 253–260, Sep. 2005, doi: 10.1016/j.ijrobp.2005.05.045.

[89] M. G. Witte, J. Van Der Geer, C. Schneider, J. V. Lebesque, and M. Van Herk, “The effects of target size and tissue density on the minimum margin required for random errors,” *Med. Phys.*, vol. 31, no. 11, pp. 3068–3079, Nov. 2004, doi: 10.1118/1.1809991.

[90] M. Guckenberger *et al.*, “Potential of image-guidance, gating and real-time tracking to improve accuracy in pulmonary stereotactic body radiotherapy,” *Radiother. Oncol.*, vol. 91, no. 3, pp. 288–295, Jun. 2009, doi: 10.1016/j.radonc.2008.08.010.

[91] S. J. Thomas, B. J. Evans, L. Harihar, H. J. Chantler, A. G. R. Martin, and S. V. Harden, “An evaluation of the mid-ventilation method for the planning of stereotactic lung plans,” *Radiother. Oncol.*, vol. 137, pp. 110–116, Aug. 2019, doi: 10.1016/j.radonc.2019.04.031.

[92] M. Velec, J. L. Moseley, L. A. Dawson, and K. K. Brock, “Dose escalated liver stereotactic body radiation therapy at the mean respiratory position,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 89, no. 5, pp. 1121–1128, Aug. 2014, doi: 10.1016/j.ijrobp.2014.04.051.

[93] E. Lens, A. Van Der Horst, E. Versteijne, G. Van Tienhoven, and A. Bel, “Dosimetric advantages of midventilation compared with internal target volume for radiation therapy of pancreatic cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 92, no. 3, pp. 675–682, Jul. 2015, doi: 10.1016/j.ijrobp.2015.02.015.

[94] C. Pinter, A. Lasso, A. Wang, D. Jaffray, and G. Fichtinger, “SlicerRT: Radiation therapy research toolkit for 3D Slicer,” *Med. Phys.*, vol. 39, no. 10, pp. 6332–6338, 2012, doi: 10.1118/1.4754659.

[95] G. C. Sharp *et al.*, “Plastimatch – An Open Source Software Suite for Radiotherapy Image Processing,” May 2010.

[96] S. Vieira, J. Stroom, K. Anderle, B. Salas, N. Pimentel, and C. Greco, “PO-0918: Validation of freeware-based mid-ventilation CT calculation for upper abdominal cancer patients,” *Radiother. Oncol.*, vol. 119, pp. S444–S445, Apr. 2016, doi: 10.1016/S0167-8140(16)32168-5.

[97] “Elastix parameters files database.” Accessed: Jul. 13, 2021. [Online]. Available: https://github.com/SuperElastix/ElastixModelZoo/tree/master/models.

[98] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, “4D XCAT phantom for multimodality imaging research,” *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, 2010, doi: 10.1118/1.3480985.

[99] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[100] J. Vandemeulebroucke, S. Rit, P. Clarysse, and D. Sarrut, “The POPI-model, a point-validated pixel-based breathing thorax model,” *Proceedings of the XVth ICCR*, 2007. https://www.creatis.insa-lyon.fr/rio/popi-model/ (accessed Jul. 13, 2021).

[101] F. Ghareeb, J. Stroom, M. Pereira, D. Boukerroui, M. J. Gooding, and C. Greco, “PD-0890 A novel approach to validate the accuracy of Mid-Position images calculated from 4DCT data,” *Radiother. Oncol.*, vol. 161, pp. S728–S729, Aug. 2021, doi: 10.1016/S0167-8140(21)07169-3.

[102] J. J. Sonke, M. Rossi, J. Wolthaus, M. van Herk, E. Damen, and J. Belderbos, “Frameless stereotactic body radiotherapy for lung cancer using four-dimensional cone beam CT guidance,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 74, no. 2, pp. 567–574, Jun. 2009, doi: 10.1016/J.IJROBP.2008.08.004.

[103] T. N. van de Lindt, G. Schubert, U. A. van der Heide, and J. J. Sonke, “An MRI-based mid-ventilation approach for radiotherapy of the liver,” *Radiother. Oncol.*, vol. 121, no. 2, pp. 276–280, Nov. 2016, doi: 10.1016/J.RADONC.2016.10.020.

[104] H.-G. Menzel, H. Schibilla, and D. Teunen, “European Guidelines on Quality Criteria for Computed Tomography,” 2004.

[105] A. Viera and J. Garrett, “Understanding interobserver agreement: the kappa statistic.,” *Fam. Med.*, 2005.

[106] R. Varadhan, G. Karangelis, K. Krishnan, and S. Hui, “A framework for deformable image registration validation in radiotherapy clinical applications,” *J. Appl. Clin. Med. Phys.*, vol. 14, no. 1, pp. 192–213, 2013, doi: 10.1120/JACMP.V14I1.4066.

[107] K. Murphy *et al.*, “Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge,” *IEEE Trans. Med. Imaging*, vol. 30, no. 11, pp. 1901–1920, 2011, doi: 10.1109/TMI.2011.2158349.

[108] J. Piper *et al.*, “Objective evaluation of the correction by non-rigid registration of abdominal organ motion in low-dose 4D dynamic contrast-enhanced CT,” *Phys. Med. Biol.*, vol. 57, no. 6, pp. 1701–1715, Mar. 2012, doi: 10.1088/0031-9155/57/6/1701.

[109] K. K. Brock, “Results of a multi-institution deformable registration accuracy study (MIDRAS),” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 76, no. 2, pp. 583–596, Feb. 2010, doi: 10.1016/J.IJROBP.2009.06.031.

[110] A. Renner, H. Furtado, Y. Seppenwoolde, W. Birkfellner, and D. Georg, “Deformable image registration with a featurelet algorithm: implementation as a 3D-slicer extension and validation,” *Med. Imaging 2016 Image Process.*, vol. 9784, p. 97844B, Mar. 2016, doi: 10.1117/12.2216863.

[111] B. D. Lucas, *Generalized image matching by the method of differences (Ph.D. dissertation)*. Dept. of Computer Science, 1984.

[112] W. Kilby, J. Sage, and V. Rabett, “Tolerance levels for quality assurance of electron density values generated from CT in radiotherapy treatment planning,” *Phys. Med. Biol.*, vol. 47, no. 9, pp. 1485–1492, May 2002, doi: 10.1088/0031-9155/47/9/304.

[113] Q. Liu, J. Liang, C. W. Stanhope, and D. Yan, “The effect of density variation on photon dose calculation and its impact on intensity modulated radiotherapy and stereotactic body radiotherapy,” *Med. Phys.*, vol. 43, no. 10, pp. 5717–5729, Oct. 2016, doi: 10.1118/1.4963207.

[114] E. C. Ford, G. S. Mageras, E. Yorke, and C. C. Ling, “Respiration-correlated spiral CT: a method of measuring respiratory-induced anatomic motion for radiation treatment planning,” *Med. Phys.*, vol. 30, no. 1, pp. 88–97, Jan. 2003, doi: 10.1118/1.1531177.

[115] H. Wang *et al.*, “Improving soft-tissue contrast in four-dimensional computed tomography images of liver cancer patients using a deformable image registration method,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 72, no. 1, pp. 201–209, Sep. 2008, doi: 10.1016/J.IJROBP.2008.04.054.

[116] A. F. Abdelnour *et al.*, “Phase and amplitude binning for 4D-CT imaging,” *Phys. Med. Biol.*, vol. 52, no. 12, p. 3515, May 2007, doi: 10.1088/0031-9155/52/12/012.

[117] R. George *et al.*, “Audio-visual biofeedback for respiratory-gated radiotherapy: impact of audio instruction and audio-visual biofeedback on respiratory-gated radiotherapy,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 65, no. 3, pp. 924–933, Jul. 2006, doi: 10.1016/J.IJROBP.2006.02.035.

[118] “Lung | Champalimaud Foundation.” Accessed: Jul. 03, 2021. [Online]. Available: https://fchampalimaud.org/clinical_areas/lung.

[119] A. L. H. Arnett *et al.*, “Long-term Clinical Outcomes and Safety Profile of SBRT for Centrally Located NSCLC,” *Adv. Radiat. Oncol.*, vol. 4, no. 2, pp. 422–428, Apr. 2019, doi: 10.1016/J.ADRO.2019.01.002.

[120] Under preparation. Appendix VIII-A.

[121] Under preparation. Appendix VIII-B.

Appendices

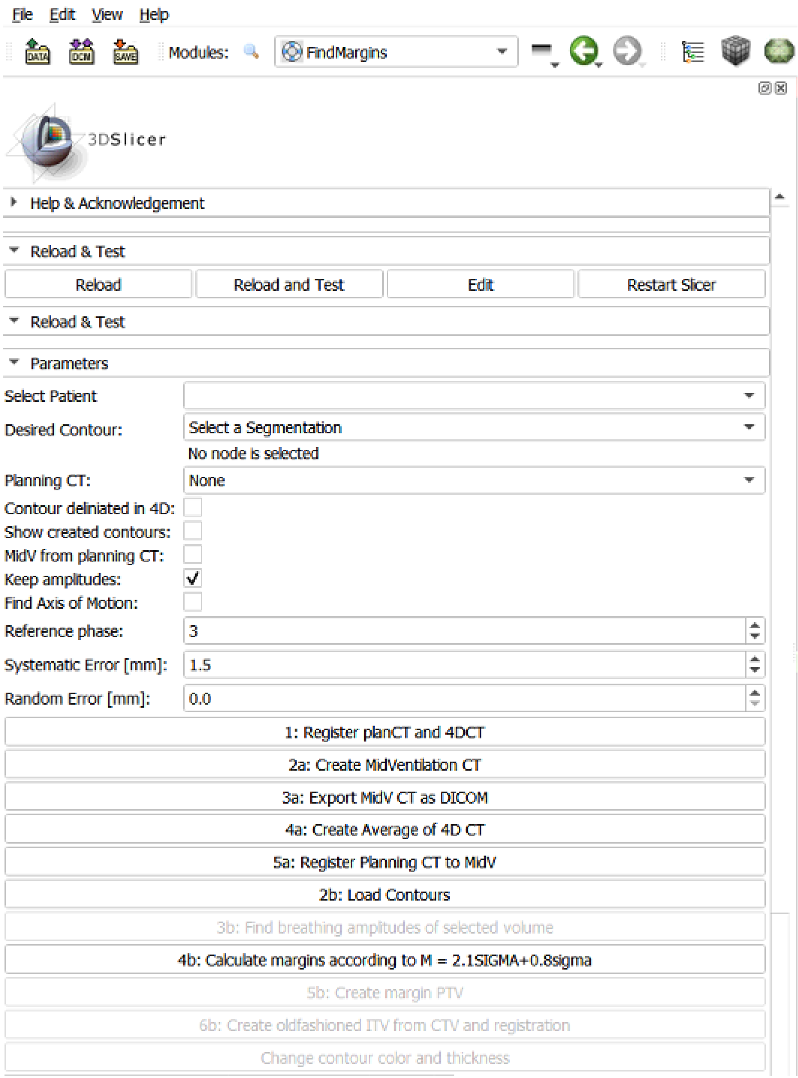
Appendix I

Lung tumour motion and respective planning margins

https://drive.google.com/file/d/1PJ2YXPk5i_5WGfXo7Gb9de20EzZjMKos/view?usp=sharing

Appendix II

The first version of the Mid-Position module software



Select Patient

Desired Contour: Select a Segmentation

No node is selected

Planning CT: None

Contour deliniated in 4D: ☐

Show created contours: ☐

MidV from planning CT: ☐

Keep amplitudes: ☑

Find Axis of Motion: ☐

Reference phase: 3

Systematic Error [mm]: 1.5

Random Error [mm]: 0.0

Appendix III

Plastimatch parameter file

```
[STAGE I]
# transformation
xform = bspline
# implementation
impl = plastimatch
# Run for at most  iterations
max_its = 200
# Smoothness term
regularization_lambda = 0.005
# B-spline grid spacing (in mm)
grid_spac = 50 50 50
# Subsample the image 4x4x2 voxels
res = 4 4 2

[STAGE II]
max_its = 100
regularization_lambda = 0.005
grid_spac = 25 25 25
res = 2 2 1
```

Appendix IV

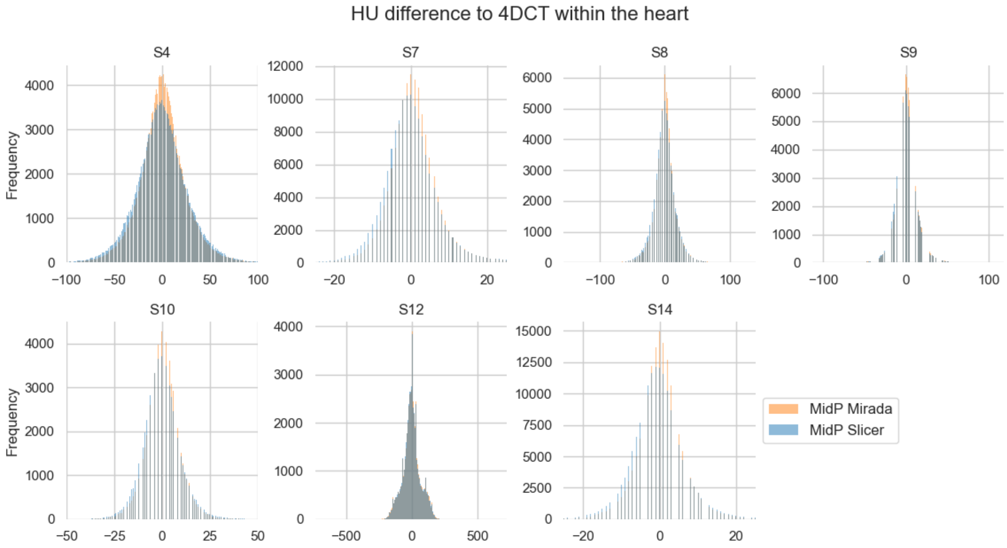
Elastix parameter file

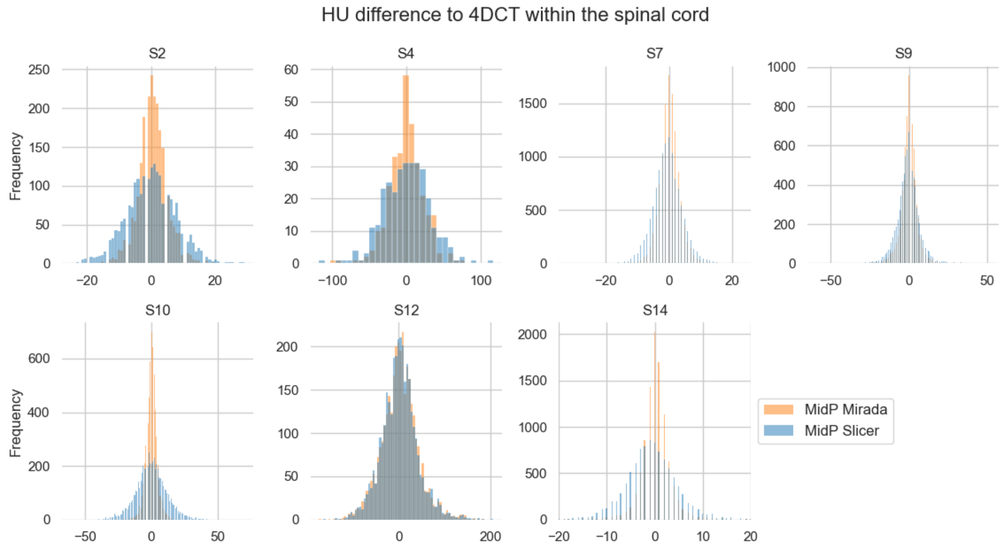
```
// Parameter file for B-spline registration (4DCT data)

// The internal pixel type, used for internal computations
(FixedInternalImagePixelType "float")
(MovingInternalImagePixelType "float")

// **************** Main Components **************************
(Registration "MultiResolutionRegistration")
(Interpolator "BSplineInterpolator")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "DefaultResampler")
(FixedImagePyramid "FixedSmoothingImagePyramid")
(MovingImagePyramid "MovingSmoothingImagePyramid")

(Optimizer "AdaptiveStochasticGradientDescent")
(Transform "BSplineTransform")
(Metric "AdvancedMattesMutualInformation")

// ***************** Transformation **************************
// The control point spacing of the bspline transformation in the finest resolution
level.
(FinalGridSpacingInPhysicalUnits 14)
(GridSpacingSchedule 8 4 2 1)

// ******************* Similarity measure *********************
// Number of grey level bins in each resolution level, for the mutual information.
(NumberOfHistogramBins 32)
(ErodeMask "false")

// ******************** Multiresolution **********************
// The number of resolutions
(NumberOfResolutions 4)

// The downsampling/blurring factors for the image pyramids.
(ImagePyramidSchedule 16 16 8  8 8 4  4 4 2  2 2 1  1 1 1 )

// ******************* Optimizer ****************************
(MaximumNumberOfIterations 500)


// **************** Image sampling **********************
// Number of spatial samples used to compute the mutual
// information (and its derivative) in each iteration.
(NumberOfSpatialSamples 2048)
```

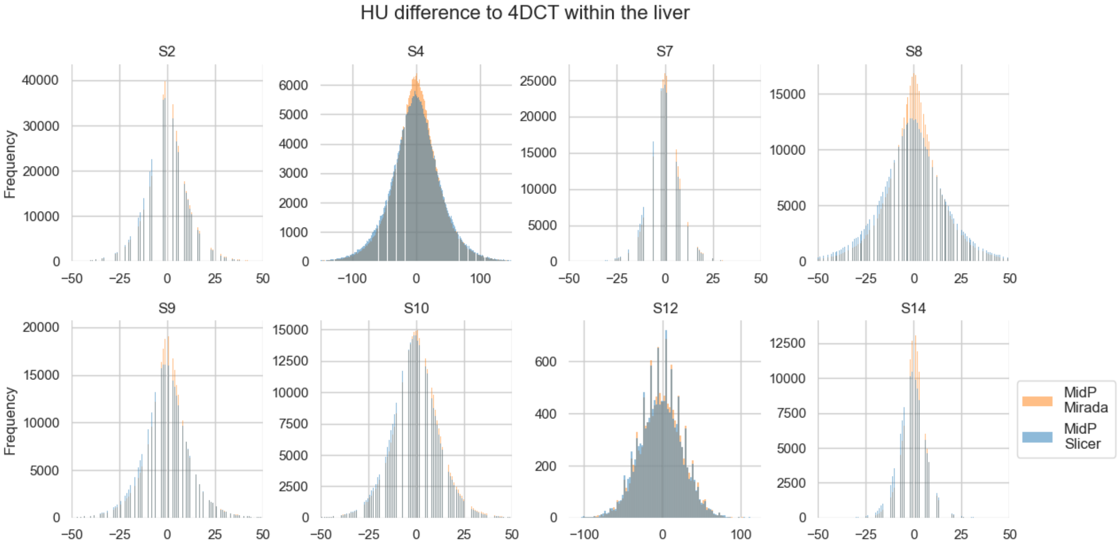
```
(NewSamplesEveryIteration "true")
(ImageSampler "Random")

// ************* Interpolation and Resampling ****************
// Order of B-Spline interpolation used during registration/optimisation.
(BSplineInterpolationOrder 3)

// Order of B-Spline interpolation used for applying the final deformation.
// 3 gives good accuracy; recommended in most cases.
(FinalBSplineInterpolationOrder 3)

//Default pixel value for pixels that come from outside the picture:
(DefaultPixelValue 0)

// Choose whether to generate the deformed moving image.
(WriteResultImage "true")

// The pixel type and format of the resulting deformed moving image
(ResultImagePixelType "short")
(ResultImageFormat "mhd")
```

Appendix V

Histograms showing the distributions of the intensity difference within selected ROIs between two Mid-Position CT scans (reconstructed by 3D Slicer and Mirada software) and the 4DCT scan






HU difference to 4DCT within the liver
S2
S4
S7
S8
S9
S10
S12
S14
Frequency
Frequency
MidP Mirada
MidP Slicer
0
10000
20000
30000
40000
1000
2000
3000
4000
5000
6000
25000
15000
5000
12500
7500
2500
200
400
600
−50
−25
25
50
−100
100

Appendix VI

Planning CT assessment form

This form should be filled in as you examine the planning CT scan for each patient

Patient ID (e.g. ‘MidP patient 1’): ____

Your initials: ____

Please confirm visually sharp reproduction of the following features:

Image quality criteria	YES	NO
Heart and great vessels		
Oesophagus		
Lung Parenchyma		
Tumour		
Carina and lymph node area		
Main and segmental bronchi		
Pleuromediastinal border		
Border between pleura and thoracic wall		
Diaphragm		

Do you have any comments about the feature reproduction in the image?

____

____

____

Please rate the quality of the image according to the following criteria:

Sharpness

☐ Blurry
☐ Sub-optimal
☐ Acceptable
☐ Good
☐ Sharpest

Image Contrast

☐ Very poor
☐ Sub-optimal
☐ Acceptable
☐ Good
☐ Excellent

Artefacts

☐ Unacceptable artefacts
☐ Many/large artefacts
☐ Some artefacts but acceptable
☐ Minimal artefacts
☐ No artefacts

Contouring acceptability

☐ Unacceptable
☐ Suboptimal
☐ Acceptable
☐ Good
☐ Excellent

Do you have any comments about the overall quality of the image?

Appendix VII

A. P-values of qualitative assessment of image sharpness, image contrast, artefacts and contouring acceptability for different planning CTs (Average CT, Mid-P CT and BH CT). P-values statistically significant are highlighted in yellow. Results are organized by institution (CCC and NKI).

Scans	Institution	Sharpness	Image Contrast	Artefacts	Contouring acceptability
Ave vs Slicer	CCC	**0.0255**	**0.0148**	**0.0308**	**0.0215**
	NKI	**0.0017**	**0.0335**	**0.0074**	**0.0008**
Ave vs Wimp	CCC	**0.0109**	**0.0074**	0.3246	**0.0260**
	NKI	**0.0032**	**0.0219**	**0.0075**	**0.0019**
Ave vs Mirada	CCC	0.0542	**0.0325**	**0.0463**	0.1201
	NKI	**0.0008**	**0.0030**	**0.0070**	**0.0010**
Slicer vs BH	CCC	N/A	N/A	N/A	N/A
	NKI	0.2482	0.1242	0.4170	0.7963
Wimp vs BH	CCC	N/A	N/A	N/A	N/A
	NKI	0.1317	0.1664	0.3173	0.4669
Mirada vs BH	CCC	N/A	N/A	N/A	N/A
	NKI	0.1797	0.0578	0.7815	0.5637

B. P-values of qualitative assessment of visually sharp reproduction of some anatomical structures in different planning CTs (Average CT, Mid-P CT and BH CT). P-values statistically significant are highlighted in yellow. Results are organized by institution (CCC and NKI).

Scans	Institution	Heart and great vessels	Oesophagus	Lung Parenchyma	Tumour	Carina and lymph node area
Ave vs Slicer	CCC	0.1797	1	**0.0253**	0.1025	1
	NKI	1	1	0.0833	0.5637	**0.0253**
Ave vs Wimp	CCC	0.2568	1	0.4795	0.4142	0.7389
	NKI	0.3173	0.5637	0.0833	0.5637	**0.0455**
Ave vs Mirada	CCC	**0.0253**	0.6547	**0.0253**	0.3173	0.7055
	NKI	0.3173	0.3173	0.1573	0.1573	0.0588
Slicer vs BH	CCC	N/A	N/A	N/A	N/A	N/A
	NKI	0.3173	1	E	0.3173	0.5637
Wimp vs BH	CCC	N/A	N/A	N/A	N/A	N/A
	NKI	E	0.5637	E	0.3173	0.1573
Mirada vs BH	CCC	N/A	N/A	N/A	N/A	N/A
	NKI	E	0.3173	0.3173	E	0.5637

Scans	Institution	Main and segmental bronchi	Pleuromediastinal border	Border between pleura and thoracic wall	Diaphragm
Ave vs Slicer	CCC	0.3173	0.4142	0.1573	0.1797
	NKI	**0.0253**	0.1025	0.3173	0.1573
Ave vs Wimp	CCC	1	0.4795	0.5637	0.3173
	NKI	0.1025	**0.0253**	0.3173	0.0588
Ave vs Mirada	CCC	0.3173	0.2568	0.1573	0.5637
	NKI	**0.0253**	**0.0082**	1	**0.0455**
Slicer vs BH	CCC	N/A	N/A	N/A	N/A
	NKI	E	0.1797	E	0.3173
Wimp vs BH	CCC	N/A	N/A	N/A	N/A
	NKI	0.3173	0.3173	E	0.5637
Mirada vs BH	CCC	N/A	N/A	N/A	N/A
	NKI	E	1	0.3173	0.1573

Abbreviations

E: Equal distribution

Appendix VIII

List of publications

This section summarizes the scientific publications produced during the course of this work.

Published abstracts

1. **M. Pereira**, J. Stroom, F. Ghareeb, D. Boukerroui, M. J. Gooding, and C. Greco, "PO-1665 Comparison of mid-position CT reconstruction systems," ESTRO 2021, Madrid, *Radiother. Oncol.*, vol. 161, pp. S1387–S1388, Aug. 2021, doi: 10.1016/S0167-8140(21)08116-0.

2. F. Ghareeb, J. Stroom, **M. Pereira**, D. Boukerroui, M. J. Gooding, and C. Greco, "PD-0890 A novel approach to validate the accuracy of Mid-Position images calculated from 4DCT data," ESTRO 2021, Madrid, *Radiother. Oncol.*, vol. 161, pp. S728–S729, Aug. 2021, doi: 10.1016/S0167-8140(21)07169-3.

Publications in progress

A. ***Validation of the Mid-Position strategy for lung cancer: a quantitative and qualitative assessment***

Authors: Edward Jackson[1], **Mariana Pereira**[2], Djamal Boukerroui[3], Beatriz Nunes[2], Nuno Pimentel[2], Margriet Kwint[1], Joep Stroom[2], Mark J. Gooding[3], Peter Remeijer[1]

[1]Netherlands Cancer Institute, Amsterdam, Netherlands
[2]Champalimaud Centre for the Unknown, Department of Radiation Oncology, Lisbon, Portugal
[3]Mirada Medical Ltd., Oxford, UK

Background and Purpose: The intra-fractional motion of lung tumours leads to larger treatment volumes by the more commonly used internal target volume (ITV) strategy in comparison to the Mid-Position (Mid-P) strategy. The purpose of this study was to validate the performance of three independent packages for the implementation of the Mid-P CT strategy.

Materials and Methods: Thirty-six 4DCT lung patient data from two institutions were used to reconstruct the Mid-P CT and estimate tumour motion amplitudes by three separate Mid-P packages. Tumour motion ranged from 1 to 17 mm. Both outputs require the use of deformable image registration (DIR). The performance of each package was assessed both quantitatively and qualitatively. The motion amplitude accuracy was quantified using ground truth estimates based on manually delineated tumour contours. Those contours were also used to evaluate whether the tissue attenuation (Hounsfield Units) is correctly represented by the reconstructed Mid-P scans. We also quantified the sharpness of the diaphragm as a measure of the image quality of the Mid-P scans. Three doctors reviewed and scored the Mid-P CTs in terms of image sharpness, image contrast, image artefacts and their acceptability for contouring using a 5-point grading scale. In addition, doctors evaluated the visibility of relevant anatomical structures on the Mid-P CTs. The Mid-P CTs were also compared to alternative planning CTs, such as average intensity projection (AIP) and breath-hold (BH) scans.

Results: We found no significant differences in the performance of the three packages in terms of their motion amplitude estimates within the tumour and throughout the lung, although the Mid-P package from CCC seems to perform slightly worse. When converted to margins, all were within 1 mm. Intensity distribution differences between Mid-P and BH scans were similar to differences found within frames of the 4DCT dataset. The results of quantifying the sharpness of the diaphragm boundary also showed that the three Mid-P methods performed similarly, with sharpness similar to a single 4DCT frame and just below that of a BH scan. For the qualitative assessment, all three Mid-P packages scored much higher than AIP scans and slightly lower than BH scans. However, there was a difference between datasets acquired at different institutions in both the quantitative and qualitative assessments.

Conclusion: Our results show that the tested Mid-P packages have a good performance on the reconstruction of a Mid-Position planning CT for radiotherapy treatment purposes. We found that the Mid-P method is robust to the choice of the registration algorithm used, and that the reconstructed Mid-P scans were significantly better than AIP scans and not inferior to BH scans.

B. *A novel approach to validate and benchmark Mid-Position software for lung tumour patients using synthetic 4D CT scans*

Authors: Firass Ghareeb[1], Edward Jackson[2], **Mariana Pereira**[1], Joep Stroom[1], Djamal Boukerroui[3], Mark J. Gooding[3], Carlo Greco[1]

[1]Champalimaud Centre for the Unknown, Department of Radiation Oncology, Lisbon, Portugal
[2]Netherlands Cancer Institute, Amsterdam, Netherlands
[3]Mirada Medical Ltd., Oxford, UK

Background and Purpose: The Mid-Position (Mid-P) CT is calculated using a deformable image registration (DIR) of 4DCT data to create a time-weighted mean position of the anatomy during the breathing cycle. In addition, the deformation vector fields (DVFs) can be used to estimate motion amplitudes for relevant ROIs. Several 4DCT validation datasets are publicly available to test DIR algorithms and therefore the accuracy of the constructed Mid-P CT. Although the accuracy of DVFs between phases was verified by experts using anatomical landmarks, these data provide a limited number of cases and are subject to human and registration accuracy. This work proposes a new approach to create a synthetic 4DCT dataset (similar to clinical data) using pre-defined DVFs derived from real patients' data. The suggested approach provides the exact Mid-P CT scan of the synthetic 4DCT dataset. This Mid-P scan in addition to the pre-defined DVFs can be used as ground truth to validate the spatial accuracy and the motion amplitudes within ROIs.

Materials and Methods: Twenty synthetic 4DCT were created based on twenty lung patients' 4DCT scans. For each scan, the reference phase (maximum exhale phase) was registered to the other nine using MATLAB Demons DIR algorithm to acquire 9 DVFs. To eliminate the registration accuracy dependency, the original phases (except the reference) were deleted and a new synthetic 4DCT was created by applying the 9 DVFs to the reference phase. Since these 9 DVFs do not have the same starting point, they cannot be averaged directly. The DVFs were inversed first, then they were averaged to a single inversed mean DVF (DVF_{mean}^{-1}), which was inversed back to create the mean DVF. The ground truth Mid-P CT dataset of the synthetic 4DCT was created by applying the mean DVF (of the 9 DVFs and 1 zero-motion reference DVF) to the reference phase. The Mid-P CT of the synthetic 4DCT was also constructed using three Mid-P applications and the constructed Mid-P CTs were individually compared to the ground truth Mid-P CT using structural similarity (SSIM) index as a perceptual metric. Twenty gross tumour volumes (GTVs) were delineated on the ground truth Mid-P CTs, the ground truth mean peak-to-peak motion amplitudes within the GTVs were calculated using the pre-defined DVFs and the results were compared to the estimated mean peak-to-peak motion amplitudes (within the same GTVs) by the three applications. Finally, an additional twenty masked 4DCT datasets (containing the masks of the GTVs deformed in 10 phases) were created similarly to the ground-truth Mid-P protocol. The acquired DVFs by each application during the generation of the Mid-P CT for the original synthetic datasets were used to generate the masked Mid-P CT and the results were compared to the ground truth masked Mid-P CT by means of Dice similarity coefficient (DSC).

Results and Conclusion: Under evaluation.