Feasibility and uncertainties of 4D dose simulation for post-treatment quality assurance in radiotherapy of moving targets

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik,
Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Thilo Sentker (geb. Sothmann)

Hamburg
2019

Gutachter der Dissertation:	Prof. Dr. Florian Grüner PD Dr. René Werner
Zusammensetzung der Prüfungskommission:	Prof. Dr. Florian Grüner PD Dr. René Werner Prof. Dr. Michael Potthoff Prof. Dr. Erika Garutti PD Dr. Dr. Thorsten Frenzel
Vorsitzender der Prüfungskommission:	Prof. Dr. Michael Potthoff
Datum der Disputation:	13.02.2020
Vorsitzender Fach-Promotionsausschusses Physik:	Prof. Dr. Günter Hans Walter Sigl
Leiter des Fachbereichs Physik:	Prof. Dr. Wolfgang Hansen
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Danksagung

Zuallererst möchte ich mich bei meinen Betreuern PD Dr. René Werner und Dr. Tobias Gauer für die Möglichkeit bedanken, meine Doktorarbeit am Universitätsklinikum Hamburg-Eppendorf (UKE) durchführen zu können. Danke für die hervorragende Betreuung und die sehr gute Zusammenarbeit in den letzten Jahren! Weiterhin gilt mein Dank Prof. Dr. Florian Grüner, welcher sich nach dem unerwarteten Tod von Prof. Dr. Wilfried Wurth ohne zu zögern bereiterklärte, die Position des Erstgutachters zu übernehmen. Den Mitgliedern der Prüfungskommission danke ich für die Teilnahme an meiner Disputation. Auch bedanken möchte ich mich bei der Universität Hamburg, die mein Promotionsvorhaben über zwei Jahre mittels eines Promotionsstipendiums finanziert hat.

Den Mitgliedern der Arbeitsgruppe Image Processing and Medical Informatics im Department für Computational Neuroscience am UKE unter der Leitung von René Werner möchte ich für die angenehme Arbeitsatmosphäre, die gute Zusammenarbeit und viele konstruktive Diskussionen danken. Besonderer Dank gebührt hierbei René Werner, Frederic Madesta und Daniel Schetelig.

Weiterer Dank gilt der am UKE für die Bereitstellung der verwendeten Daten und die Ermöglichung von Messungen bzw. das Aufnehmen von Bilddaten an den Bestrahlungs- und bildgebenden Geräten. Hervorzuheben ist hierbei die Unterstützung während zum Teil langer Messabende durch Tobias Gauer, Jonathan Waschkewitz und Frederic Madesta.

Abschließend möchte ich noch meinen großen Dank für die Unterstützung während meines Studiums und meiner Promotion an meine Eltern, Angelika und Dierk, meiner Frau Kathrin und meinen Sohn Thorben aussprechen. Ohne Euch hätte ich das Alles nicht geschafft!

Abstract

Quality assurance in 4D radiotherapy is an essential process to verify that the dose delivered to a patient is sufficient to achieve pre-treatment defined goals, which are typically the total tumor eradication and optimal sparing of healthy tissue and organs. However, there are currently no appropriate tools available to account for the dynamical nature of free patient breathing in combination with complex dose delivery techniques. The focus of the present thesis lies therefore within a specific clinical context application: the development of a framework for retrospective quality assurance in 4D radiotherapy of lung and liver metastases treated by volumetric modulated arc therapy (VMAT). Achieving this requires following key steps to be taken:

(1) Development of a patient-specific image- and model-based 4D (3D + time) dose accumulation framework for the highly dynamic VMAT dose delivery technique.
(2) Application of respective framework to real patient data in order to investigate the correlation of during-treatment motion, its interplay with VMAT dose delivery dynamics and observed local recurrence of lung/liver metastases after treatment.
(3) Investigation of simulation robustness, accuracy and potential uncertainty sources and implementation of uncertainty propagation.

The basis of step (1) is combining the modeling of the dynamic VMAT dose delivery, which employs the variation of gantry speed, dose rate and collimator leaf positions, and the patient-specific internal structure motion. However, internal structure motion information is routinely not acquired during dose delivery. Thus, a dedicated modeling approach has to be utilized to estimate the internal patient motion, enabling the simulation of motion-affected dose distributions. For step (2), the patient-specific 4D-simulated dose distributions are computed and compared to pre-treatment planned (reference-)dose distributions. The estimated deviations (underdosages) are analyzed and correlated to information about the clinical outcome. A potential linkage is found, which to some extent demonstrate the dose simulation to be reliable. Despite this result, dose simulation uncertainties and impacting parameters as well as the general accuracy are extensively investigated in step (3). Limitations encountered during phantom-based verification measurements motivated to further improve the 4D dose simulation framework by introducing an uncertainty propagation scheme and re-implementing the actual dose calculation utilizing gold standard Monte Carlo dose simulations. It is concluded that the consideration of individual patient motion variability during dose delivery in combination with VMAT dose accumulation for quality assurance in 4D radiotherapy is feasible.

Kurzfassung

Die Qualitätssicherung (QS) in der 4D-Strahlentherapie stellt sicher, dass vor der Patientenbehandlung definierte Ziele, d. h. typischerweise die irreparable Tumorschädigung und die optimale Schonung von gesundem Gewebe und Organen, erreichbar sind. Eine Berücksichtigung der freien Patientenatmungsdynamik während einer Behandlung mit komplexer Bestrahlungstechnik im Sinne einer QS ist jedoch mit zurzeit verfügbaren Methoden nicht möglich. Der Schwerpunkt dieser Arbeit liegt daher auf einer spezifischen klinischen Anwendung: die Entwicklung eines Frameworks für die retrospektive QS in der 4D-Strahlentherapie von Lungen- und Lebermetastasen, behandelt durch volumetrisch modulierte Bogenbestrahlung (VMAT). Folgende Arbeitsschritte sind nötig:

(1) Entwicklung eines patientenspezifischen bild- und modellbasierten 4D (3D + Zeit)-Dosisakkumulationsframeworks für die dynamische VMAT-Bestrahlungstechnik.
(2) Anwendung des entsprechenden Frameworks auf reale Patientendaten sowie die Untersuchung der Korrelation zwischen Bewegung während der Behandlung, ihr Zusammenspiel mit der VMAT-Bestrahlungsdynamik und das lokale Wiederauftreten von Lungen-/Lebermetastasen nach der Behandlung.
(3) Analyse von Simulationsrobustheit, -genauigkeit und potenziellen Unsicherheitsquellen sowie Implementierung einer Unsicherheitsfortpflanzung.

Die Grundlage von Schritt (1) ist die Kombination aus der Modellierung der dynamischen VMAT-Technik (variierende Gantry-Geschwindigkeit, Dosisleistung und Kollimatorlamellenposition) und Informationen über die patientenspezifische interne Bewegung. Allerdings wird diese während der Dosisapplikation routinemäßig nicht erfasst. Ein spezieller Modellierungsansatz ist daher für die interne Schätzung der Bewegung nötig, um eine Simulation von bewegungsbeeinflussten Dosisverteilungen zu ermöglichen. Für Schritt (2) werden die patientenspezifischen 4D-simulierten Dosisverteilungen mit den vor der Behandlung geplanten (Referenz-)Dosisverteilungen verglichen. Die resultierenden Abweichungen werden analysiert und mit Informationen über den klinischen Ausgang korreliert. Die festgestellte Korrelation demonstriert zumindest zum Teil die Zuverlässigkeit der Dosissimulation. Trotz dieses Ergebnisses werden in Schritt (3) die Unsicherheiten der Dosissimulation und mögliche Einflussgrößen sowie die allgemeine Simulationsgenauigkeit untersucht. Bei phantombasierten Verifikationsmessungen identifizierte Limitierungen motivierten, das 4D-Dosissimulationframework durch die Einführung einer Unsicherheitsfortpflanzung und die Implementierung einer eigenständigen Dosisberechnung (Goldstandard Monte Carlo-Simulation) zu verbessern. Die Resultate belegen die prinzipielle Möglichkeit der Berücksichtigung von der individuellen Bewegungsvariabilität in einer VMAT-Dosisakkumulation zur QS in der 4D-Strahlentherapie.

Content

Chapter 1

Introduction

In 2018, cancer was among the globally leading causes of death, as reported by the World Health Organization; latest studies even suggest that cancer actually is the leading cause of death in high-income countries [1, 2]. According to recent publications, incidences and mortality of cancer are rapidly growing, reflecting the higher life expectancy and exponential population growth [3–5]. Over the whole year 2018, there were approximately 18 million new cancer cases with approximately 8.5 million deaths. Of all cancer cases, lung and liver cancer show worldwide high incidence rates with 11.6% and 4.7%, respectively. Further, corresponding mortality rates of 18.4% (lung cancer) and 8.2% (liver cancer) are substantial. This substantiates that especially lung and liver cancer are among the most difficult treatable cancer types [4].

About 80% of all lung tumor cases are nowadays treated interdisciplinary by a combination of surgery, chemo- and radiotherapy or solely radiotherapy [6]. For liver tumor patients, the most common treatment option is surgery. However, in the last years the advances in technology turned radiotherapy into an equally adequate treatment option, with relatively high local control rates and the advantage of providing a noninvasive procedure [7].

The basic idea of radiotherapy is to use ionizing radiation to deposit a specific amount of energy inside a target volume to irreparably damage and, as a consequence, destroy the malignant cancer cells. At the same time, radiation-based side effects in surrounding healthy tissue and organs should be minimized. Therefore, a perfect radiotherapy treatment would essentially solely irradiate the identified target volume, applying the prescribed dose homogeneously, while no energy is deposited around it. In reality, this is physically hardly possible. However, new external beam radiotherapy treatment techniques, e. g. intensity modulated radiation therapy (IMRT) and especially volumetric modulated arc therapy (VMAT), allow for minimized margins around the target volume and optimized sparing of healthy tissue and organs [8, 9].

A standard radiotherapy procedure consists of three major processes, independent of the utilized treatment technique: CT imaging, treatment planning and treatment delivery, as sketched and described in more detail in Fig. 1.1. Note that the described workflow is based on the treatment of lung and liver cancer patients at the University Medical Center Hamburg-Eppendorf (UKE). Prior to the actual radiotherapy processes,

the individual patient diagnosis is conducted by dedicated imaging techniques by the radiology department. After cancer diagnosis by appropriate imaging methods, e. g. magnetic resonance imaging (MRI) and positron emission tomography (PET), computed tomography (CT) images are acquired to correctly represent the patient anatomy. These images are subsequently used for treatment planning, i. e. contouring of healthy organs as well as target volumes and for computing a dose distribution that fulfills given dose constraints. Eventually, the planned dose distribution is applied to the patient using one of the available radiotherapy treatment technique according to clinical guidelines.

Unfortunately, acquired CT images, which are the basis of all following treatment steps, only represent the patient geometry at a specific point in time (3D CT) or at best at consecutive points in time (time resolved 3D CT, i. e. 4D CT) during image acquisition. Additionally, they are usually acquired days or even weeks before the actual treatment. Thus, patient motion, setup errors and physiological processes like respiration and cardiac pulsation introduce uncertainties that directly lead to localization and shape deviations in the patient geometry during treatment compared to the planning CT patient geometry [10]. To account for these types of deviations and uncertainties during treatment, the International Commission on Radiation Units and Measurements (ICRU) established guidelines that define general radiotherapy treatment standards. Within those guidelines, uncertainties in localization and shape of radiotherapy-relevant structures are considered by introducing appropriately sized safety margins around corresponding treatment volumes. Definitions of primary volumes as established in the latest ICRU report [11] are briefly summarized in Table 1.1.

Within the planning CT image, the visible tumor volume can directly be identified and segmented. This macroscopic tumor volume is called gross tumor volume (GTV) and is the basis of further target volume contouring steps. CT image-based uncertainties, for instance blurring of tumor edges caused by patient motion and the possibility of existing microscopic spread around the GTV that is not necessarily apparent in CT images, have to be considered. Therefore, the clinical target volume (CTV) is defined, which aims to include subclinical target tissue besides the GTV. Applying the prescribed dose to this volume is the primary goal of a radiotherapy treatment process. At the same time, however, healthy organs and tissue, so-called organs at risk (OAR), have to be contoured and specific dose constraints[1] have to be considered during treatment planning.

For the treatment of tumors that can be subject to motion caused by e. g. patient respiration as seen in lung and liver cancer patients, the individual tumor motion has

[1] Dose constraints differ from organ to organ and are specified in corresponding literature and radiotherapy guidelines. For example, not more than 10%/33% of a healthy lung/liver should receive a dose above 20 Gy/15 Gy [12, 13].

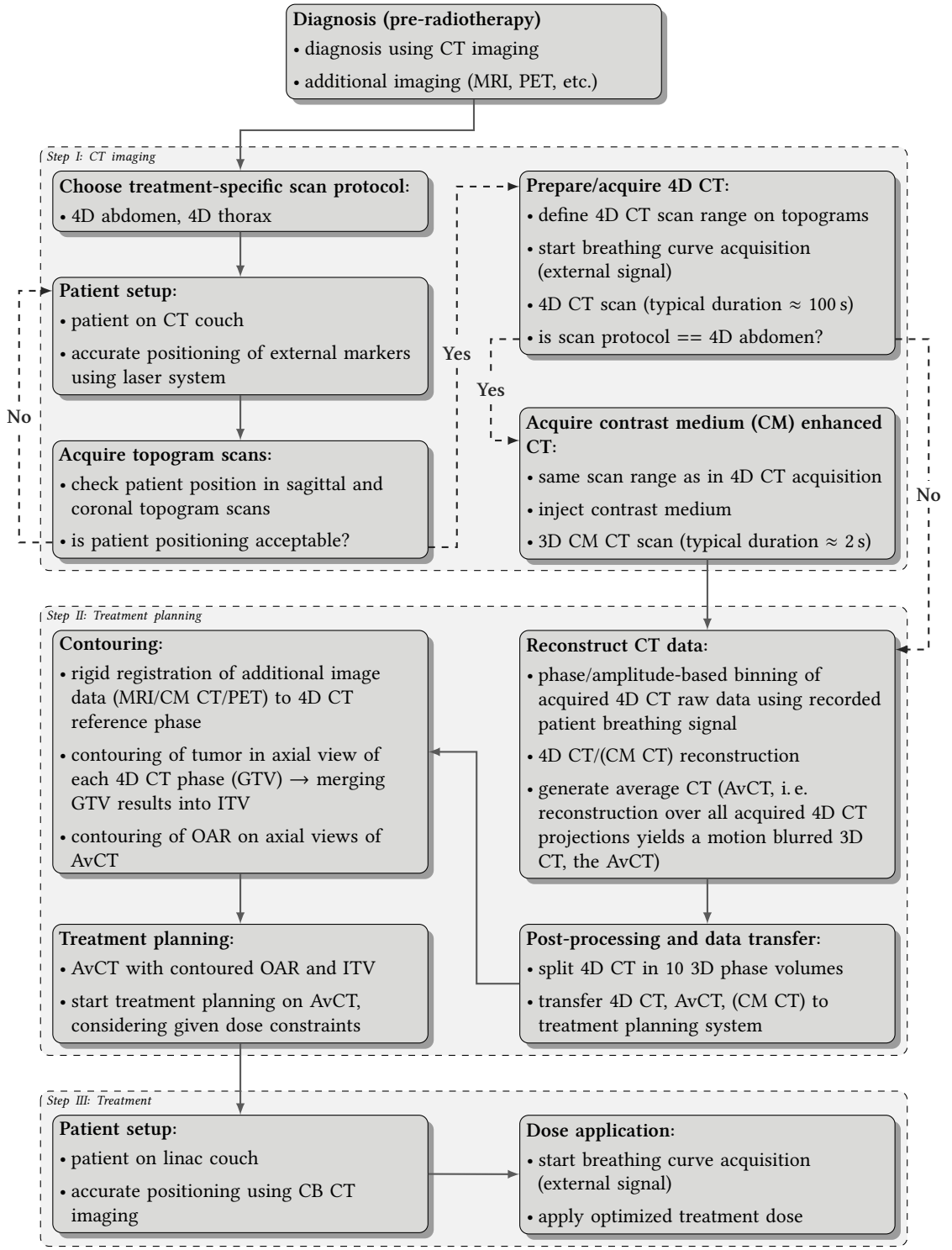


Figure 1.1.: 4D radiotherapy treatment process (with focus on technical aspects) of lung and liver cancer patients as executed at the UKE. After pre-radiotherapy conducted diagnosis, 4D CT image acquisition (step I), treatment planning (step II) and treatment delivery (step III) is performed. Uncertainties in patient positioning and breathing variability (step I and III) as well as delineation of the target volume and OAR (step II) directly impacts the treatment success.

to be considered while treatment. This is emphasized by tumor motion amplitudes of up to several centimeters [14]. A standard approach is to use acquired time resolved 4D CT data (cf. Fig. 1.1) to segment the tumor volume, i. e. the GTV, in each 4D CT breathing phase and compute an union over all phase specific GTV to generate a motion encompassing safety margin. This so-called internal margin is added to the CTV, and the total volume referred to as internal target volume (ITV). Eventually, an usually uniform margin of a few millimeters is applied to the ITV to account for patient setup errors in each treatment fraction and defines the planning target volume (PTV). This volume assures that the prescribed dose is correctly and sufficiently delivered to the target volume. A substantial reduction of this margin and setup uncertainties, respectively, is possible by acquiring a static cone beam CT (CBCT) image before treatment and matching CBCT and planning CT reference image [15]. More sophisticated radiotherapy techniques even predict and track the tumor motion during treatment [16] and can therefore further reduce especially the internal margin. Overall high costs and long treatment duration, however, are disadvantages and reasons why tumor tracking is rarely applied. Therefore, in this thesis the approach of tumor tracking is not considered. Instead, the focus is on the widely used standard radiotherapy approach, i. e. no real-time motion compensation during radiotherapy treatment.

Nowadays, complex and fast treatment techniques are frequently used. The popular VMAT treatment employs rotation of the beam source around the patient and simultaneously modulates the field form as well as particle fluence (i. e. the number of particles incident on a sphere of cross sectional area) to allow for fast and high precision radiotherapy treatment [17]. Treatment parameters for this kind of approach are usually manually pre-defined and afterwards, under consideration of dose constraints, inversely optimized by the treatment planning system (TPS). For dose application, the principle of so-called stereotactic ablative radiotherapy (SABR) or stereotactic body radiation therapy (SBRT) is commonly utilized. Both approaches are following the concept of applying high doses with high precision and accuracy in a small number of fractions (1 to 5 fractions, i. e. hypofractionation) [18, 19] to exploit radiobiology effects (α/β ratio[2]) inside the target volume [21]. However, when treating tumors that are subject to breathing induced motion, e. g. lung and liver tumors, the accuracy of such radiotherapy treatment techniques is potentially reduced due to uncertainties introduced by intra- (during a treatment fraction) and interfractional (between treatment fractions) patient breathing variability. The patient breathing irregularity during the actual radiotherapy treatment fractions can

[2] The α/β ratio is a tissue/organ-dependent measure for the curvature of a cell survival curve defined by the linear-quadratic formula [20]. For higher α/β ratios the tissue/organ is less susceptible to the effect of the fractionation scheme, i. e. the hazard of long-term damages due to fractionation of the total dose is reduced.

Table 1.1.: Primary ICRU treatment volumes as defined in ICRU report 91 [11].

RT treatment volumes	**ICRU-definition**
Gross Tumor Volume (GTV)	the gross demonstrable extent and location of the "tumor"
Clinical Target Volume (CTV)	a volume of tissue that contains a demonstrable GTV and/or subclinical target tissue at a probability considered relevant for therapy
Internal Target Volume (ITV)	CTV plus an "internal margin", taking into account motion-related uncertainties in size, shape, and position of the CTV within the patient
Planning Target Volume (PTV)	surrounds the CTV typically with a margin, which takes into account both the internal and the setup (external) uncertainties
Organs At Risk (OAR)	organs that, when irradiated, could result in significant morbidity, and thus influence treatment planning

therefore be, to some extent, the cause for uncertainties in dose application and thus negatively impact the treatment success. To emphasize this, an example of breathing variability of four real patient cases is given in Fig. 1.2. Here, the recorded breathing curve during 4D CT imaging can be interpreted as reference breathing information, as it is the basis of planning 4D CT reconstruction. The comparison between reference respiration signal and respiration during treatment illustrates the level of intra- and interfractional patient breathing variability. Despite the high complexity of the patient respiration, the signal is usually recorded as simple one-dimensional breathing information, representing solely the anterior-posterior (AP) motion amplitude alterations of the chest wall.

To reduce the impact of patient respiration, treatment methods that use common motion management strategies can be employed. These strategies are either based on completely stopping the tumor motion by guiding the patient breathing or irradiating the motion encompassing target volume. A straightforward approach to stop the tumor motion is to instruct the patients to hold their breath in a specific respiratory phase, e. g. maximal inhalation. The treatment plan is generated on this specific 4D CT phase image and therefore the radiation should perfectly hit the target if the tumor position can be reproduced during breath hold at treatment. This so-called deep inspiration breath hold (DIBH) approach has the advantage of reduced dose inside OAR but requires optimal

patient collaboration, compliance, sufficient pulmonary reserve and longer treatment times [22]. If DIBH is not applicable, e. g. the patient has no sufficient pulmonary reserve, the spontaneous breathing gating is an alternative. Here, the patient breathing is recorded during treatment and at a specific time window, where the tumor position is known, the radiation is applied. Similar to DIBH, the inspiration phase is favorable. Advantages are again reduced dose in the OAR[3], but patient collaboration, regular breathing patterns and longer treatment times are needed [22]. When none of the mentioned techniques is applicable, a common treatment approach is to let the patient breath freely and irradiate the motion encompassing target volume (ITV + setup error margin, i. e. PTV), with the advantage of fast treatment times and less requirements on patient breathing. However, exposure of healthy tissue and organs is higher and unfavorable interactions between patient respiration variability and treatment technique are possible. As described above, beam parameters during VMAT treatment are dynamically changed to allow for best possible application of dose distributions. This means, the total homogeneous dose distribution inside the target volume is generated by adding up all heterogeneous dose distributions per beam position. The general concept of planning the corresponding dose delivery process (treatment planning) is usually performed on a motion-blurred 3D CT volume reconstructed from all acquired 4D CT projections, i. e. the so-called average CT (AvCT), in combination with the generated ITV. That is, the motion dynamics are not explicitly considered during the planning process. However, as the CTV is assumed to move inside the ITV, a homogeneous dose distribution would sufficiently hit the tumor with the prescribed dose. Unfortunately, this assumption is not valid at all times and underdosages in the target volume are possible that can impact the treatment success. Mainly, two different motion-related effects are responsible for dose deviations if the treatment is performed under free patient breathing. Firstly, patient respiration amplitudes can be larger during treatment than during CT imaging. This results in possibly larger tumor motion amplitudes. Thus the CT-based ITV is to small to ensure sufficient irradiation of the target volume. Secondly, an unfavorable interplay between tumor motion and dynamically changing beam parameters, the so-called interplay effect, can occur. Here, the tumor moves inside the ITV, but due to continuously modulation of field openings, the tumor can be partly, or in total, in low dose areas. In this case, solely healthy tissue and not the target receives the described dose. Thus, the target is potentially receiving less dose than planned [19, 23]. Over the course of many fractions, this effect is likely to average out [24], but the current trend to use extremely hypofractionated treatment schemes (up to a single fraction) potentially increases the impact of this effect

[3]Typically the lung and cardiac dose is reduced due to lung expansion and smaller PTV [22].

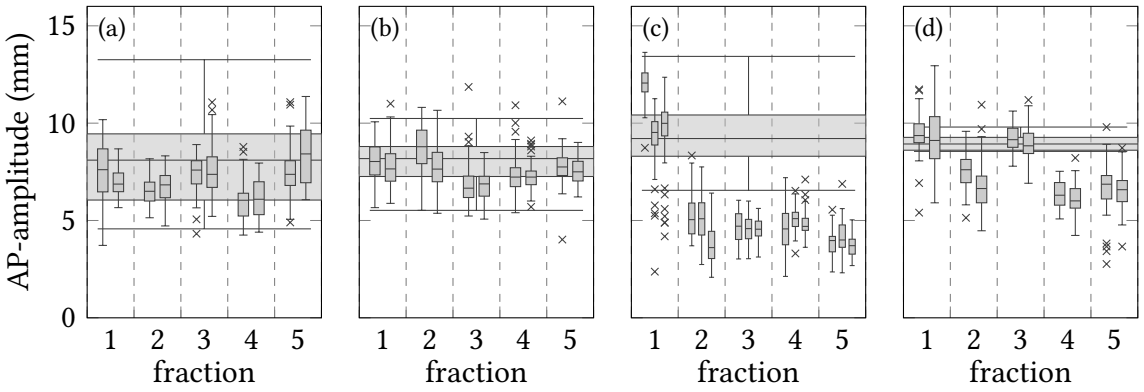


Figure 1.2.: Illustration of chest wall AP-amplitude variability during radiotherapy treatment using box plots for four selected in-house patient data sets. The reference respiration signal recorded during CT imaging (large box) is compared to the breathing signal acquisitions during the individual treatment fraction (small boxes). A fraction consists of two (a, b and d) or three (c) VMAT arcs. Figure (a) and (b) show good accordance between reference and treatment signal. In (c), the amplitude signal during treatment is constant after fraction 1 but of about factor two smaller than the reference signal, and (d) shows high interfractional respiratory variability and large differences compared to the reference signal.

on treatment success [19, 25]. To account for and address the interplay effect, appropriate quality assurance approaches for 4D radiotherapy have to be implemented.

For stationary target volumes, i. e. for 3D radiotherapy, quality assurance is in general straightforward, as solely setup errors and changes in target shape and size have to be considered. However, understanding motion-induced differences between planned and delivered dose over the course of a 4D radiotherapy treatment requires a combination of patient-specific motion data and information about the dose delivery process.

1.1. Quality assurance in 4D radiotherapy

In general, quality assurance in radiotherapy is a tool to quantify and monitor uncertainties and errors introduced by, for instance, treatment planning, treatment device performances and dose application [26]. This directly aims at increasing the probability of identifying deviations, i. e. underdosages, and possible accidents before they actually occur. Further, identification of errors in dose application after each treatment fraction by adequate quality assurance-based dose accumulation could be helpful to allow for an appropriate response, e. g. adapting the treatment plan of the next fraction or treatment of possible radiation-based side effects [27]. Applicable tools are commonly referred to as 4D dose accumulation, simulation, calculation or reconstruction.

The basic principle of 4D dose accumulation schemes is usually a weighted summation over simulated dose distributions based on different (breathing) states of the time-dependent patient geometry [28]. Here, the extraction of motion information from

the planning 4D CT by application of deformable image registration (DIR) between one reference phase and all other phases is often used to acquire necessary information about dynamics of the internal patient geometry. The deformable registration between 4D CT phases yields motion fields that can be applied to subsequently deform corresponding and originally planned 3D dose distributions. This mapping of dose distributions into the patient geometry of the reference image allows combining the deformed doses by weighted summation [29–31]. However, these methods neglect information about dynamics of dose delivery and actual patient breathing during treatment as only pre-treatment acquired information (e. g. 4D CT and planned 3D dose) are used to predict the applied dose. More advanced approaches suggest therefore to not only assign the 3D dose distribution but individual dose segments or even monitor units (MU) to the phases of the planning 4D CT [31, 32]. To further address inter-fractional motion differences, pre-treatment fraction acquired 4D CBCT images can be utilized to update the motion information extracted from the corresponding 4D CT data sets [33, 34]. Unfortunately, all these approaches are solely based on internal patient motion information about single respiratory cycles as represented by the available 4D images. Information about the patients' actual breathing pattern during dose delivery, i. e. intrafractional respiratory variability, and their interplay with the dynamical dose delivery process have so far not been taken into account.

1.2. Aims and contributions of this work

Current standards in radiotherapy outline the problematic of 4D radiotherapy treatment of moving targets treated by complex techniques that make use of high radiation doses and hypofractionation. These techniques are nevertheless popular, mainly because treatment times are short, requirements to the patients are low and, most importantly, local control rates are comparable, or even higher, in contrast to 3D conformal radiotherapy approaches [35–39]. Standard quality assurance in 4D radiotherapy by applying tools developed for 3D radiotherapy is, however, hardly possible. Therefore, the general aim of the present thesis is to develop and implement a 4D dose simulation framework that allows for retrospective 4D quality assurance of real patient treatments under consideration of influencing parameters during the 4D radiotherapy process with focus on technical aspects delineated in Fig. 1.1. Such a framework directly depends on uncertainties in utilized patient data and evaluation/computation tools applied. Hence, the framework should include some sort of uncertainty propagation mechanism. This means, a systematic analysis of error sources and uncertainties during 4D imaging and the influence on subsequent processes has to be performed. To allow for a quantification and verification of such uncertainties, appropriate measurement setups are developed and used to generate

phantom-based image data. In doing so, error chains resulting from, e. g., image artifacts need to be understood and verified by suitable simulation approaches and conception of metrological methods. Achieved insights ideally enable an estimation and assessment of potential error sources and their dosimetrical impact on resulting dose distributions in 4D radiotherapy. Note that the focus of this thesis is on investigating the uncertainties of technical aspects of the dose delivery process. Consideration of uncertainties introduced by human interaction, e. g. contouring of radiotherapy relevant structures by radiologists or patient positioning before treatment, is beyond the scope of this thesis.

The contemplated methodological procedure can be characterized as a highly interdisciplinary work, where the disciplines of physics and computer science within a medical application context are represented. In the following, the individual contributions to this work are listed.

Physics contribution: The basis of this thesis are the physical principles of photon-matter interactions and corresponding simulation approaches like e. g. Monte Carlo simulations for estimation of delivered doses. In particular, X-ray imaging by computed tomography and dose delivery methods utilized in radiotherapy are fundamental components of this thesis. Identification and evaluation of uncertainties within those processes is mandatory to allow for an analysis of the proposed dose accumulation scheme with regard to its accuracy and parameters affecting it. The obtained knowledge is applied to design appropriate experimental measurement setups for e. g. dose measurement and image data acquisition. Reasonable interpretations and analysis of achieved results are performed, specifically under consideration of an uncertainty propagation through the dose accumulation pipeline.

Computer science contribution: The implementation of the 4D dose simulation framework is one of the main contributions to this discipline. More specifically, this includes the application and modification of existing tools like image registration frameworks and correspondence modeling tools. Further, the pre- and post-processing steps performed on medical image data are primary procedures in computer science. In addition, the utilization of deep learning to allow for an implementation of a CNN-based DIR algorithm and the application of Monte Carlo-based dose computation algorithms can also be associated to this contribution.

Medicine contribution: The development and implementation of the proposed dose accumulation scheme within a medical application context, i. e. as a quality assurance tool in 4D radiotherapy, is one of the main medical contributions. This includes the identification of potential patient-specific motion-related causes of local metastasis

recurrence. Further, the developed dose accumulation scheme is retrospectively applied on real patient data sets. A novel concept of using corresponding clinical endpoints to allow for a proof-of-principle investigation is employed.

This thesis represents, to the best of my knowledge, the first attempt to identify motion management and treatment failure by a respective patient-specific 4D dose accumulation including information about the clinical outcome. In turn, a critical assessment of the robustness of the dose accumulation framework is necessary. This includes analyzing the influence of uncertainties of its main building blocks, the correspondence model and the underlying non-linear registration approach.

1.3. Thesis structure

The structure of the thesis is as follows. In Chapter 2, physical fundamentals regarding photon beam radiotherapy as well as particle transport by applying Monte Carlo simulations are presented. In addition, the principle of photon beam generation, modulation and application using a medical linear accelerator is introduced. Thereafter, the principle of the primarily utilized image modality, i. e. the CT, and general image processing in radiotherapy in combination with deep learning is described in Chapter 3. Based upon the physical and imaging related theoretical background, Chapter 4 introduces methods developed and applied in this thesis that are essential for a 4D quality assurance framework, focusing on correspondence-model based 4D dose accumulation as well as deep learning-based image registration. Further, the utilized patient data sets are introduced. Afterwards, results of the performed experiments and simulations using the developed methodological approaches are presented. Basis of this chapter are the results published in peer-reviewed journals/proceedings, as detailed in subsequent section. Eventually, achieved results are discussed and put in context of current literature, an outlook to further possible research is given and the thesis is concluded.

Publications:

Results of this thesis are mostly published in peer-reviewed journals/proceedings and written in the form of a cumulative thesis. That is, each section in Chapter 5 (except Section 5.7) is based on one publication. Its structure follows the guidelines of the University of Hamburg MIN-doctoral degree regulations for a cumulative/interdisciplinary dissertation. Bylaws for safeguarding good scientific practice and avoiding scientific misconduct provided by the Deutsche Forschungsgemeinschaft and the University of Hamburg have been followed. The composition of manuscripts was led by corresponding first author(s)

of each article. Prior to publication, all authors approved on the manuscript to be submitted. Additional contributions of each author to the research study are indicated by the following notation:

[1] conception and supervision of research study
[2] experiments and data acquisition
[3] theoretical calculations and simulations
[4] data analysis
[5] data interpretation
[6] writing the manuscript

Section 5.1: T. Sothmann[1,2,3,4,5,6], T. Gauer[2,5], M. Wilms[3,5], R. Werner[1,3,5,6]. Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT. *Phys Med Biol*, 62(23):9001–9017, 2017.

Section 5.2: N. Mogadas[2,3,4,5,6], T. Sothmann[1,2,3,4,5,6] (shared first authorship), T. Knopp[1], T. Gauer[1,5], C. Petersen[1], R. Werner[1,5,6]. Influence of deformable image registration on 4D dose simulation for extracranial SBRT: A multi-registration framework study. *Radiother Oncol*, 127(2):225–232, 2018.

Section 5.3: T. Sothmann[1,2,3,4,5,6], T. Gauer[2,5], R. Werner[1,3,5]. Influence of 4D CT motion artifacts on correspondence model-based 4D dose accumulation. In: *Proc SPIE*, Vol. 10576, 105760F. International Society for Optics and Photonics, SPIE, 2018.

Section 5.4: T. Sentker[1,2,3,4,5], F. Madesta[1,2,3,4,5] and R. Werner[1,5,6]. GDL-FIRE^{4D}: Deep learning-based fast 4D CT image registration. In: *Lect Notes Comput Sc*, 765–773. Springer, 2018.

Section 5.5: T. Sothmann[1,2,3,4,5], T. Gauer[1,2,5] and R. Werner[1,3,5,6]. 4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters. *PLoS One*, 12(2):e0172810, 2017.

Section 5.6: T. Sentker[1,2,3,4,5,6], F. Madesta[1,3,4,5] and R. Werner[1,5]. Patient-specific 4D Monte Carlo dose accumulation using correspondence model-based motion prediction. In: *Proc SPIE*, Vol. 10951, 1095109. International Society for Optics and Photonics, SPIE, 2019.

Reproduction of publications was approved by IOP Publishing Limited (journal: *Physics in Medicine and Biology*), Elsevier (journal: *Radiotherapy & Oncology*), SPIE (proceeding: *Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*) and Springer (proceeding: *Medical Image Computing and Computer-Assisted Intervention*). The included *PLoS One* article is published openly accessible.

Chapter 2

Physical principles of radiotherapy

The fundamental physical principles of radiotherapy are presented in this chapter. First, to understand how and why energy is deposited in a medium, interactions of photons with matter are briefly explained. Subsequently, the quantification of energy deposition caused by scatter and collision effects, i. e. dosimetry, is explained for photon beams. The estimation of such effects without performing dose measurements using Monte Carlo simulations to predict the particle transport is described in the following section. Finally, beam generation and modulation of a medical linear accelerator is explained.

2.1. Interactions of photons with matter

Photon beams with small wavelengths and therefore higher energies have a very high and material-dependent potential to penetrate matter. However, while traveling through matter, the number of photons decreases exponentially along the incident direction due to absorption and scatter effects. The theoretical background of the following sections is based on the textbooks of T. Buzug [40], W. Demtröder [41, 42] and H. Reich [43].

Assuming a monochromatic photon beam penetrates a homogeneous medium, i. e. the linear attenuation coefficient $\mu(\kappa) = \mu$ is constant along κ, the change in photon beam intensity I can be described as a homogeneous and ordinary linear, first order differential equation with constant coefficients

$$\frac{\mathrm{d}I}{I(\kappa)} = -\mu \, \mathrm{d}\kappa. \tag{2.1}$$

After integration of both sides

$$\int \frac{\mathrm{d}I}{I(\kappa)} = -\mu \int \mathrm{d}\kappa \tag{2.2}$$

and using $\int \frac{\mathrm{d}I}{I(\kappa)} = \ln I(\kappa)$

$$\ln I(\kappa) = -\mu\kappa + C \tag{2.3}$$

we get

$$I(\kappa) = \exp(-\mu\kappa + C) = \exp(-\mu\kappa) \cdot \exp(C). \tag{2.4}$$

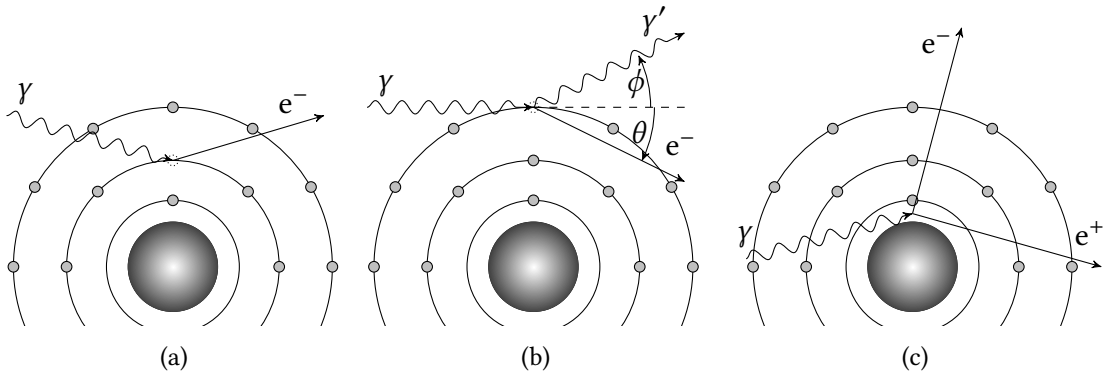


Figure 2.1.: Schematic illustration of photon-matter interactions, which primarily occur in radiotherapy processes. (a) Photoelectric effect. (b) Inelastic Compton scattering. (c) Pair production. Figure inspired by [40].

For the initial condition $I(0) = I_0$, the special solution for Eq. (2.4) is obtained as

$$I(\kappa) = I_0 \cdot \exp(-\mu\kappa), \tag{2.5}$$

which is known as Lambert-Beer's law. Major physical mechanisms that lead to attenuation of photon beams with energy ranges used in radiotherapy are the photoelectric effect, Compton scattering, pair production and Rayleigh scattering. The total linear attenuation coefficient μ for photon beams in matter can be divided into three fundamental attenuation coefficients defined by the individual cross sections σ of the interaction processes, i. e. into the cross section for the photoelectric effect σ_{pe}, the total scattering effects $\sigma_{\text{s,tot}}$ and the pair production σ_{pp},

$$\mu = \left(\sigma_{\text{pe}} + \sigma_{\text{s,tot}} + \sigma_{\text{pp}}\right) \cdot {}^{N_\text{A}}/_{M} \tag{2.6}$$

multiplied by the ratio of Avogadro constant N_A and molar mass M. The cross section $\sigma_{\text{s,tot}}$ can further be expressed by a coherent scattering term (i. e. Rayleigh scattering) σ_{rs} and incoherent scattering term (i. e. Compton scattering) σ_{cs}, leading to

$$\mu = \left(\sigma_{\text{pe}} + \sigma_{\text{rs}} + \sigma_{\text{cs}} + \sigma_{\text{pp}}\right) \cdot {}^{N_\text{A}}/_{M}. \tag{2.7}$$

Physical processes contributing to Eq. (2.7) are described subsequently and the influence of individual attenuation coefficients regarding the total attenuation is illustrated. Eventually, the local energy deposition to matter is described.

2.1.1. Photoelectric effect

The photoelectric effect describes the interaction of a photon with an atom where as a result an electron near the core is ejected and the atom ionized, as visualized in Fig. 2.1 (a).

This effect is the dominating interaction process of materials/compounds with low atomic numbers Z and low-energy photons (photon energies < 30 keV) [43]. The kinetic energy of the emitted electron T_e is the energy difference between the photon $h\nu$ and the binding energy E_b of the electron

$$T_e = h\nu - E_b, \tag{2.8}$$

where h is the Planck constant and ν the frequency of the incident photon. As the atom is afterwards ionized, characteristic X-rays or shell electrons, so-called Auger electrons, are emitted during transition into its basic state. Highest photoelectric effect cross sections σ_{pe} are achieved for K-electrons in matter with high density, resulting in a proportionality to the fourth or fifth power of Z and an inverse proportionality to the third power of the incident photon energy E_γ

$$\sigma_{pe} \propto \rho \frac{Z^n}{A} \cdot E_\gamma^{-3}. \tag{2.9}$$

For tissue-equivalent matter and in the dominant section of the photoelectric effect, index n is in the range of 4 to 4.6. For high photon energies, i. e. $E_\gamma \gg E_b(K)$, the photoelectric effect is proportional to Z^5/E_γ.

2.1.2. Compton scattering

The inelastic scattering between a photon and a weakly bound, i. e. quasi-free, electron, e. g. a valence electron in the outer shell of an atom, is called Compton scattering and illustrated in Fig. 2.1 (b). The incident photon is scattered from its initial trajectory by an angle ϕ due to the collision with the electron. In this process, in contrast to the photoelectric effect, the photon loses parts of its kinetic energy (inelastic scattering), which is passed to the electron. The electron leaves the atom under an angle θ. Using the law of conservation of energy

$$T_e = h\nu - h\nu' = E - E' \tag{2.10}$$

and conservation of momentum p along the incident direction

$$h\nu = \frac{E}{c} = \frac{E' \cos\phi}{c} + p\cos\theta \tag{2.11}$$

or perpendicular to it

$$\frac{E' \cos\phi}{c} = p \sin\phi \tag{2.12}$$

the photon energy

$$h\nu' = h\nu \cdot \frac{1}{1 + \varepsilon \cdot (1 - \cos\phi)} \tag{2.13}$$

and wavelength variation

$$\Delta\lambda = \frac{h}{m_e c} \cdot (1 - \cos\phi) \tag{2.14}$$

can be determined by means of the relativistic kinematics, where $\varepsilon = {}^{h\nu}/_{m_e c^2}$, m_e is the mass of the electron and c is the speed of light. Coherence between scattering angle of the photon and the electron is defined as

$$\cot\phi = (1 + \varepsilon) \cdot \tan({}^{\theta}/_{2}). \tag{2.15}$$

For small ε, the scattering angle θ can be in a range of $-\pi$ to $+\pi$. However, the probability of forward and backward scattering is equal and at the same time two times higher than sideways scattering. At higher photon energies, photons and electrons scatter primarily in forward direction. Compton scattering is the dominant interaction between photons and electrons for an incident photon energy range of 0.2 MeV to 10 MeV and matter with atomic numbers smaller than ten. The cross section σ_{cs} of the Compton effect is defined by the Klein-Nishina formula [44] and reads for very high energies ($E_\gamma \gg m_e c^2$)

$$\begin{aligned} \sigma_{\mathrm{cs}} &= \pi \cdot r_e^2 \cdot Z \cdot \frac{m_e}{E_\gamma}\left[\ln\left(\frac{2E_\gamma}{m_e c^2}\right) + \frac{1}{2}\right] \\ &\propto \frac{Z}{E_\gamma}. \end{aligned} \tag{2.16}$$

2.1.3. Pair production

For photon energies $E_\gamma > 2m_e c^2$ the pair production is besides the photoelectric effect an additional possible absorption process. Here, the photon annihilates to an electron-positron pair in the coulomb field of the core of an atom, as shown in Fig. 2.1 (c). Both the electron and the positron have a rest mass of m_e and a rest energy of $m_e c^2 = 0.511\,\mathrm{MeV}$. The remaining energy is divided between positron (E_+) and electron (E_-)

$$h\nu - 2m_e c^2 = E_+ + E_-. \tag{2.17}$$

The positron is subject to the same particle-matter interactions as the electron, however, after collision with an electron, both particles annihilate and the energy of $E = 2m_e c^2$ (if both particles come together at rest) is released in form of (most of the time) two photons with same energy $E = m_e c^2$ and opposing direction. The cross section σ_{pp} for the pair production

$$\sigma_{\mathrm{pp}} \propto Z^2 \cdot \ln E_\gamma \tag{2.18}$$

first increases logarithmically with the photon energy E_γ, before it becomes constant for photon energies of $E_\gamma \gg m_e c^2$.

2.1.4. Rayleigh scattering

Coherent scattering or Rayleigh scattering mainly occurs for incident photon energies $h\nu < E_b$. The incident photons excite the electrons of atoms in matter to forced oscillations (so-called dipole antenna or Hertzian dipole), which then emit photons with the same frequency ν as the incident photons. No energy is lost by the photon as it only transfers momentum to the atom and is afterwards scattered. The cross section σ_{rs} for Rayleigh scattering is proportional to ν^4. As long as the wavelength of the photon is large compared to the diameter of the atom, the elastically scattered parts of the incident photon beam can be added up coherently. Magnitude of the scattered wave is then proportional to Z and the intensity proportional to Z^2.

2.1.5. Total mass attenuation

More fundamental than the total linear attenuation coefficient defined in Eq. (2.7) is the mass attenuation coefficient as it is independent of the actual mass density as well as physical state of the absorber. The total mass attenuation reads

$$\frac{\mu}{\rho} = \frac{1}{\rho}(\sigma_{pe} + \sigma_{rs} + \sigma_{cs} + \sigma_{pp}) \cdot {}^{N_A}/_{M}, \tag{2.19}$$

with ρ being the density of the absorber material.

Figure 2.2 shows the functional behaviour of the mass attenuation coefficients regarding the attenuation in water for beforehand described photon matter interactions as a function of photon energy; associated data is being obtained from the XCOM database [45]. In a wide photon energy range from 30 keV to 30 MeV, Compton scattering is the dominant interaction between photons and matter and has the largest contribution to the total mass attenuation. For smaller (< 30 keV) and for higher photon energies (> 30 MeV), other interaction processes like the photoelectric effect and the pair production, respectively, are the prevailing effects. Rayleigh scattering plays only a secondary role in comparison to the other interaction processes. The total mass attenuation depends, as described previously, directly on the considered material. Thus, for materials with higher Z, Compton scattering, the in water over a wide range dominant effect, is limited to a much smaller energy interval. This is mainly caused by the strong dependence of the pair production on the atomic number and its onset at energies above $2m_e c^2 = 1.02$ MeV. However, in radiotherapy absorber with low Z (e. g. tissue with $Z < 10$) are common, indicating that

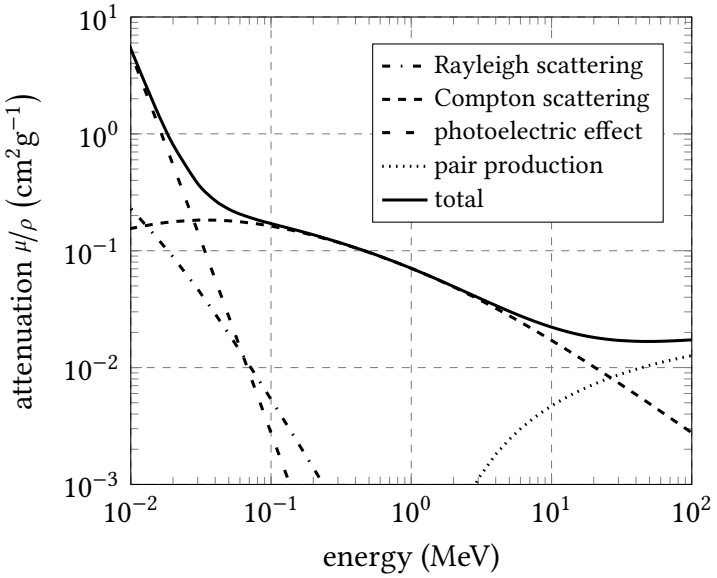


Figure 2.2.: Total mass attenuation coefficient as well as individual attenuation coefficients for Rayleigh scattering, Compton scattering, photoelectric effect and pair production in water as a function of the incident photon energy and normalized to corresponding absorber density ρ. Data is obtained from the XCOM database [45].

the Compton scattering is one of the most important interactions of photon beams and tissue in radiotherapy.

For photon beam dosimetry, it is of particular interest of how much energy is transferred to kinetic energy of secondary electrons (cf. Section 2.1.6). To this end, the mass energy transfer μ_{tr} coefficient can be defined as

$$\frac{\mu_{\mathrm{tr}}}{\rho} = \frac{\mu\langle T\rangle}{\rho h\nu} \tag{2.20}$$

with $\langle T\rangle$ being the expectation value of the energy converted to kinetic energy of secondary electrons in an interaction. Further, a mass energy absorption coefficient can be defined that describes the energy loss of electrons to secondary photons

$$\frac{\mu_{\mathrm{en}}}{\rho} = (1-g)\frac{\mu_{\mathrm{tr}}}{\rho} \tag{2.21}$$

where g is the fraction of the kinetic energy lost to photons during the complete slowing down of secondary electrons.

2.1.6. Energy deposition in matter

The local energy deposition of a photon beam that traverses through matter is primarily executed by emitted electrons resulting from photon-matter interactions. Here, the kinetic

energy of the emitted electrons is gradually transferred to electrons of the tissue mainly by inelastic scattering with orbital electrons, i. e. ionization and excitation of atoms. During their path trough matter, the emitted electrons can further create δ-electrons that, again, have an considerable range of their own. Further, generation of bremsstrahlung in the electromagnetic field of the atom core as well as orbital electrons contribute to the energy transfer. Elastic scattering in the field of the core and inelastic scattering with the atomic core have only a minor impact on the energy loss of electrons. To characterize the kinetic energy loss $\mathrm{d}T$ of electrons as they travel distance $\mathrm{d}x$ in a medium, the stopping power quantity $S = {}^{\mathrm{d}T}/_{\mathrm{d}x}$ is normally employed. More commonly, however, is to express the distance with respect to the mass per unit area of the material with density ρ

$$\frac{S}{\rho} = \frac{\mathrm{d}T}{\rho \mathrm{d}x}. \tag{2.22}$$

The stopping power combines the energy loss due to ionization and excitation of atoms, i. e. the collision stopping power S^{col}, and the generation of bremsstrahlung, i. e. the radiation stopping power S^{rad}

$$\frac{S}{\rho} = \frac{S^{\mathrm{col}}}{\rho} + \frac{S^{\mathrm{rad}}}{\rho}. \tag{2.23}$$

Here, the energy of the incident electrons directly define the relative proportion of both interaction processes onto the total energy loss. S^{col} of a material is described by the Bethe-Bloch equation, modified to be applicable for electrons

$$\frac{S^{\mathrm{col}}}{\rho} = 2\pi r_{\mathrm{e}}^2 N_{\mathrm{e}} \frac{m_{\mathrm{e}}c^2}{\beta^2} \left[\ln \frac{T_{\mathrm{e}}^2 (T_{\mathrm{e}} + 2m_{\mathrm{e}}c^2)}{2m_{\mathrm{e}}c^2 I^2} + \frac{{}^{T_{\mathrm{e}}^2}/_8 - (2T_{\mathrm{e}} + m_{\mathrm{e}}c^2) m_{\mathrm{e}}c^2 \ln 2}{(T_{\mathrm{e}} + m_{\mathrm{e}}c^2)^2} + 1 - \beta^2 - \delta \right] \tag{2.24}$$

where r_{e} is the classical radius of the electron, $N_{\mathrm{e}} = N_{\mathrm{A}}({}^{Z}/_{A_{\mathrm{r}}})$ with N_{A} being the Avogadro constant and A_{r} the atomic weight of the material, $\beta = {}^{v_{\mathrm{e}}}/_{c}$ with v_{e} as the speed of the electron and δ the density correction term. An approximation for S^{rad} is given by the ICRU report 37 [46] as

$$\frac{S^{\mathrm{rad}}}{\rho} = \sigma_0 \frac{N_{\mathrm{A}}}{A_{\mathrm{r}}} Z^2 (T_{\mathrm{e}} + m_{\mathrm{e}}c^2) \langle B \rangle, \tag{2.25}$$

where $\sigma_0 = ({}^{1}/_{137}) \left({}^{e^2}/_{m_{\mathrm{e}}c^2}\right)^2$ and $B = B({}^{h\nu}/_{T_{\mathrm{e}}})$ is a slowly varying function with an average for $T_{\mathrm{e}} \ll m_{\mathrm{e}}c^2$ of $\langle B \rangle = {}^{16}/_{3}$.

2.2. Photon beam dosimetry

Dosimetry in radiotherapy is an important tool to allow quantifying the energy that radiation emits due to the above described photon-matter interactions when traversing

through matter and generation of secondary particles, mainly electrons. The alterations in physical and chemical properties of the irradiated medium are measurable and often proportional to the absorbed dose and can be used to quantify the applied dose. As biological effects directly depend on the absorbed dose, estimation and measurement of dose distributions are important in radiotherapy. The following section therefore concentrate on briefly explaining the general concepts of dosimetry and is based on the textbooks of P. Mayles [47], H. Reich [43] and the International Atomic Energy Agency (IAEA) [48].

2.2.1. Energy imparted, absorbed dose and KERMA

In general, the dose can be seen as a quantity with respect to a point that is spatially and temporally derivable. Further, the point dose is related to a mass element of the irradiated material. That is, for the experimental determination of that dose, the mass element has to be sufficiently dimensioned, as otherwise the absorbed energy will be a stochastic variable[1]. With regard to this, a quantity of stochastic nature, the energy imparted ε in a reference volume, is defined. The definition reads

$$\varepsilon = R_{\text{in}} - R_{\text{out}} + \sum Q \tag{2.26}$$

with R_{in} being the sum of energies of all charged and uncharged ionizing particles entering the volume, R_{out}, similar to R_{in}, only considers particles that leave the volume and $\sum Q$ gives the sum of changes of the rest mass energy of nuclei and elementary particles in any nuclear transformations that occur in the volume. In R_{in} and R_{out}, rest mass energies are excluded. The expectation value of ε, the mean imparted energy $\langle\varepsilon\rangle$ is a non-stochastic quantity.

Under a biological perspective, the central quantity in radiotherapy is the specific energy $\mathfrak{e}$, defined as quotient of local imparted energy and the mass m of the absorbing volume

$$\mathfrak{e} = \frac{\varepsilon}{m}. \tag{2.27}$$

The reason for this is that the microscopic scale in the area of cells or even smaller parts of cells make it necessary to consider the stochastic nature of the energy absorption; this is only adequately possible by applying the specific energy $\mathfrak{e}$. Biological effects $\mathcal{E}_{\text{biol}}$ are therefore directly dependent on $\mathfrak{e}$, i. e. $\mathcal{E}_{\text{biol}} = \mathcal{E}_{\text{biol}}(\mathfrak{e})$. In radiotherapy, however, radiation effects are commonly specified by the macroscopic and measurable quantity of

[1] Dose as a stochastic quantity, i. e. with spatial, temporal and spectral aspects of the stochastic nature of the energy deposition processes, is generally dealt with under the heading of microdosimetry, cf. [49, 50].

the absorbed dose, D. Correlation between $\mathfrak{e}$ and D and the impact on biological effects can be given as

$$\mathcal{E}_{\text{biol}}(D) = \int \mathcal{E}_{\text{biol}}(\mathfrak{e}) f(\mathfrak{e}, D)\, \mathrm{d}\mathfrak{e} \tag{2.28}$$

where $f(\mathfrak{e}, D)$ is the spatial distribution of the dose. That is, biological effects are not only dependent on the dose but also on the distribution pattern of the microscopic energy deposition.

The absorbed dose with unit Gray (Gy, $1\,\text{Gy} = 1\,\text{Jkg}^{-1}$) can be derived from the mean of the specific energy

$$\langle \mathfrak{e} \rangle = \int_0^\infty \mathfrak{e} \cdot f(\mathfrak{e})\, \mathrm{d}\mathfrak{e} \tag{2.29}$$

as its boundary value of the mean energy absorption in a small mass element

$$D = \lim_{m \to 0} \langle \mathfrak{e} \rangle. \tag{2.30}$$

However it is more common to define D as the quotient of the mean energy imparted $\langle \varepsilon \rangle$ delivered to matter with mass dm

$$D = \frac{\mathrm{d}\langle \varepsilon \rangle}{\mathrm{d}m} = \frac{\mathrm{d}\langle \varepsilon \rangle}{\rho \mathrm{d}V}. \tag{2.31}$$

In radiotherapy, the absorbed dose is one of the most important quantities because it directly represents the energy per unit mass that remains in matter to produce any effects attributable to the radiation. Additionally, it is relevant to directly or indirectly ionizing radiation as well as to ionizing radiation sources distributed within the absorbing medium. However, as the absorbed dose is being deposited by secondary charged particles, it is not possible to relate the absorbed dose directly to the particle fluence or energy fluence

$$\Phi = \frac{\mathrm{d}N}{\mathrm{d}a}, \qquad \Psi = \frac{\mathrm{d}R}{\mathrm{d}a}, \tag{2.32}$$

respectively, of a field of indirectly ionizing radiation. Here, dN denotes the number of particles and dR the energy incident on a sphere of cross sectional area da.

Quantification of radiation fields is usually done by the kinetic energy released per unit mass (KERMA). KERMA describes the first order of energy transfer for indirect ionizing radiation and is defined as

$$K = \frac{\mathrm{d}E_{\text{tr}}}{\mathrm{d}m} = \frac{\mathrm{d}E_{\text{tr}}}{\rho \mathrm{d}V} \tag{2.33}$$

with dE_{tr} as the sum of the initial kinetic energies of all the charged ionizing particles released by uncharged ionizing particles. Similar to the total stopping power defined in Eq. (2.23), KERMA can be divided into a collision and a radiation part, i. e. $K = K_{\text{col}} + K_{\text{rad}}$.

Computation of KERMA in a specific medium at a point in space with fluence Φ of photons with energy $h\nu$ is possible by using the relationship defined in Eq. (2.20)

$$K = \left(\frac{\mu_{\text{tr}}}{\rho}\right) h\nu\Phi = \left(\frac{\mu_{\text{tr}}}{\rho}\right) \Psi. \tag{2.34}$$

As a direct measurement of the absorbed dose is usually not possible, the correlation to the measurable KERMA quantity is exploited. Under the so-called charged particle equilibrium (CPE), i. e. the extent of the measurement device is greater than the maximum range of the secondary electrons generated in the detector medium, it follows that $\varepsilon = E_{\text{tr}}$ and thus the absorbed dose in a medium D_{med} is equal to $K_{\text{col,med}}$. Using the relationship

$$K_{\text{col,med}} = \left(\frac{\mu_{\text{en}}}{\rho}\right) \Psi \tag{2.35}$$

it follows for monoenergetic energies and CPE at a depths of interest d that

$$D_{\text{med},d} = \left(\frac{\mu_{\text{en}}}{\rho}\right)_{\text{med}} \Psi_{\text{med},d}. \tag{2.36}$$

2.3. Monte Carlo simulation of particle transport

The basic principle in radiotherapy is to plan a dose distribution that is going to be applied to a patient to hit and destroy a previously identified target volume, as described in Chapter 1. For given irradiation conditions, e. g. particle type, field sizes and forms, energy and beam directions, a transport equation has to be solved under consideration of corresponding boundary conditions. Usually, coupled integro-differential equations[2] that describe the electromagnetic shower are applied. However, they are only analytically solvable by approximation under simplifying assumptions. Thus, results of currently employed dose calculation methods in radiotherapy contain uncertainties of about 3% to 4%. For specific irradiation conditions, like dose calculation near tissue inhomogeneities or the usage of complex and small field forms, uncertainties can be even larger [51, 52]. This motivates application of the Monte Carlo method in radiotherapy treatment planning, as this is the only known method that can be applied for any energy range of interest. Such simulations achieve planned dose distributions with highest possible accuracy by employing fundamental physical principles of particle transport, i. e. interaction processes as described in Section 2.1. Remaining uncertainties are statistical fluctuations, as a Monte Carlo simulation with no uncertainty would take an infinite amount of time. A trade-

[2] In integro-differential equations, integrals and derivatives of a function are involved.

off between an acceptable simulation uncertainty and the respective computational simulation duration has to be made [53, 54].

Thus, when boundary conditions are sufficiently known, the transport equation is solvable with arbitrary accuracy by applying Monte Carlo simulations. Each particle track in a geometry of interest is simulated. That is, an incident electron or photon, with all its interactions with matter and resulting generation of secondary particles, is described from its entry into the geometry until the incident particle exits the geometry or is completely absorbed. Therefore, interaction probabilities defined by corresponding cross sections have to be known and estimated by accordingly distributed random variables.

Assuming that x is such a random variable in interval $[a, b]$ with a probability density function, i. e. cross section, and $f(x)$ with $f(x) \geq 0$ describing the distribution, then the distribution function is defined as

$$F(x) = \int_a^x f(x')\,\mathrm{d}x' \tag{2.37}$$

with $a \leq x \leq b$ and $F(b) = 1$. $F(x)$ is monotonically growing in interval [0,1], i. e. an ordinary generator of random variables would generate uncorrelated uniformly distributed variables ξ in that interval. For $\xi = F(x)$ or $x = F^{-1}(\xi)$, where F^{-1} is the inverse function of F, x is distributed according to $f(x)$. Thus, arbitrary distributed random variables x can be generated out of uniformly distributed random variables ξ.

For a photon, the distance to an interaction can then be estimated by the Monte Carlo method as

$$s = -\frac{1}{\mu}\ln(1-\xi), \tag{2.38}$$

where μ is the attenuation coefficient for photons with a given energy, as defined in Eq. (2.7). This estimation can be directly derived by the probability distribution function for interaction distances, i. e.

$$f(s) = \mu \exp(-\mu s) \tag{2.39}$$

where $0 \leq s \leq \infty$. Generation of secondary particles can afterwards be estimated by corresponding differential cross sections and resulting particle trajectories are further traced. In each volume element (voxel), the deposited energy is computed and accumulated. The number of simulated particle histories N defines the statistical accuracy of the simulation. A particle history, also referred to as case or shower, describes the transport of one initial particle, its interaction processes and during this procedure potentially produced secondary particles until all particles are absorbed or leave the geometry under consideration. Depending on the desired statistical accuracy, which usually decreases with $N^{-0.5}$, long computation times may be necessary [53, 55].

As described in Section 2.1, there are four primarily important photon-matter interaction processes that have to be considered in Monte Carlo simulations by using corresponding cross sections. The energy loss of electrons, as they traverse matter, can be described by two basic processes. First, the at high electron energies dominant mechanism of bremsstrahlung radiation transfers energy back to photons, leading to a coupling of the electron and photon radiation fields. Secondly, for low energies, the inelastic collisions with atomic electrons are dominant. This leads to excitation and ionization of the atoms along the paths of the particles. Further, elastic collisions occur at a high rate and cause a frequent change in electron direction.

For the estimation of electron trajectories and interactions, it is nearly impossible to use the same Monte Carlo simulation approach as for photons. The reason for this is that photons in an energy range of $10\,\text{keV} \leq E_\gamma \leq 40\,\text{MeV}$ have a relatively large mean free path of about 20 cm in radiotherapy-relevant low Z materials and hence only experience a few interactions [47]. Fast electrons or charged particles in general, however, are subject to a high number of collisions of about 10^3 to 10^4 in the process of slowing down [55, 56]; an event-by-event simulation of the electron transport is computationally not feasible. An approach to circumvent this difficulty is the condensed history technique. The main idea of this method is to condense large numbers of subsequent transport and collision processes to one single step [57]. More specifically, in each step, the cumulative effect of the individual interactions is considered. This is achieved by sampling the change of the particle's direction of motion, energy and position at the end of the step from appropriate multiple scattering distributions. The condensed history technique is based on and justified by the observation that solely a few of all electron interactions cause a considerable directional change or energy loss. The commonly utilized so-called class II implementation of the condensed history method differentiates between events with low energy loss and catastrophic collision events. Here, a catastrophic event can be an inelastic collision or a bremsstrahlung process with high energy transfer. Between two catastrophic events exactly one condensed history step with multiple scattering occurs.

For the electron transport in heterogeneous geometries, the interfaces between different materials have to be taken into account. The simulation of an electron close to an interface with another material requires to consider the curved electron trajectory. That is, parts of the electron trajectory may be in this different material and thus differs from the actual simulated trajectory, as schematically visualized in Fig. 2.3. The common approach to address this problem is to utilize a boundary crossing algorithm. Whenever an electron comes closer to a boundary than a defined minimum distance, the Monte Carlo simulation is not longer using the condensed history technique but simulates single elastic scattering effects. Thus, a potential boundary crossing can be correctly estimated.

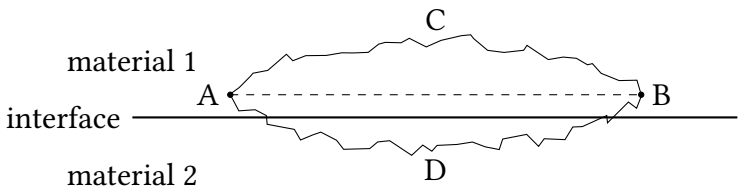


Figure 2.3.: Problem of electron transport in heterogeneous geometries when applying the condensed history technique. The electron transport is simulated between A and B (dashed line, both points in material 1). For the indicated electron trajectory C in material 1, no problems occur. However, parts of the electron trajectory indicated by line D may be in material 2. Thus, the entire particle history is potentially affected. Figure modified from [58].

The efficiency of Monte Carlo simulations can generally be improved by the application of variance reduction techniques. That is, instead of increasing the number of simulated histories, N, the variance σ is reduced by constructing a new Monte Carlo problem with the same answer as the original one but with a lower σ. Typical and in this thesis applied variance reduction techniques regarding the particle transport in a medical linear accelerator are range rejection, bremsstrahlung splitting and Russian Roulette.

Range rejection The electron range is checked against the distance to the nearest boundary on every step. Whenever the estimated range is shorter than the distance to the boundary, the electron is terminated and its energy deposited in the current region. As the electron could emit a bremsstrahlung photon, which potentially is able to leave the region, this technique is only executed if the electron's energy is below an energy threshold.

Bremsstrahlung splitting This technique employs the splitting of bremsstrahlung interactions, i. e. each bremsstrahlung event creates an arbitrary number of bremsstrahlung photons with their weight suitably reduced. The energy of the electron creating this event is decreased by the energy given off by one of the generated photons. Even if this is in violation of conservation of energy on an individual interaction basis the resulting fluctuations in energy loss for electrons and expectation values for photon energy and angular distributions are correct. The gain in simulation efficiency is high as simulating the photon transport is fast and this technique makes optimal use of each electron track.

Russian Roulette Particles generated in a class of events that are of little interest for the general simulation result as they, for instance, would never reach the geometry of interest, motivate the application of the Russian Roulette technique; with a given probability the low-interest particles are eliminated. Increasing the weight of the surviving particles by the inverse of that probability ensures that the result is unbiased.

2.4. Linear accelerators in radiotherapy: Beam generation and modulation

Generation of external beam radiation applied in radiotherapy is generally conducted by employing medical linear accelerators, where electrons are accelerated by an synchronized electromagnetic field. Typically, the high-frequency field oscillates at a vibration frequency of 2856 MHz or 2998 MHz (S-band), which corresponds to a wavelength of 0.15 m and 0.1 m in vacuum, respectively. The electrons to be accelerated are produced by an electron gun and synchronized with the pulsed electromagnetic wave injected into a waveguide. If the accelerator operates in photon-mode, the accelerated electrons will be focused onto a thick target, having a high atomic number (e. g. tungsten). In doing so, bremsstrahlung radiation is produced, which subsequently is focused onto the isocenter.[3] Subsequently, the radiation field is formed by dedicated collimators.

In the following, functionality of primary components of a medical linear accelerator are explained in detail. This section builds on theoretical basics given in books of the IAEA [48], P. Mayles [47] and W. Schlegel [59].

2.4.1. Microwave power sources

Generation of microwave pulses needed for electron acceleration is realized by magnetrons (low and medium energy machines) or klystrons (high energy machines). The magnetron is a radiofrequency (RF) self oscillator. It has a cylindrical structure with a centrally placed cathode. Thermionic electrons that are emitted from the cathode are attracted toward the surrounding concentric anode by the positive anode potential. The anode itself is comprised of an array of cylindrical cavities. A static magnetic field is applied parallel to the cathode axis that leads to a complex cycloidal electron trajectory around the cathode. Interaction of electrons with the RF cavities and influences of space charge forces yields rotating bunches of electrons. This process creates an oscillating electric field in the resonant cavities that reduces the kinetic energy of the moving electrons. The energy is transferred to the oscillating cavities and RF power can be extracted.

In contrast to the magnetron, the klystron is not a RF oscillator but a RF amplifier. An electron gun is used to generate an electron beam that is passed to the RF cavities. In the first cavity, the buncher cavity, electrons are bunched by an applied RF signal (velocity modulation). In the following, similarly designed, resonator cavity, the bunched electrons induce RF oscillations. Repeating this process, the RF power is build up, i. e. electron

[3] If the linear accelerator operates in electron-mode, the bremsstrahlung radiation target is moved out of the beam path. As in this thesis only the more commonly applied photon-mode is employed, this chapter focuses on the description of a medical linear accelerator operating in photon-mode.

energy of the initial beam is further reduced and transferred to the RF power. Eventually, the RF power can be extracted in the last cavity and propagated to the accelerating waveguide.

2.4.2. Accelerating waveguide

The accelerating waveguide is a metallic (copper) pipe, separated into cylindrical cavities by irises. Injected RF waves are used to generate strong magnetic and electric fields in each cavity by induction currents in the surface of the cavity walls. Thus, electrons are accelerated by induced fields along the guide axis. Two distinct types of guides exist, the travelling waveguide, where the RF wave is propagated along the guide axis, and the standing waveguide, where the wave is stationary.

The first part of the travelling waveguide is a buncher unit that synchronizes space, phase and velocity of the injected electrons as they acquire energy from the electromagnetic field. At resonance, the charge distribution in a travelling waveguide changes in sync with the microwave frequency. Electron bunches that move in the same direction as the microwave propagation are accelerated, where energy gain of electrons results in a relativistic mass increase.

In a standing waveguide, microwaves are reflected with a $\pi/2$-phase change back into the waveguide. The superposition of reflected and injected microwave build up a standing wave. In every second pipe segment, an oscillation antinode is generated, in intermediate segments oscillation nodes occur. Electrons can only be accelerated in segments with oscillation nodes. However, simultaneously to the electrons passing through the node, maximum negative electromagnetic field intensity has to be present.

2.4.3. Beam steering and focusing

In Fig. 2.4, the beam steering and focusing in a medical linear accelerator is schematically illustrated. The electron beam coming from the accelerating waveguide is not directly applicable for radiotherapy treatment. In general, the waveguide of a medical linear accelerator is not vertically arranged, i. e. the beam is not yet pointing to the isocenter. Thus, inside the linear accelerator head, the beam has to be bent. In modern therapy devices, a magnetic bending system is used to bend the beam by 270°. The magnetic field strength increases for larger electron orbits, i. e. electrons with higher energy experience higher magnetic fields. Thus, the inhomogeneous magnetic field acts like an achromatic lens. The electrons, independent of their energy, hit the same focal spot with minimum spatial extent after passing through the bending system.

For a linear accelerator operating in photon beam mode, the now narrow focused beam of electrons has to be converted to a photon beam. This is achieved by shooting the electrons onto a thick target with high atomic number (usually tungsten) to generate ultrahard bremsstrahlung radiation. Afterwards, a first beam collimation is performed by a static primary collimator. The bremsstrahlung radiation has higher intensities in the direction of the initial electron beam as on the field edges. Thus, a flattening filter is inserted in the beam to homogenize the photon energies, i. e. compensate for the lack of scatter at the edge of the field. The flattening filter is circularly symmetric with a profile with decreasing thickness towards the edges.

An important device placed beneath the flattening filter is the monitoring system, consisting of a thin layer of ionization chambers. Here, the total beam profile is measured to monitor the beam intensity. Further, beam homogeneity and symmetry are controlled by employing an appropriate sectioning of chambers. Subsequently, the lower part of the accelerator head[4] begins with secondary collimators or so-called jaws. These are utilized to define the maximum field size. In modern medical linear accelerators, the jaws are automatically controlled and moved during dose application, providing an optimal beam shielding. The actual field forming, i. e. adapting the spatial extent of the beam, is afterwards done by a more flexible collimator type, a multileaf collimator (MLC). As the name suggests, a MLC consists of a variety of thin (1.6 mm to 3 mm) tungsten leaves that are arranged opposing each other pairwise. During treatment, each leaf is independently computer-controlled allowing the MLC to form almost arbitrary radiation fields. This beam modulation is the basis of complex treatment techniques like e. g., the VMAT technique that is mainly applied in this thesis. A more detailed explanation of VMAT is therefore given subsequently.

2.4.4. Beam modulation and dose application by VMAT

The VMAT treatment technique is a highly dynamical dose delivery approach that employs a continuous target irradiation while the gantry rotates around the patient. As VMAT is an operating mode, conventional medical linear accelerators with standard MLC are able to use this technique. During VMAT, the dynamical modulation of gantry rotation speed, dose rate and MLC leaf positions is employed to achieve optimal dose distributions. The continuous irradiation during the VMAT treatment allows to deliver the dose in arcs and thus in a short period of time (in the order of minutes) [60]. The dynamic source motion, i. e. the gantry rotation around the target, is defined by a finite number of static beams or so-called control points (CP, the smallest segment of a VMAT arc). For each

[4] A more detailed illustration of the lower part of the accelerator head can be found in Fig. 4.4.

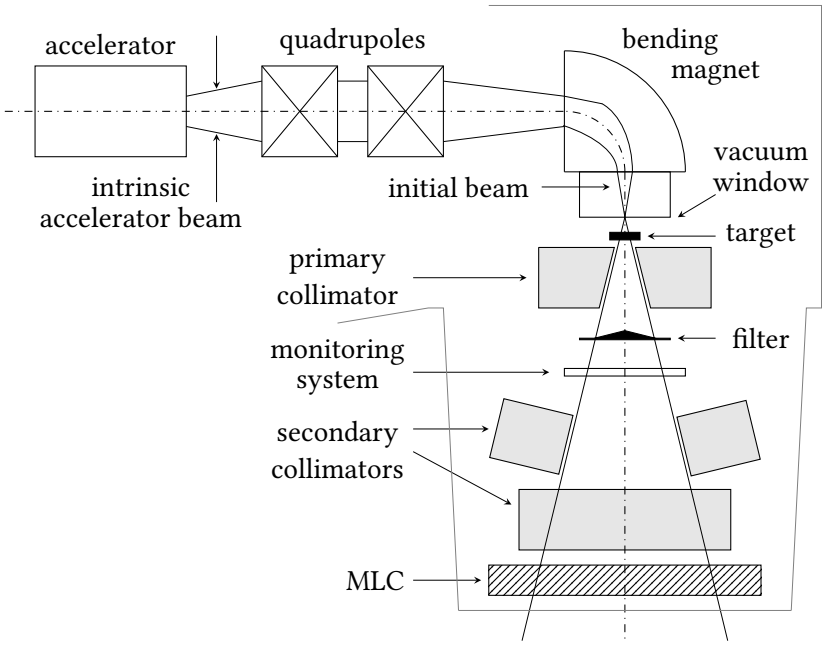


Figure 2.4.: Illustration of beam steering and focusing in a medical linear accelerator.

CP, the source position and instantaneous MLC configuration is given. Before treatment, a dedicated TPS is used to optimize the planned dose by randomly and simultaneously varying the MLC shapes, the dose rate and the gantry rotation speed for every CP. During dose delivery, MLC leaves and the gantry move linearly in between each CP. Physical restrictions of machine components have to be considered during treatment optimization. More specifically, a gantry rotation speed restriction to a maximum of $4.8° \, s^{-1}$ limits the duration of a full arc to a minimum of 75 s [61]. Further, collimator leaves in individual MLC rows must be able to finish their transition from the previous CP to the next CP (maximum leaf speed of $2.5 \, cm \, s^{-1}$) and a maximum dose rate of $600 \, MU \, min^{-1}$ has to be taken into account. As the in one CP delivered number of MU is variable, the gantry will move at maximum speed and apply the dose with a reduced dose rate when less than 2.08 MU are going to be delivered per degree. When the dose rate is at maximum, the gantry will slow down when a larger amount of MU is to be delivered. Note that these numbers differ from manufacturer to manufacturer and refer to the medical linear accelerator applied in this thesis.

Depending on the TPS, the dose to be delivered can be computed in total, i. e. for the total VMAT arc, or individually for subbeams of specific angular size (summation over all subbeam doses results in the total VMAT arc dose). The minimum angular size of a subbeam is defined by the TPS (for the TPS applied in this thesis a minimum angular size of $\delta\alpha \geq 2.3°$ is allowed), where a subbeam contains of at least two CP to consider MLC and gantry motion during dose calculation.

Chapter 3

Imaging and image processing in 4D radiotherapy

The basis of radiotherapy treatment planning and quality assurance is medical imaging and image processing. For dose calculation, solely CT data sets provide necessary information about the voxel-wise electron density in a medium. Therefore, this chapter gives an overview of the principle of CT imaging, followed by an outline of image processing in the context of radiotherapy treatment planning images, with the focus lying on the concepts of image registration and deep learning.

3.1. Principle of computed tomography

Modern CT scanner employ a rotating X-ray source with fan beam geometry and a detector opposing to the beam source to produce cross sectional images of an object, as schematically shown in Fig. 3.1 (a). Material-specific attenuation of the X-ray beam intensity that is measured by the detector after traversing through the considered slice of the object can be established with Eq. (2.5) by integration along the considered X-ray beam direction

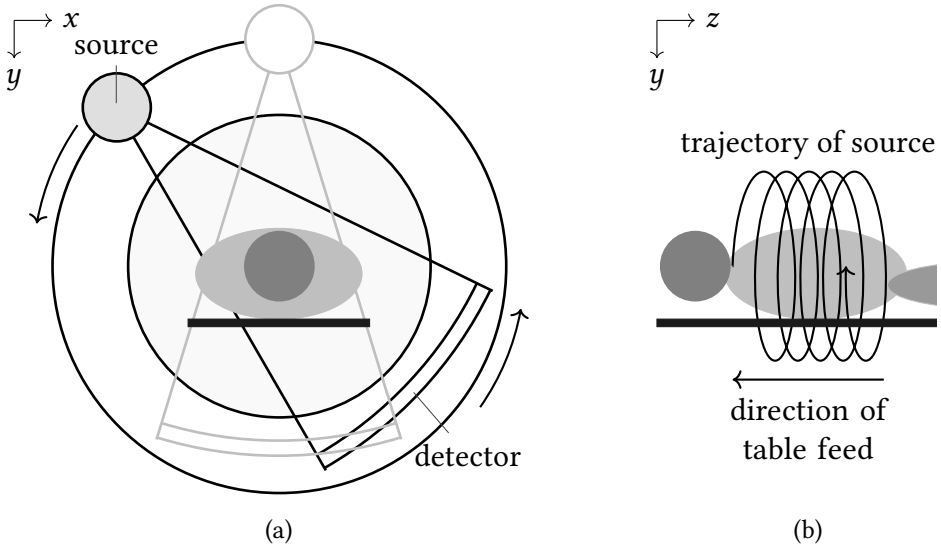
$$I = I_0 \cdot \exp\left(-\int_s \mu(\kappa)\,\mathrm{d}\kappa\right), \tag{3.1}$$

with the integral being a line integral along the beam direction s. As CT imaging is performed via a rotating X-ray source/detector, a variety of intensity values is measured from different angles. By applying the corresponding projection integral

$$-\ln\left(\frac{I}{I_0}\right) = \int \mu(\kappa)\,\mathrm{d}\kappa \tag{3.2}$$

CT imaging aims at reconstructing the spatial distribution of the attenuation coefficients of a considered slice based on projection data. For this task different reconstruction algorithms are available. A commonly utilized method is the filtered backprojection [40].

As the detector is composed of four up to 128 rows of detector elements (depending on the CT manufacturer and type), it is generally possible to acquire more than one image slice of the investigated object for each CT table position. In general, however,



(a) (b)

Figure 3.1.: Schematic of the principle of modern CT scanner. (a) For acquisition of a cross sectional image, the X-ray source with fan beam geometry and the detector rotate around an object. (b) The acquisition of larger volumes is generally realized by applying a spiral CT scan. Here, the same rotation of the source/detector as in (a) is utilized, but at the same time a continuous table feed through the CT tube is necessary, to yield the spiral trajectory in the patient coordinate system. Figure based on [62].

the reconstructed CT image has a slice thickness, i. e. image spacing in z-direction, of 0.6 mm to 3 mm. The straightforward method to acquire even larger volumes would be to move the CT table in z-direction after each acquisition until the total field of view to be investigated is imaged (step & shoot CT scanning). This, however, takes a relatively long time and e. g. patient motion while imaging could lead to artifacts between stacks of slices in the reconstructed volume. Therefore, a commonly applied method is the so-called spiral CT scan, where the CT table continuously but slowly moves in z-direction through the CT tube while at the same time cross sectional images are obtained; see Fig. 3.1 (b) for a visualization of the approach.

To obtain a gray scale image after image reconstruction, the 3-dimensional distribution of μ is converted to a dimensionless scale in relation to the attenuation coefficient of water

$$\text{CT-value}\ (\mu) = \frac{\mu - \mu_{H_2O}}{\mu_{H_2O}} \cdot 1000. \tag{3.3}$$

In medical practice, the so-called Hounsfield scale, named after G. Hounsfield, uses $2^{12} = 4096$ possible Hounsfield (HU) values in the range of -1024 to $+3071$. According to the HU definition, HU values for water and lung are 0 and -1000, respectively. For soft tissue, HU values are primarily in the range of -100 to $+100$ and HU values above $+700$ are representing bones. Visual inspection of obtained images is usually done by applying the windowing technique. Here, high contrast is achieved by selecting the HU

value range of interest, characterized in terms of a window center C as well as a window width W and visualized by a gray level scale with 256 values. Values below the chosen window are shown in black, values above in white. In this work, applied windows are with ($C = -600$, $W = 1500$) for lung and ($C = 50$, $W = 400$) for liver/tissue.

3.1.1. CT with temporal resolution – 4D CT

3D CT imaging of moving objects, e. g. internal patient geometry that moves due to patient respiration, is inevitably subject to motion artifacts due to discrepancies in the acquired projection data. Those deviations directly influence the reconstruction process and thus the resulting 3D CT image of the patient. Depending on utilized imaging parameters and internal motion magnitudes as well as direction of internal motion, different artifact types (double structures, motion blurring of contour edges, etc.) with varying severity can occur.

In 4D radiotherapy, as explained in the introduction of this thesis, a common approach to tackle this issue is to consider the temporal dimension while imaging, i. e. to acquire 3D + t (4D) CT data sets. The basis of 4D CT imaging is to record the external breathing signal of the patient during CT scanning by applying for instance an infrared camera mounted on the wall of the treatment room in combination with an infrared reflecting marker block that is placed on the patients' chest wall. Synchronization between CT acquisition and breathing signal allows to assign breathing phases to corresponding CT projections. As the acquisition time of a 4D CT ($\approx$ 100 s) is usually much larger then an average patient breathing cycle ($\approx$ 4 s), the reconstruction process in 4D CT imaging makes use of combining projection data of equal breathing phases of different breathing cycles. Thus, a 4D CT does not describe the internal patient motion for the whole acquisition process, but solely the internal patient motion for one average breathing cycle during 4D CT imaging. This means, dividing each patient breathing signal cycle into n_{ph} phases yields a 4D CT data set consisting of n_{ph} 3D CT volumes where each volume is reconstructed using only a fraction of all available projections. Motivated by the compromise of sufficiently covering the average breathing cycle and at the same time providing enough projections per phase image to provide an adequate image quality after reconstruction, a commonly chosen value for n_{ph} is ten. Positioning of phases, so-called binning, is usually done by a phase- or amplitude-based approach. The phase-based binning divides each breathing cycle into temporal equidistantly spaced bins, as shown in Fig. 3.2 (a). In Fig. 3.2 (b), the amplitude-based binning is illustrated. Here, bins are equidistantly spaced in amplitude, i. e. for the shown example at 0%, 20%, 40%, 60%, 80% and 100% of the amplitude of one breathing cycle for in- and expiration. Deviations of

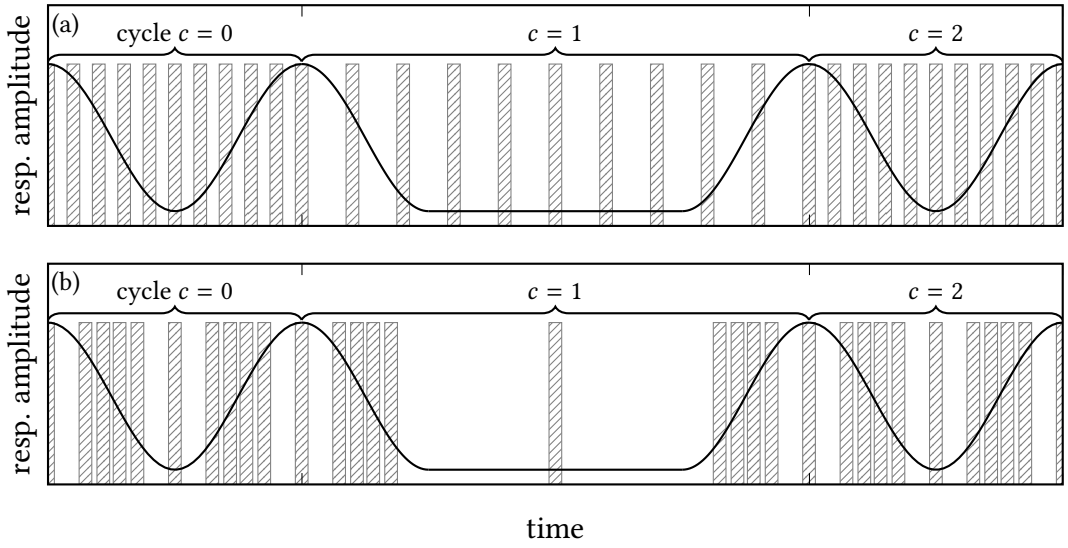


Figure 3.2.: Comparison of binning methods in 4D CT reconstruction for $n_{\text{ph}} = 10$. The bin width is defined in utilized CT protocol and directly dependent on the table feed during scanning. (a) The phase-based binning divides each breathing cycle into temporal equidistantly spaced bins. For breathing cycles with large breathing periods, e. g. cycle $c = 1$, bins cover the total cycle but are placed in areas with small or even no change in breathing amplitudes. (b) Division of each breathing cycle into n_{ph} bins using amplitude-based binning, where bins are equidistantly spaced in amplitude. For this approach, binning in cycle $c = 1$ cover areas with change in amplitude but projections acquired small or even no change in breathing amplitude are not used in CT reconstruction.

phase-based binning and amplitude-based binning becomes clear for breathing cycle $c = 1$. Due to the equidistant spacing in time, when applying phase-based binning, areas with small or even no change in breathing amplitude are sampled and combined with bins in other breathing cycles. However, for the amplitude-based binning, areas with change in breathing amplitude are correctly covered but projections acquired during small or even no change in breathing amplitude are discarded and not used in CT reconstruction. For both approaches, image artifacts may be present after reconstruction. Hence, commonly occurring artifact types are discussed in following section.

3.1.2. Image artifacts in 4D CT data

Typical image artifacts that occur in 4D CT data sets are illustrated in Fig. 3.3. The so-called double structure artifacts (cf. left part of Fig. 3.3) are caused by a faulty bin sorting during image reconstruction. If phase-based binning is used, the example in Fig. 3.2 (a) could directly lead to this kind of artifacts, as bins for z-positions for $c = 1$ are not correctly reflecting the internal patient geometry as seen in remaining breathing cycles ($c = 0$ and $c = 2$). Thus, slices are reconstructed that do not match the physiological state represented by neighbouring axial slices. However, double structure artifacts can

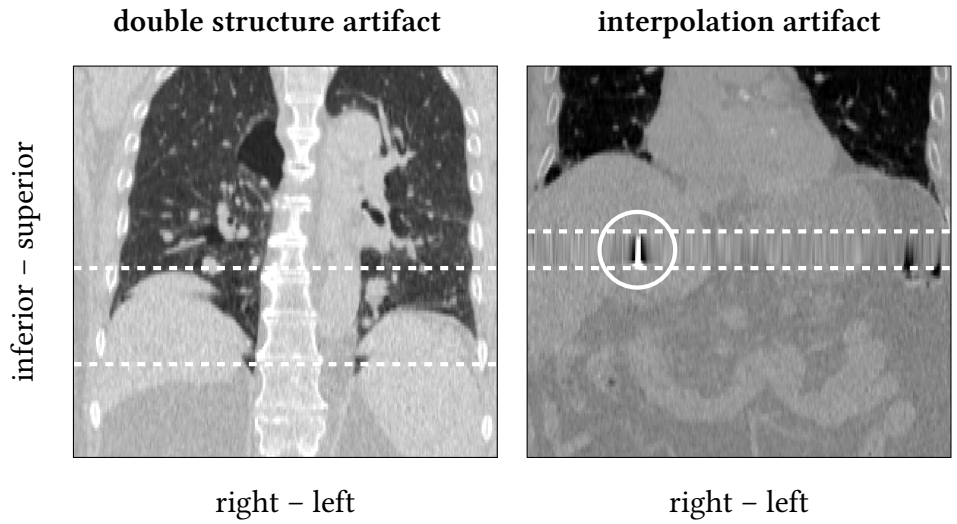


Figure 3.3.: Typical image artifacts in 4D CT data sets. Double structure artifacts (left) caused by a faulty bin sorting while image reconstruction, caused by e. g. an irregular patient breathing during imaging. Interpolation artifacts (right) occur when information between slices is missing and missing information is then compensated while image reconstruction by interpolation between adjacent z-slices.

also occur for amplitude-based binning. Here, a high variability in patient breathing amplitude leads to a erroneous correspondence between internal and external motion and in consequence to a combination of bins representing different physiological states.

Amplitude-based binning as shown in Fig. 3.2 (b) can further cause so-called interpolation artifacts (cf. right part of Fig. 3.3), as the CT scan does not stop during the illustrated breathing pause in cycle $c = 1$ and therefore projection data for, e. g., inspiration phases is missing for the investigated z-positions. Missing information is compensated during reconstruction by interpolation between z-slices that were successfully reconstructed. Such artifacts can also occur for phase-based binning when the spacing between bins, i. e. the breathing period of the patient, is too large.

Generally, amplitude-based binning is more robust than phase-based binning, as investigated and shown in e. g. [63, 64]. Therefore, in-house acquired image data is solely reconstructed by applying standard amplitude-based binning. More advanced CT imaging and reconstruction methods currently developed, i. e. an intelligent 4D CT sequence scanning approach [65] or an optimized projection binning algorithm [66], are unfortunately not yet available for present clinical CT scanners.

3.2. Biomedical image processing

In medical imaging in a broader sense, the general aim of image processing is to support the patient therapy or diagnosis by computer-based identification, distinction, analyses

and visualization of image objects of interest. For this kind of tasks, image processing encompasses a variety of different methods and algorithms, applied individually or in combination. Especially algorithms for e. g. image registration, segmentation, pattern detection and recognition as well as classification are important methods that are commonly utilized. Nowadays, complex image analysis tasks are often solved by machine learning methods, as such methods are fast and achieve similar or even better results compared to standard algorithmic approaches. In the last years, especially deep learning, a subfield of machine learning, has come to high popularity in (biomedical) image analysis. The advantage of deep learning is that high-level, abstract features from raw data are considered by learning and introducing task-specific representations that are expressed in terms of other, simpler representations, i. e. complex concepts are represented by a combination of simpler concepts. Deep learning in biomedical image processing is, however, commonly utilized for image segmentation and classification tasks. The application in non-linear image registration problems can only be rarely found.

In this thesis, non-linear image registration is of particular interest, because it is one of the fundamental image processing tasks to extract necessary patient motion information out of 4D CT image data sets. To understand the idea of non-linear image registration, a more detailed description is given in the subsequent section. Thereafter, the concept of deep learning is introduced, as a new methodological approach to allow for a deep learning-based DIR is proposed in this thesis (cf. Section 4.3 and Section 5.4).

3.2.1. Non-linear image registration

Image registration is a fundamental field of research in medical image processing to allow for analyses of, for instance, patient images of different imaging modalities. To do so, images usually located in different coordinate spaces have to be transformed into one joint coordinate space to enable a direct comparison and visualization of images (cf. for example mentioned registration between imaging modalities in Fig. 1.1). Further, image registration can be applied to image data that uses the same coordinate space. However, differences between images exist also, e. g., in 4D CT data sets due to patient respiration. Here, characterization and extraction of internal patient motion information using computed transformations are of interest. The theory of this section is based on the books of H. Handels [67], J. Modersitzki [68] and R. Werner [62].

The registration process itself is based on an image pair, a reference image R and a template image T with $R, T : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$, and aims at finding transformation $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ so that similarity between transformed template image $T \circ \varphi$ and the reference image becomes highest under consideration of a desired smoothness of φ. The

application of an appropriate distance measure $\mathcal{D}$ used as image dissimilarity measure allows to express the registration problem as a minimization problem of a joint functional $\mathcal{J}$

$$\mathcal{J}[\varphi] := \mathcal{D}[R, T, \varphi] + \alpha \mathcal{S}[\varphi] \xrightarrow{\varphi} \min, \tag{3.4}$$

with $\alpha \in \mathbb{R}^+$ denoting a positive weighting factor that controls registration smoothness $\mathcal{S}$. The regularizer $\mathcal{S}$ is utilized to penalize unrealistic transformations that can occur in the ill-posed[1] registration problem and further aims at generating a smooth transformation. Thus, the type of $\mathcal{S}$ directly influences the registration result and has to be selected according to the registration task and prior knowledge. The sought transformation φ is generally split into identity function id as trivial part and a displacement u, i. e. $\varphi = \mathrm{id} + u$. After successful registration, the transformation φ projects corresponding image information in template T and reference image R onto each other. The obtained displacement field u is then a d-dimensional vector field that entirely describes the transformation φ.

To solve the non-linear optimization problem defined in Eq. (3.4), numerical optimization approaches, i. e. iterative approximations, have to be applied. A common choice is the method of gradient descent. However, various different strategies to solve optimizing problems exist and are also applicable to the registration problem[2]. All optimizers use a similar approach where a series of guesses from an initial starting position is taken and the sought solution iteratively approached. Here, it is important that the starting position is sufficiently close for the algorithm to converge to the correct answer. The registration algorithm can then compute the similarity between both images for the starting point using the pre-defined distance measure $\mathcal{D}$. Afterwards, the registration process proceeds by applying a small transformation to the template image and again computes the cost function between both images. Progression towards an optimal registration is then achieved by seeking transformations that decrease the distance measure until a minimum is found. Usually, stop criteria are defined, i. e. the registration process stops if a desired image similarity is achieved or the change of the computed distance measure over some iterations is negligible.

Distance measures primarily employed in the registration algorithms used in this thesis are the sum of squared differences (SSD) and the normalized cross correlation (NCC). The sum of squared differences of intensity values of template and reference image is a standard similarity measure in monomodal image registration. Especially for registration

[1] A problem is well-posed if it has a solution, the solution is unique and depends continuously on the data. If not, the problem is ill-posed [68].

[2] Other approaches are e. g. the finite difference method, Newton's method, the Quasi-Newton method, the Gauss–Newton algorithm and the Levenberg–Marquardt algorithm.

of images that are acquired with the same scanner, registration using SSD as distance measure is a good choice, because image intensities are directly comparable. The SSD distance measure is defined as

$$\mathcal{D}^{\mathrm{SSD}}[R, T, \varphi] = \frac{1}{2} \int_{\Omega} \Big(R(x) - (T \circ \varphi)(x) \Big)^2 \, \mathrm{d}x. \tag{3.5}$$

However, for registration of multimodal images, i. e. CBCT and CT images of the same patient as seen in radiotherapy treatment planning, the SSD distance measure is not suitable due to its direct dependency on comparable image intensities. The NCC measure that only assumes a linear correlation between reference and template image

$$\mathcal{D}^{\mathrm{NCC}}[R, T, \varphi] = 1 - \frac{\langle R, T \circ \varphi \rangle^2}{\|R\|^2 \, \|T \circ \varphi\|^2} \tag{3.6}$$

with the applied norm being the Euclidean norm (L2 norm) is a better approach for this specific task. Similarity between images with different intensity ranges can be expressed. The NCC distance measure is also applicable for monomodal image registration.

Solely minimizing the distance functional $\mathcal{D}$ is an ill-posed problem. Thus, a regularization functional $\mathcal{S}$ is added that depends convexly on derivatives of the transformation. Implausible transformations like containing singularities, surjective mappings or non-physiological behaviour can, as a consequence, be penalized. The common elastic regularizer $\mathcal{S}^{\mathrm{elas}}$ is based on the application of the linear-elastic model onto the displacement field u. More descriptively, the elastic regularizer is motivated by the assumption that the objects being imaged deform elastically. That is, deforming one object by an external force, to maximize the similarity to a second object, is counteracted by an internal force given by the linear-elastic model. The linear-elastic regularization functional is defined as

$$\mathcal{S}^{\mathrm{elas}}[u] = \int_{\Omega} \frac{\eta}{4} \sum_{i,k=1}^{d} \left(\frac{\partial u_k}{\partial x_i} + \frac{\partial u_i}{\partial x_k} \right)^2 + \frac{\lambda}{2} (\nabla u)^2 \, \mathrm{d}x, \tag{3.7}$$

with the Lamé parameters $\lambda \in \mathbb{R}$ and $\eta \in \mathbb{R}$ describing the elasticity properties of the partial derivatives of the displacement function u. Depending on the registration task, Lamé parameters have to be adequately chosen. Another regularization approach is followed by the diffusive regularization functional $\mathcal{S}^{\mathrm{diff}}$. Here, the aim is to prevent strong variations in the sought displacement field by integrating over the quadratic gradient of the displacement field in each dimension, i. e.

$$\mathcal{S}^{\mathrm{diff}}[u] = \frac{1}{2} \sum_{l=1}^{d} \int_{\Omega} \|\nabla u_l(x)\|^2 \, \mathrm{d}x. \tag{3.8}$$

Large deviations between neighbouring vector components are therefore suppressed during registration and a component-wise smoothing of the displacement field is achieved. A closely related and commonly employed regularization approach is the Gaussian regularization, i. e. a component-wise Gaussian smoothing of the displacement field.

3.2.2. Deep learning

In (biomedical) image analysis, deep learning, a sub-field of machine learning, is nowadays often applied. The reason for the currently high popularity of deep learning is diverse. Most important are, however, the increasing amount of training data available and the improvements in computer hardware and software as well as the resulting high accuracy in solving increasingly complicated tasks. This section is mainly based on the books of Goodfellow et al. [69] and Zhou et al. [70].

The basis of deep learning is the application of neural networks, a family of models that try to mimic the learning process in human brains but are generally not designed to be realistic models of biological function. In its simplest form, a feed-forward neural network[3] consists of an input $\{v_i\}_{i=1}^{D}$, trainable weights $\{w_i\}_{i=1}^{D}$, a bias b and an output y. For a given observation $\boldsymbol{v} \in \mathbb{R}^D$ and an activation function $f(\cdot)$, y results from the weighted sum of the inputs

$$y(\boldsymbol{v};\Theta) = f\left(\sum_{i=1}^{D} v_i w_i + b\right) = f\left(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{v} + b\right). \tag{3.9}$$

Here, $\Theta = \{\boldsymbol{w}, b\}$ is a parameter set and $\boldsymbol{w}$ a connection weight vector. The non-linear activation function $f(\cdot)$ is in general chosen to be a function that is applied element-wise. A typical choice and in modern neural networks the default setting is the rectified linear unit (ReLU) or variants of it. Using a pre-activation variable x that is determined by the weighted sum of the inputs, i. e. $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{v} + b$, ReLU activation is defined as

$$f_{\mathrm{ReLU}}(x) = \max\{0, x\}. \tag{3.10}$$

If more than one output is desired, i. e. in a multi-output task, the model defined in Eq. (3.9) can be extended to

$$y_k(\boldsymbol{v};\Theta) = f\left(\sum_{i=1}^{D} v_i W_{k,i} + b_k\right) = f\left(\boldsymbol{w}_k^{\mathsf{T}}\boldsymbol{v} + b_k\right), \tag{3.11}$$

[3] Feed-forward networks are the quintessential deep learning models and the basic concept of most network architectures. If feedback connections are included in feed-forward networks, they are generally referred to as recurrent neural networks, which are often applied for speech recognition.

where $\{y_k\}_{k=1}^{K}$ defines multiple outputs, $W_{k,i}$ is the corresponding connection weight between v_i and y_k and $\Theta = \left\{\boldsymbol{W} \in \mathbb{R}^{D\times K}, \boldsymbol{b} \in \mathbb{R}^{K}\right\}$. Both introduced network configurations are, however, single-layer neural networks, as solely one layer, i. e. the output layer, exists besides the visible input layer. The output of such single-layer networks is limited to linear combinations of the input, even if a non-linear activation function is used. To overcome this limitation, any number of hidden layers, i. e. layer between input and output layer, can be added to make the network deeper[4]. For instance, a two layer network with M hidden units can be written as

$$y_k(\boldsymbol{v};\Theta) = f^{(2)}\left(\sum_{j=1}^{M}\left[\left(W_{k,j}^{(2)} + b_k^{(2)}\right) f^{(1)}\left(\sum_{i=1}^{D}\left(W_{j,i}^{(1)} v_i + b_j^{(1)}\right)\right)\right]\right) \tag{3.12}$$

with the superscript being the layer index and

$$\Theta = \left\{\boldsymbol{W}^{(1)} \in \mathbb{R}^{D\times M}, \boldsymbol{W}^{(2)} \in \mathbb{R}^{M\times K}, \boldsymbol{b}^{(1)} \in \mathbb{R}^{M}, \boldsymbol{b}^{(2)} \in \mathbb{R}^{K}\right\}. \tag{3.13}$$

Depending on the task, the architecture of the neural network has to be designed. This includes not only choosing the number of layers that the network should contain but also the number of units in each layer and how the layers are connected to each other. To train such a network, the optimizer, the loss function and the form of the output units have to be defined and gradients of complicated functions computed. Thus, a brief overview of the more complex convolutional neural networks that are generally more suitable for medical image data applications is given subsequently before the actual network training utilizing backpropagation is introduced. Eventually, regularization approaches for deep learning are explained.

Deep convolutional neural networks

In (biomedical) image analysis tasks, spatial information about neighboring pixels or voxels is often necessary to achieve the desired output. Feed-forward networks, as introduced in previous section, need an input in vector form, i. e. spatial information is lost when applied for input data with dimensionality greater than one. Convolutional neural networks (CNNs), however, that, as already suggested by its name, contain convolutional layers, are designed for processing data with grid-like topology. That is, image features are extracted by convolutional operations performed in the hidden layers of a network.

A convolutional layer uses trainable kernels to detect image features in the input feature maps at different spatial positions. For instance, the kernel $\boldsymbol{K}_{i,j}^{(l)}$ defines connection

[4]This justifies the name deep learning.

weights between feature maps i and j at layers $l-1$ and l, respectively. More specifically, the activation $\boldsymbol{A}_j^{(l)}$ solely depends on a contiguous subset of units in the following layer feature maps $\boldsymbol{A}_j^{(l-1)}$, i. e.

$$\boldsymbol{A}_j^{(l)} = f\left(\sum_{i=1}^{M^{l-1}} \boldsymbol{A}_i^{(l-1)} * \boldsymbol{K}_{i,j}^{(l)} + b_j^{(l)}\right), \tag{3.14}$$

where the convolutional operation is denoted by $*$. Usually, the spatial extend of the applied kernel $\boldsymbol{K}$ is chosen to be smaller than the input. For instance, a kernel $\boldsymbol{K} \in \mathbb{R}^{3\times3}$ is able to extract meaningful features, e. g. edges, of an input image using only nine parameters. Thus, fewer parameters have to be stored reducing the memory requirements, improving statistical efficiency and the overall computational time is decreased.

After a convolutional layer, the feature maps are in general further modified by applying pooling layers to reduce the spatial size. A pooling layer can more specifically be seen as a summary over the responses over a whole neighborhood. For instance, the max/average pooling operation reports the maximum/average output within a rectangular neighborhood. However, pooling layers not only reduce spatial size but also achieve translation invariance over small spatial shifts in the input.

Backpropagation and network training

Assume a set of observations $\boldsymbol{v}_i$ and corresponding targets or labels $\boldsymbol{y}_i$, i. e. the data set can be denoted as $\{\boldsymbol{v}_i, \boldsymbol{y}_i\}_{i=1}^N$, for neural network training. A loss function $\mathcal{L}(\boldsymbol{y}_i, \tilde{\boldsymbol{y}}_i)$ must be defined that compares the predicted output $\tilde{\boldsymbol{y}}_i$ and the given target $\boldsymbol{y}_i$ using a metric. For image-to-image transformation tasks, image intensity-based loss functions are commonly used. A typically choice for neural network training is, similar to the SSD defined in Eq. (3.5), the mean of the squared error (MSE) loss function

$$\mathcal{L}(\boldsymbol{y}_i, \tilde{\boldsymbol{y}}_i) = \frac{1}{N}||\boldsymbol{y}_i - \tilde{\boldsymbol{y}}_i||_2^2 = E. \tag{3.15}$$

Here, N denotes the number of available samples and E the discrepancy or error. A gradient descent-based optimization is usually performed to minimize the given error function by updating parameters of the network iteratively. The computation of the corresponding gradient, i. e. ∇E, has to be performed in an algorithmic and efficient way for millions of parameters in a neural network. For instance, the recently proposed EfficientNet, which achieves much better accuracy and efficiency than other state of the art networks, contains up to 66 millions of parameters [71]. The common and best performing choice is to calculate the gradient by means of error backpropagation [72].

Error backpropagation makes use of the chain rule that allows to propagate the errors in the output layer computed by the loss function back to the input layer. That is, the derivative of error E regarding the parameters for the lth layer, i. e. $\boldsymbol{W}^{(l)}$, in a L-layer neural network can be denoted as

$$\frac{\partial E}{\partial \boldsymbol{W}^{(l)}} = \frac{\partial E}{\partial \boldsymbol{a}^{(L)}} \frac{\partial \boldsymbol{a}^{(L)}}{\partial \boldsymbol{a}^{(L-1)}} \cdots \frac{\partial \boldsymbol{a}^{(l+2)}}{\partial \boldsymbol{a}^{(l+1)}} \frac{\partial \boldsymbol{a}^{(l+1)}}{\partial \boldsymbol{a}^{(l)}} \frac{\partial \boldsymbol{a}^{(l)}}{\partial \boldsymbol{z}^{(l)}} \frac{\partial \boldsymbol{z}^{(l)}}{\partial \boldsymbol{W}^{(l)}}. \tag{3.16}$$

Here, activation and pre-activation vectors of layer l are denoted by $\boldsymbol{a}^{(l)}$ and $\boldsymbol{z}^{(l)}$, respectively. As $\boldsymbol{a}^{(L)} = \tilde{\boldsymbol{y}}$, the partial derivative of E, $\frac{\partial E}{\partial \boldsymbol{a}^{(L)}}$, directly corresponds to the error computed by the loss function. The gradient estimation of a given loss function regarding the parameter $\boldsymbol{W}^{(l)}$ is performed by using the back-propagated error from the output layer. That is, $\frac{\partial \boldsymbol{a}^{(k+1)}}{\partial \boldsymbol{a}^{(k)}}$ for $k = l, l+1, \ldots, L-1$ and $\frac{\partial \boldsymbol{a}^{l}}{\partial \boldsymbol{z}^{l}} \frac{\partial \boldsymbol{z}^{l}}{\partial \boldsymbol{W}^{l}}$ is applied where $\frac{\partial \boldsymbol{a}^{(k+1)}}{\partial \boldsymbol{a}^{(k)}}$ can be computed by

$$\frac{\partial \boldsymbol{a}^{(k+1)}}{\partial \boldsymbol{a}^{(k)}} = \frac{\partial \boldsymbol{a}^{(l+1)}}{\partial \boldsymbol{z}^{(l+1)}} \frac{\partial \boldsymbol{z}^{(l+1)}}{\partial \boldsymbol{a}^{(l)}}. \tag{3.17}$$

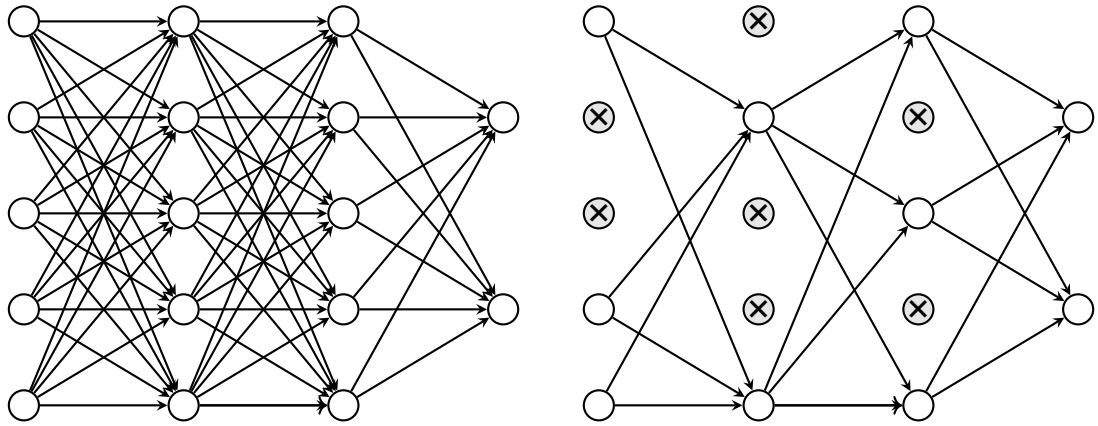
Doing this for all layers of the neural network allows to update the network parameter set $\boldsymbol{W}$ and thus to train the network.

Regularization approaches

The performance of a neural network, trained and tested on a specific data set, when applied to a new, previously unseen data set, is often reduced compared to the initially achieved performance during testing. Methods that improve generalizability of the network are commonly referred to as regularization approaches. A fundamental method applied in deep learning is data set augmentation [73].

Data set augmentation is an effective technique, which can be described as a way to create additional data for network training achieved by applying geometrical and value transformations to the input data. More specifically, for high dimensional image data with an enormous variety of factors of variation as input data, even a small rigid translation of a few pixels can introduce a high level of generalizability. Commonly used transformations, applied individually or in combination, are flipping, rotating, random cropping, zooming, injecting synthetic noise and image intensity modifications. For a supervised image-to-image transformation task it is mandatory to apply the geometrical transformations used to augment the training data to corresponding ground truth data during network training.

Another simple and helpful technique is the dropout method [74]. The underlying idea is to randomly drop non-output neurons with a specified probability during network training, as visualized in Fig. 3.4. That is, only a randomly chosen subset of neurons is



(a) (b)

Figure 3.4.: Schematic visualization of the dropout method. (a) Initial network with fully connected layers. (b) Application of dropout randomly selects non-output neurons to be dropped during a training iteration. The number of connections is reduced.

trained during one network training iteration. In particular overfitting[5] to the training data during model training, especially for a limited training data set, is reduced. Further, dropout greatly improves the generalizability for large neural networks.

In the initially introduced form of the dropout approach, dropout was solely performed during training. However, Yang et al. [75] and references therein proposed that a deterministic CNN architecture is easily modified to be probabilistic by using dropout. More specifically, dropout that were active during training is also enabled during model prediction to obtain a probabilistic network output by a Bayesian approximation [76].

[5] A network is overfitted when it is closely fitted to the input data. More specifically, the network performs extremely well on the input data but poor on previously unseen inputs, i. e. the generalizability is lost.

Chapter 4

Methods and materials

In this chapter, utilized materials and applied methods are explained. The first three sections focus on describing individually developed/applied and combined methods to predict internal patient motion, using this information for 4D dose accumulation, and finally on how to apply deep learning in the context of image registration to allow for an uncertainty estimation in 4D dose simulation. Eventually, patient data sets used for the evaluation of the different approaches are briefly introduced. Corresponding results are given in the following chapter.

4.1. Prediction of internal patient motion

During radiotherapy treatment, internal patient motion is usually not acquired by during-treatment imaging[1]. However, for a retrospective analysis of, e. g., the applied dose, the internal patient motion information during treatment is one of the main sources of uncertainty and of particular importance and interest (cf. Chapter 1). Usually, only limited information in the form of external breathing signals are acquired during dose delivery (here: an external breathing signal that represents motion of the patient surface). An approach to estimate the internal patient motion during treatment is to employ a pre-treatment trained correspondence model. The general concept of correspondence modeling is to establish a functional relationship between external breathing signal measurements and the internal structure motion as represented by, e. g., 4D CT data sets (cf. Section 3.1.1). More specifically, correspondence modeling can be understood as a regression task that makes use of during-treatment acquired external breathing signals to predict the internal patient motion. In this thesis, the correspondence model approach of Wilms et al. [77] was used as the basis for internal motion prediction and further extended/adapted to the specific problem at hand. The general approach is visualized in Fig. 4.1 and subsequently described.

[1] Tumor tracking during radiotherapy is possible and performed by specialized radiotherapy systems. Such systems rely on X-ray acquisitions of the patient during treatment to update/check a pre-treatment built patient-specific correspondence model, i. e. an internal patient motion signal in form of a time series of X-ray images is available. As described in Chapter 1, disadvantages like e. g. high costs, increased imaging dose and long treatment duration times are reasons why this technologies are rarely applied.

For each patient, an individual correspondence model is built. The patient 4D treatment planning CT is a series of 3D CT images. Each image represents a different breathing phase of the patient, i. e.

$$(I_i)_{i\in\{1,\dots,n_{\text{ph}}\}}, I_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R}. \tag{4.1}$$

Here, $i \in \{1, \dots, n_{\text{ph}}\}$ is a breathing phase and n_{ph} the total number of 3D CT images in the 4D CT data set. In agreement with all data sets employed in this thesis (later introduced in Section 4.4), a typical choice for n_{ph} is 10. During 4D CT imaging, an external respiratory motion signal of the patients' chest wall is acquired and utilized for reconstruction. The breathing signal values, assigned to the phase images for reconstruction purposes, analogously read

$$(\zeta_i)_{i\in\{1,\dots,n_{\text{ph}}\}}, \zeta_i \in \mathbb{R}^{n_{\text{ind}}} \tag{4.2}$$

with n_{ind} being the dimensionality of the breathing (indicator) signal. In addition to the external motion information, internal motion information has to be available to establish the sought correspondence model. The extraction of internal motion information is therefore performed by computing DIR-based motion fields that represent the respiratory motion of the internal structures of the patient. Assuming an arbitrary breathing phase $i_0 \in \{1, \dots, n_{\text{ph}}\}$ being selected as reference phase, and the corresponding CT image I_{i_0} as reference image during registration, the registration process results in a series of transformations

$$(\varphi_i)_{i\in\{1,\dots,n_{\text{ph}}\}}, \varphi_i : \Omega \to \Omega. \tag{4.3}$$

In this thesis, the selected reference phase is a mid-expiration phase (i. e. $i_0 = 3$) to allow a more robust extrapolation in both in- and exhalation direction for the during treatment external motion acquisitions beyond the range of the 4D CT-related regressor signals and respective model-based motion field estimation.

Analogously to the splitting of transformations into an identity function (trivial part) and a displacement as described in Section 3.2.1, motion fields with regard to Eq. (4.3) read

$$(u_i)_{i\in\{1,\dots,n_{\text{ph}}\}}, u_i : \Omega \to \mathbb{R}^3, \tag{4.4}$$

with $u_i = \varphi_i - id$. The breathing signal information $(\zeta_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$ in combination with the respective motion fields $(u_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$ can now be correlated to allow for correspondence model training. In the following, measurements of the breathing signal and computed motion fields are interpreted as multivariate random variables.

The random variable for the motion fields, now denoted as U_i, combines the vector components of the field u_i for all $n_1 n_2 n_3$ voxel in Ω to one column vector of length

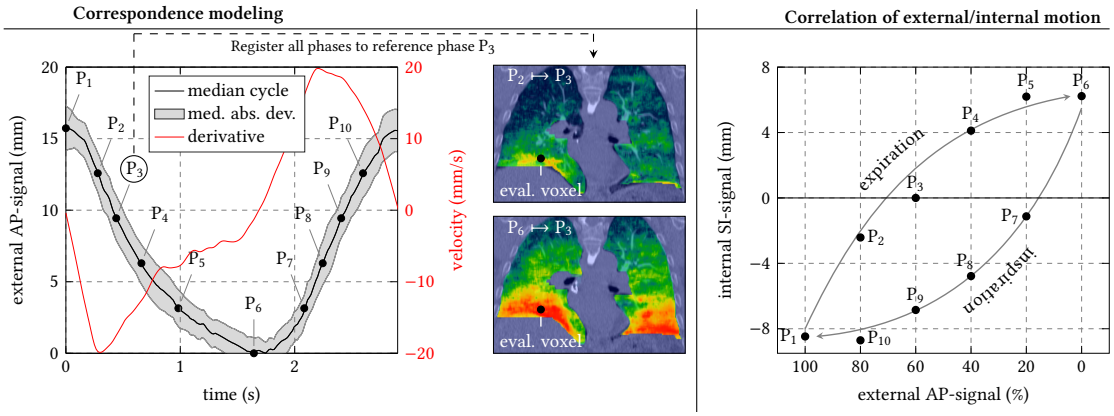


Figure 4.1.: Concept of correspondence modeling. Left: The functional relationship between external breathing signal measurements and internal motion information is established by multivariate regression-based correlation of breathing signal information (mean AP breathing cycle and its temporal derivative) and corresponding motion vector fields obtained by DIR with respect to the mid-expiration phase P_3. Right: Illustration of the relationship between motion vector SI components of the voxel highlighted in the middle column and the breathing signal values for the patient's 10-phase 4D planning CT. Differences between inspiration and expiration (hysteresis) highlight the need to incorporate additional information (here: temporal derivative of breathing signal). Figure based on [78].

$m = 3n_1n_2n_3$, i. e. $\boldsymbol{U}_i \in \mathbb{R}^m$. The correspondingly defined random variable of the breathing signal measurements is denoted as $\boldsymbol{Z}_i$. The random variables of the motion fields are combined to a mean-centered observation or regressand as

$$\boldsymbol{U} = \left(\boldsymbol{U}_1 - \langle \boldsymbol{U} \rangle, \ldots, \boldsymbol{U}_{n_{\mathrm{ph}}} - \langle \boldsymbol{U} \rangle\right) \in \mathbb{R}^{m \times n_{\mathrm{ph}}} \tag{4.5}$$

with $\langle \boldsymbol{U} \rangle = 1/n_{\mathrm{ph}} \sum_{i=1}^{n_{\mathrm{ph}}} \boldsymbol{U}_i$ being the mean motion field. Analogously, the breathing signal measurements are combined and form the regressor, i. e.

$$\boldsymbol{Z} = \left(\boldsymbol{Z}_1 - \langle \boldsymbol{Z} \rangle, \ldots, \boldsymbol{Z}_{n_{\mathrm{ph}}} - \langle \boldsymbol{Z} \rangle\right) \in \mathbb{R}^{n_{\mathrm{ind}} \times n_{\mathrm{ph}}}, \tag{4.6}$$

with $\langle \boldsymbol{Z} \rangle = 1/n_{\mathrm{ph}} \sum_{i=1}^{n_{\mathrm{ph}}} \boldsymbol{Z}_i$ being the mean breathing signal. The aim of a multivariate regression is now to estimate the correlation between regressor and regressand by

$$\boldsymbol{U} = \boldsymbol{B} \cdot \boldsymbol{Z}, \tag{4.7}$$

where the model parameters are represented by the system matrix $\boldsymbol{B} \in \mathbb{R}^{m \times n_{\mathrm{ind}}}$. Application of an ordinary least-squares regression approach yields an estimation of $\boldsymbol{B}$ as

$$\boldsymbol{B} = \arg\min_{\boldsymbol{B}'} \operatorname{tr}\left[(\boldsymbol{U} - \boldsymbol{B}'\boldsymbol{Z})\,(\boldsymbol{U} - \boldsymbol{B}'\boldsymbol{Z})^{\mathsf{T}}\right] = \boldsymbol{U}\boldsymbol{Z}^{+}, \tag{4.8}$$

where $\boldsymbol{Z}^{+} = \boldsymbol{Z}^{\mathsf{T}} \left(\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}\right)^{-1}$ is the Moore-Penrose pseudoinverse of $\boldsymbol{Z}$.

Assuming a suitable data set, the training of the correlation between breathing motion signal and motion fields allows for determination of the system matrix or estimator $\boldsymbol{B}$. For an additional measurement of the breathing signal, i. e. $\hat{\zeta}_i = \hat{\boldsymbol{Z}} \in \mathbb{R}^{n_{\text{ind}}}$, corresponding motion fields $\hat{u}$ and associated $\hat{\boldsymbol{U}}$ can be predicted by the sought correspondence model – i. e. the assumed relationship between breathing signal measurements and internal motion fields – as

$$\hat{\boldsymbol{U}} = \langle \boldsymbol{U} \rangle + \boldsymbol{B} \left(\hat{\boldsymbol{Z}} - \langle \boldsymbol{Z} \rangle \right). \tag{4.9}$$

In this thesis, the available breathing signal measurements of in-house acquired patient data sets are recorded by the Real-time Position Management (RPM) system of the treatment device manufacturer (Varian). This system uses an infrared camera mounted on the wall of the treatment room in combination with an infrared reflecting marker block that is placed on the patients' chest wall. The acquired signal is usually one-dimensional, i. e. only the AP motion component of the marker block is recorded. However, to allow for modeling of hysteresis behaviour (cf. right part of Fig. 4.1) typically induced by physiological differences between expiration and inspiration phases, the respective temporal derivative of ζ_i, i. e. $\partial_t z_i$ is additionally employed. Thus, the regressor is two-dimensional.

4.2. 4D dose accumulation

The term dose accumulation, as briefly described in Chapter 1, refers to computational approaches that incorporate information about the patient-specific breathing dynamics into the dose calculation process during treatment planning. More specifically, a motion-affected representation of the dose distribution to be delivered to the patient is sought and estimated by the simulation approach. The simulation itself can be understood as weighted summation of dose distributions that correspond to different breathing states of the time-dependent patient geometry [28]. Retrospective quality assurance in form of dose accumulation in radiotherapy of moving tumors is of particular interest to assure that a sufficient dose is applied to the target. As stated in Chapter 1, this is necessary to enable an appropriate response like an adaptation of the treatment plan, if underdosages occurred. However, the estimation of the actually applied dose does not only depend on the utilized treatment technique, i. e. the dose delivery process, but also on the patient-specific, respiration-induced motion and deformation of internal structures. In the previous section, an approach to estimate the internal patient motion during radiotherapy treatment by application of solely an external motion signal was introduced. This information can now be utilized in 4D dose accumulation algorithms to

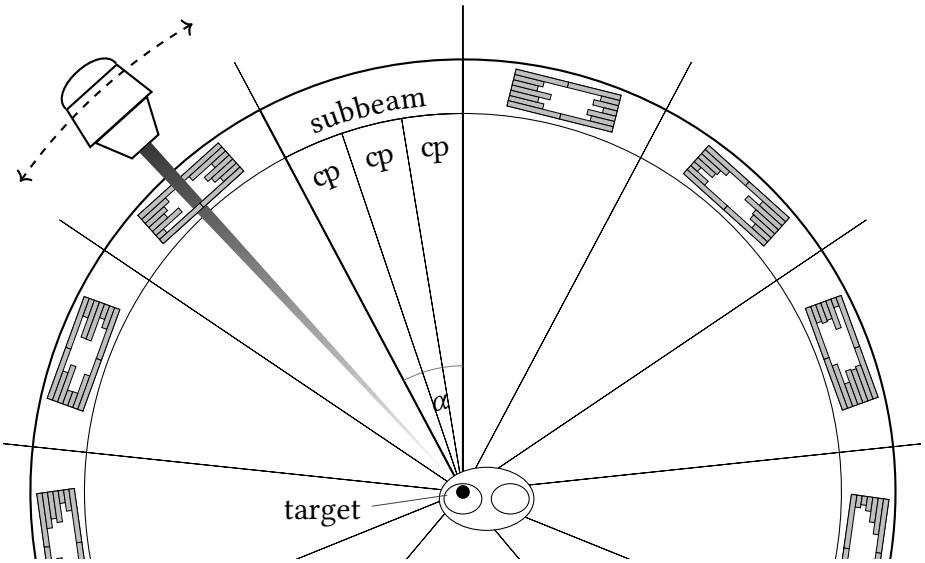


Figure 4.2.: Principle of VMAT dose accumulation. The medical linear accelerator rotates around the target, which is usually located in the isocenter. During rotation, the field form and fluence is adapted so that the target receives the prescribed dose and, at the same time, OAR are spared as good as possible. The VMAT arc can be divided by the TPS into segments of size $\alpha = 2.3°$, whereas each segment contains 2 to 5 CP that define necessary information about field form, rotation speed and dose rate. For dose accumulation without re-calculation of the dose distribution, the dose of each subbeam segment computed by the TPS is exported and subsequently used. The more advanced approach, utilizing dose re-calculation by application of Monte Carlo simulations, extracts the dose delivery information of all CP of the patient treatment plan, which are then used for dose calculation and simulation.

compute the dose delivered to the patient. In this thesis, in-house treated patients were exclusively irradiated by VMAT. Thus, the dose accumulation scheme proposed in this thesis is primarily developed and optimized for this technique. However, it can be easily modified to allow dose accumulation for the nowadays, especially for moving targets, scarcer used IMRT or conformal radiotherapy treatment (CRT) techniques.

Whereas for IMRT and CRT the treatment plan is composed of a specific number of irradiation fields with different but small number of manually (in general by a medical physicist) selected directions, angles and field forms to achieve an optimal dose distribution, the VMAT treatment is a more dynamic process and therefore more complex to model. More specifically, the VMAT technique employs continuous rotation around the target geometry, while simultaneously field form and fluence is dynamically modulated, as described and illustrated in Section 2.4.4 and Fig. 4.2. To simplify modeling of the dose delivery process and consequently the dose accumulation approach, one can assume that the dynamic process is in general a series of static irradiation fields with changing parameter settings. Thus, dividing a VMAT arc into angular segments with corresponding pre-computed dose segments is the straightforward approach to allow for a first implementation of the here proposed 4D dose accumulation scheme. Eventually,

the scheme is enhanced by introducing a Monte Carlo-based simulation approach. Both approaches are visualized and compared in Fig. 4.3 and subsequently in detail described.

Analogously to the previous section, a patient 3D planning CT image $I_{i_0} : \Omega \rightarrow \mathbb{R}$ out of the 4D CT series $\left(I_1, \ldots, I_{n_{\text{ph}}}\right) = (I_i)_{i \in \{1,\ldots,n_{\text{ph}}\}}$ is assumed to be the basis of the dose accumulation. Further, let $\varphi : \Omega \times \mathcal{T} \rightarrow \mathbb{R}^3$ be the position $\varphi(x,t)$ of a voxel $x \in \Omega$ of the planning CT image I_{i_0} at time $t \in \mathcal{T} = [0, T) \subset \mathbb{R}$. In a very general notation, the dose rate of one treatment fraction reads

$$\dot{D} : \Omega \times \mathcal{T} \subset \mathbb{R} \rightarrow \mathbb{R}_+. \tag{4.10}$$

Then, the interplay of patient motion and the dynamical dose delivery process can be expressed as

$$\begin{aligned} D_{\text{4D}}(x) &= \int_{\mathcal{T}} \dot{D}\left(\varphi(x,t), t\right) \, \mathrm{d}t \\ &\approx \sum_{t \in \tilde{\mathcal{T}}} \dot{D}\left(\varphi(x,t), t\right) \, \Delta t \\ &= \sum_{t \in \tilde{\mathcal{T}}} D_t\left(\varphi(x,t)\right) \end{aligned} \tag{4.11}$$

with $D_{\text{4D}} : \Omega \rightarrow \mathbb{R}_+$ being the sought dose distribution. In Eq. (4.11), an equidistant sampling is indicated by the numerically required temporal discretization. Here, a sampling period of Δt and sampling points of $t \in \tilde{\mathcal{T}} = \{1/2\,\Delta t, 3/2\,\Delta t, \ldots\}$, $\tilde{\mathcal{T}} \subset \mathcal{T}$ are assumed. The integral dose delivered in the time interval $[t - 1/2\,\Delta t, t + 1/2\,\Delta t)$ is thus $D_t : \Omega \rightarrow \mathbb{R}_+$.

Information about internal patient motion during radiotherapy treatment is usually not available, i. e. the time-dependent position $\varphi(x,t)$ of the inner voxels and structures is unknown. The straightforward and simple approach of common dose accumulation approaches is therefore to ignore the actual dynamics during dose delivery. Instead, the DIR estimated motion fields, resulting from the registration process between one reference CT phase and all remaining 4D CT phases, are applied to warp the dose. Then, Eq. (4.11) can be modified to

$$D_{\text{4D CT Sim}}(x) = \frac{1}{n_{\text{ph}}} \sum_{i=1}^{n_{\text{ph}}} D(\varphi_i(x)) \tag{4.12}$$

with the initially planned 3D dose distribution $D : \Omega \rightarrow \mathbb{R}_+$.

However, as described in the previous section, a breathing signal measurement of the external patient motion is continuously acquired during dose delivery. Therefore, the assumption that the planned dose delivery process can be appropriately discretized in time

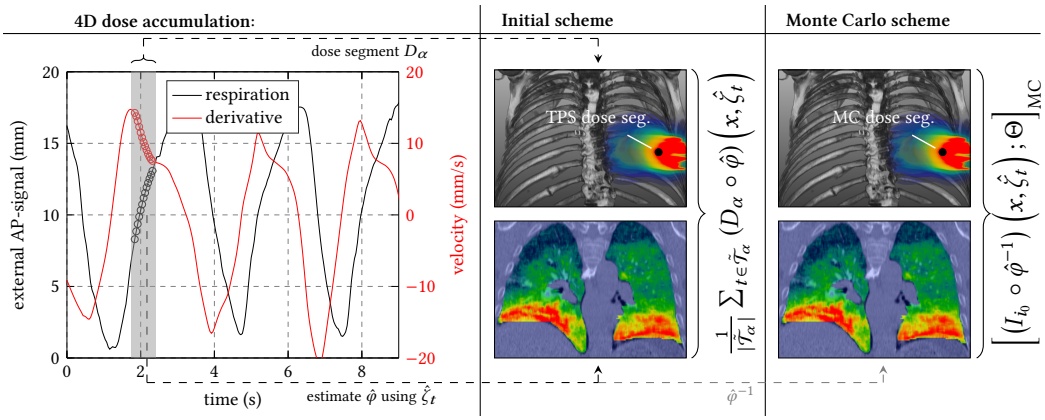


Figure 4.3.: General concepts of proposed dose accumulation schemes. Left: For each during-treatment breathing signal measurement $\hat{\zeta}_t$, a motion field $\hat{\varphi}$ is derived model-based. Middle: For the initial dose accumulation scheme, $\hat{\varphi}$ is used to deform the temporally corresponding planned VMAT (TPS-based) dose segment D_α, resulting in a motion-affected dose segment. Weighted summation of all motion-affected dose segments results in the sought dose $D_{4\mathrm{D}}$. Right: For the Monte Carlo-based dose accumulation scheme, the inverse of $\hat{\varphi}$ is computed and used to warp the reference image I_{i_0} to yield the virtual moving image $\hat{I}$. Dose calculation is then performed by Monte Carlo simulation with respect to corresponding beam parameters Θ. Note that the individual Monte Carlo-simulated dose segment $D_{\alpha_{\mathrm{MC}}}$ has to be warped back into the reference frame by applying $\hat{\varphi}$ before the summation over all dose segments can be conducted.

allows to correlate each individual breathing signal to a planned dose segment delivered at the time of measurement. In addition, the correspondence model can be applied to derive internal motion fields that correspond to the external signal measurement. That is, the integration of the correspondence model formation and the acquired breathing signal into Eq. (4.11) is possible. By application of $\hat{\zeta} : \tilde{\mathcal{T}} \rightarrow \mathbb{R}^2$ and $\hat{\varphi} = id + \hat{u}$, the correspondence model-based 4D dose simulation for a single fraction can be written as

$$D_{4\mathrm{D}}(x) \approx \sum_{t\in\tilde{\mathcal{T}}} D_t\left(\hat{\varphi}\left(x,\hat{\zeta}_t\right)\right) = \sum_{t\in\tilde{\mathcal{T}}} D_t\left(x+\hat{u}\left(x,\hat{\zeta}_t\right)\right), \tag{4.13}$$

with $\hat{\zeta}_t = \hat{\zeta}(t)$. As a radiotherapy patient treatment in general comprises multiple treatment fractions, Eq. (4.13) can be extended to

$$D_{4\mathrm{D}}^{\mathrm{total}}(x) = \sum_{\mathrm{fx}} D_{4\mathrm{D}}^{\mathrm{fx}}(x) = \sum_{\mathrm{fx}}\sum_{t\in\tilde{\mathcal{T}}_{\mathrm{fx}}} D_t\left(x+\hat{u}\left(x,\hat{\zeta}_t\right)\right), \tag{4.14}$$

with the irradiation fraction index denoted by fx and the temporal sampling points used to compute the sought integral fraction dose distribution as $\tilde{\mathcal{T}}_{\mathrm{fx}} \subset \mathbb{R}$. In common radiotherapy treatment processes, a plan adaption between fractions to account for e. g. changes in patient geometry is only sparsely conducted. This means, $\tilde{\mathcal{T}}_{\mathrm{fx}}$ is equal to $\tilde{\mathcal{T}}$ for all treatment fractions. In consequence, the only impact on deviations between the individual fraction dose distributions $D_{4\mathrm{D}}^{\mathrm{fx}}$ is due to the individual patient motion patterns

for each fraction. More precisely, differences in the fraction-wise dose accumulation are solely attributable to the observations $\hat{\zeta}_t$.

The proposed dose accumulation scheme in Eq. (4.14) is in its current form valid for arbitrary dose delivery techniques. However, the goal is to simulate the dose delivery of VMAT treatments. Thus, VMAT-specific characteristics, especially the dose delivery during continuous gantry rotation around the patient and field form modulation as visualized in Fig. 4.2, have to be considered. Further, the dose delivery in one fraction is normally conducted by applying a small number of VMAT arcs (e. g. for a fraction divided into two arcs, the gantry would rotate in arc one from start to end position and in arc two vice versa). For simplification of this process, the temporal discretization of the dose delivery process introduced in Eq. (4.13) can be re-parameterized. Here, it is beneficial and equivalent to replace t for each planned arc by the gantry rotation angle $\alpha \in [0°, 360°)$. Similar to $\tilde{\mathcal{T}}$, the sampling points of the gantry angle range $\mathcal{A}_{\text{ax}} \subset [0°, 360°)$ of the respective VMAT arc then read $\tilde{\mathcal{A}}_{\text{ax}} = \{1/2\,\Delta\alpha, 3/2\,\Delta\alpha, \dots\}$ with ax denoting the arc index. Following Eq. (4.13) and applying introduced re-parametrization yields

$$D_{\text{4D}}(x) \approx \sum_{\text{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\text{ax}}} D_\alpha \left(x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)\right). \tag{4.15}$$

Here, the voxel x of the reference CT image I_{i_0} in the angle range $[\alpha - 1/2\Delta\alpha, \alpha + 1/2\Delta\alpha)$ receives a dose of $D_\alpha \left(x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)\right)$. The position of voxel x is determined by the correspondence model estimation using the breathing signal measurement $\hat{\zeta}_\alpha$. It follows the predicted location of x as $x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)$.

Compared to the minimum achievable degree of VMAT arc discretization of 2.3° and the maximum gantry rotation speed of $4.8°\text{s}^{-1}$ (see Section 2.4.4 for details), the breathing signal acquisition rate with 25 Hz is relatively high. To account for this, each dose segment is weighted by $\tilde{\mathcal{T}}_\alpha^{-1}$, the inverse of the breathing signal measurement duration during dose delivery at gantry angle α. The implemented single fraction accumulation scheme then reads

$$\begin{aligned} D_{\text{4D}}(x) &\approx \sum_{\text{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\text{ax}}} \frac{1}{|\tilde{\mathcal{T}}_\alpha|} \sum_{t \in \tilde{\mathcal{T}}_\alpha} D_\alpha \left(x + \hat{u}\left(x, \hat{\zeta}_t\right)\right) \\ &= \sum_{\text{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\text{ax}}} \frac{1}{|\tilde{\mathcal{T}}_\alpha|} \sum_{t \in \tilde{\mathcal{T}}_\alpha} \left(D_\alpha \circ \hat{\varphi}\right)\left(x, \hat{\zeta}_t\right). \end{aligned} \tag{4.16}$$

Eventually, summing up the 4D-simulated dose distributions of the individual treatment fractions results in the total 4D dose distribution:

$$D_{\text{4D}}^{\text{total}}(x) = \sum_{\text{fx}} D_{\text{4D}}^{\text{fx}}(x) = \sum_{\text{fx}} \sum_{\text{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\text{ax}}} \frac{1}{|\tilde{\mathcal{T}}_\alpha|} \sum_{t \in \tilde{\mathcal{T}}_\alpha} \left(D_\alpha \circ \hat{\varphi}\right)\left(x, \hat{\zeta}_t\right). \tag{4.17}$$

4.2.1. Monte Carlo-based 4D dose accumulation

In the previous section, an algorithm to retrospectively compute a motion-affected VMAT dose distribution based on a correspondence model estimated patient breathing-induced internal motion during dose delivery was introduced. However, this approach still relies on the pre-treatment, TPS-based planned dose distribution, optimized on the reference CT image I_{i_0}. Angular dose segments are extracted from the planned 3D dose and correspondingly warped by application of estimated transformations $\hat{\varphi}$. That is, density changes within the internal patient geometry induced by respiratory motion are not considered during dose simulation. Further, as stated before, the angular resolution is limited to 2.3°, which approximately, depending on the gantry rotation speed[2], corresponds to a typical temporal spacing between adjacent segments of 0.5 s to 2 s (see Section 2.4.4 for details). Moreover, the extracted dose segments are computed based on dose calculation algorithms integrated into the TPS, which rely on analytically solving a transport equation under simplifying assumptions and therefore introduce dose uncertainties of 3% to 4% (cf. Section 2.3). An improvement of the algorithm proposed in Eq. (4.16) is therefore possible by improving the accuracy of the computed dose D_α and, at the same time, reducing the size of each angular segment. More specifically, the patient-specific treatment plan RT_{plan}, generated on the corresponding reference phase I_{i_0} of the 4D CT series, is divided into its smallest possible segments, i. e. CP (cf. Section 2.4.4). For a specific time point t_{CP}, the beam parameters Θ (e. g. gantry angle, leaf positions, beam energy etc.) that are mandatory for the radiotherapy treatment machine can be extracted:

$$RT_{\text{plan}}(t_{\text{CP}}) : t_{\text{CP}} \mapsto \Theta. \tag{4.18}$$

Furthermore, the information available in Θ about the treatment beam for each CP can be directly employed to allow for a CP-based Monte Carlo dose simulation. However, to consider density changes due to internal patient motion during treatment, the Monte Carlo dose simulation has to be conducted on the virtual moving image $\hat{I}(t_{\text{CP}})$. More specifically, the estimated transformation $\hat{\varphi}$ for a breathing signal measurement $\hat{\zeta}$ at time point t_{CP} defines the mapping of $\hat{I}(t_{\text{CP}})$ to I_{i_0}, where $\hat{I}(t_{\text{CP}})$ is assumed to represent the patient geometry at t_{CP}. This means, utilizing the inverse of $\hat{\varphi}$, i. e. $\hat{\varphi}^{-1}$, to warp I_{i_0}

[2]The maximum gantry rotation speed is $4.8°\text{s}^{-1}$, which is only rarely reached. The reason is the dose defined in the treatment plan that has to be delivered in a specific angular segment. How fast this is achieved is directly limited by the maximum possible dose application rate. For instance, for a number of 5 MU to be delivered in a 1° segment with a dose rate of 600 MU min^{-1}, i. e. 10 MU s^{-1}, the gantry rotation speed has to be reduced to $2°\text{s}^{-1}$.

results into the sought virtual moving image $\hat{I}$. Employing this strategy, the Monte Carlo simulation $D_{\alpha_{\mathrm{MC}}}$ reads

$$D_{\alpha_{\mathrm{MC}}}(x) = \left[\left(I_{i_0} \circ \hat{\varphi}^{-1} \right) \left(x, \hat{\zeta}_t \right) ; \Theta \right]_{\mathrm{MC}} \tag{4.19}$$

Here, the Monte Carlo simulation $[\cdot]_{\mathrm{MC}}$ is executed on the virtual moving image $\hat{I} = I_{i_0} \circ \hat{\varphi}^{-1}$ with respect to corresponding beam parameters Θ. Using the dose accumulation approach formalized in Eq. (4.17), the total 4D Monte Carlo dose simulation scheme can be expressed as

$$D_{\mathrm{4D}}^{\mathrm{total}}(x) = \sum_{\mathrm{fx}} \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} \frac{1}{|\tilde{\mathcal{T}}_\alpha|} \sum_{t \in \tilde{\mathcal{T}}_\alpha} \left(D_{\alpha_{\mathrm{MC}}} \circ \hat{\varphi} \right) \left(x, \hat{\zeta}_t \right) . \tag{4.20}$$

As illustrated and described in Fig. 4.2 and Section 2.4.4, an angular segment of size 2.3° consists of 2 CP to 5 CP. Consequently, the temporal resolution of the proposed Monte Carlo-based dose accumulation scheme is about 2 to 5 times higher compared to the initially introduced approach, resulting in a temporal spacing between adjacent CP of 0.1 s to 1 s. Further, exploiting the warped reference CT image allows for Monte Carlo dose computation on the patient geometry as it was actually irradiated during treatment – and thereby explicitly accounts for density changes during treatment resulting from patient breathing motion.

Monte Carlo simulations

Monte Carlo simulations were conducted using the EGSnrc/DOSXYZnrc user code [55]. To allow for an accurate Monte Carlo simulation, the geometry of the treatment machine originally employed for dose delivery has to be modelled as precise as possible. In this thesis, all patients considered for dose accumulation were treated by SBRT with the same linear accelerator (Varian TrueBeam linear accelerator) at the UKE. Information about the beam geometry and materials are directly available in the corresponding manual, provided by the machine manufacturer. Additionally, phase space files that describe type, energy, position and direction of particles at a given plane in the beam geometry were available and used for initialization of the treatment beam above the dynamically changeable accelerator geometry. This corresponds to the EGSnrc/DOSXYZnrc source 20, i. e. a phase space source through a dynamically library with multiple variable geometry settings[3]. In consequence, solely the lower part of the accelerator head (cf. Section 2.4.3) had to be modelled. An illustration of the modelled components is given in Fig. 4.4.

[3]The phase space file was provided by the linear accelerator manufacturer (Varian).

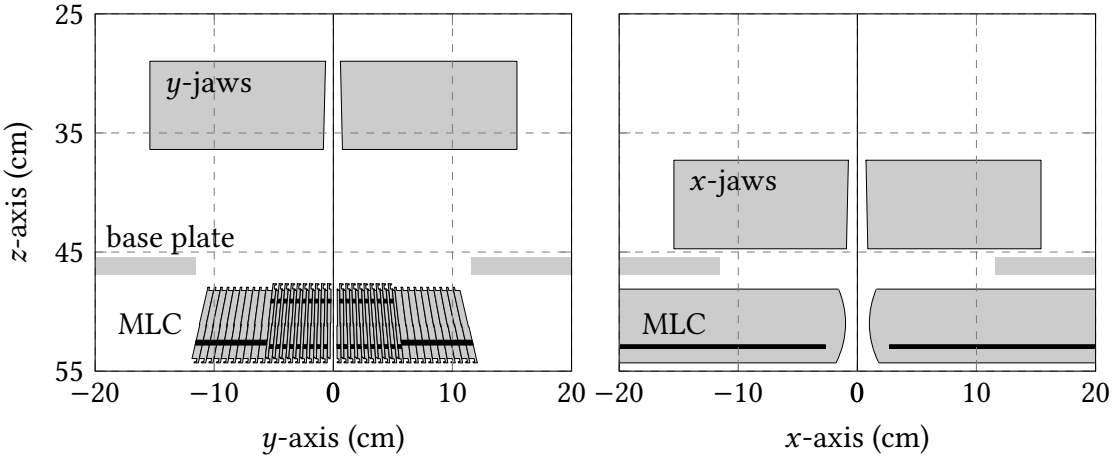


Figure 4.4.: Schematic of accelerator head geometry modeled in EGSnrc. As the corresponding phase space, located at $z = 0$ cm and defined by the linear accelerator manufacturer, is available, the upper part of the accelerator head, e. g. bending magnets, exit window, primary collimators and flattening filter, does not have to be modeled. As shown in the figure for yz (left part of figure) and xz-axis (right part of figure) the components that have to be considered during Monte Carlo simulation are the y and x-jaws, the base plate and eventually the field forming device, the MLC.

The dose distribution for one CP is Monte Carlo-simulated using 5×10^5 histories and a 2 mm isotropic dose grid size. As each treatment plan consisted of 300 to 400 CP, the total number of simulated particles ranged between 2×10^8 and 3×10^8 per treatment fraction. Corresponding simulation uncertainties, which directly depend on the number of simulated histories, were below 1% for all simulations. For the explicit modeling of MLC leaf motion, leaf positions for a CP are, similar to the real dose delivery process, linearly interpolated between CP specific leaf starting and ending position.

Absolute dose calibration

In general, the output of a Monte Carlo dose simulation consists solely of a relative dose distribution with a given dose uncertainty for each voxel. For the utilized EGSnrc/-DOSXYZnrc user code the output is normalized by an estimate of the number of particles incident from the original, non-phase space source. Consequently, an absolute dose calibration has to be applied to allow for a quantification of dose deviations between 4D-simulated, 3D-planned and 2D-measured dose distributions. An absolute dose calibration requires actual dose measurements under standard conditions of the respective treatment machine. Here, the dose is measured in different depths along the central axis (z-direction) inside a water phantom using a dedicated and calibrated ionization chamber to determine the sought calibration factor. The same setup geometry is reproduced and simulated utilizing the implemented Monte Carlo method. This allows for a correlation of measured absolute dose at a specific z-position $D_{z,\text{chamber}}$, delivered by a defined number

of MU, and simulated relative dose at the same z-position $D_{z,\mathrm{MC}}$. Standard definition for calibration measurements is that 100 MU ≡ 100 cGy at dose maximum applies. The relationship between simulation and measurement can then be exploited to calibrate the simulation, i. e. the simulated absolute dose at (x, y, z)-position, D_{xyz}, is defined as

$$D_{xyz} = \frac{D_{z,\mathrm{MC}}}{D_{z,\mathrm{chamber}}} \cdot \frac{1\,\mathrm{cGy}}{1\,\mathrm{MU}} \cdot \mathrm{MU}_{\mathrm{tot}}, \tag{4.21}$$

with $\mathrm{MU}_{\mathrm{tot}}$ being the total number of applied MU during treatment. The calibration factor, however, is solely valid for the beam energy used for the measurements, i. e. calibration has to be performed independently for each beam energy.

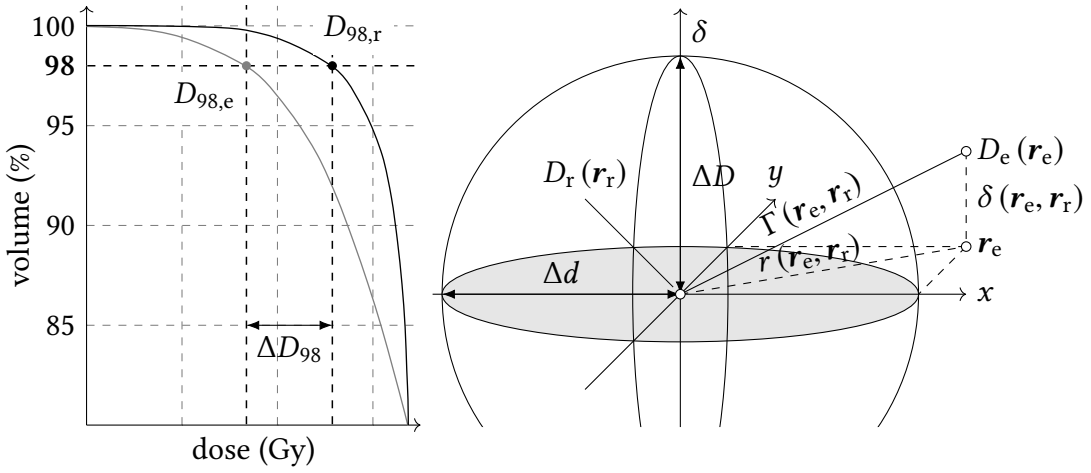
4.2.2. Evaluation of dose deviations

The evaluation of dose deviations between a reference and a dose distribution to be evaluated, D_{r} and D_{e}, respectively, can be performed by a variety of metrics. For visualization purposes, an often in this thesis applied and trivial approach is to simply compute and show a dose difference distribution, i. e. $\Delta D = D_{\mathrm{e}} - D_{\mathrm{r}}$. In areas with small dose gradients, this evaluation strategy is reasonable; for large dose gradients, however, high dose differences can occur that are not necessarily relevant. Here, the usage of dose volume/area histogram illustrations for specific structures of interest is preferable, as the spatial distribution of dose values is not considered. That is, the dose difference for a specific amount of volume/area, e. g. 98%, can be determined. This metric is mainly applied in this thesis. The general concept is visualized in Fig. 4.5 (a). Another method to evaluate the similarity between dose distributions is to compute the spatial distance between a reference $\boldsymbol{r}_{\mathrm{r}}$ and the nearest point in the dose distribution to be evaluated $\boldsymbol{r}_{\mathrm{e}}$ with exactly the same dose value. Here, however, areas with small dose gradients can be problematic. A combination of both, dose difference and spatial distance, is more beneficial. This concept is established by the Gamma index evaluation method (cf. Fig. 4.5 (b)) [79] and applied in this work. Criteria for a maximum allowed dose difference ΔD and the spatial distance Δd are user selected (chosen criteria are mainly optimized on the given analysis and defined by radiotherapy guidelines [80]) and used to define an ellipsoid around $\boldsymbol{r}_{\mathrm{r}}$ given by

$$1 = \sqrt{\frac{r^2\left(\boldsymbol{r}, \boldsymbol{r}_{\mathrm{r}}\right)}{\Delta d^2} + \frac{\delta^2\left(\boldsymbol{r}, \boldsymbol{r}_{\mathrm{r}}\right)}{\Delta D^2}}, \tag{4.22}$$

where $\delta\left(\boldsymbol{r}, \boldsymbol{r}_{\mathrm{r}}\right)$ and $r\left(\boldsymbol{r}, \boldsymbol{r}_{\mathrm{r}}\right)$ are the dose difference and spatial distance between reference and an arbitrary position, respectively. The Gamma criterion can now be defined using the right side of Eq. (4.22)

$$\gamma\left(\boldsymbol{r}_{\mathrm{r}}\right) = \min\left\{\Gamma\left(\boldsymbol{r}_{\mathrm{e}}, \boldsymbol{r}_{\mathrm{r}}\right)\right\} \forall \left\{\boldsymbol{r}_{\mathrm{e}}\right\} \tag{4.23}$$



(a) (b)

Figure 4.5.: Dose difference evaluation metrics. (a) Visualization of ΔD_{98} for exemplary dose volume histograms. The dose in Gray applied to 98% of the volume of interest is evaluated for the reference dose distribution ($D_{98,r}$) as well as for the dose to be evaluated ($D_{98,e}$). Subtracting $D_{98,e}$ from $D_{98,r}$ yields ΔD_{98}. (b) Geometrical illustration of the Gamma index. The combination of dose difference ΔD and spatial distance Δd forms an ellipsoid around the reference position $D_r(\boldsymbol{r}_r)$, where the surface of the ellipsoid defines the acceptance criterion. $\delta(\boldsymbol{r}_e, \boldsymbol{r}_r)$ and $r(\boldsymbol{r}_e, \boldsymbol{r}_r)$ are the dose difference and spatial distance between reference and the value to be evaluated, respectively. If the Euclidean distance $\Gamma(\boldsymbol{r}_e, \boldsymbol{r}_r)$ is with respect to chosen values for ΔD and Δd less or equal than 1, the Gamma evaluation passes.

where

$$\Gamma(\boldsymbol{r}_e, \boldsymbol{r}_r) = \sqrt{\frac{r^2(\boldsymbol{r}_e, \boldsymbol{r}_r)}{\Delta d^2} + \frac{\delta^2(\boldsymbol{r}_e, \boldsymbol{r}_r)}{\Delta D^2}}. \tag{4.24}$$

Here, $\delta(\boldsymbol{r}_e, \boldsymbol{r}_r)$ and $r(\boldsymbol{r}_e, \boldsymbol{r}_r)$ are the dose difference and spatial distance between the reference and the value to be evaluated, respectively. If the resulting value for $\gamma(\boldsymbol{r}_r)$ is less or equal than 1, the computation passes, otherwise not. After evaluation of every value, a total γ-passing rate, i. e. how many percent of all dose points pass the evaluation, can be calculated. Here, higher total γ-passing rates mean higher similarity between dose distributions. Typical values for quality assurance in radiotherapy for ΔD and Δd criteria are 3% of the dose of the reference point and 3 mm spatial distance, respectively.

4.3. Deep learning-based image registration

The traditional approach of DIR algorithms, as described in Section 3.2.1, is to find a transformation that maximizes the similarity of reference and transformed template image. Strategies for estimation of an optimal transformation are in general based on iterative optimization schemes, hence standard DIR is typically time consuming and a risk of getting stuck in local minima exists. The computation time further directly depends on the desired registration accuracy, i. e. user defined stopping conditions and parameter

settings, and can therefore also be relatively fast. Typical computation times (using default registration parameter settings) for frameworks applied in this thesis are approximately 15 min for the registration of two images. Time-critical applications like intra-operative registration for e. g. patient setup control during radiotherapy treatment, is due to the long run-times of current DIR frameworks not feasible; a response time of a few seconds is desirable. Thus, there exist a strong effort in related research community to reduce the computational time of DIR by improving registration techniques. The application of deep learning, allowing an efficient and general CNN-based implementation of DIR, is an option to accelerate the registration task [75].

Again and similar to previous sections, assume a patient's 4D CT series $(I_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$ with corresponding reference image I_{i_0} given. The computation of motion vector fields $(\varphi_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$ between reference image $I_{i_0} \equiv I_{\text{R}}$ and all remaining phase images of the 4D CT data set as template images $I_i \equiv I_{\text{T}}$ is in a traditional DIR setting, as formalized in Eq. (3.4), achieved by

$$\varphi_i = \arg \min_{\varphi_i^* \in C^2[\Omega]} \mathcal{J}\left[I_{\text{R}}, I_{\text{T}}; \varphi_i^*\right]. \tag{4.25}$$

Implementation of the traditional DIR approach into a standard CNN-based DIR scheme is not directly possible, as supervised training of a neural network requires some sort of training tuples as input (cf. Section 3.2.2). In the field of CNN-based DIR of thoracic and abdominal 4D CT data, a sufficiently large number of patient data sets applicable for network training have to be present to allow for a general deployment of the model after training. Assuming a database consisting of real patient 4D CT data sets, i. e. n_{pat} 4D CT data sets, each set can be denoted as training tuple $\left(I_i^p, I_j^p, \varphi_{ij}^p\right)$, $i, j \in \{1, \dots, n_{\text{ph}}\}$, $p \in \{1, \dots, n_{\text{pat}}\}$. The transformation $\varphi_{ij}^p = id + u_{ij}^p$ is yielded by deformable registration of the phase images I_i and I_j of patient p. The applied network should then learn the correlation between the input images $\left(I_i^p, I_j^p\right)$ and the vector field u_{ij}^p during network training. However, the number of voxels in a typical 4D CT phase image is about $512 \cdot 512 \cdot 159 \approx 41.7 \cdot 10^6$, i. e. in combination with the corresponding vector field it is currently computationally not feasible to directly feed the described training tuple into a CNN or GPU memory [81]. Nevertheless, a standard approach to allow for network training is to utilize only parts of the input image. For instance, patch-based approaches that divide the input into smaller 3-dimensional volumes, are commonly applied in CNN-based classification and segmentation tasks. However, in Fig. 4.6, the disadvantage of this method for estimating a vector field is visualized. Image intensity information in patches can be sparse, especially when extracted from the lung. A correct matching between patches for registration is consequently hardly possible and the predicted motion vector field lacks accuracy. Therefore, a slab-based approach is proposed, exploiting the main

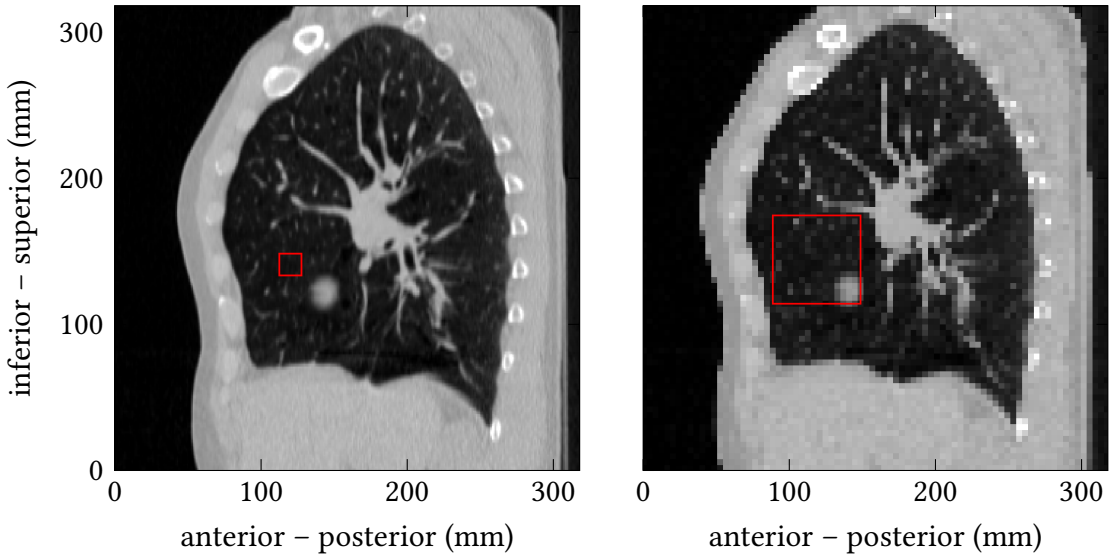


Figure 4.6.: Example of missing image intensity and structure information in patches extracted from a lung CT. Left: CT image slice with pixel spacing of $1 \times 1\,\mathrm{mm}^2$. The patch (red square) with size $15 \times 15\,\mathrm{pixel}^2$ (i. e. $15 \times 15\,\mathrm{mm}^2$) contains almost no structure information. Right: The same image as on the left but resampled to a pixel spacing of $4 \times 4\,\mathrm{mm}^2$. Even for larger patches (same patch center point as on the left) with size $15 \times 15\,\mathrm{pixel}^2$ (i. e. $60 \times 60\,\mathrm{mm}^2$) structure information is not sufficiently present for CNN-based DIR.

internal patient motion direction in 4D CT image data sets, i. e. SI and AP direction, as well as to some extent anatomical context in the lateral plane. More specifically, a reformulation of the above mentioned training tuple to a slab-based training sample is introduced. Expressing the sagittal slice of I at x-position $\hat{x}$ by the restriction of image I to $\Omega_{\hat{x}} = \{(x, y, z) \in \Omega \mid x = \hat{x}\}$ yields $I|_{\hat{x}} := I|_{\Omega_{\hat{x}}}$. An image slab can then be seen as multiple neighbouring sagittal slices adjacent to sagittal slice $I|_{\hat{x}}$. This means, the restriction of I to $\Omega_{[\hat{x}_1,\hat{x}_2]} = \{(x, y, z) \in \Omega \mid \hat{x}_1 \leq x \leq \hat{x}_2\}$, denoted as $I|_{[\hat{x}_1,\hat{x}_2]}$, is an image slab consisting of sagittal slices $\hat{x}_1, \ldots, \hat{x}_2$ of I. The reformulation of the initially mentioned training tuple reads then

$$\left(I_i^p|_{[x-2,x+2]}, I_j^p|_{[x-2,x+2]}, u_{ij}^p|_x\right), \quad x \in \{x_{\min}, \ldots, x_{\max}\}, \tag{4.26}$$

where the whole set of sagittal slices of I are covered and two adjacent slices in $\pm x$-direction around $I|_{\hat{x}}$ are employed to provide contextual information.

Before the actual network training using the proposed slab-based approach, image intensities were rescaled to $[0, 1]$ to enable a faster and better network convergence. In addition, the resolution was resampled to an isotropic resolution of 2 mm to reduce the general image size. Cropping or zero-padding was used to achieve identical image dimensions and thus simplify image processing. Further, application of Otsu thresholding [82] allowed to compute individual patient background masks that were applied to mask,

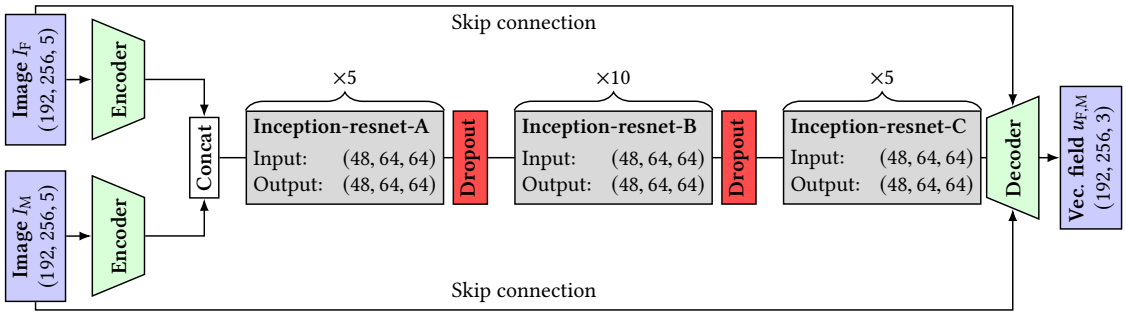


Figure 4.7.: Applied CNN architecture for deep learning-based DIR. Moving and fixed input slabs are encoded and subsequently fed into the modified Inception-ResNet-v2. Skip connections from the input to the decoder part of the network are used in order to facilitate information flow, i. e. reconstructing image contours. From publication [83].

i. e. set to zero, the non-patient background intensity values. Unintended suppression of small displacements during CNN training is avoided by a voxel-level z-transformation of x-, y- and z-displacement components, i. e. the network learns normalized 3D-vectors for the individual voxels of sagittal slices. After pre-processing, the training tuples for n_{pat} read $(\tilde{I}_i^p|_{[x-2,x+2]}, \tilde{I}_j^p|_{[x-2,x+2]}, \tilde{u}_{ij}^p|_x)$. Before training, all slabs were shuffled to eliminate unintended bias between slabs of same patient 4D CT data sets. Training and testing of the network (85%/15% train/test split, respectively) was conducted using 69 in-house acquired abdominal and thoracic 4D CT data sets. A more detailed description of the patient data is given in Section 4.4.4. Corresponding pseudo ground truth[4] motion vector fields are computed by three open source DIR frameworks utilizing default parameter settings and are applied in a plug-and-play manner. More specifically, three different versions of the network corresponding to each DIR algorithm were trained. The accuracy of each variant was evaluated by the target registration error (TRE), computed by means of the landmarks publicly available for the DIRLAB (cf. Section 4.4.1) and CREATIS Section 4.4.2) data.

The chosen network architecture, as illustrated in Fig. 4.7, was a modified Inception-ResNet-v2 [84], which was found to be more robust for the given registration task compared to a standard U-Net [85] architecture. To allow for a deeper network structure the encoder part of a pre-trained CT autoencoder was used to reduce the image slab dimensions by 60%. An autoencoder is a neural network that learns efficient data representations (encoding). That is, an autoencoder is trained by attempting to output its input. Smaller hidden layers, i. e. layers with less parameters than the input/output layer, force the autoencoder to use dimensional reduction to eliminate noise and reconstruct the inputs. Using such an encoding for CT data is reasonable as air surrounding the patient

[4]Ground truth information for DIR of clinical data sets is not existing, i. e. estimated motion vector fields are approximate and rely on visible features and applied registration framework/approach. Thus, the term pseudo ground truth is used.

or air inside the lung can be efficiently encoded. Probabilistic dropout, as introduced in Section 3.2.2, in deeper network layers leads to an intrinsic DIR representation. At the same time, overfitting during training is prevented. Applying the trained network iteratively, i. e. the trained network is cascaded, results in an improved coverage of large motion patterns. Repeating the motion prediction with enabled dropout during testing allowed to compute the sought motion field as the mean of the sampled predicted fields. Additionally, corresponding voxel-wise variances can be interpreted as local registration uncertainty estimates. The applied loss function was the MSE loss, cf. Eq. (3.15). ReLu, as introduced in Eq. (3.10), was used as activation function.

4.4. Data sets

In this work, 4D data sets of four different facilities were used. The open source 4D CT lung data sets described in Section 4.4.1 and Section 4.4.2, DIRLAB and CREATIS, respectively, were included as they contain landmarks (prominent anatomical points like vasculature bifurcations, manually selected by medical experts) that are necessary to allow for a quantitative evaluation of registration results. Further, an open source 4D MRI liver data set, as described in Section 4.4.3, was utilized to enable a similar analyses in low contrast areas. Finally, in-house acquired lung and liver 4D CT data sets (cf. Section 4.4.4) were used for 4D dose accumulation and model training for CNN-based DIR. Details of primarily employed 4D CT data sets are given in Table 4.1 and Table 4.2.

4.4.1. DIRLAB 4D CT data

The open source DIRLAB repository consists of ten 4D CT lung data sets and is provided by the University of Texas M.D. Anderson Cancer Centers (Houston, USA); cf. `https://www.dir-lab.com`. Each data set is comprised of 10 3D CT phase images, defining the internal motion during CT imaging for one average breathing cycle. Further, a variety of corresponding anatomical landmark positions inside the lung for the first five phase images of each 4D CT were defined and provided by a thoracic imaging expert. That is, the DIRLAB data set can not only be employed to quantify DIR results (comparison of landmark positions in reference and deformed template image by application of predicted transformation) but also to analyze the accuracy of estimated motion fields by e. g. correspondence model based motion prediction (cf. Section 4.1). For further details about the 4D CT data of the DIRLAB data set see Table 4.1 and [86].

Table 4.1.: 4D CT lung data sets of open data repositories DIRLAB [86] and CREATIS [87]. The displacement is defined as mean Euclidean distance of all available landmarks (n_{LM}) of each individual 4D CT data set between end inspiration and end expiration.

Data		Pat.	Size (voxel)	Spacing (mm^3)	n_{LM}	Displacement (mm)
DIRLAB	lung	01	$256 \times 256 \times 94$	$0.97 \times 0.97 \times 2.5$	300	4.01 ± 2.91
		02	$256 \times 256 \times 112$	$1.16 \times 1.16 \times 2.5$	300	4.65 ± 4.09
		03	$256 \times 256 \times 104$	$1.15 \times 1.15 \times 2.5$	300	6.73 ± 4.21
		04	$256 \times 256 \times 99$	$1.13 \times 1.13 \times 2.5$	300	9.42 ± 4.81
		05	$256 \times 256 \times 106$	$1.10 \times 1.10 \times 2.5$	300	7.10 ± 5.14
		06	$512 \times 512 \times 128$	$0.97 \times 0.97 \times 2.5$	300	11.10 ± 6.98
		07	$512 \times 512 \times 136$	$0.97 \times 0.97 \times 2.5$	300	11.59 ± 7.87
		08	$512 \times 512 \times 128$	$0.97 \times 0.97 \times 2.5$	300	15.16 ± 9.11
		09	$512 \times 512 \times 128$	$0.97 \times 0.97 \times 2.5$	300	7.82 ± 3.99
		10	$512 \times 512 \times 120$	$0.97 \times 0.97 \times 2.5$	300	7.63 ± 6.54
CREATIS	lung	01	$512 \times 512 \times 141$	$0.97 \times 0.97 \times 2.0$	100	6.34 ± 2.94
		02	$512 \times 512 \times 169$	$0.97 \times 0.97 \times 2.0$	100	14.00 ± 7.17
		03	$512 \times 512 \times 170$	$0.88 \times 0.88 \times 2.0$	100	7.67 ± 5.03
		04	$512 \times 512 \times 187$	$0.78 \times 0.78 \times 2.0$	100	7.33 ± 4.86
		05	$512 \times 512 \times 139$	$1.17 \times 1.17 \times 2.0$	100	7.09 ± 5.08
		06	$512 \times 512 \times 161$	$1.17 \times 1.17 \times 2.0$	100	6.68 ± 3.67

4.4.2. CREATIS 4D CT data

Similar to the DIRLAB data set, the CREATIS data is an open source repository that provides 4D CT data as well as corresponding anatomical landmarks in the inhale and exhale frame. The data is made available by the Léon Bérard Cancer Center & CREATIS Laboratory of the University of Lyon (France); cf. `https://www.creatis.insa-lyon.fr/rio/popi-model`. Here, six lung 4D CT patient data sets are available. Further details are described in Table 4.1 and [87].

4.4.3. 4D MRI

Landmarks to quantify image registration accuracy in lung CT data, as provided in the DIRLAB and CREATIS repositories, are generally not available in the low-to-no contrast CT areas of tissue, especially the liver. MR images, however, contain high contrast in tissue and liver but have usually no temporal resolution, i. e. are only available in 3D[5]. Fortunately, one open source 10-phase 4D MRI data set is provided at `https://www.vision.ee.ethz.ch/~organmot/` and described in [89]. To allow for a similar evaluation of registration accuracy for liver registration compared to the DIRLAB and

[5]Currently there is no 4D-MRI technique available for clinical application. The present generation of scanners is constrained by the limited frequency at which full 3D volumes can be acquired [88].

CREATIS data, an adaption of the data set had to be implemented. Inside the liver of the 4D MRI data set that represents one breathing cycle

$$\left(I_i^{\mathrm{MRI}}\right)_{i\in\{1,\ldots,10\}}, I^{\mathrm{MRI}} : \Omega^{\mathrm{MRI}} \subset \mathbb{R}^3 \rightarrow \mathbb{R}, \tag{4.27}$$

with clearly visible inner-liver structures, 20 corresponding landmarks (i. e. prominent anatomical points like vasculature bifurcations) were selected by an expert in each individual MRI phase image. By reducing the structure-to-background contrast inside the liver a series of masked 4D MRI sets were generated to mimic the real challenge of liver registration in 4D CT images. Application of manually segmented liver masks, denoted by

$$\left(M_i^{\mathrm{MRI}}\right)_{i\in\{1,\ldots,10\}}, M^{\mathrm{MRI}} : \Omega^{\mathrm{MRI}} \rightarrow \mathbb{R} \tag{4.28}$$

with the median value $M_i^{\mathrm{MRI}}(x)$ of liver voxels in I_i^{MRI} and zero elsewhere, allowed to combine mask and original MRI data

$$\left(I_i^{\mathrm{MRI},\alpha}\right)_{i\in\{1,\ldots,10\}}, M^{\mathrm{MRI},\alpha} : \Omega^{\mathrm{MRI}} \rightarrow \mathbb{R} \tag{4.29}$$

with

$$I_i^{\mathrm{MRI},\alpha}(x) = \begin{cases} (1-\alpha) I_i^{\mathrm{MRI}}(x) + \alpha \left[M_i^{\mathrm{MRI}}(x) + n(x)\right] & \text{if } M_i^{\mathrm{MRI}}(x) \neq 0 \\ I_i^{\mathrm{MRI}}(x) & \text{else} \end{cases} \tag{4.30}$$

and form test data sets. Here, $n(x)$ denotes Gaussian noise with an expectation value of zero and a variance based on common liver voxel intensity distributions in 4D CT data re-scaled to MRI data intensity ranges. Thus, a CT-like noise distribution inside the liver of the $I_i^{\mathrm{MRI},\alpha}$ data is generated. Variation of α in Eq. (4.30) led to MRI data sets with different noise levels that were subsequently used for evaluation of image registration accuracy by application of ground truth landmarks. In Fig. 4.8 the MRI-liver is exemplary shown for $\alpha = 0\%$ and $\alpha = 90\%$ and compared to a CT-liver.

4.4.4. In-house acquired 4D CT data

In total, 69 in-house acquired liver and lung 4D CT data sets were utilized in this thesis. Patient data sets that contained planning 4D CT, breathing signal data for all treatment fractions, VMAT treatment plan, planned dose distribution and information about clinical outcome (local metastasis recurrence: yes/no), were selected for the dose accumulation studies (see Section 4.2); the eventually chosen five lung and five liver patients (cf. Table 4.2) had six and nine treated metastases, respectively, with two cases of local

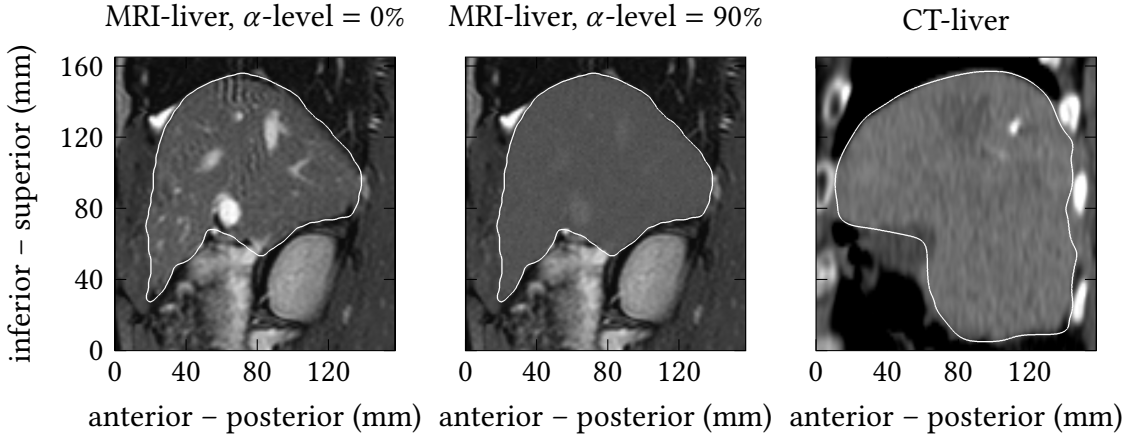


Figure 4.8.: MRI-liver with different α-level compared to a CT-liver. Left: Visibility of liver structures in the original MRI data. Middle: MRI-liver with α-level of 90%. Right: CT phase image of a 4D CT image sequence for comparison. Structures inside the liver are hardly detectable for the MRI data with $\alpha = 90\%$ and the CT data. From publication [90].

Table 4.2.: In-house acquired 4D CT data sets of liver and lung cancer patients. Parameters ⟨Span⟩ and ⟨Period⟩ denote the mean peak-to peak amplitude and the mean period of the respiratory signal recorded during 4D CT acquisition, respectively.

Data		Pat.	Size (voxel)	Spacing (mm^3)	⟨Span⟩ (mm)	⟨Period⟩ (s)
In-house	liver	01	$512 \times 512 \times 117$	$0.98 \times 0.98 \times 2.0$	5.5 ± 2.7	5.3 ± 1.8
		02	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	8.3 ± 1.6	5.2 ± 0.5
		03	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	15.4 ± 1.4	2.9 ± 0.3
		04	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	15.3 ± 1.9	6.6 ± 1.1
		05	$512 \times 512 \times 150$	$0.98 \times 0.98 \times 2.0$	4.6 ± 0.5	3.6 ± 0.5
	lung	06	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	6.5 ± 3.5	5.5 ± 1.8
		07	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	8.4 ± 0.8	5.7 ± 0.4
		08	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	14.7 ± 4.0	5.7 ± 0.9
		09	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	7.0 ± 1.1	2.4 ± 0.2
		10	$512 \times 512 \times 159$	$0.98 \times 0.98 \times 2.0$	6.7 ± 2.4	3.3 ± 0.8

metastasis recurrence each. For the development of the deep learning based image registration framework (cf. Section 4.3), all 69 patient cases were used. Here, the patient 4D CT data was utilized to train and test the proposed neural network.

All 4D CT are recorded by the same scanner (Siemens Somatom AS Open CT) in standard spiral 4D CT scanning mode in a period of time of about 18 months (March 2014 to September 2015). 4D CT image reconstruction of acquired projection data was done by amplitude-based binning, as described in Section 3.1.1.

Chapter 5

Results

Chapters 2 to 4 outlined the physical principles and medical imaging-related fundamentals in 4D radiotherapy relevant for this thesis. In this chapter, the results obtained by applying the introduced materials and methods (cf. Chapter 4) are presented. As the results have been published in peer-reviewed journals in respective fields of research, each section (except Section 5.7) is based on one publication and a preceding summary of the article. In Section 5.7, additional and not yet published results are presented. A brief overview of the corresponding sections is given in the following.

Section 5.1: Correspondence model-based 4D dose simulation [PMB 2017]
In this section, the focus is on illustrating the combination of internal patient motion prediction and dose accumulation to allow for implementation of a 4D dose accumulation algorithm for complex radiotherapy treatments.

Section 5.2: Image registration in 4D dose simulation [R&O 2018]
Correspondence modeling as utilized in the previous section requires the extraction of internal patient motion out of 4D CT data sets using DIR (cf. Section 4.1). The assumption that the applied registration algorithm and respective uncertainties therein highly impacts the dose accumulation output is investigated in this section.

Section 5.3: Dose simulation in the presence of image artifacts [SPIE 2018]
4D CT planning data is the basis of all following dose accumulation steps. Often occurring image artifacts in 4D CT data sets are assumed to be an uncertainty source as they might impact the registration and, as a result, the dose accumulation accuracy. This is analyzed in this section.

Section 5.4: Deep learning-based deformable image registration [MICCAI 2018]
A new registration tool based on deep learning is developed and introduced. The framework allows to estimate the DIR uncertainty, which opens up the possibility to compute voxel-wise dose confidence intervals.

Section 5.5: Accuracy of 4D dose simulation [PLOS 2017]
The validation and verification of the proposed 4D dose accumulation approach by comparing dose simulations and dose measurements is presented. Here, affecting parameters and limitations of the dose simulation scheme are demonstrated.

Section 5.6: Monte Carlo-based 4D dose simulation [SPIE 2019]

To achieve best possible dose accumulation accuracy, i. e. further reduce deviations between measurements and simulations, the accumulation scheme is enhanced by re-implementing the algorithm to allow for actual dose re-calculation using Monte Carlo simulations.

Section 5.7: Error propagation in 4D dose simulation [Not yet published]

The estimation of dose confidence intervals by registration uncertainty propagation is performed.

5.1. Correspondence model-based 4D dose simulation

Dose accumulation algorithms are important tools for quality assurance in 4D radiotherapy, as mentioned in Chapter 1. Current approaches, i. e. standard 4D CT-only dose accumulation schemes, are, however, not able to account for individual internal patient motion during dose delivery. Thus, interplay effects due to unfavourable interaction of dynamic and complex dose delivery techniques and respiration-induced tumor motion can not be quantified and further evaluated. In this section, a first approach to allow for an accurate estimation of such effects is proposed. Further, predicted underdosages were correlated to corresponding clinical outcome information (i. e. was the treatment successful). The results are based on the following publication:

T. Sothmann, T. Gauer, M. Wilms, R. Werner. Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT. *Phys Med Biol*, 62(23):9001–9017, 2017.

In the respective study, correspondence modeling, i. e. the correlation of external and internal patient motion information as described in Section 4.1, was utilized. The representation of intrafractional patient motion variability by breathing signal-steered interpolation and extrapolating of the DIR motion fields was combined with 4D dose simulation (cf. Section 4.2 for the general simulation approach). Further, the proposed correspondence model-based dose accumulation method was applied to clinical patient data to 1) retrospectively simulate the delivered dose for a population of 10 real patient treatments and 2) correlate computed underdosages (dose differences between planned and 4D-simulated doses) to known clinical endpoints, i. e. occurrence of local metastasis recurrence or successful treatment (no occurrence). For each patient in the investigated patient cohort (see Table 1 [PMB 2017][1] and Table 4.2 for additional patient treatment,

[1] For readability/identification purposes, references to tables, figures and equations of included publications comprise information about corresponding journal and year of publication in square brackets.

motion and data set information), an individual correspondence model was built using the corresponding planning 4D CT and the simultaneously acquired surrogate breathing curve. Here, DIR between one reference 4D CT phase and all remaining phases yielded the necessary motion vector fields that were correlated by multivariate regression to the external breathing signal and its temporal derivative, forming the basis of a regression based correspondence model training. During dose delivery, the external patient breathing signal has been recorded by the same technical device used during 4D CT imaging. Discretization of the dose delivery process allowed correlating a specific breathing signal measurement to the corresponding planned dose segment. Applying the beforehand built correspondence model, an internal motion field that corresponds to each external signal measurement was derived. Deforming all dose segments with the estimated motion fields and weighted summation over all deformed dose segments yielded the sought motion-affected dose distribution for one treatment fraction (summation over all fractions yielded the total motion-affected dose distribution). Further, to illustrate advantages regarding dose estimation accuracy, a standard 4D CT-only dose accumulation scheme was implemented and compared to the proposed approach.

Results of the proposed and the standard 4D dose accumulation approach are summarized in Table 2 [PMB 2017]. Target, i. e. GTV, dose coverage was quantified by differences ΔD_{98} of $D_{98,\mathrm{Sim}}$ and $D_{98,\mathrm{Plan}}$, whereas D_{98} is the dose that is received by 98% of the target volume. It was noticeable that for patients with local metastasis recurrence the proposed 4D dose simulation approach estimated larger negative total and fraction-wise ΔD_{98} values compared to metastases without recurrence. Respective results for the standard 4D CT-based simulation scheme showed a potential underestimation of motion effects, i. e. only small ΔD_{98} values. This was somehow expected as information about motion variability was in this approach not taken into account. Further, a fraction-wise investigation of dose deviations was not possible for the standard approach. In conclusion, the introduced correspondence model-based 4D dose accumulation method illustrates the possibility of linking estimated underdosages to local metastasis recurrence, and thus offers the potential to explain motion-related errors during treatment planning and dose delivery. Implementation into the clinical workflow of this tool for quality assurance purposes after treatment planning and/or after the individual treatment session is desirably. A possible application would be a treatment plan adaption after the first treatment session if respiration-induced underdosages of target volumes are identified. However, feasibility, dose simulation accuracy and uncertainties of the proposed scheme have to be further investigated before a clinical application can be considered.

In terms of shortcomings apart from the small patient cohort size that was available for the current feasibility study, it has to be noted that the correspondence modeling used

for 4D dose simulation was based on only pre-treatment 4D CT image data. Further, the extraction of internal patient motion out of the available 4D CT data sets was performed using an open source non-linear registration framework, optimized for 4D CT lung registration. Performance in low-to-no contrast areas, e. g. the liver, remains to be evaluated. This directly motivated the investigation of the influence of different DIR algorithms on the 4D dose simulation accuracy, as extensively analyzed in Section 5.2. Also, often occurring 4D CT image artifacts, as described in Section 3.1.2, are uncertainties that explicitly influence registration and with that dose simulation accuracy, i. e. the impact of typical 4D CT artifacts had to be and was therefore analyzed in Section 5.3. The limited VMAT arc discretization level, as introduced by the treatment planning system, restricts the temporal resolution, and was consequently evaluated in Section 5.5, including verification measurements. Nevertheless and despite those uncertainties, the here presented results highlight the potential of correspondence model-based 4D VMAT dose simulation.

Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT

T Sothmann[1,2], T Gauer[2], M Wilms[3] and R Werner[1]

[1] Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2] Department of Radiotherapy and Radio-Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3] Institute of Medical Informatics, University of Lübeck, Lübeck, Germany

E-mail: t.sothmann@uke.de

Abstract. The purpose of this study is to introduce a novel approach to incorporate patient-specific breathing variability information into 4D dose simulation of volumetric arc therapy (VMAT)-based stereotactic body radiotherapy (SBRT) of extracranial metastases. Feasibility of the approach is illustrated by application to treatment planning and motion data of lung and liver metastasis patients.

The novel 4D dose simulation approach makes use of a regression-based correspondence model that allows representing patient motion variability by breathing signal-steered interpolation and extrapolation of deformable image registration motion fields. To predict the internal patient motion during treatment with only external breathing signal measurements being available, the patients' internal motion information and external breathing signals acquired during 4D CT imaging were correlated. Combining the correspondence model, patient-specific breathing signal measurements during treatment and time-resolved information about dose delivery, reconstruction of a motion variability-affected dose becomes possible.

As a proof of concept, the proposed approach is illustrated by retrospective 4D simulation of VMAT-based SBRT treatment of ten patients with 15 treated lung and liver metastases and known clinical endpoints for the individual metastases (local metastasis recurrence yes/no). Resulting 4D-simulated dose distributions were compared to motion-affected dose distributions estimated by standard 4D CT-only dose accumulation and the originally (i. e. statically) planned dose distributions by means of GTV D_{98} indices (dose to 98% of the GTV volume). A potential linkage of metastasis-specific endpoints to differences between GTV D_{98} indices of planned and 4D-simulated dose distributions was analyzed.

Keywords: Respiratory Motion, 4D Dose Simulation, Dose Accumulation, Correspondence Model

1. Introduction

Extracranial radiotherapy of metastatic disease is nowadays commonly treated by stereotactic body radiotherapy (SBRT), exploiting highly conformal treatment planning to achieve equivalent biologically effective doses (BED) in only a few fractions compared to former standard 3D conformal RT fractionation (hypofractionation with 1-10 fractions vs. 20-40 fractions) (Lartigau 2011, Sahgal *et al* 2012, Katz *et al* 2007). Despite reported local control (LC) rates of 55% to 100% at 2-3 years (Timmerman *et al* 2010, Alongi *et al* 2012, Rieber *et al* 2016, Høyer *et al* 2012, Chang *et al* 2011, Andratschke *et al* 2016), especially SBRT of lung and liver metastases is still considered challenging due to e. g. respiration-induced target volume deformation and tumor motion amplitudes up to several centimeters (Keall *et al* 2006). Further, intra- and inter-fractional breathing variability as well as the use of only a single planning (4D) computed tomography (CT) image without treatment plan adaption before the individual treatment fractions (Korreman 2012, Ge *et al* 2013) leads to uncertainties and differences between planned and actually delivered SBRT dose distributions. The assumption underlying our study is that a better understanding of the uncertainties and differences could potentially allow us to explain (a fraction of) the up to 40% missing LC rates.

Understanding motion-induced differences between planned and delivered dose over the course of treatment requires combining patient-specific motion data and information about the dose delivery process. Related tools are referred to as 4D dose calculation, simulation or reconstruction. Generally speaking, 4D dose simulation can be understood as a weighted summation of dose distributions that correspond to different (breathing) states of the time-dependent patient geometry (Milz *et al* 2014). In the most common form, the 3D CT phase images of the patient's planning 4D CT are considered representative for the patient geometry during treatment. Deformable image registration (DIR) is applied between a reference breathing phase image and the other 4D CT phase images. The resulting motion fields are directly applied to deform the originally planned 3D dose distribution. The deformed dose distributions represent the dose corresponding to the different breathing phases, but mapped into the patient geometry of the reference image. They are finally combined by weighted summation (Sarrut 2006, Werner *et al* 2012, Velec *et al* 2011).

Deforming the entire 3D dose distribution as computed by the treatment planning system, however, neglects the dynamics of dose delivery. It has therefore also been suggested to assign individual dose segments of the treatment plan (for intensity-modulated radiotherapy) or even the individual monitor units to the phases of the planning 4D CT (Ehrbar *et al* 2016, Werner *et al* 2012). This allows studying so-called interplay effects between respiratory and dose delivery dynamics, but still ignores intra- and inter-fractional respiratory variability. Especially addressing inter-fraction motion differences, some groups further proposed taking into account 4D cone beam CT (CBCT) data acquired prior to the individual treatment: For each fraction, the inhale CT of the planning 4D CT was registered to the inhale image of the pre-fraction 4D CBCT to account for, e. g., inter-fraction setup variations, followed by registration of the inhale 4D CBCT phase to other 4D CBCT phase images to represent fraction-specific motion patterns (Velec *et al* 2012, Samavati *et al* 2016). However, the resulting motion fields were again directly applied to warp the planned dose distribution. Information about the patient's actual breathing patterns during dose delivery was not taken into account; intra-fraction motion variability and

interplay thereof with dose delivery dynamics were still neglected.

On the contrary, correspondence modeling, i. e. to correlate the patient's internal motion information and external breathing signals acquired during 4D imaging, has attracted great interest over the past years. Related studies primarily aimed at estimating internal patient motion during treatment, with only external breathing signal measurements being available (Wilms *et al* 2014, Fassi *et al* 2014, McClelland *et al* 2013). Thus, compared to motion representation by only a series of motion fields as obtained by DIR in a 4D image sequence, correspondence modeling allows representing intra-fractional patient motion variability by breathing signal-steered interpolation and extrapolating of the DIR motion fields with high accuracy (Wilms *et al* 2014).

From our perspective, combining correspondence modeling and 4D dose simulation appears to be a natural next step. The present study describes three main and novel contributions in that direction:

- We present a first correspondence model-based approach to 4D-simulate the delivered dose of SBRT treatments. The proposed scheme explicitly models the interplay of intra-fraction patient motion and the dynamics of volumetric arc therapy (VMAT) dose delivery.
- The approach is applied to clinical patient data and VMAT treatment plans, using real patient-specific breathing signal measurements observed during dose delivery.
- For all patients and metastases, clinical endpoints were known (local metastasis recurrence: yes/no). Trying to identify potential patient-specific motion-related causes of local metastasis recurrence, clinical outcome information and differences between 4D-simulated and the planned gross tumor volume (GTV) dose distributions were evaluated.

The study builds on and combines established and well-evaluated methodological approaches: For correspondence modeling, we apply our linear regression approach detailed in (Wilms *et al* 2014) and our open-source DIR framework evaluated in (Werner *et al* 2014). Technical feasibility of the conducted temporal discretization of the VMAT dose delivery process was further analyzed in a motion phantom pre-study described in (Sothmann *et al* 2017). For ease of readability, related methodical aspects are nevertheless briefly introduced in the subsequent Sec. 2, before describing our materials and experiments and corresponding results (Sec. 3-4). Despite or precisely because of its novelty, our correspondence model-based 4D dose simulation approach is still prone to different sources of uncertainties. These are detailed in Sec. 5, accompanied by identification of next steps to reduce them.

2. Methods: Correspondence model-based 4D dose simulation

The concept of correspondence model-based 4D dose simulation is illustrated in Fig. 1, with the individual parts being subsequently explained. Nomenclature and used symbols are based on (Wilms *et al* 2014, Werner *et al* 2014, Sothmann *et al* 2017) and the technique for correspondence modeling upon standard approaches from the literature as reviewed in (McClelland *et al* 2013).

2.1. Correspondence modeling: correlating internal and external motion information

For each patient, the 4D treatment planning CT represents a series of 3D CT phase images, i. e.

$$(I_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ I_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$$

Here, $i \in \{1,\dots,n_{\mathrm{ph}}\}$ indicates the breathing phase of a 3D CT image and n_{ph} the total number of breathing phases of the 4D CT. Analogously, corresponding external breathing signal measurements that are assigned to the images read

$$(\zeta_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ \zeta_i \in \mathbb{R}^d$$

with the most common choice being $d = 2$ and specifically

$$(\zeta_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ \zeta_i = (z_i, \partial_t z_i)^T \in \mathbb{R}^2;$$

cf. (Low *et al* 2005, Wilms *et al* 2014). Now, given the breathing phase-specific images I_i and the breathing signal measurements ζ_i for $i \in \{1,\dots,n_{\mathrm{ph}}\}$, correspondence modeling aims at representing a patient-specific functional relationship between the breathing signal measurements and respiratory motion of the internal structures (target volumes, organs at risk). In contrast to the low-dimensional breathing signals, the internal motion is usually represented by dense vector fields, i. e. by breathing-induced voxel-wise displacement vectors with respect to a reference breathing phase / voxel position. These displacement fields are to be estimated by non-linear registration. Thus, assume an arbitrary breathing phase $i_0 \in \{1,\dots,n_{\mathrm{ph}}\}$ to be selected as reference phase (here: $i_0 = 3$, denoting mid-expiration) and the corresponding CT image I_{i_0} as fixed image during registration. Then, the registration process results in a series of transformations

$$(\varphi_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ \varphi_i : \Omega \to \Omega$$

with φ_{i_0} = identity map (id) and corresponding motion fields

$$(u_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ u_i : \Omega \to \mathbb{R}^3$$

with $u_i = \varphi_i - id$ (i. e. $u_{i_0} = (0;0;0)^T$).

In combination with the breathing signal measurements $(\zeta_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}}$, these motion fields $(u_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}}$ form the basis of a regression-based correspondence model training; see Fig. 1, top-left. For ease of readability, let therefore the breathing signal measurements and motion fields be interpreted as random variables $\mathbf{Z}_i$ ($\equiv \zeta_i$) and $\mathbf{U}_i \in \mathbb{R}^{3m}$ with m denoting the number of voxels of the reference phase image I_{i_0}.

Then, the assumed relationship between breathing signal measurements and internal motion fields – i. e. the correspondence model – is defined as a multivariate linear relationship, i. e. by

$$\hat{\mathbf{U}} = \overline{\mathbf{U}} + \mathbf{B}\left(\hat{\mathbf{Z}} - \overline{\mathbf{Z}}\right) \tag{1}$$

$\hat{\mathbf{Z}} \in \mathbb{R}^2$ represents a breathing signal observation and $\hat{\mathbf{U}} \in \mathbb{R}^{3m}$ the corresponding and sought motion field ($\overline{\mathbf{U}} = 1/n_{\mathrm{ph}} \sum_i^{n_{\mathrm{ph}}} \mathbf{U}_i$ and analogously $\overline{\mathbf{Z}} = 1/n_{\mathrm{ph}} \sum_i^{n_{\mathrm{ph}}} \mathbf{Z}_i$). The coefficient matrix $\mathbf{B} \in \mathbb{R}^{3m\times 2}$ is computed in an ordinary least-squares regression approach, i. e.

$$\mathbf{B} = \arg\min_{\mathbf{B}'} \mathrm{tr}\left[(\mathbf{U} - \mathbf{B}'\mathbf{Z})\,(\mathbf{U} - \mathbf{B}'\mathbf{Z})^T\right] = \mathbf{U}\mathbf{Z}^+ \tag{2}$$

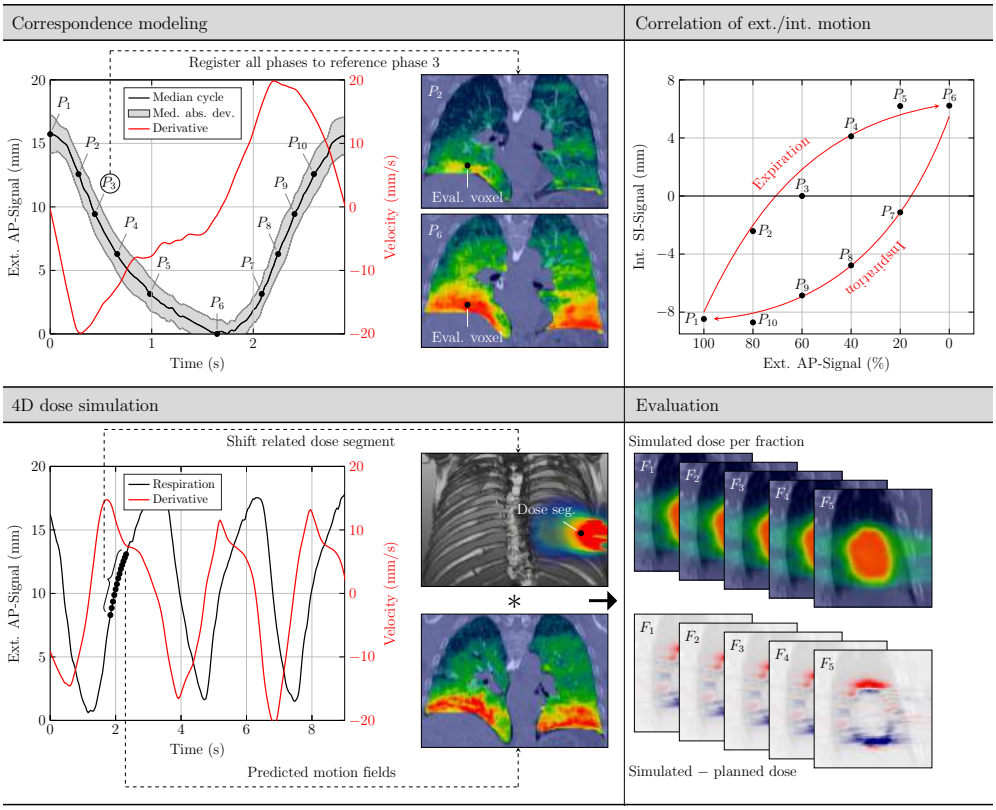


Figure 1. Concept of correspondence model-based 4D dose simulation. Top: To establish a functional relationship between external breathing signal measurements and internal motion patterns, breathing signal information (left; in our case the anterior-posterior (AP) component of the Varian Real-Time Position Management (RPM) system and its temporal derivative) corresponding to the RT planning 4D CT phase images were correlated (by multivariate regression) to motion fields obtained by DIR with respect to the mid-expiration phase P3. Top-right: relationship between motion vector superior-inferior (SI) components of the voxel highlighted in the middle column and the breathing signal values for the patient's 10-phase 4D planning CT. Differences between inspiration and expiration (hysteresis) highlight the need to incorporate additional information (here: temporal derivative of breathing signal). Bottom: For each breathing signal measurement, a motion field is computed by means of the correspondence model. This field is used to deform the temporally corresponding planned VMAT dose segment, resulting in a motion-affected dose segment. Summing up all motion-affected dose segments for, e. g., the individual treatment fractions affords the opportunity to compare the planned and the estimate of the actually delivered treatment fraction dose distributions, taking into account the observed patient-specific breathing and related irregularity information.

with

$$\mathbf{Z} = \left(\mathbf{Z}_1 - \overline{\mathbf{Z}}, \ldots, \mathbf{Z}_{n_{\mathrm{ph}}} - \overline{\mathbf{Z}}\right)$$
$$\mathbf{U} = \left(\mathbf{U}_1 - \overline{\mathbf{U}}, \ldots, \mathbf{U}_{n_{\mathrm{ph}}} - \overline{\mathbf{U}}\right)$$

as mean-centered observations and

$$\mathbf{Z}^{+} = \mathbf{Z}^{T} \left(\mathbf{Z}\mathbf{Z}^{T}\right)^{-1}$$

as Moore-Penrose pseudoinverse. Subsequently, $\hat{\mathbf{U}}$ and the motion field $\hat{u} : \Omega \times \mathbb{R}^2 \to \mathbb{R}^3$ that results by application of Eq. (1) for a given $\hat{\mathbf{Z}}$ and $\hat{\zeta}$, respectively, are used as equivalent terms.

2.2. Combining correspondence modeling and 4D VMAT dose simulation

Having established a patient-specific functional relationship between external and internal motion data, retrospective correspondence model-based 4D dose simulation is conceptually straightforward (Fig. 1, bottom):

Throughout the course of treatment, a breathing signal has to be recorded (using the same technical device exploited during model formation; otherwise, applicability of the trained correspondence model is questionable) in parallel to the dose delivery process, and the respective temporal information has to be correlated. Assuming that the planned dose delivery process can be appropriately discretized in time, the latter means that each individual breathing signal measurement can be correlated to a planned dose segment delivered at the time of measurement. In addition, the correspondence model allows deriving internal motion fields that correspond to the external signal measurement. Thus, deforming the planned dose segment by means of the motion fields yields an estimate of the dose actually delivered, taking into account the patient's state of breathing at the time of measurement.

Mathematically speaking, let $\dot{D} : \Omega \times \mathcal{T} \subset \mathbb{R} \to \mathbb{R}_{+}$ be the time-dependent dose rate during a single treatment fraction. Then, generally speaking, the dynamical dose delivery process, its interplay with patient motion and the resulting dose distribution $D_{\mathrm{4D}} : \Omega \to \mathbb{R}_{+}$ can be expressed as

$$\begin{aligned} D_{\mathrm{4D}}(x) &= \int_{\mathcal{T}} \dot{D}(\varphi(x,t),t)\, dt \\ &\approx \sum_{t \in \tilde{\mathcal{T}}} \dot{D}(\varphi(x,t),t)\, \Delta t \\ &= \sum_{t \in \tilde{\mathcal{T}}} D_t(\varphi(x,t)) \end{aligned} \tag{3}$$

with $\varphi : \Omega \times \mathcal{T} \to \mathbb{R}^3$ as the position $\varphi(x,t)$ of voxel $x \in \Omega$ of the reference phase CT I_{i_0} at time $t \in \mathcal{T} = [0;T) \subset \mathbb{R}$ (i. e. the representation of the respiratory-induced inner structure motion during dose delivery of length T). The numerically required temporal discretization in Eq. (3) indicates an equidistant sampling of the dose delivery process, with the sampling points $t \in \tilde{\mathcal{T}} = \{1/2\,\Delta t; 3/2\,\Delta t; \dots\}$, $\tilde{\mathcal{T}} \subset \mathcal{T}$, and a sampling period of Δt. Consequently, $D_t : \Omega \to \mathbb{R}_{+}$ denotes the dose applied during the time interval $[t - 1/2\,\Delta t; t + 1/2\,\Delta t)$.

As described in the introduction, for common dose accumulation approaches, the actual dynamics of the dose delivery process are ignored and the DIR-estimated motion fields directly applied for dose warping,

$$D_{\mathrm{4D\ CT\ sim}}(x) = \frac{1}{n_{\mathrm{ph}}} \sum_{i=1}^{n_{\mathrm{ph}}} D(\varphi_i(x)) \tag{4}$$

with $D : \Omega \to \mathbb{R}_+$ as the planned 3D dose distribution (Werner *et al* 2012).

Now, to not only account for dose delivery dynamics but also the during treatment-acquired patient-specific breathing signal information about breathing variability, the breathing signal and the established correspondence model from Eq. (1) are integrated into Eq. (3). With $\hat{\varphi} = id + \hat{u}$ and $\hat{\zeta} : \mathcal{T} \to \mathbb{R}^2$, correspondence model-based 4D dose simulation for a single fraction finally reads

$$D_{4\mathrm{D}}(x) \approx \sum_{t \in \tilde{\mathcal{T}}} D_t\left(\hat{\varphi}\left(x, \hat{\zeta}_t\right)\right) = \sum_{t \in \tilde{\mathcal{T}}} D_t\left(x + \hat{u}\left(x, \hat{\zeta}_t\right)\right) \tag{5}$$

where $\hat{\zeta}_t = \hat{\zeta}(t)$. To cover multiple treatment fractions, Eq. (5) directly extends to

$$D_{4\mathrm{D}}^{\mathrm{total}}(x) = \sum_{\mathrm{fx}} D_{4\mathrm{D}}^{\mathrm{fx}}(x) = \sum_{\mathrm{fx}} \sum_{t \in \tilde{\mathcal{T}}_{\mathrm{fx}}} D_t\left(x + \hat{u}\left(x, \hat{\zeta}_t\right)\right)$$

with the outer summation accumulating the individual fraction dose distributions and $\tilde{\mathcal{T}}_{\mathrm{fx}} \subset \mathbb{R}$ being the sampling points of dose delivery during the respective fraction. Note that without treatment plan adaptation it can be assumed that $\tilde{\mathcal{T}}_{\mathrm{fx}} = \tilde{\mathcal{T}}$ for all treatment fractions. Thus, in the general case, differences between the $D_{4\mathrm{D}}^{\mathrm{fx}}$ result from patient motion variability and differences of the $\hat{\zeta}_t$ observed during dose delivery.

2.2.1. Temporal discretization of VMAT dose delivery Equation (5) is valid for arbitrary dose delivery techniques. Aiming at 4D dose simulation for VMAT dose delivery, VMAT-specific characteristics have to be accounted for. As illustrated in Fig. 2, the dose is delivered during continuous gantry rotation around the patients, covering a relatively wide range of gantry angles and a small number of arcs. From the perspective of discretization of the dose delivery process, this means that without loss of generality Eq. (5) can be equivalently re-parameterized by replacing the temporal variable t for each planned arc by the gantry rotation angle $\alpha \in [0°; 360°)$:

$$D_{4\mathrm{D}}(x) \approx \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} D_\alpha\left(x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)\right) \tag{6}$$

with $\tilde{\mathcal{A}}_{\mathrm{ax}} = \{1/2\,\Delta\alpha; 3/2\,\Delta\alpha; \dots\}$ as a discretized version of the gantry angle range $\mathcal{A}_{\mathrm{ax}} \subset [0°; 360°)$ of the respective VMAT arc (Sothmann *et al* 2017). Thus, similar to the afore-mentioned explanations, $D_\alpha\left(x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)\right)$ represents the dose applied during the angle range $[\alpha - 1/2\Delta\alpha; \alpha + 1/2\Delta\alpha)]$ to voxel x of the reference CT image I_{i_0}, which, according to the respiratory signal measurement $\hat{\zeta}_\alpha$ and the correspondence model, is located at $x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)$. Further assuming that the breathing signal acquisition rate is usually very high compared to gantry rotation time and achievable degree of VMAT arc discretization, the finally implemented single fraction accumulation scheme was

$$\begin{aligned} D_{4\mathrm{D}}(x) &\approx \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} \frac{1}{|\tilde{\mathcal{T}}_\alpha|} \sum_{t \in \tilde{\mathcal{T}}_\alpha} D_\alpha\left(x + \hat{u}\left(x, \hat{\zeta}_t\right)\right) \\ &= \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} \frac{1}{|\tilde{\mathcal{T}}_\alpha|} \sum_{t \in \tilde{\mathcal{T}}_\alpha} \left(D_\alpha \circ \hat{\varphi}\right)\left(x, \hat{\zeta}_t\right) \end{aligned} \tag{7}$$

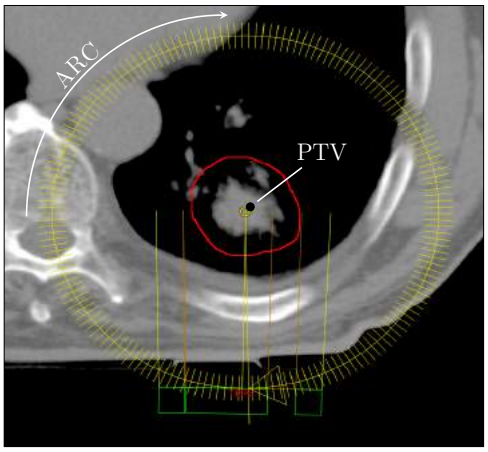


Figure 2. VMAT dose delivery. Illustration of VMAT dose delivery and temporal discretization of the delivery process: During VMAT treatment, the gantry rotates continuously around the patient, with dose rate, gantry speed and multi-leaf collimator leaf positions varying to simultaneously maximize organs at risk sparing and target dose coverage (shown: planning target volume, PTV). For 4D dose simulation, these so-called VMAT arcs were discretized into small arc segments (indicated by the yellow dashes) and the dose planned to be delivered for each segment was used for dose accumulation.

with $\tilde{\mathcal{T}}_\alpha$ denoting the breathing signal measurements during dose delivery at gantry angle α and the respective dose segment. The total 4D-simulated dose after all treatment fractions was computed by summing up the 4D-simulated dose distributions of the individual fractions.

3. Materials and Experiments

3.1. Patient cohort and treatment planning

To include a patient into this retrospective study, planning 4D CT, breathing signal data for all treatment fractions and information about clinical outcome (local metastasis recurrence: yes/no) after SBRT had to be present. Further, we aimed for a patient cohort with homogeneous BED of more than 75 Gy; lower BED was, besides inappropriate motion management, concluded to cause worse LC rates for extracranial metastasis and especially lung and liver SBRT (Van den Begin *et al* 2014).

In total, five liver patients with nine treated metastases and five lung patients with six treated metastases were selected. All patients were treated by VMAT (RapidArc, Varian, USA). 3D CBCT image-guided patient setup correction was applied before each treatment fraction. The dose was delivered in "safety gating" mode: During treatment, a breathing signal (using the Real-Time Position Management (RPM) system, Varian, USA) was acquired and the linac beam was switched off if the online signal exceeded amplitude thresholds defined beforehand by means of the 4D CT breathing signal records.

Antecedent VMAT treatment planning was based on a ten-phase 4D CT (image spatial resolution: $0.98\times0.98\times$ 2 mm; Siemens Definition AS+, Siemens, Germany; breathing signal acquisition: RPM system, Varian, USA) and an internal target

volume (ITV) concept (GTV delineation on each phase; union of phase GTVs = ITV; ITV + 4 mm isotropic safety margin = PTV; treatment planning system: Eclipse 13, Varian); segmentation of organs at risk and dose optimization (AAA/Acuros) was performed on a temporal average CT derived from the ten-phase 4D CT. Primary planning goal was to describe a minimum dose of 90% of the planned dose to the PTV while considering normal tissue constraints. Further detailed information on treatment planning as well as patient-specific information on metastasis motion and recurrence as observed in the 4D CT planning data are summarized in Table 1.

3.2. Design of our proof-of-concept study

3.2.1. Data preparation and correspondence modeling 4D CT data used for RT treatment planning in our clinic are reconstructed using local amplitude-based binning (LAB) of the CT projection data: The individual breathing cycles of the breathing signal, i. e. the AP component of the RPM system, acquired during CT scanning are extracted and the breathing cycle-specific minimum and maximum signal values determined. These values define the *local*, i. e. the breathing cycle-specific 0 % and 100 % amplitude values, which are assigned as phase information to the respective CT projection data. Separately for inspiration and expiration, but again based on only the breathing cycle-specific signal values, phase values between 0 % and 100 % are assigned to projection data acquired in between. Finally, based on the assigned phase information, the CT images at the desired breathing phases are reconstructed. As a consequence, the phase values (0 %, 20 % inspiration, . . .) of the reconstructed images do not refer to a *single* breathing signal value but a series of values (in detail: as many values as breathing cycles existing in the recorded breathing signal). To obtain the 1-to-1 relationship between breathing signal value and the reconstructed CT images required for correspondence model formation, we re-parameterized the image phase information. The re-parameterization process consisted of the following steps:

(i) extraction of the individual breathing cycles (similar to LAB),

(ii) normalization of all breathing cycles to a median cycle length, and

(iii) computation of median signal values for each temporal sampling point of the time interval defined by (ii).

The original percentage amplitude values assigned to the images were then replaced by the respective signal values of the median breathing cycle, which are referred to as z_i ($i \in \{1, \ldots, n_{\mathrm{ph}}\}$). Together with the temporal derivative values $\partial_t z_i$ at the z_i positions of the median breathing cycle, they form the breathing signal measurements $\zeta_i = (z_i, \partial_t z_i)^T$ that are used as two-dimensional regressors to build the sought correspondence model.

To compute the corresponding regressands u_i, the open source insight toolkit variational registration framework (Werner *et al* 2014) was applied for DIR of the 4D CT phase images (distance measure: active normalized sum-of-squared differences forces; diffusion regularization; non-diffeomorphic setting; number of levels of multi-resolution scheme = 4; maximum number of iterations per level = 800; registration time step 1.0; regularization weighting factor = 0.5). These settings resulted in subvoxel registration accuracy (Werner *et al* 2014), which was assumed to be a sufficient basis for regression-based correspondence model training. As DIR reference phase, we selected a mid-expiration phase to allow more robust extrapolation in both in- and exhalation direction for the during treatment-RPM measurements beyond the

Table 1. Patient treatment and motion information. Met. = metastasis ID (for several metastases/patient, the second digit denotes the individual metastasis); Loc. = location according to Couinaud classification for liver segments (Couinaud 1957) and a respective classification for lung segments (Agur & Dalley 2009); GTV motion extracted from planning 4D CT; BED refers to the ITV; LC = local control in months; Gray row = local metastasis recurrence (*=deceased).

	Met.	Loc.	GTV Motion [mm]			Fractionation Scheme					BED [Gy]	Energy [MV]	GTV [cm^3]	LC [mos.]
			LR	AP	SI	# Frac.	Dose [Gy]	# Arcs	Length Arc [°]	Length Frac. [s]				
Liver	1	5	0	4	12	8	6	2	250	110	111.6	15	24.1	17
	2.1	7	4	3	11	5	11	3	180	309	111.6	15	0.8	8
	2.2	8	2	4	12	5	11	3	180	309	111.8	15	3.8	8
	2.3	3	1	3	8	5	11	3	180	309	110.6	15	1.4	8
	3.1	8	1	2	6	5	11	3	180	330	116.4	15	3.9	7
	3.2	7	2	2	1	5	11	3	180	330	116.6	15	2.3	7
	3.3	4	0	0	8	5	11	3	180	330	117.6	15	0.8	7
	4	7	1	5	19	5	11	2	210	242	115.8	15	16.8	13
	5	8	4	3	10	5	11	2	190	225	109.6	6	18.7	5
Lung	6	l.3	0	2	4	5	11	2	220	264	112.7	6	2.4	6*
	7.1	r.6	1	3	8	5	11	2	360	450	106.7	6	0.7	8
	7.2	l.2	0	4	7	5	11	2	360	450	109.1	6	0.6	8
	8	r.9	2	3	9	5	11	2	145	334	107.8	6	1.0	3*
	9	l.6	0	1	6	5	11	2	180	474	108.3	6	2.2	18
	10	l.3	0	0	1	5	11	2	180	329	116.0	6	12.7	17

range of the 4D CT-related RPM regressor signals and respective model-based motion field estimation.

3.2.2. 4D dose simulation For each patient, two different 4D dose simulation approaches were pursued: a standard 4D CT-based dose simulation and a 4D simulation using the proposed correspondence model-based approach.

Standard 4D CT-based simulation refers to Eq. (4), with the dose D being the planned 3D VMAT dose as computed and optimized by means of the temporal average CT of the ten-phase planning 4D CT. The transformations φ_i and respective motion fields were computed by non-linear registration between the individual phase images and the reference phase as described above.

For correspondence model-based 4D dose simulation, the planned 3D VMAT dose distribution was divided into dose segments of the minimal size that could be achieved by our treatment planning system (Eclipse 13). Thus, the resulting dose segments D_α still referred to the dose computed by means of the average CT and covered 2.3° angular intervals each. Depending on gantry speed, this corresponded to Δt_{Dose}-values between 1 s and 2.5 s. The breathing signal acquisition (using the Varian RPM system) had a frequency of 25 Hz, i. e. $\Delta t_{\mathrm{RPM}} = 0.04\,s$. With respect to Eq. (7), this means that for each individual treatment fraction, the originally planned dose segments D_α were deformed by the 25-63 correspondence model-based estimated motion fields derived using the RPM signal measurements acquired during the actual delivery of D_α. The deformed dose segments were then accumulated by weighted summation, resulting in the sought estimation of motion-affected dose distributions for the individual treatment fractions and the overall dose.

3.2.3. Experiments and analysis of 4D-simulated dose distributions Given the pilot and proof-of-concept character of the present study, its design was chosen as simple as possible. 4D dose simulation was performed for all ten patients as described above. Resulting dose distributions were compared to the original, i. e. the statically planned dose distributions, with the focus being GTV dose coverage. In agreement with previous studies on motion-related dose effects during VMAT treatment (Stambaugh *et al* 2013), deviations of D_{98} (dose to 98% of the GTV volume) of 4D-simulated and planned dose distributions were analyzed per fraction (only correspondence model-based 4D dose simulation) and in total (both standard and correspondence model-based 4D dose simulation); the values were linked to the known clinical endpoints of the individual metastases (local recurrence: yes/no). Accounting for the small population size, the results are presented in a descriptive way.

4. Results

The differences $\Delta\mathrm{D}_{98}$ between D_{98} of the retrospectively simulated 4D dose distributions, $\mathrm{D}_{98,\mathrm{Sim}}$, and the originally planned dose indices $\mathrm{D}_{98,\mathrm{Plan}}$ for the individual treatment fractions (only correspondence model-based 4D dose simulation) and for the total dose (i. e. dose after all treatment fractions) are summarized in Table 2. Bold entries underline maximum underdosages per fraction and metastasis when comparing $\mathrm{D}_{98,\mathrm{Sim}}$ estimated by correspondence model-based 4D dose simulation to the originally planned $\mathrm{D}_{98,\mathrm{Plan}}$. Differences between the $\Delta\mathrm{D}_{98}$ values for the individual treatment fractions as well as differences between the sum of the single

fraction values and the total dose ΔD_{98} values are due to patient-specific inter-fraction motion differences and motion variability; its consideration during 4D dose simulation is the major strength of the proposed correspondence model-based approach. Standard 4D CT-based dose accumulation (last column of Table 2) does not allow integrating information about motion variability; a fraction-by-fraction 4D dose simulation and analysis of motion variability-induced effects is not feasible.

From a clinical perspective, it is noteworthy that for correspondence model-based 4D dose simulation, metastases with local recurrence show higher negative total dose ΔD_{98} compared to metastases without recurrence (see especially the liver metastases 1 and 3.3). Furthermore, the fraction-wise ΔD_{98} values are consistently negative for all treatment fractions in the case of metastasis recurrence, but vary between the fractions (see again metastasis 1). Again, respective insight can not be achieved by standard 4D CT-based dose accumulation. Moreover, potential underdosages as estimated by 4D CT-based dose accumulation for metastases with local recurrence are much smaller than underdosages computed by the correspondence model-based 4D dose simulation, suggesting a potential underestimation of the motion effects. In contrast, at least for most metastases without recurrence, motion variability and breathing signal-steered 4D dose simulation led to fraction-wise ΔD_{98} values fluctuating around zero for correspondence model-based 4D dose simulation and similar total dose values for both 4D dose simulation approaches applied.

Table 2. GTV dose coverage of treatment plan and retrospectively 4D-simulated dose distribution. GTV dose coverage is quantified by differences ΔD_{98} of $D_{98,Sim}$ and $D_{98,Plan}$. For the proposed correspondence model-based 4D dose simulation, the data is shown separately for each patient treatment plan fraction and for the accumulated fraction dose distributions (= total dose). Standard 4D CT-based dose accumulation (last column) does not allow for estimation of fraction-specific effects; the shown data refers to the estimated total dose. For metastasis 1, i. e. the only treatment plan with 8 treatment fractions, results for fraction 6, 7 and 8 are shown in a separate line (indicated by the notation 'Frac. 1/6' etc.). Gray row = local metastasis recurrence.

	Met.	**$\Delta D_{98} = D_{98,Sim} - D_{98,Plan}$ (Gy)**						
		Frac. 1/6	**Frac. 2/7**	**Frac. 3/8**	**Frac. 4**	**Frac. 5**	**Total**	***4D CT sim***
Liver	1	− 0.13	− 0.21	− 0.18	− **0.49**	− 0.18		
		− 0.16	− 0.15	− 0.14			− 1.49	− *0.13*
	2.1	+ 0.05	− 0.04	− **0.27**	− 0.07	− 0.15	− 0.25	+ *0.07*
	2.2	− 0.02	− 0.06	− **0.07**	+ 0.04	− 0.06	+ 0.29	+ *0.07*
	2.3	− **0.12**	+ 0.04	+ 0.15	− 0.06	+ 0.01	+ 0.33	+ *0.07*
	3.1	− **0.02**	+ 0.01	+ 0.04	+ 0.02	+ 0.01	+ 0.10	+ *0.06*
	3.2	+ 0.03	+ 0.02	+ 0.01	+ 0.01	+ 0.03	+ 0.10	+ *0.02*
	3.3	− 1.58	− **3.11**	− 2.85	− 2.78	− 2.91	− 13.28	− *0.92*
	4	− **0.25**	− 0.11	− 0.13	− 0.07	− 0.13	− 0.18	+ *0.06*
	5	− 0.24	− 0.11	− **0.25**	− 0.12	− 0.11	− 0.71	+ *0.11*
Lung	6	− **0.01**	+ 0.06	+ 0.07	+ 0.09	+ 0.05	+ 0.50	+ *0.07*
	7.1	+ 0.03	− 0.08	+ 0.01	− **0.16**	+ 0.02	− 0.08	+ *0.05*
	7.2	− 0.02	− 0.07	− **0.11**	− 0.06	− 0.06	− 0.21	+ *0.01*
	8	− 0.03	− **0.34**	+ 0.02	− 0.01	+ 0.14	+ 0.44	− *0.01*
	9	− **0.26**	− 0.03	− 0.12	− 0.02	− 0.20	− 0.47	+ *0.10*
	10	− 0.01	− 0.01	− 0.01	− 0.01	− **0.02**	− 0.03	+ *0.01*

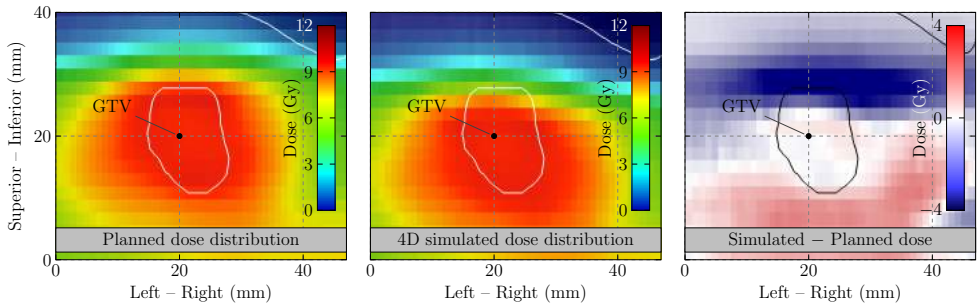


Figure 3. Correspondence model-based 4D dose simulation for metastasis 3.3. Planned (left) and retrospectively 4D-simulated (middle) dose distributions and corresponding dose difference distribution (right) for fraction 2 of metastasis 3.3.

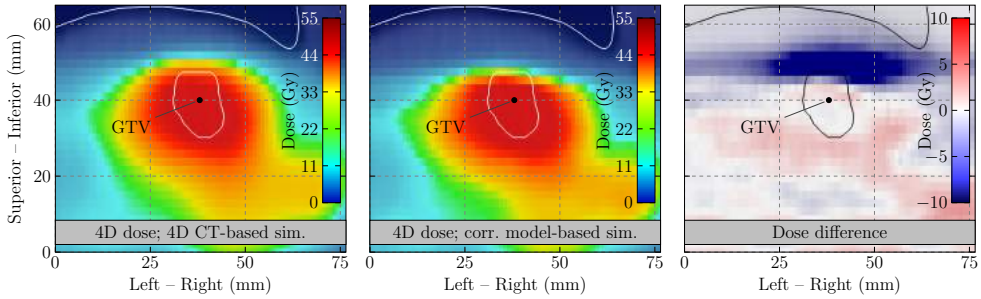


Figure 4. 4D-simulated total dose comparison for metastasis 3.3. Comparison of total dose distributions computed by standard 4D CT-based (left) and correspondence model-based 4D dose simulation (middle) and the difference thereof (right).

Aiming at a more detailed understanding of especially the given numbers for metastasis 3.3, Fig. 3 visualizes the deviations between planned and correspondence model-based 4D-simulated dose distributions in and around the GTV. The dose difference in Fig. 3, right, illustrates a GTV underdosage of up to 4 Gy in the superior part of the metastasis, i. e. the entire planned dose distribution was primarily shifted in inferior direction due to unfavorable internal patient motion. The analysis of the patient breathing signals acquired during the treatment fractions revealed a mean peak-to-peak RPM amplitude during treatment of only about half the 4D CT amplitude (cf. supplementary data table), with the signal values further being close to the lower CT-based safety gating limit. Thus, compared to the planning situation, the GTV probability of presence during treatment was significantly larger in the inferior ITV part than planned. In addition, the GTV-to-ITV safety margin for this specific metastasis appeared slightly too tight, which aggravates the aforementioned effects. However, estimated effects of the seemingly too small GTV-to-ITV margin were much smaller when using a standard 4D CT-based dose accumulation and, consequently, neglecting motion variability. For illustration, Fig. 4 shows the total dose distributions as obtained by correspondence model-based and standard 4D CT-based 4D dose simulation and the difference thereof.

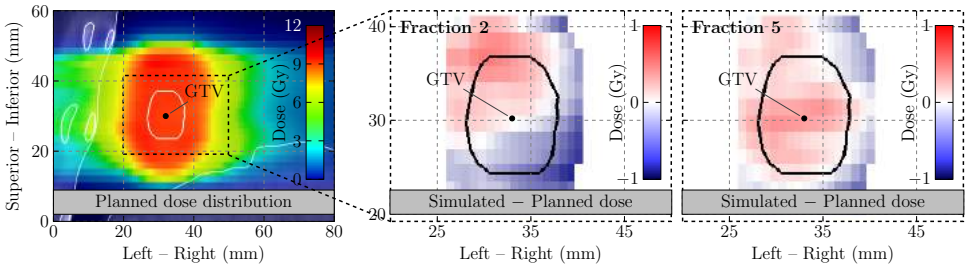


Figure 5. Correspondence model-based 4D dose simulation for metastasis 8. Planned dose distribution for metastasis 8 (left) and dose difference between retrospectively 4D-simulated and planned dose distributions for fractions 2 (middle) and 5 (right).

As another example to illustrate the influence of inter-fraction motion variability as well as the capability of correspondence model-based 4D dose simulation to reflect it, in Fig. 5, differences between planned and retrospective dose simulation results are shown for metastasis 8 and fractions 2 and 5. While no underdosage could be observed for fraction 5 (see also Table 2: $\Delta D_{98} = +\,0.14\,\mathrm{Gy}$), the difference image reveals voxel-wise underdosage of up to 0.8 Gy inside the GTV for fraction 2 ($\Delta D_{98} = -\,0.34\,\mathrm{Gy}$). In total, differences observed for the individual fractions not only average out over the course of treatment but result in a slightly positive total dose ΔD_{98} value.

5. Discussion and Conclusion

In this paper, we introduced a novel correspondence model-based 4D dose simulation approach to include breathing signal information acquired during treatment into retrospective 4D dose reconstruction. In contrast to common DIR-based dose accumulation methods like (Samavati *et al* 2016, Velec *et al* 2012, Werner *et al* 2012), the proposed approach allows consideration of patient-specific breathing variability by correlating observed internal and external breathing information. This, in turn, promises to yield a more realistic estimation of the dose actually delivered to the patient during, e. g., individual treatment sessions.

Having modeled the interplay of breathing-induced patient-specific target motion and the dynamics of the VMAT dose delivery process, we retrospectively simulated delivered motion-affected dose distributions for ten patients with 15 liver and lung metastases to analyze the effect of motion variability on hypofractionated SBRT treatment. Interestingly, metastasis recurrence seemed to be linked to negative deviations of the correspondence model-based 4D-simulated and originally, i. e. statically planned VMAT dose distributions that indicated motion-affected GTV underdosage. In addition, respective 4D dose simulation indicated inappropriate dimensioning of the ITV and breathing variability-related GTV underdosage for one of the metastases. Comparison of correspondence model-based 4D dose simulation to standard DIR- and 4D CT-based dose accumulation reveals that for the liver metastases the aforementioned linkage between the estimated deviations between planned and 4D-simulated dose distributions and the clinical endpoints is much weaker for standard dose accumulation; for the lung metastases, it is no longer observable.

Thus, although based on a small number of patients, our proof of concept study on

correspondence model-based 4D VMAT dose simulation and the correlation of motion-affected dose indices (here: D_{98}) to clinical endpoints illustrates the potential of the approach to possibly explain motion-related errors during treatment planning and dose delivery. This motivates further testing of the accuracy of the correspondence model-based simulation and integrating it as a quality assurance tool after treatment planning and/or after the individual treatment sessions. In addition, early identification of respiration-induced underdosage of target areas after the first treatment fraction(s) could be utilized to modify the treatment plan for the remaining fractions, e. g. by safety margin adaptation and introduction of local boosts. Thus, besides being used as a quality assurance tool, correspondence model-based 4D dose simulation could also provide valuable information for adaptive SBRT schemes of extracranial metastases.

However, before aiming at clinical application of correspondence model-based 4D dose simulation, remaining uncertainties and their impact on the dose simulation accuracy have to be analyzed. To begin with, our proof-of-concept study and correspondence modeling was based on only pre-treatment 4D CT image data. Thus, inter-fraction variations of the relationship between the external surrogate signal and internal motion (McClelland *et al* 2011) have not been accounted for so far. Acquisition of pre-fraction 4D CBCT instead of 3D CBCT images and a related pre-treatment updating of the correspondence model could solve this problem. Similarly, this would help to reduce uncertainties that are due to limited reproducibility of, in our case, the patient-specific positioning of the RPM marker block during the different treatment fractions. Due to the large amount of streaking artifacts in current clinical 4D CBCT data, DIR of CBCT phase images is, nevertheless, a challenge by itself and requires implementation of sophisticated solutions (Brehm *et al* 2013). Moreover, it should be noted that current 4D CT and 4D CBCT images represent so-called retrospectively sorted image data. This means that the actual raw data (i. e. the projection data) is acquired during a period that covers multiple breathing cycles of the patient. Image reconstruction then implicitly assumes that no motion variability (precisely: no *inter*-cycle variability, see definition in (McClelland *et al* 2013)) existed during data acquisition. In turn, motion variability during data acquisition leads to motion artifacts in the reconstructed images (Werner *et al* 2017); strictly speaking, no 1-to-1 matching between the reconstructed 3D images and breathing signal values exists (cf. Sec. 3.2.1). Although this limitation does not only apply to the present study but to the majority of existing studies on correspondence modeling, it leads to uncertainties during correspondence modeling; related motion artifacts affect, for instance, motion fields estimation by DIR in the 4D CT data, which compromises correspondence modeling accuracy. A further source of uncertainty is the correspondence model-based extrapolation of motion fields that is applied for modeling inter-cycle motion variability for surrogate measurements beyond the surrogate data used for model calibration. Extrapolation capabilities of the model are currently evaluated for only limited scenarios (Wilms *et al* 2014); extrapolation for surrogate data further beyond the model calibration data could lead to unrealistic motion fields. As future work and with increasing availability of related image data and algorithms, it would therefore be interesting to quantify and reduce these uncertainties using sorting- and artifact-free 4D image data (Thomas *et al* 2014) or techniques for correspondence modeling directly on unsorted/raw data (McClelland *et al* 2017).

Refining the VMAT arc discretization level (currently not possible due to treatment planning system restrictions) would further increase simulation accuracy; however and as shown in our motion phantom-based pre-study (Sothmann *et al* 2017),

the effects can be expected to be rather small. Analogously, the pursued approach of computing the dose for all angular VMAT segments based on the patient's average CT (instead of, e. g., simulated breathing signal measurement-specific CT images) introduces uncertainties, but these were already shown to be small (Milz *et al* 2014). Furthermore, the applied open-source non-linear registration framework has only been extensively evaluated for (4D CT) lung registration (Werner *et al* 2014) and (magnetic resonance) brain normalization (Ehrhardt *et al* 2015). For both applications it achieved high accuracy; nevertheless, its accuracy and behavior when applied to motion estimation of low to no-contrast organs such as the liver remains to be analyzed.

Finally and from a clinical perspective, the size of the currently small patient collection considered in the present study has to be significantly increased to verify our initial observations. Furthermore, refining information about the location of metastasis recurrence (this study: textual description whether metastasis recurrence occurred or not) would provide more detailed insight. Ideally, follow-up image information and voxel-wise recurrence delineation in, e. g., positron-emission tomography images (Van den Begin *et al* 2014) would be registered to the planning CT data and correlated on a voxel level to the 4D-simulated dose distributions.

Nevertheless and despite all necessary future work especially from the perspective of clinical application, the focus of the current study was and still is to present a correspondence model-based approach for 4D VMAT simulation that allows integrating breathing signal information about motion variability and to illustrate its feasibility and potential by correlating 4D-simulated dose distributions to clinical endpoint data. In our opinion, the results of our proof of concept study highlight that correspondence model-based 4D VMAT dose simulation offers the prospect of gaining further insight into respiratory motion and motion variability-related error sources in VMAT-based SBRT of extracranial metastases.

References

Agur A M & Dalley A F 2009 *Grant's atlas of anatomy* Lippincott Williams & Wilkins.

Alongi F, Arcangeli S, Filippi A R, Ricardi U & Scorsetti M 2012 Review and uses of stereotactic body radiation therapy for oligometastases. *Oncologist* **17**, 1100–1107.

Andratschke N *et al* 2016 Oc-0445: Patterns of care and outcome analysis of sbrt for liver metastases-a degro database initiative *Radiotherapy and Oncology* **119**, S208.

Brehm M, Paysan P, Oelhafen M & Kachelrieß M 2013 Artifact-resistant motion estimation with a patient-specific artifact model for motion-compensated cone-beam ct *Med Phys* **40**(10), 101913.

Chang D T *et al* 2011 Stereotactic body radiotherapy for colorectal liver metastases: a pooled analysis. *Cancer* **117**, 4060–4069.

Couinaud C 1957 *Le foie: études anatomiques et chirurgicales* Masson & Cie.

Ehrbar S, Lang S, Stieb S, Riesterer O, Stark L S, Guckenberger M & Klöck S 2016 Three-dimensional versus four-dimensional dose calculation for volumetric modulated arc therapy of hypofractionated treatments. *Z Med Phys* **26**, 45–53.

Ehrhardt J, Schmidt-Richberg A, Werner R & Handels H 2015 Springer Berlin Heidelberg Berlin, Heidelberg chapter P14, pp. 209–214.

Fassi A, Schaerer J, Fernandes M, Riboldi M, Sarrut D & Baroni G 2014 Tumor tracking method based on a deformable 4d ct breathing motion model driven by an external surface surrogate. *Int J Radiat Oncol Biol Phys* **88**, 182–188.

Ge J, Santanam L, Noel C & Parikh P J 2013 Planning 4-dimensional computed tomography (4DCT) cannot adequately represent daily intrafractional motion of abdominal tumors *Int J Radiat Oncol Biol Phys* **85**(4), 999–1005.

Høyer M, Swaminath A, Bydder S, Lock M, Méndez Romero A, Kavanagh B, Goodman K A, Okunieff P & Dawson L A 2012 Radiotherapy for liver metastases: a review of evidence. *Int J Radiat Oncol Biol Phys* **82**, 1047–1057.

Katz A W, Carey-Sampson M, Muhs A G, Milano M T, Schell M C & Okunieff P 2007 Hypofractionated stereotactic body radiation therapy (SBRT) for limited hepatic metastases. *Int J Radiat Oncol Biol Phys* **67**, 793–798.

Keall P J *et al* 2006 The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys* **33**(10), 3874–3900.

Korreman S S 2012 Motion in radiotherapy: photon therapy *Phys Med Biol* **57**(23), R161–91.

Lartigau E 2011 Stereotactic body radiotherapy *BMJ* **343:d4286**.

Low D A, Parikh P J, Lu W, Dempsey J F, Wahab S H, Hubenschmidt J P, Nystrom M M, Handoko M & Bradley J D 2005 Novel breathing motion model for radiotherapy. *International journal of radiation oncology, biology, physics* **63**, 921–929.

McClelland J R, Hawkes D J, Schaeffter T & King A P 2013 Respiratory motion models: a review. *Med Image Anal* **17**, 19–42.

McClelland J R, Hughes S, Modat M, Qureshi A, Ahmad S, Landau D B, Ourselin S & Hawkes D J 2011 Inter-fraction variations in respiratory motion models. *Physics in medicine and biology* **56**, 251–272.

McClelland J R *et al* 2017 A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images. *Physics in medicine and biology* **62**, 4273–4292.

Milz S, Wilkens J J & Ullrich W 2014 A dose error evaluation study for 4D dose calculations *Phys Med Biol* **59**(21), 6401–15.

Rieber J *et al* 2016 Influence of institutional experience and technological advances on outcome of stereotactic body radiation therapy for oligometastatic lung disease. *Int J Radiat Oncol Biol Phys* .

Sahgal A *et al* 2012 The Canadian Association of Radiation Oncology scope of practice guidelines for lung, liver and spine stereotactic body radiotherapy. *Clin Oncol (R Coll Radiol)* **24**, 629–639.

Samavati N, Velec M & Brock K K 2016 Effect of deformable registration uncertainty on lung SBRT dose accumulation. *Med Phys* **43**, 233.

Sarrut D 2006 Deformable registration for image-guided radiation therapy *Z Med Phys* **16**(4), 285–97.

Sothmann T, Gauer T & Werner R 2017 4d dose simulation in volumetric arc therapy: Accuracy and affecting parameters. *PloS one* **12**, e0172810.

Stambaugh C, Nelms B E, Dilling T, Stevens C, Latifi K, Zhang G, Moros E & Feygelman V 2013 Experimentally studied dynamic dose interplay does not meaningfully affect target dose in vmat sbrt lung treatments *Med Phys* **40**(9).

Thomas D, Lamb J, White B, Jani S, Gaudio S, Lee P, Ruan D, McNitt-Gray M & Low D 2014 A novel fast helical 4d-ct acquisition technique to generate low-noise sorting artifact-free images at user-selected breathing phases. *International journal of radiation oncology, biology, physics* **89**, 191–198.

Timmerman R *et al* 2010 Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA* **303**, 1070–1076.

Van den Begin R, Engels B, Gevaert T, Duchateau M, Tournel K, Verellen D, Storme G & De Ridder M 2014 Impact of inadequate respiratory motion management in SBRT for oligometastatic colorectal cancer *Radiother Oncol* **113**(2), 235–239.

Velec M, Moseley J L, Craig T, Dawson L A & Brock K K 2012 Accumulated dose in liver stereotactic body radiotherapy: positioning, breathing, and deformation effects. *Int J Radiat Oncol Biol Phys* **83**, 1132–1140.

Velec M, Moseley J L, Eccles C L, Craig T, Sharpe M B, Dawson L A & Brock K K 2011 Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration. *International journal of radiation oncology, biology, physics* **80**, 265–272.

Werner R, Ehrhardt J, Schmidt-Richberg A, Albers D, Frenzel T, Petersen C, Cremers F & Handels H 2012 Towards accurate dose accumulation for Step-&-Shoot IMRT: Impact of weighting schemes and temporal image resolution on the estimation of dosimetric motion effects *Z Med Phys* **22**(2), 109–22.

Werner R, Hofmann C, Mücke E & Gauer T 2017 Reduction of breathing irregularity-related motion artifacts in low-pitch spiral 4d ct by optimized projection binning *Radiat Oncol* **12**(1), 100.

Werner R, Schmidt-Richberg A, Handels H & Ehrhardt J 2014 Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study *Phys Med Biol* **59**(15), 4247–60.

Wilms M, Werner R, Ehrhardt J, Schmidt-Richberg A, Schlemmer H P & Handels H 2014 Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy. *Phys Med Biol* **59**, 1147–1164.

5.2. Image registration in 4D dose simulation

As motivated and illustrated in the previous section, the registration process in the introduced 4D dose accumulation scheme is crucial. The natural next step, after implementation of the general simulation framework, was therefore to investigate the impact of different registration algorithms on correspondence model building and subsequently using the generated models for correspondence model-based 4D dose accumulation. Results of this evaluation were published in the following journal article:

N. Mogadas[†], T. Sothmann[†], T. Knopp, R. Werner. Influence of deformable image registration on 4D dose simulation for extracranial SBRT: A multi-registration framework study. *Radiother Oncol*, 127(2):225–232, 2018.

To investigate the influence of different DIR algorithms, a set of registration algorithms had to be chosen. Properties of those algorithms were defined as 1) open source availability and 2) high-ranked in the initial EMPIRE10 challenge[2] or 3) described to have been applied in exactly the given context. In the end, six common DIR frameworks, as introduced in Table 1 [R&O 2018], were considered to be applied for correspondence model-based 4D dose simulation. Additionally, correspondence model accuracy and image registration accuracy were evaluated using a widely accepted benchmark image data base, i. e. DIRLAB, and a 4D MRI data set (see Section 4.4.1 and Section 4.4.3 for information about the utilized 4D image data). Parameterization of DIR algorithms was selected as published by the corresponding developers during the initial EMPIRE10 challenge. Distance measures primarily applied were the SSD as defined in Eq. (3.5) or subforms of it and subforms of the NCC, cf. Eq. (3.6). Mainly used regularization approaches were elastic and diffusive regularization as defined in Eq. (3.7) and Eq. (3.8), respectively, as well as Gaussian regularization (see Section 3.2.1 for more details).

In the first experimental part of this study, the DIR accuracy inside the lung and the liver was evaluated using the DIRLAB data base and the 4D MRI data set, respectively. Here, anatomical landmarks inside the 4D CT images were utilized to compute the TRE. The TRE is defined as the mean Euclidean distance between the reference landmark positions (i. e. landmarks in the fixed image) and landmark positions in the deformed image (i. e. landmarks in the moving image after applying achieved registration result). This was analyzed for the registration of end-inspiration to end-expiration phase images and vice versa (phase 00 $\leftrightarrow$ phase 50). Further, registration accuracy between phase 20 (reference phase for correspondence model building) and all other phase images was eval-

[†]Shared first authorship.

[2]The EMPIRE10 challenge is a comprehensive inter-institutional evaluation study for registration algorithms [91].

uated using the same principle. Vector field smoothness (analysis of the transformation Jacobian determinant[3]) and intensity-based similarity before and after registration were determined. As second part, a leave-one-out (LOO) evaluation was applied to analyze the accuracy of the trained correspondence model (cf. Fig. 5.1 for the applied LOO approach). As third part, correspondence model-based 4D dose simulation for the in-house patient data (cf. Section 4.4.4) was performed. Results of this study showed that three (VarReg, ANTS, Elastix) out of the six evaluated DIR frameworks achieved similarly high accuracy for lung registration, as shown in the upper part of Table 2 [R&O 2018]. For liver registration using the MRI data set, VarReg, Elastix and NiftyReg performed best (see Table E.5 [R&O 2018 supp. materials]). As VarReg was already used for internal motion extraction in Section 5.1 and was, for both anatomical cases, among the most accurate DIR approaches, an additional analysis of the influence of the degree of regularization was performed that, however, did not noticeably influenced DIR accuracy in the liver (cf. Table E.6 [R&O 2018 supp. materials]).

The LOO evaluation of correspondence model-based prediction of vector fields showed that higher registration accuracy leads to higher correspondence model accuracy, as supported by values in Table D.4 [R&O 2018 supp. materials] (intensity-based similarity measures) and lower part of Table 2 [R&O 2018], with again best performance in terms of lowest errors of VarReg, ANTS and Elastix.

In Table 3 [R&O 2018], results for the third experiment are summarized. The influence of DIR on the subsequent dose accumulation is clearly visible for the investigated liver cases by means of deviations between computed ΔD_{95} values. Deviations between ΔD_{95} values for lung cancer patients are smaller but nevertheless existing.

The achieved results suggest that, especially for the lung cases, accuracy of DIR algorithms directly impact correspondence modeling accuracy, but in the same time did not necessarily lead to relevant differences of 4D-simulated dose distributions and related dose indices. In contrast, results for the liver metastasis cases showed that the chosen DIR framework highly impacts the correspondence model-based dose accumulation output due to application of purely intensity-driven DIR in low-to-no contrast areas. Having a closer look at the liver metastases results, it can, for instance, be seen that computed vector fields of ANTS and VarReg are in well agreement, while results of the other registration approaches somewhat differ. The cause remains unclear. However, due to the

[3] In general, a smaller standard deviation of the transformation Jacobian determinant value, i. e. a smoother vector field, is for a comparable registration accuracy physiological more plausible. However, this depends on the registration task. For instance, sliding motion on lung borders should be considered by the registration algorithm to achieve highest registration accuracy. Homogeneous smoothing of the vector field will, however, result in locally reduced registration accuracy in these regions as the estimated motion is physiological implausible [92].

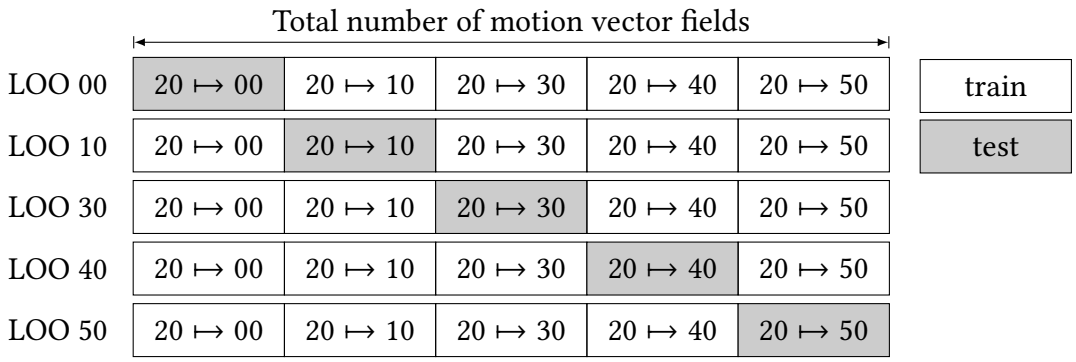


Figure 5.1.: Visualization of performed leave-one-out (LOO) evaluation. Correspondence models were trained with the motion vector fields and corresponding breathing signal information for a phase $n \in \{00, 10, \ldots, 50\}$ left out (indicated as gray cell) and the predicted field for the left-out phase was evaluated. Note that phase 20 is the reference phase and thus cannot be left out.

absence of ground truth data and the small sample size, it also remains unclear, which simulation reveals reality best. For further evaluation of 4D-simulated dose distributions, additional information of e. g. follow-up imaging could provide helpful information to allow for correlation of voxel-wise recurrence delineation and subsequent 4D dose simulation results. Unfortunately, respective information is currently not routinely available. Thus, computed dose deviations for the investigated liver cases leave an impression of unpredictability. The unknown impact of 4D CT image artifacts on DIR amplifies this impression. A first investigation of an image artifact in one of the patients 4D CT data sets near the target volume suggests that the influence can be under certain circumstances severe (cf. Figure F.3 [R&O 2018 supp. materials]). Therefore, in the following section, the general influence of typical 4D CT image artifacts on the proposed dose accumulation scheme is analyzed. Afterwards an approach to quantify DIR uncertainties is presented (Section 5.4).

Influence of deformable image registration on 4D dose simulation for extracranial SBRT: a multi-registration framework study

Nik Mogadas[a,*], Thilo Sothmann[a,b,*], Tobias Knopp[c], Tobias Gauer[b], Cordula Petersen[b], René Werner[a]

[a]*Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[b]*Department of Radiotherapy and Radio-Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[c]*Section for Biomedical Imaging, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

Abstract

Background and Purpose: To evaluate the influence of deformable image registration approaches on correspondence model-based 4D dose simulation in extracranial SBRT by means of open source deformable image registration (DIR) frameworks.
Material and Methods: Established DIR algorithms of six different open source DIR frameworks were considered and registration accuracy evaluated using freely available 4D image data. Furthermore, correspondence models (regression-based correlation of external breathing signal measurements and internal structure motion field) were built and model accuracy evaluated. Finally, the DIR algorithms were applied for motion field estimation in radiotherapy planning 4D CT data of five lung and five liver lesion patients, correspondence model formation, and model-based 4D dose simulation. Deviations between the original, statically planned and the 4D-simulated VMAT dose distributions were analyzed and correlated to DIR accuracy differences.
Results: Registration errors varied among the DIR approaches, with lower DIR accuracy translating into lower correspondence modeling accuracy. Yet, for lung metastases, indices of 4D-simulated dose distributions widely agreed, irrespective of DIR accuracy differences. In contrast, liver metastases 4D dose simulation results strongly vary for the different DIR approaches.
Conclusions: Especially in treatment areas with low image contrast (e. g. the liver), DIR-based 4D dose simulation results strongly depend on the applied DIR algorithm, drawing resulting dose simulations and indices questionable.

Keywords: Deformable Image Registration, 4D Dose Simulation, Dose Accumulation, Correspondence Modeling

Introduction

Stereotactic body radiotherapy (SBRT) provides an effective and widely used treatment option for lung or liver cancer patients [1, 2, 3]. However, respiration-induced motion of target structures and organs at risk (OAR) of up to several centimeters [4] and additional structure deformation [5] pose challenges and have to be accounted for during treatment planning. In current clinical 4D radiation therapy (RT) workflows, a respiration-correlated CT (4D CT) is usually acquired before therapy; the 4D CT and derived data like tMIP (temporal maximum intensity projection) or average CT images are then mainly used to define the internal target volume (ITV) [6]. In addition, target and OAR dynamics represented by the planning 4D CT have frequently been described to be incorporated into dose distribution computation and optimization, i. e. to explicitly account for motion effects by means of 4D dose simulations or dose accumulations [7, 8].

The standard approach to extract target and OAR dynamics from the 4D CT is to apply deformable image registration (DIR). The resulting motion fields between a pre-selected reference phase CT and the other 4D CT phase images are then – either directly or indirectly, making additional use of motion modeling – used to deform the planned dose

*Equal contribution
Email address: t.sothmann@uke.de (Thilo Sothmann)

Table 1: Overview of the considered deformable image registration frameworks and algorithms. CC = cross correlation, (N)SSD = (normalized) sum of squared differences, NMI = normalised mutual information, NC = normalized correlation, MSD = mean squared differences. Computation times refer to DIR of two liver-4D CT phase images (image size: $512 \times 512 \times 159$ voxel; CPU: Intel Xeon e5-1630, 3.50 GHz; 32 GB RAM).

DIR framework		**Version**	**DIR details**	**Similarity measure**	**Computation time**
ANTS	[26]	2.2	symmetric B-spline-based reg.	CC	$\approx$ 300 min
VarReg	[27]	4.11	non-parametric demons-type reg.	NSSD	$\approx$ 180 min
DIRART	[24]	1.0	Horn-Schunck optical flow	SSD	$\approx$ 15 min
NiftyReg	[28]	master	B-spline-based registration	NMI	$\approx$ 30 min
Elastix	[29, 30]	4.8	B-spline-based registration	NC	$\approx$ 120 min
Plastimatch	[31]	1.6.4	B-spline-based registration	MSD	$\approx$ 60 min

or dose segments to obtain the sought 4D dose distribution [9, 10]; dose accumulation accuracy therefore depends on the DIR-based motion estimation accuracy. Evaluation studies of DIR accuracy in 4D CT data can meanwhile be found for a series of registration algorithms, frameworks and commercial programs, such as Morpheus [11, 12, 13], MIM Software [9, 10], Pinnacle3 [14, 15], Elastix [16] or ITK-based registration approaches [17, 18, 19]. Comprehensive inter-institutional evaluation studies like EMPIRE10 [20] and MIDRAS [21] as well as open data repositories for evaluation purposes (e. g. DIRLAB data [22, 23]) also exist. However, most evaluation studies are focused on lung registration only; transferability of respective accuracy statements to registration in low contrast image areas like liver or other soft tissues remains unclear. Moreover, similar registration accuracy values are not necessarily related to similar properties of computed transformations [19]; the influence of different transformation models on and a propagation of potential errors into 4D dose simulation is widely neglected.

From our perspective, joint investigation of DIR performance and DIR influence on 4D dose simulation is essential to obtain a better understanding of related uncertainties; and (for the best of our knowledge), the present study is the first to present such data in a multi-framework setting. Six DIR algorithms of different common open source DIR frameworks that were either high-ranked in the initial EMPIRE10 challenge or described to be applied in exactly the given context were considered. Using 4D CT data and treatment plans of ten patients with lung and liver metastases, all registration algorithms were applied for correspondence model-based 4D dose simulation. Furthermore, image registration accuracy as well as DIR-based correspondence modeling accuracy were evaluated by means of freely available 4D image data. The DIR-dependent 4D dose distributions were compared to analyze the influence of the different DIR algorithms, and DIR-related dose differences were correlated to DIR accuracy measures.

Material and Methods

This section first describes the DIR algorithms, the patient collective, treatment planning and the principles of correspondence model-based 4D dose simulation. The description is followed by explanation of the performed experiments and the strategy to evaluate DIR and correspondence modeling accuracy.

DIR algorithms

In total, six different open source DIR frameworks were considered for this study. Five of the frameworks are wide-spread general purpose DIR frameworks; the choice of the specific DIR algorithms and their parameterization was, however, motivated by high ranks in the initial EMPIRE10 challenge [20]. The sixth framework was chosen as it has been described to be specifically designed for RT applications [24] and has, e. g., already been applied for a phantom-based experimental dose warping evaluation study [25]. The specific algorithm in the present study was the best performing one according to [25]. Methodical details like transformation models and distance measures significantly vary among the DIR algorithms. A brief overview is given in Table 1. For reproducibility purposes, example registration scripts used for the present study can be found as part of the supplemental materials.

Patient collective and treatment planning

Ten cancer patients treated by volumetric arc therapy (VMAT) were included into this retrospective study: five patients with six lung metastases and five patients with nine liver metastases. For each patient, a ten-phase 4D CT

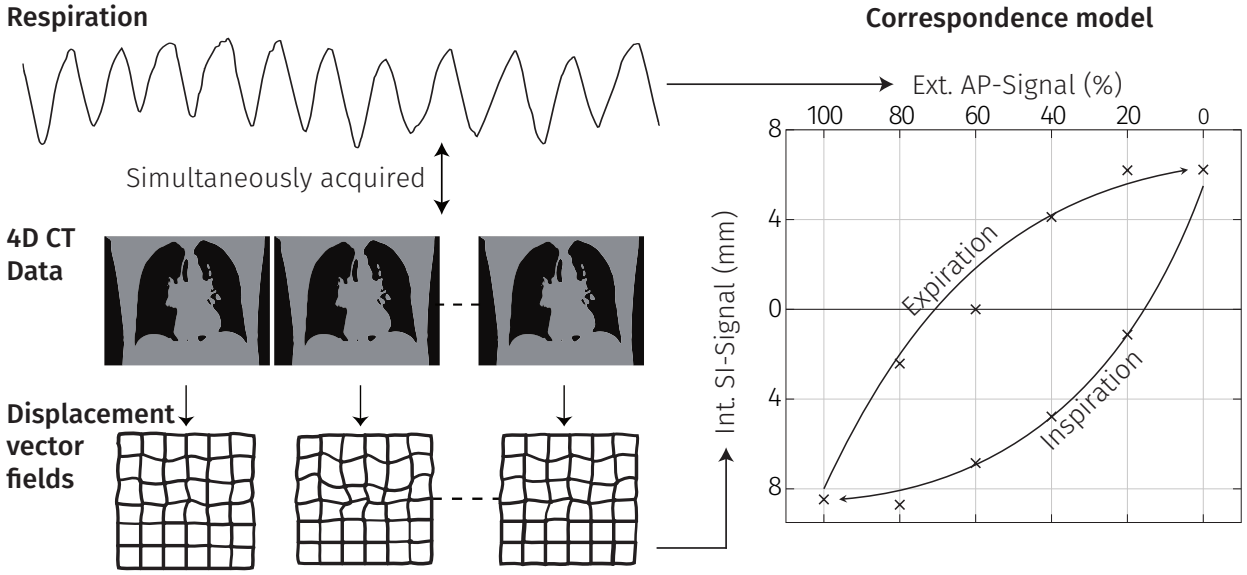


Figure 1: Process of correspondence modeling. A ten-phase 4D CT data and external breathing signal measurements are simultaneously acquired. Internal structure motion information is extracted using DIR between a reference 4D CT phase and the remaining nine phases, yielding individual displacement vector fields (DVFs). Multivariate linear regression of DVFs and corresponding breathing signal information establishes a functional relationship between internal motion and external breathing measurements. In the present study, both the AP component of the RPM system and the respective temporal derivative are combined to a two-dimensional regressor to allow for modeling of hysteresis behavior (illustrated for an individual voxel and the superior-inferior (SI) component of the DFVs). Adapted from [32].

with spatial resolution of $0.98 \times 0.98 \times 2$ mm was acquired (Siemens Definition AS Open CT; RPM system, Varian Medical Systems). VMAT treatment planning was based on an ITV concept: gross tumor volume (GTV) delineation on each 4D CT phase image; combination of all phase GTVs = ITV; ITV + 4 mm safety margin = planning target volume (PTV); dose calculation and optimization on average CT (Eclipse 13, Varian Medical Systems; prescribed minimum dose surrounding the PTV ranged from 48 and 55 Gy). For each metastasis, clinical outcome information (local metastasis recurrence: yes/no) was available.

Principles of correspondence model-based 4D dose simulation

Standard 4D dose simulation consists of three main steps: estimation of the motion fields between a reference CT data set of the patient and the individual (in our case ten) 3D phase CTs of the patient's planning 4D CT by DIR; deformation of the planned 3D dose distribution(s) according to the DIR-estimated motion fields; weighted accumulation of the resulting deformed dose distributions to obtain the sought 4D dose distribution.

Solely based on the planning 4D CT image information, this approach neglects patient-specific breathing irregularity during treatment and related effects on the 4D-simulated dose. To mitigate this issue, current work on correspondence modeling and dose accumulation were combined. Here, correspondence modeling refers to correlating respiratory breathing signal measurements to DIR-estimated motion fields in order to derive a functional relationship between an easy-to-acquire breathing signal and internal structure motion. In the present study, the relationship between RPM measurements and internal motion fields were modeled by multivariate linear regression. To establish the model, DIR was performed between a mid-expiration phase CT (= reference image) and the other phases of the patient's planning 4D CT. To simulate direct "plug-and-play" application of the open source DIR frameworks and algorithms, unmasked DIR was used. The resulting motion fields were considered the regressands; corresponding RPM measurements performed during 4D CT data acquisitions served as regressors. A schematic illustration of the correspondence modeling process is given in Figure 1.

After establishing the model, correspondence model-based VMAT 4D dose simulation is a conceptually straightforward extension of standard dose accumulation. The applied approach is based on [33]. VMAT dose delivery takes place during continuous gantry rotation, with varying gantry speed, dose rate and multileaf collimator (MLC) posi-

tioning. For 4D dose simulation, each so called VMAT arc is divided into small angular segments (2.3° to 3°). In addition, RPM signals were acquired during dose delivery for all patients. Thus, briefly speaking, the during-treatment RPM measurements were temporally correlated to the respective VMAT dose segments. For each individual breathing signal measurement, corresponding internal motion fields (displacement vector fields, DVFs) were computed using the patient-specifically trained correspondence model and applied to deform the temporally related dose segment. The sought 4D simulated VMAT dose distribution results as a weighted summation over all deformed dose segments (weighting factor = 1/[number of breathing signal measurements during dose delivery]).

Making use of the underlying concepts and nomenclature of [33, 19, 34], a detailed description of correspondence model-based dose simulation can be found in the supplemental materials, Suppl. A.

Experiments and evaluation strategy

Experiments were divided into two main parts: (1) evaluation of DIR and correspondence modeling accuracy for the different DIR algorithms; (2) correspondence model-based 4D dose simulation and subsequent analysis of dose parameters to quantify motion effects on the dose distribution.

Part (1) was based on freely available 4D image data. First, ten 4D CT data sets provided by the DIRLAB [22, 23] were used. The 4D CT sets consist of ten phase images (denoted by phases 00, 10, ..., 90). Further, corresponding anatomical landmarks inside the lungs were defined by a thoracic imaging expert in the phase images 00, 10, ..., 50. By means of the landmarks, DIR accuracy is quantified by the target registration error (TRE). The TRE is defined as mean Euclidean distance between the registration target image landmark positions and the positions of the warped landmarks of the registration reference image. In agreement with literature on DIR accuracy evaluation [19], the TRE for end-inspiration to end-expiration DIR and vice versa ($00 \leftrightarrow 50$) was determined. As, however, phase 20 served as reference phase during corresponding modeling, the TRE of DIR with the phase 20 image as registration reference image was also evaluated for all remaining phases with existing landmarks (i.e. phases 00, 10, and 30-50) serving as target images. In addition and to provide a more comprehensive picture of accuracy and properties of the computed transformations, common intensity-based similarity measures (mean squared differences, MSD; normalized cross correlation, NCC; normalized mutual information, NMI) before and after DIR and transformation plausibility/smoothness measures were evaluated. To avoid bias due to DIR algorithm design, all similarity measures contained in Table 1 were considered. Transformation plausibility and smoothness evaluation was based on analysis of the transformation Jacobian determinant. Further explanations are given in the supplemental materials, Suppl. B.

The DIRLAB data, however, do not allow evaluating registration accuracy in the liver due to low image contrast and hardly visible inner-liver structures in 4D CT data. To, nevertheless, obtain an understanding of the inner-liver DIR performance and the ability of the algorithms to provide plausible liver motion fields, a separate test scenario was designed using a publicly available liver 4D-MRI data set with clearly visible liver structures [35] (see also `www.vision.ethz.ch/4dmri`) that allowed for identification of inner-liver landmarks and computation of TRE values for liver DIR. To mimic a 4D CT-like low contrast situation, the contrast inside the liver was synthetically reduced and the influence of the reduced contrast on DIR accuracy and transformation properties evaluated. Details can be found in the supplemental materials, Suppl. E.

Eventually, for each DIRLAB data set, correspondence models were built by means of the DVFs (all ten phases registered to phase 20) of the different DIR algorithms. External breathing signal measurements are, however, not available for the DIRLAB data. Following Wilms et al., simulated thorax-belt data was extracted from the 4D CT image data and a leave-one phase-out (LOO) evaluation performed [34]: Correspondence models were trained with the DVF and breathing signal information for a phase $n \in \{00, 10, \ldots, 90\}$ left out and the predicted field for the left-out phase evaluated. Due to the absence of landmarks for the inspiration phases $n \in \{60, \ldots, 90\}$, TRE values were only determined for the predicted expiration DVFs. To also capture accuracy for inspiration DVF prediction, the intensity-based similarity measures introduced before were additionally evaluated for all LOO phases $n \in \{00, 10, \ldots, 90\}$.

Part (2) refers to correspondence model-based 4D dose simulation for the described patient collective. For each patient and DIR algorithm, a correspondence model was built and applied for 4D dose simulation. Simulated 4D dose distributions (i. e. dose distributions that are assumed to account for patient-specific motion effects) were compared to the original dose distribution as planned and optimized by means of the average CT. Dose distribution differences were quantified by ΔD_{95}, defined as the difference of $D_{95,\text{4D-sim}}$ and $D_{95,\text{plan}}$ with D_{95} as the dose received by 95% of the GTV. For patients with > 1 metastases, differences were evaluated for the individual metastases.

Table 2: DIR and correspondence model accuracy for the different DIR algorithms, evaluated by means of the DIRLAB data. All data is given as mean ± standard deviation of the ten DIRLAB data sets. For extreme phase registration (00 ↔ 50), DIR was performed with both phase 00 and phase 50 as reference image; results were averaged. $\varnothing_n$ denotes average values over the 20 ↦ n experiments. Bold numbers highlight best results in each category.

Algorithm	**TRE, registration (mm)**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	**2.4 ± 1.3**	1.5 ± 0.4	1.6 ± 0.3	1.6 ± 0.4	**1.7 ± 0.5**	**1.9 ± 0.6**
VarReg	2.5 ± 1.3	**1.3 ± 0.3**	**1.5 ± 0.3**	**1.5 ± 0.4**	**1.7 ± 0.4**	**1.9 ± 0.4**
DIRART	5.1 ± 2.0	3.1 ± 1.0	2.6 ± 0.9	2.2 ± 0.7	3.0 ± 1.1	3.7 ± 1.5
NiftyReg	4.6 ± 2.9	2.6 ± 1.1	2.3 ± 0.8	2.0 ± 0.7	2.6 ± 1.3	3.0 ± 1.5
Elastix	2.6 ± 1.2	1.7 ± 0.7	1.7 ± 0.5	1.6 ± 0.5	1.9 ± 0.7	2.2 ± 1.0
Plastimatch	3.8 ± 1.9	2.1 ± 0.7	2.1 ± 0.6	2.2 ± 0.9	2.2 ± 0.9	2.5 ± 1.1
w/o reg.	8.5 ± 3.3	4.4 ± 1.4	3.2 ± 1.1	2.5 ± 1.1	3.9 ± 1.8	4.9 ± 2.4
	TRE, LOO model-based prediction of DVF 20 ↦ n (mm)					
	$\varnothing_n$	n = 00	n = 10	n = 30	n = 40	n = 50
ANTS	2.1 ± 0.3	2.1 ± 0.9	**2.0 ± 0.5**	**1.9 ± 0.6**	**2.1 ± 0.8**	2.3 ± 0.7
VarReg	**2.0 ± 0.3**	**2.0 ± 0.9**	**2.0 ± 0.5**	**1.9 ± 0.6**	**2.1 ± 0.8**	**2.2 ± 0.7**
DIRART	3.0 ± 0.8	3.4 ± 1.2	2.6 ± 0.8	2.2 ± 0.8	3.1 ± 1.2	3.7 ± 1.5
NiftyReg	2.7 ± 0.6	2.9 ± 1.3	2.5 ± 0.9	2.2 ± 0.8	2.8 ± 1.3	3.3 ± 1.6
Elastix	2.2 ± 0.4	2.2 ± 1.1	2.2 ± 0.6	**1.9 ± 0.7**	2.2 ± 1.0	2.4 ± 1.0
Plastimatch	2.4 ± 0.4	2.5 ± 1.1	2.3 ± 0.7	2.0 ± 0.7	2.5 ± 1.0	2.8 ± 1.1

Results

DIR and correspondence model accuracy

The results of the lung landmark-based DIR accuracy evaluation by means of the DIRLAB data are summarized in Table 2. For extreme phase registration 00 ↔ 50 (left column of Table 2), ANTS [mean TRE = (2.4 ± 1.3) mm], VarReg [TRE = (2.5 ± 1.3) mm], and Elastix [TRE = (2.6 ± 1.2) mm] show similar accuracy; accuracy of the other DIR algorithms was lower. Comparable observations hold true for registration with reference phase 20 (i.e. estimation of DVFs 20 ↦ n with $n \in \{00, 10, 30, 40, 50\}$). Due to smaller landmark distances before registration, the TRE values and differences between the DIR algorithms are, however, smaller than for 00 ↔ 50 registration. The TRE-based split of the DIR algorithms into two groups – ANTS, VarReg, and Elastix as more accurate compared to the other algorithms – is further supported by the evaluation of the intensity-based similarity measures after registration (supplemental materials Suppl. C and Table C.1). Consideration of the mean Jacobian determinant as transformation plausibility measure did not allow further insight into the DIR algorithm performance. It can, however, be seen that high DIR accuracy of VarReg is associated with lowest smoothness across all tested DIR algorithms. Yet, as a similar statement is not applicable for ANTS, a more general statement about the relationship between DIR accuracy and smoothness cannot be made (supplemental materials Suppl. C and Table C.2 and C.3).

The results of the investigation of the behavior of the DIR algorithms in image areas with low image contrast, like the liver in 4D CT data, and the respective 4D MRI experiments are summarized in the supplemental materials, Suppl. E. Overall, extreme phase registration TRE values for the 4D MRI experiments were lowest for VarReg, Elastix and NiftyReg (landmark distances before registration: 14.7 mm; after DIR in MRI data with synthetically reduced, CT-like inner-liver contrast: < 3.0 mm; after DIR in original MRI: between 2.0 mm and 2.6 mm). ANTS and Plastimatch show an approximately 1.5 mm higher TRE; DIRART with the chosen parameterization failed for the liver registration (TRE > 9.0 mm for MRI data with CT-like inner-liver contrast). Similar to the DIRLAB data evaluation, TRE observations are supported by the intensity-based similarity measures. Statements regarding the transformation Jacobian determinant evaluation agree with respective observations for the DIRLAB data. Focusing on VarReg (among the most accurate DIR approach for both DIRLAB and MRI data, plus the DIR approach our group has most experience with [19]), the MRI data DIR results also illustrate that varying the degree of regularization did

Table 3: Differences in GTV dose coverage of the planned and retrospectively 4D simulated dose distribution, separately quantified by differences ΔD_{95} between $D_{95,4D\text{-}sim}$ and $D_{95,plan}$ for the individual metastases (met.). Confirmed local metastasis recurrence is indicated by an asterisk. Negative values of ΔD_{95} mean that $D_{95,plan}$ was larger than $D_{95,4D\text{-}sim}$, indicating deterioration of GTV coverage when accounting for the patient-specific motion information. Bold values indicate smallest and largest ΔD_{95} values per metastasis.

	Met.	$\Delta D_{95} = D_{95,4D\text{-}sim} - D_{95,Plan}$ (Gy)					
		ANTS	**VarReg**	**DIRART**	**NiftyReg**	**Elastix**	**Plastimatch**
Lung met. patients	1	+ **0.47**	+ 0.45	+ 0.37	+ 0.45	+ 0.44	+ **0.28**
	2.1	– **0.13**	– 0.12	– 0.12	– 0.12	+ **0.02**	– 0.12
	2.2*	– 0.28	– **0.30**	– 0.13	– 0.12	– **0.02**	– 0.05
	3	+ 0.35	+ 0.39	– **7.85**	+ **0.47**	+ 0.26	+ 0.25
	4*	– 0.51	– 0.52	– **0.40**	– 0.50	– **0.56**	– 0.50
	5	– 0.09	– 0.06	– **0.05**	– **0.05**	– 0.09	– **0.12**
Liver met. patients	6*	– 2.73	– 2.65	– 0.32	– **0.19**	– **8.40**	– 6.96
	7.1	– 0.49	– 0.47	– **0.30**	– 0.39	– 1.57	– **1.78**
	7.2	+ 0.17	+ 0.17	+ 0.18	+ **0.19**	– 3.40	– **15.65**
	7.3	+ 0.26	+ **0.32**	+ **0.19**	+ 0.24	+ 0.24	+ 0.23
	8.1	+ **0.14**	+ 0.12	+ **0.10**	+ 0.13	+ 0.13	+ 0.13
	8.2	+ 0.17	+ 0.11	+ **0.07**	+ 0.09	+ **0.20**	+ 0.12
	8.3*	– 12.21	– 13.17	– **4.73**	– 9.10	– 26.49	– **29.09**
	9	– 0.43	– 0.30	– 0.16	– **0.09**	– **1.48**	– 0.34
	10	– 0.69	– 0.85	– 0.40	– **0.17**	– **0.96**	– **0.96**

* = local metastasis recurrence confirmed

not noticeably influence DIR accuracy for the different contrast levels, suggesting robustness of respective liver-CT DIR results.

The lower part of Table 2 contains the accuracy evaluation results of the different leave-one phase-out correspondence models performed by means of the DIRLAB data. Similar to the DIRLAB 4D CT DIR evaluation, lowest TRE values were obtained for VarReg, Elastix and ANTS, indicating that higher DIR accuracy leads to higher correspondence model accuracy – and, vice versa, low(er) DIR accuracy to inaccurate correspondence models. Consideration of the intensity-based similarity measures for additional accuracy evaluation (supplemental materials Suppl. D) again supports the TRE-related observations for the predicted expiration phase DVFs; similar similarity measure values for prediction of inspiration and expiration DVFs further reveals the potential of the models to capture potential motion differences between inspiration and expiration.

Correspondence model-based 4D dose simulation

Separately for each metastasis, the results of the DIR-specific correspondence model-based 4D dose simulation and DIR-specific ΔD_{95} values, respectively, are summarized in Table 3.

For lung metastases (upper half of the table), ΔD_{95} for the different DIR algorithms show high accordance. However, for metastasis 3, the DIRART ΔD_{95} clearly differs from the other values (DIRART: $\Delta D_{95} = -7.85$ Gy; range of ΔD_{95} values for other DIR algorithms: + 0.25 Gy to + 0.47 Gy). To better understand the causes, registrations between the extreme phases 00 and 50 of the respective 4D CT data set were re-run for all algorithms. The resulting DVFs were used to warp the GTV of the phase 50 image (GTV_{50}). Ideally, the warped GTV (GTV_{50}^{warp}) would be identical to the 00 phase GTV (GTV_{00}). The overlap between GTV_{50}^{warp} and GTV_{00} is evaluated by Dice coefficient,

$$\text{Dice} = \frac{2 \cdot \left| GTV_{50}^{warp} \cap GTV_{00} \right|}{\left| GTV_{50}^{warp} \right| + |GTV_{00}|}$$

and the results are visualized in Figure 2. While ANTS, VarReg, NiftyReg and Elastix showed high Dice values $\geq 70\%$, DIRART led to zero overlap; the corresponding ΔD_{95} value was due to registration failure (here: during registration matching of the metastasis border to the diaphragm border). However, the also imperfect Plastimatch

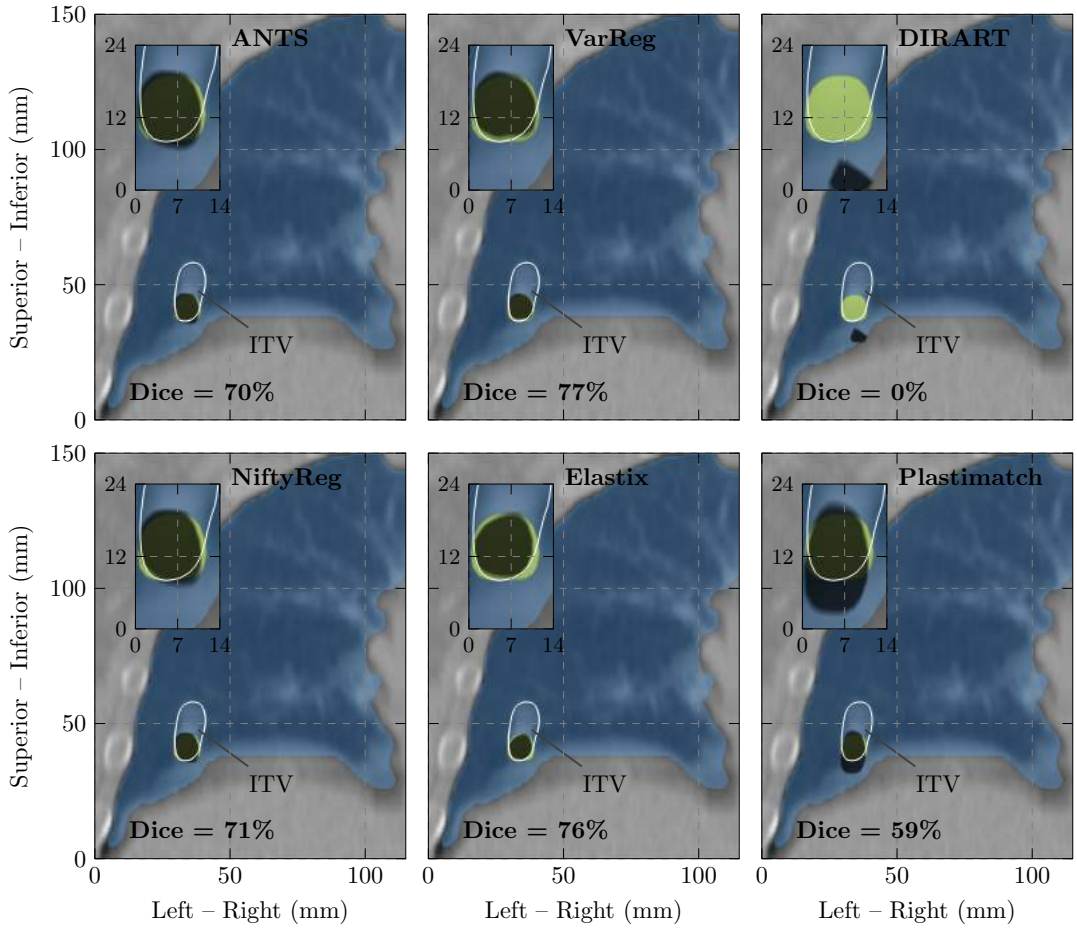


Figure 2: Visualization of extreme phase registration results for metastasis 3: The phase 50 GTV was warped by the respective DVF (black structure in ITV = warped GTV_{50}) and compared to the reference GTV in phase 00 (yellow structure in ITV = reference GTV_{00}). Ideally, the warped GTV_{50} should cover the entire yellow GTV_{00}; the given numbers are the Dice overlap coefficients of the two structures.

registration (Dice coefficient of 59%) did, in the current case, not influence ΔD_{95} when compared to the more accurate DIR algorithms.

Compared to the lung cases, ΔD_{95} values for the liver metastases are more diverging. For instance, Elastix and Plastimatch show much higher negative ΔD_{95} values for cases 6, 7.1, 7.2 and 8.3 than the other DIR algorithms (cf. lower half of Table 3). Again to better understand this behavior, DIR-DVFs between phases 20 and 50 as well as 4D simulated dose distributions for metastasis 6 are illustrated in Figure 3. The location of the GTV is also visualized. It can be seen that the magnitude of the DVFs as well as the lateral motion direction strongly differ between the DIR algorithms. While, for instance, NiftyReg estimates almost no GTV motion and, consequently, ΔD_{95} is close to zero, Elastix and Plastimatch lead to large motion vectors and high negative ΔD_{95} values. Respective effects can also be seen by means of the 4D simulated dose distribution: Elastix and Plastimatch lead to lower dose areas in the superior part of the GTV after 4D dose simulation; for NiftyReg, the GTV appears to be enclosed by a high isodose.

Similar observations can be made for the other liver cases with large ΔD_{95} differences between the DIR algorithms. Yet, judgement of the DIR-specific dosimetric motion effects and their plausibility based on, e.g., the aforementioned liver DIR accuracy results appears not feasible: With respect to the 4D MRI experiments, Elastix, NiftyReg, as well as VarReg (metastasis 6 motion amplitudes somewhere between Elastix and NiftyReg) were all among the most accurate DIR approaches. Having a deeper look, the small NiftyReg-estimated motion amplitudes inside the liver appear, however, not realistic and can potentially be associated to an interpolation artifact in the GTV area that distracted the

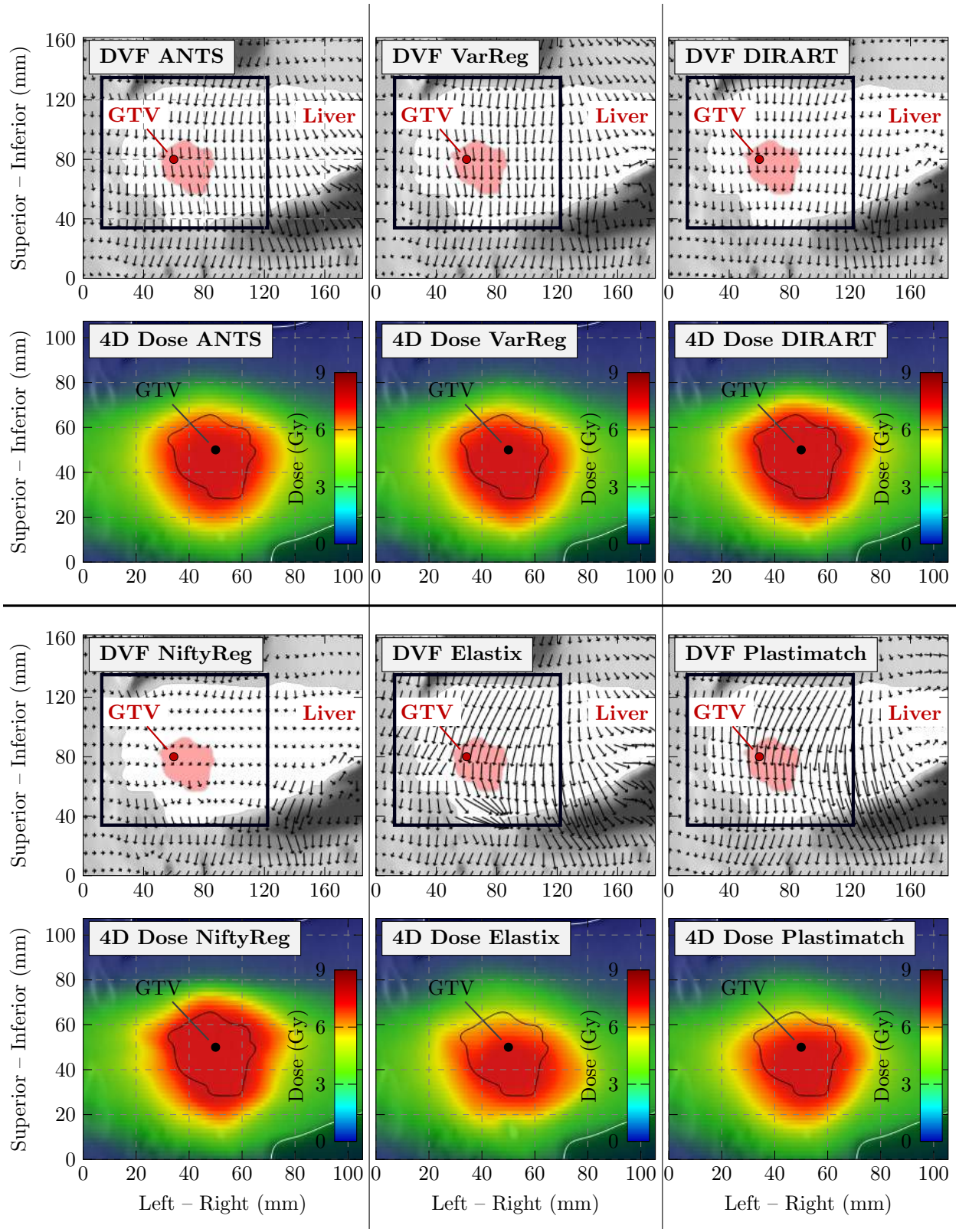


Figure 3: Visualization (coronal views) of the estimated motion field between phases 50 and 20 and DIR-specific 4D simulated dose distributions (one fraction) for metastasis 6. Magnitude and direction (superior-inferior and left-right) of motion field are illustrated by arrow length and angle.

algorithm (see Suppl. E, Fig. E.3). Further, the very high ΔD_{95} values for Elastix appear questionable; such values would have likely led to local metastasis recurrence – which was not observed for this patient. But even such in-depth analysis does not always help; with certainty, it can only be stated that the DIR algorithm has a significant impact on the correspondence model-based dose accumulation process, especially for liver tumours.

Discussion

The present study addressed the influence of DIR techniques on 4D dose simulation for extracranial (here: lung and liver) SBRT and potential interrelation of 4D dose simulation results and DIR accuracy. To the best of our knowledge, this is the first study that explicitly addresses the pipeline character of current 4D dose simulation solutions and comprehensively investigated the role of DIR algorithms and respective uncertainties therein. To foster reproducibility, only common open source DIR frameworks and freely available, established algorithms for motion field estimation in 4D CT data were used; corresponding registration scripts are provided as supplemental materials.

The study builds on correspondence model-based 4D dose simulation [8]. Most 4D dose simulation approaches directly and solely utilize DVFs extracted from the patient's planning 4D CT [9, 10], in some studies enriched with 4D cone beam CT information acquired prior to the individual SBRT fractions [13], for dose mapping purposes. In contrast, correspondence model-based 4D dose simulation allows integrating (external) breathing signal information about breathing irregularity during treatment into the dose accumulation process. In terms of the aforementioned pipeline character, it, however, introduces an additional step between DIR and dose accumulation; therefore, the correlation between registration and correspondence model accuracy was also analyzed.

The experiments for the *lung metastasis cases* illustrated

- that, in the sense of an error propagation, low(er) DIR accuracy in terms of high(er) TRE values directly translates into low(er) correspondence modeling accuracy,
- that, however, quantitative differences in DIR and correspondence model accuracy of the algorithms (as determined using the DIRLAB benchmark data) do not necessarily lead to relevant differences of 4D-simulated dose distributions and related dose indices, and
- that the 4D-simulated dose distributions and dose indices widely agree among the DIR approaches, rendering the results somehow trustworthy.

Still, the outlier described in the results section (DIRART-computed dose for metastasis 3) *is* associated to the DIR algorithm with highest TRE values for the DIRLAB data. General conclusions regarding the interrelation between robustness (frequency of total registration failure) and benchmark data-derived TRE values can, however, not be drawn yet, due to the small sample size of the present study.

In contrast, the results for the *liver metastasis cases*, respective 4D dose simulation and related large differences of the ΔD_{95} values clearly illustrate issues regarding dose accumulation for tumors in low contrast image areas like the liver by means of (like in the current study) purely intensity-driven DIR. Focusing on the ΔD_{95} values, ANTS and VarReg seem to be in well agreement, while results of the other registration approaches somewhat differ. The cause remains unclear: While VarReg was among the most accurate approaches for the 4D MRI-based low contrast structure registration test scenario, ANTS was not. In addition, any attempts to correlate differences in ΔD_{95} with specific characteristics of the considered DIR algorithms also failed: Neither similar transformation models (cf. the B-spline-based DIR algorithms ANTS and Elastix) nor similar DIR similarity measures led to similar DIR accuracy and/or ΔD_{95} values.

In view of the divergence of the estimated liver motion patterns, further consideration of additional DIR types and especially the class of biophysical and finite element modeling-based registration approaches appears promising [36, 37]. Such DIR approaches have already been shown to outperform purely intensity-based DIR for liver registration in the MIDRAS study [21]; in contrast to the DIR algorithms applied in the current study, they are, however, usually not available as open source and "plug-and-play" tools.

Yet, even when incorporating potentially more reliable FEM-based DIR approaches for liver motion estimation, assessment of 4D dose simulation reliability still suffers from the lack of a detailed ground truth to compare the estimated dosimetric effects to. The known outcome (local metastasis recurrence: yes/no) exploited in the current study

only allowed for plausibility considerations for individual cases. At this, analysis of follow-up image information like voxel-wise recurrence delineation in, e. g., positron-emission tomography images [38] and subsequent correlation to 4D-simulated dose distributions could provide helpful information to evaluate the 4D dose simulation results. Such information is, however, at the moment not routinely available; therefore, aforementioned differences for liver cases leave (from our perspective) an impression of uncertainty. Continuing with remaining sources of uncertainty, it should also be noted that the applied 4D dose simulation allowed for incorporation of information about patient-specific breathing irregularity acquired during treatment; due to the absence of online imaging, potential changes of the relationship between the used external breathing signal and the actual internal structure dynamics [39, 40] were, however, neither known nor accounted for.

Thus, taken together: The reported observations raise doubts regarding reliability of CT-based liver tumor 4D dose simulations. In our opinion and more general for tumors in low contrast areas, current open source DIR frameworks should not be considered ready for "plug-and-play" use for 4D dose accumulation; respective application of DIR by inexperienced users is questionable.

Conflict of interest statement

None.

Acknowledgements

This work was in parts funded by the University of Hamburg (scholarship to TS). Further, we would like to thank Jamie McClelland and Felix Bragman (Centre for Medical Image Computing, University College London) for their assistance concerning NiftyReg and Nicholas Tustison (University of Virginia) for providing the Empire10 challenge parameters for ANTS.

References

[1] Lartigau E. Stereotactic body radiotherapy. BMJ (Clinical research ed) 2011;343:d4286.
[2] Sahgal A, Roberge D, Schellenberg D et al. The Canadian Association of Radiation Oncology scope of practice guidelines for lung, liver and spine stereotactic body radiotherapy. Clin Oncol 2012;24(9):629–639.
[3] Katz AW, Carey-Sampson M, Muhs AG, Milano MT, Schell MC, Okunieff P. Hypofractionated stereotactic body radiation therapy (SBRT) for limited hepatic metastases. Int J Radiat Oncol Biol Phys 2007;67:793–798.
[4] Keall PJ, Mageras GS, Balter JM et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med Phys 2006;33:3874–3900.
[5] Chan MKH, Werner R, Ayadi M, Blanck O. Comparison of 3D and 4D monte carlo optimization in robotic tracking stereotactic body radiotherapy of lung cancer. Strahlenther Onkol 2015;191(2):161–71.
[6] Hof H, Rhein B, Haering P, Kopp-Schneider A, Debus J, Herfarth K. 4D-CT-based target volume definition in stereotactic radiotherapy of lung tumours: comparison with a conventional technique using individual margins. Radiother Oncol 2009;93:419–423.
[7] Rosu M, Hugo GD. Advances in 4D radiation therapy for managing respiration: part II - 4D treatment planning. Z Med Phys 2012;22(4):272–80.
[8] Sothmann T, Gauer T, Wilms M, Werner R. Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT. Phys Med Biol 2017;62(23):9001–9017.
[9] Ehrbar S, Lang S, Stieb S et al. Three-dimensional versus four-dimensional dose calculation for volumetric modulated arc therapy of hypofractionated treatments. Z Med Phys 2016;26(1):45–53.
[10] Valdes G, Lee C, Tenn S et al. The relative accuracy of 4D dose accumulation for lung radiotherapy using rigid dose projection versus dose recalculation on every breathing phase. Med Phys 2016;.
[11] Samavati N, Velec M, Brock KK. Effect of deformable registration uncertainty on lung SBRT dose accumulation. Med Phys 2016;43:233.
[12] Swaminath A, Massey C, Brierley JD et al. Accumulated delivered dose response of stereotactic body radiation therapy for liver metastases. Int J Radiat Oncol Biol Phys 2015;93:639–648.
[13] Velec M, Moseley JL, Craig T, Dawson LA, Brock KK. Accumulated dose in liver stereotactic body radiotherapy: positioning, breathing, and deformation effects. Int J Radiat Oncol Biol Phys 2012;83:1132–1140.
[14] Admiraal MA, Schuring D, Hurkmans CW. Dose calculations accounting for breathing motion in stereotactic lung radiotherapy based on 4D-CT and the internal target volume. Radiother Oncol 2008;86:55–60.
[15] Ehler ED, Tomé WA. Lung 4D-IMRT treatment planning: an evaluation of three methods applied to four-dimensional data sets. Radiother Oncol 2008;88:319–325.
[16] Yu H, Zhang Sx, Wang Rh, Zhang Gq, Tan Jm. The feasibility of mapping dose distribution of 4DCT images with deformable image registration in lung. Biomed Mater Eng 2014;24:145–153.

[17] Flampouri S, Jiang SB, Sharp GC, Wolfgang J, Patel AA, Choi NC. Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4D-CT data and Monte Carlo simulations. Phys Med Biol 2006;51:2763–2779.
[18] Rao M, Wu J, Cao D et al. Dosimetric impact of breathing motion in lung stereotactic body radiotherapy treatment using intensity modulated radiotherapy and volumetric modulated arc therapy [corrected]. Int J Radiat Oncol Biol Phys 2012;83:e251–e256.
[19] Werner R, Schmidt-Richberg A, Handels H, Ehrhardt J. Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study. Phys Med Biol 2014;59:4247–4260.
[20] Murphy K, Van Ginneken B, Reinhardt JM et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Trans Med Imaging 2011;30(11):1901–1920.
[21] Brock KK, Deformable Registration Accuracy Consortium . Results of a multi-institution deformable registration accuracy study (MIDRAS). Int J Radiat Oncol Biol Phys 2010;76(2):583–96.
[22] Castillo R, Castillo E, Guerra R et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys Med Biol 2009;54(7):1849–70.
[23] Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T. Four-dimensional deformable image registration using trajectory modeling. Phys Med Biol 2010;55:305–327.
[24] Yang D, Brame S, El Naqa I et al. Technical note: DIRART–a software suite for deformable image registration and adaptive radiotherapy research. Med Phys 2011;38:67–77.
[25] Yeo UJ, Taylor ML, Supple JR et al. Is it sensible to "deform" dose? 3D experimental validation of dose-warping. Med Phys 2012;39:5065–5072.
[26] Tustison NJ, Cook PA, Klein A et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 2014;99:166–179.
[27] Schmidt-Richberg A, Werner R, Handels H, Ehrhardt J. A flexible variational registration framework. Insight J 2014;.
[28] Modat M, Ridgway GR, Taylor ZA et al. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed 2010;98:278–284.
[29] Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 2010;29:196–205.
[30] Shamonin DP, Bron EE, Lelieveldt BPF et al. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. Front Neuroinf 2013;7:50.
[31] Shackleford JA, Kandasamy N, Sharp GC. On developing B-spline registration algorithms for multi-core processors. Phys Med Biol 2010;55:6329–6351.
[32] McClelland JR, Hawkes DJ, Schaeffter T, King AP. Respiratory motion models: a review. Medical image analysis 2013;17:19–42.
[33] Sothmann T, Gauer T, Werner R. 4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters. PLoS One 2017;12:e0172810.
[34] Wilms M, Werner R, Ehrhardt J, Schmidt-Richberg A, Schlemmer HP, Handels H. Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy. Phys Med Biol 2014;59:1147–1164.
[35] von Siebenthal M, Székely G, Gamper U, Boesiger P, Lomax A, Cattin P. 4d mr imaging of respiratory organ motion and its variability. Phys Med Biol 2007;52(6):1547–64.
[36] Velec M, Moseley JL, Svensson S, Hårdemark B, Jaffray DA, Brock KK. Validation of biomechanical deformable image registration in the abdomen, thorax, and pelvis in a commercial radiotherapy treatment planning system. Med Phys 2017;44(7):3407–3417.
[37] Werner R. Biophysical modeling of respiratory organ motion. In: Ehrhardt J, Lorenz C, editors. 4D Modeling and Estimation of Respiratory Motion for Radiation Therapy. Springer-Verlag; 2013,.
[38] Van den Begin R, Engels B, Gevaert T et al. Impact of inadequate respiratory motion management in SBRT for oligometastatic colorectal cancer. Radiother Oncol 2014;113(2):235–9.
[39] Ionascu D, Jiang SB, Nishioka S, Shirato H, Berbeco RI. Internal-external correlation investigations of respiratory induced motion of lung tumors. Med Phys 2007;34(10):3893–903.
[40] Poels K, Depuydt T, Verellen D et al. Improving the intra-fraction update efficiency of a correlation model used for internal motion estimation during real-time tumor tracking for sbrt patients: fast update or no update? Radiother Oncol 2014;112(3):352–9.

Supplementary materials

Influence of deformable image registration on 4D dose simulation for extracranial SBRT: a multi-registration framework study

Nik Mogadas[a,*] , Thilo Sothmann[a,b,*], Tobias Knopp[c], Tobias Gauer[b], Cordula Petersen[b], René Werner[a]

[a]*Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[b]*Department of Radiotherapy and Radio-Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[c]*Section for Biomedical Imaging, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[*]Equal contribution

Suppl. A. Correspondence model-based 4D dose simulation

While the methods part of the main manuscript primarily aimed at a brief, comprehensible description of correspondence model-based 4D VMAT dose simulation, this supplementary information provides an in-depth description thereof that allows reimplementing the applied approach. Nomenclature and symbols are based on [1, 2, 3, 4].
For each patient, the 4D treatment planning CT is a series of 3D CT images, with each image representing a different breathing phase of the patient, i. e.

$$(I_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ I_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$$

with $i \in \{1,\dots,n_{\mathrm{ph}}\}$ as the breathing phase and n_{ph} as the total number of 3D CT images in the 4D CT data set. Analogously, the breathing signal values that are assigned to the images for reconstruction purposes read

$$(\zeta_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ \zeta_i \in \mathbb{R}^d$$

with d as dimensionality of the breathing signal. In the current study, the image data consisted of $n_{\mathrm{ph}} = 10$ phases, and ζ_i was two-dimensional with $\zeta_i = (z_i, \partial_t z_i)^T$ and $z_i \in \mathbb{R}$ as the anterior-posterior (AP) component measurements of the Varian RPM system, which is routinely used for 4D CT reconstruction and safety gating purposes at our clinic. Further, $\partial_t z_i \in \mathbb{R}$ denotes the temporal derivative of the RPM AP values, evaluated at z_i.

Suppl. A.1. Correspondence model formation

Image registration. The first step to establish the sought correspondence model is to compute the motion fields that represent the respiratory motion of the internal structures of interest (organs at risk, target volumes). Therefore, assume an arbitrary breathing phase $i_0 \in \{1,\dots,n_{\mathrm{ph}}\}$ to be selected as reference phase (here: $i_0 = 3$, denoting mid-expiration) and the corresponding CT image I_{i_0} as fixed image during registration. Then, the registration process results in a series of transformations

$$(\varphi_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ \varphi_i : \Omega \to \Omega$$

and respective motion fields

$$(u_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}},\ u_i : \Omega \to \mathbb{R}^3$$

with $u_i = \varphi_i - id$. The motion fields $(u_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}}$ and the respective breathing signals $(\zeta_i)_{i\in\{1,\dots,n_{\mathrm{ph}}\}}$ form the basis of subsequent linear regression-based correspondence model training.

Multivariate regression. For ease of readability and similar to [1], breathing signal measurements and motion fields are in the following interpreted as random variables $\mathbf{Z}_i$ ($\equiv \zeta_i$) and $\mathbf{U}_i \in \mathbb{R}^{3m}$ with m denoting the number of voxels of the reference phase image I_{i_0}. Then, the sought correspondence model – i. e. the assumed relationship between breathing signal measurements and internal motion fields – is defined by

$$\hat{\mathbf{U}} = \overline{\mathbf{U}} + \mathbf{B}\left(\hat{\mathbf{Z}} - \overline{\mathbf{Z}}\right) \tag{A.1}$$

with $\hat{\mathbf{U}} \in \mathbb{R}^{3m}$ the motion field for a breathing signal observation $\hat{\mathbf{Z}} \in \mathbb{R}^2$. $\overline{\mathbf{U}} = 1/n_{\mathrm{ph}} \sum_i^{n_{\mathrm{ph}}} \mathbf{U}_i$ and $\overline{\mathbf{Z}} = 1/n_{\mathrm{ph}} \sum_i^{n_{\mathrm{ph}}} \mathbf{Z}_i$ denote the mean motion field and breathing signal, respectively.
Using an ordinary least-squares regression approach, the coefficient matrix $\mathbf{B} \in \mathbb{R}^{3m\times 2}$ can be computed by

$$\mathbf{B} = \arg\min_{\mathbf{B}'} \mathrm{tr}\left[(\mathbf{U} - \mathbf{B}'\mathbf{Z})(\mathbf{U} - \mathbf{B}'\mathbf{Z})^T\right] = \mathbf{U}\mathbf{Z}^+ \tag{A.2}$$

with

$$\mathbf{Z} = \left(\mathbf{Z}_1 - \overline{\mathbf{Z}}, \ldots, \mathbf{Z}_{n_{\text{ph}}} - \overline{\mathbf{Z}}\right)$$
$$\mathbf{U} = \left(\mathbf{U}_1 - \overline{\mathbf{U}}, \ldots, \mathbf{U}_{n_{\text{ph}}} - \overline{\mathbf{U}}\right)$$

as mean-centered observations and

$$\mathbf{Z}^{+} = \mathbf{Z}^{T}\left(\mathbf{Z}\mathbf{Z}^{T}\right)^{-1}$$

as Moore-Penrose pseudoinverse. Subsequently, $\hat{\mathbf{U}}$ and the motion field $\hat{u} : \Omega \times \mathbb{R}^2 \rightarrow \mathbb{R}^3$ that results by application of (A.1) for a given $\hat{\mathbf{Z}}$ and $\hat{\zeta}$, respectively, are used as equivalent terms.

Suppl. A.2. Correspondence model-based 4D dose simulation

The time-dependent dose rate of a treatment fraction, $\dot{D} : \Omega \times \mathcal{T} \subset \mathbb{R} \rightarrow \mathbb{R}_+$, can be used to express the dynamical dose delivery process, its interplay with patient motion and the resulting dose distribution $D_{4\text{D}} : \Omega \rightarrow \mathbb{R}_+$ via

$$\begin{aligned} D_{4\text{D}}(x) &= \int_{\mathcal{T}} \dot{D}(\varphi(x,t),t)\ dt \\ &\approx \sum_{t\in\tilde{\mathcal{T}}} \dot{D}(\varphi(x,t),t)\ \Delta t \\ &= \sum_{t\in\tilde{\mathcal{T}}} D_t(\varphi(x,t)) \end{aligned} \tag{A.3}$$

with $\varphi : \Omega \times \mathcal{T} \rightarrow \mathbb{R}^3$ representing the position $\varphi(x,t)$ of voxel $x \in \Omega$ of the reference phase CT image I_{i_0} at time $t \in \mathcal{T} = [0;T) \subset \mathbb{R}$. The numerically required temporal discretization in (A.3) indicates an equidistant sampling of the dose delivery process, with the sampling points $t \in \tilde{\mathcal{T}} = \{{}^1\!/_2\,\Delta t; {}^3\!/_2\,\Delta t; \ldots\}$, $\tilde{\mathcal{T}} \subset \mathcal{T}$, and a sampling period of Δt. Thus, $D_t : \Omega \rightarrow \mathbb{R}_+$ is the integral dose delivered during the time interval $[t - {}^1\!/_2\,\Delta t; t + {}^1\!/_2\,\Delta t)$.

The time-dependent position $\varphi(x,t)$ of the inner voxels and structures is unknown during dose delivery; however, during gated treatment, a breathing signal is continuously acquired. This signal and the established correspondence model (A.1) are now integrated into (A.3). With $\hat{\varphi} = id + \hat{u}$ and $\hat{\zeta} : \mathcal{T} \rightarrow \mathbb{R}^2$, correspondence model-based 4D dose simulation for a single fraction finally reads

$$D_{4\text{D}}(x) \approx \sum_{t\in\tilde{\mathcal{T}}} D_t\left(\hat{\varphi}\left(x,\hat{\zeta}_t\right)\right) = \sum_{t\in\tilde{\mathcal{T}}} D_t\left(x + \hat{u}\left(x,\hat{\zeta}_t\right)\right) \tag{A.4}$$

where $\hat{\zeta}_t = \hat{\zeta}(t)$. Multiple treatment fractions are covered by directly extending (A.4) to

$$D_{4\text{D}}^{\text{total}}(x) = \sum_{\text{fx}} D_{4\text{D}}^{\text{fx}}(x) = \sum_{\text{fx}} \sum_{t\in\tilde{\mathcal{T}}_{\text{fx}}} D_t\left(x + \hat{u}\left(x,\hat{\zeta}_t\right)\right).$$

Here, fx is the fraction index and $\tilde{\mathcal{T}}_{\text{fx}} \subset \mathbb{R}$ are the temporal sampling points used to compute the sought integral fraction dose distribution. Without plan adaptation between the individual treatment fractions, one can assume $\tilde{\mathcal{T}}_{\text{fx}} = \tilde{\mathcal{T}}$ for all treatment fractions. Consequently, differences between the individual fraction dose distributions $D_{4\text{D}}^{\text{fx}}$ are solely attributable to differences of the patient's motion patterns (more precisely: of the observations $\hat{\zeta}_t$) during the treatment fractions.

Suppl. B. DIR algorithm evaluation: similarity and transformation plausibility/smoothness measures

In addition to the landmark-based target registration error (TRE, see Material and Methods part of the main manuscript), the following measures were evaluated:

Intensity-based similarity measures, to be optimized during registration. All measures were evaluated using respective subclasses of the *ImageToImage* class of the open source toolkit *Insight Segmentation and Registration Toolkit* (ITK, see `https://itk.org`).

- *Mean squared intensity differences* (MSD); should be as small as possible and ideally zero.
- *Normalized cross correlation* (NCC); in particular, (−1)*NCC is minimized, i.e. −1 is the smallest possible and ideal value.
- *Normalized mutual information* (NMI); similar to NCC, −1 is the smallest possible and ideal value.

Transformation plausibility/smoothness measures, also implemented using ITK.

- *Plausibility / Change of volume*: Computed as mean Jacobian determinant $\overline{|\nabla\varphi|} = 1/|\Omega| \sum_{x\in\Omega} \nabla\varphi\,(x)$. Evaluated for a specific structure (here: lung or liver), $\overline{|\nabla\varphi|}$ should resemble the ratio of the structure volumes in the images to be registered.
- *Smoothness*: Computed as standard deviation of the Jacobian determinant of the voxels inside the structure of interest (here: lung or liver).

Suppl. C. Additional results: DIRLAB data

The results of the evaluation of the intensity-based similarity measures before and after registration of the DIRLAB data can be found in Table C.1. The results of the evaluation of transformation plausibility and smoothness are summarized in Table C.2 and Table C.3. All measures are evaluated only in the lungs (= the structure of interest in the DIRLAB data).

Table C.2. Transformation plausibility, assessed by means of the mean Jacobian determinant inside the corresponding lung mask and DIR-computed transformations. The respective ratio of the volume of the lung masks of the phase images serves as ground truth. All data is given as mean ± standard deviation of the ten DIRLAB data sets.

Algorithm	**Mean Jacobian**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	0.88 ± 0.04	1.06 ± 0.02	1.04 ± 0.01	0.97 ± 0.01	0.94 ± 0.02	0.93 ± 0.03
VarReg	0.86 ± 0.05	1.07 ± 0.02	1.05 ± 0.02	0.97 ± 0.01	0.94 ± 0.03	0.92 ± 0.03
DIRART	0.86 ± 0.05	1.07 ± 0.02	1.05 ± 0.02	0.96 ± 0.01	0.94 ± 0.03	0.92 ± 0.03
NiftyReg	0.89 ± 0.03	1.05 ± 0.01	1.04 ± 0.01	0.97 ± 0.01	0.95 ± 0.01	0.90 ± 0.02
Elastix	0.86 ± 0.05	1.07 ± 0.02	1.05 ± 0.02	0.97 ± 0.01	0.94 ± 0.03	0.92 ± 0.03
Plastimatch	0.87 ± 0.04	1.07 ± 0.02	1.04 ± 0.01	0.97 ± 0.01	0.94 ± 0.02	0.93 ± 0.03
Mask volume ratio	0.87 ± 0.05	1.07 ± 0.02	1.05 ± 0.01	0.97 ± 0.02	0.94 ± 0.03	0.92 ± 0.04

Table C.1. Accuracy of the different DIR algorithms, evaluated by means of the DIRLAB data. Similarity of reference image and warped moving image is evaluated using MSD, NCC and NMI inside the lungs. All data is given as mean ± standard deviation of the ten DIRLAB data sets.

Algorithm	**MSD**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	9660 ± 5790	4147 ± 1967	3470 ± 1386	3187 ± 1753	4292 ± 2493	5082 ± 3279
VarReg	4903 ± 3220	2824 ± 1778	2550 ± 874	2211 ± 1011	2635 ± 1230	2966 ± 1564
DIRART	13423 ± 5906	7910 ± 3145	5652 ± 2421	4938 ± 2347	7306 ± 3414	8615 ± 4442
NiftyReg	19014 ± 15609	9310 ± 5194	6958 ± 3484	5983 ± 4136	9177 ± 7213	10729 ± 8849
Elastix	6936 ± 4333	4420 ± 2375	3645 ± 1729	3129 ± 1682	4156 ± 2643	4816 ± 3253
Plastimatch	13943 ± 6920	7263 ± 2871	5671 ± 2147	5043 ± 2824	6781 ± 3538	8040 ± 4623
w/o reg.	87808 ± 29415	22518 ± 7890	14107 ± 6457	17232 ± 9753	33586 ± 15985	43276 ± 21098
Algorithm	**NCC**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	−0.967 ± 0.018	−0.983 ± 0.007	−0.985 ± 0.006	−0.987 ± 0.006	−0.984 ± 0.008	−0.982 ± 0.009
VarReg	−0.979 ± 0.010	−0.988 ± 0.004	−0.989 ± 0.004	−0.991 ± 0.004	−0.989 ± 0.005	−0.988 ± 0.005
DIRART	−0.944 ± 0.018	−0.967 ± 0.011	−0.976 ± 0.011	−0.976 ± 0.011	−0.970 ± 0.011	−0.966 ± 0.013
NiftyReg	−0.940 ± 0.040	−0.961 ± 0.021	−0.970 ± 0.015	−0.977 ± 0.014	−0.966 ± 0.027	−0.964 ± 0.027
Elastix	−0.973 ± 0.013	−0.981 ± 0.010	−0.984 ± 0.008	−0.987 ± 0.007	−0.984 ± 0.010	−0.982 ± 0.011
Plastimatch	−0.949 ± 0.023	−0.969 ± 0.013	−0.975 ± 0.011	−0.980 ± 0.010	−0.974 ± 0.013	−0.970 ± 0.016
w/o reg.	−0.807 ± 0.057	−0.899 ± 0.045	−0.937 ± 0.039	−0.939 ± 0.038	−0.898 ± 0.047	−0.881 ± 0.051
Algorithm	**NMI**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	−0.20 ± 0.04	−0.27 ± 0.03	−0.29 ± 0.03	−0.30 ± 0.04	−0.27 ± 0.04	−0.26 ± 0.05
VarReg	−0.25 ± 0.04	−0.31 ± 0.03	−0.32 ± 0.03	−0.33 ± 0.04	−0.31 ± 0.04	−0.30 ± 0.04
DIRART	−0.13 ± 0.03	−0.19 ± 0.04	−0.23 ± 0.04	−0.24 ± 0.04	−0.20 ± 0.04	−0.18 ± 0.04
NiftyReg	−0.14 ± 0.06	−0.19 ± 0.05	−0.21 ± 0.04	−0.25 ± 0.06	−0.21 ± 0.07	−0.20 ± 0.07
Elastix	−0.21 ± 0.04	−0.26 ± 0.04	−0.27 ± 0.03	−0.29 ± 0.04	−0.26 ± 0.04	−0.25 ± 0.04
Plastimatch	−0.15 ± 0.03	−0.21 ± 0.04	−0.23 ± 0.04	−0.26 ± 0.04	−0.22 ± 0.04	−0.21 ± 0.04
w/o reg.	−0.04 ± 0.02	−0.09 ± 0.04	−0.14 ± 0.05	−0.17 ± 0.04	−0.10 ± 0.04	−0.08 ± 0.04

Table C.3. Transformation smoothness, evaluated by means of the DIRLAB data. The vector field smoothness inside the lung is calculated as the standard deviation of the transformation Jacobian determinant of the lung voxels. All data is given as mean ± standard deviation of the ten DIRLAB data sets.

Algorithm	**Std Jacobian**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	0.10 ± 0.02	0.10 ± 0.03	0.09 ± 0.02	0.07 ± 0.02	0.08 ± 0.02	0.08 ± 0.02
VarReg	0.37 ± 0.09	0.28 ± 0.05	0.24 ± 0.03	0.20 ± 0.04	0.24 ± 0.04	0.26 ± 0.05
DIRART	0.22 ± 0.06	0.12 ± 0.03	0.09 ± 0.02	0.09 ± 0.03	0.13 ± 0.04	0.15 ± 0.05
NiftyReg	0.11 ± 0.03	0.06 ± 0.02	0.04 ± 0.01	0.04 ± 0.02	0.05 ± 0.02	0.06 ± 0.02
Elastix	0.16 ± 0.04	0.14 ± 0.02	0.12 ± 0.03	0.10 ± 0.03	0.11 ± 0.02	0.12 ± 0.03
Plastimatch	0.10 ± 0.03	0.07 ± 0.02	0.05 ± 0.01	0.04 ± 0.02	0.05 ± 0.02	0.06 ± 0.02

Suppl. D. Additional results: DIRLAB data-based correspondence model

Table D.4. Intensity similarity measure-based evaluation of the accuracy of the leave-one-phase-out (LOO) correspondence models, evaluated using the DIRLAB data. Similarity of the left-out phase image that was warped by the DIR-specific model-based predicted motion field and the phase 20 image is evaluated using MSD, NCC and NMI inside the lungs. All data given as mean ± standard deviation of the ten DIRLAB data sets.

Algorithm	**MSD**								
	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50	20 ↦ 60	20 ↦ 70	20 ↦ 80	20 ↦ 90
ANTS	8624 ± 5517	7352 ± 2658	7207 ± 4570	8501 ± 6362	8672 ± 6670	10520 ± 8375	8851 ± 4825	9158 ± 5919	7549 ± 2852
VarReg	8271 ± 5546	7294 ± 2614	6520 ± 4232	7069 ± 5700	6864 ± 5740	8551 ± 7553	7191 ± 3619	8385 ± 5746	6994 ± 2481
DIRART	11265 ± 5593	8529 ± 3368	7653 ± 4323	10168 ± 5674	11148 ± 6481	11413 ± 6686	10288 ± 4900	10343 ± 4669	10361 ± 3433
NiftyReg	10956 ± 5993	8549 ± 3622	8294 ± 5099	11514 ± 7419	13014 ± 9462	14228 ± 10411	11525 ± 6853	10693 ± 5962	9927 ± 4471
Elastix	8791 ± 5670	7649 ± 2632	7117 ± 4477	8051 ± 6072	8118 ± 6213	9896 ± 7858	8344 ± 4277	9188 ± 6099	7730 ± 2867
Plastimatch	10277 ± 5375	8244 ± 3050	7990 ± 4611	10076 ± 6371	10841 ± 6905	12251 ± 8072	10315 ± 5525	10542 ± 6183	9331 ± 3531
w/o reg.	22518 ± 7890	14107 ± 6457	17232 ± 9753	33586 ± 15985	43276 ± 21098	44189 ± 26939	33250 ± 22972	20578 ± 9344	20661 ± 6707
Algorithm	**NCC**								
	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50	20 ↦ 60	20 ↦ 70	20 ↦ 80	20 ↦ 90
ANTS	−0.963 ± 0.021	−0.969 ± 0.011	−0.972 ± 0.015	−0.969 ± 0.019	−0.970 ± 0.017	−0.964 ± 0.024	−0.965 ± 0.021	−0.963 ± 0.027	−0.968 ± 0.014
VarReg	−0.965 ± 0.021	−0.970 ± 0.011	−0.975 ± 0.014	−0.973 ± 0.017	−0.975 ± 0.015	−0.968 ± 0.023	−0.970 ± 0.018	−0.965 ± 0.027	−0.970 ± 0.012
DIRART	−0.952 ± 0.022	−0.964 ± 0.014	−0.970 ± 0.016	−0.960 ± 0.019	−0.958 ± 0.018	−0.956 ± 0.021	−0.958 ± 0.021	−0.957 ± 0.022	−0.956 ± 0.014
NiftyReg	−0.953 ± 0.026	−0.964 ± 0.015	−0.968 ± 0.019	−0.959 ± 0.024	−0.957 ± 0.025	−0.953 ± 0.029	−0.956 ± 0.028	−0.957 ± 0.026	−0.958 ± 0.019
Elastix	−0.962 ± 0.022	−0.968 ± 0.012	−0.972 ± 0.015	−0.970 ± 0.019	−0.971 ± 0.017	−0.964 ± 0.024	−0.966 ± 0.020	−0.962 ± 0.028	−0.967 ± 0.014
Plastimatch	−0.956 ± 0.023	−0.965 ± 0.014	−0.969 ± 0.016	−0.963 ± 0.021	−0.962 ± 0.020	−0.957 ± 0.025	−0.960 ± 0.025	−0.957 ± 0.028	−0.960 ± 0.016
w/o reg.	−0.899 ± 0.045	−0.937 ± 0.039	−0.939 ± 0.038	−0.898 ± 0.047	−0.881 ± 0.051	−0.886 ± 0.060	−0.903 ± 0.059	−0.920 ± 0.039	−0.910 ± 0.035
Algorithm	**NMI**								
	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50	20 ↦ 60	20 ↦ 70	20 ↦ 80	20 ↦ 90
ANTS	−0.20 ± 0.06	−0.21 ± 0.04	−0.23 ± 0.06	−0.21 ± 0.06	−0.22 ± 0.06	−0.20 ± 0.07	−0.20 ± 0.05	−0.20 ± 0.06	−0.20 ± 0.03
VarReg	−0.21 ± 0.06	−0.22 ± 0.04	−0.24 ± 0.06	−0.23 ± 0.07	−0.24 ± 0.07	−0.22 ± 0.07	−0.22 ± 0.05	−0.22 ± 0.06	−0.21 ± 0.03
DIRART	−0.16 ± 0.05	−0.19 ± 0.04	−0.21 ± 0.05	−0.17 ± 0.05	−0.16 ± 0.05	−0.16 ± 0.05	−0.17 ± 0.04	−0.17 ± 0.04	−0.16 ± 0.03
NiftyReg	−0.17 ± 0.06	−0.20 ± 0.04	−0.22 ± 0.06	−0.18 ± 0.06	−0.18 ± 0.07	−0.17 ± 0.07	−0.17 ± 0.05	−0.18 ± 0.05	−0.17 ± 0.04
Elastix	−0.20 ± 0.05	−0.21 ± 0.03	−0.23 ± 0.06	−0.22 ± 0.06	−0.22 ± 0.06	−0.19 ± 0.07	−0.20 ± 0.05	−0.20 ± 0.06	−0.20 ± 0.03
Plastimatch	−0.18 ± 0.05	−0.20 ± 0.04	−0.22 ± 0.05	−0.19 ± 0.06	−0.18 ± 0.06	−0.17 ± 0.06	−0.18 ± 0.05	−0.18 ± 0.05	−0.18 ± 0.03
w/o reg.	−0.09 ± 0.04	−0.14 ± 0.05	−0.17 ± 0.04	−0.10 ± 0.04	−0.08 ± 0.04	−0.09 ± 0.05	−0.11 ± 0.05	−0.12 ± 0.04	−0.10 ± 0.03

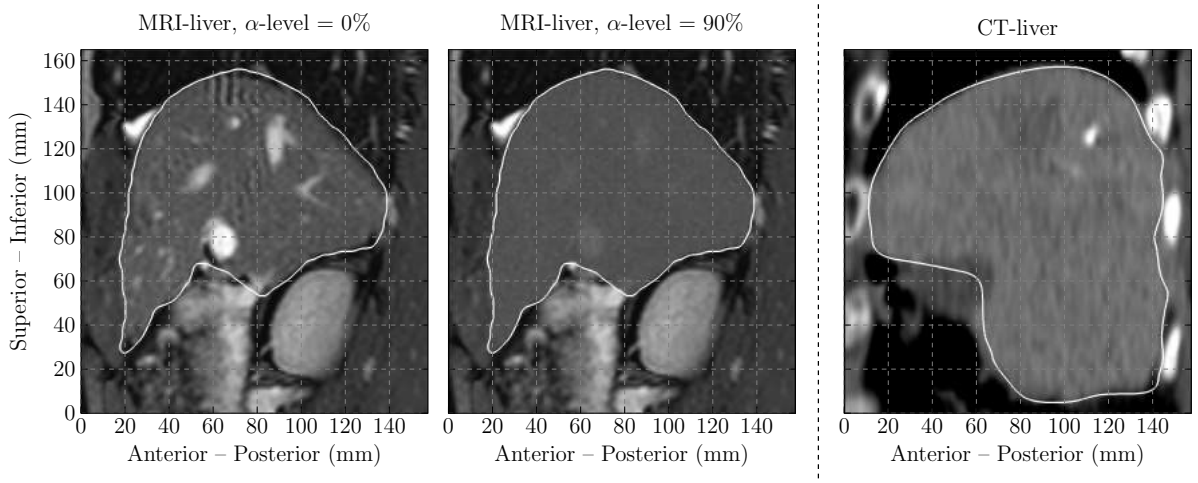


Figure E.1. Visibility of liver structures in the original MRI data (left), with an α-level of 90% (middle) and in a CT phase image of a 4D CT image sequence. By eye, structures inside the liver are hardly detectable for MRI data with $\alpha = 90\%$ and CT data.

Suppl. E. 4D MRI-based evaluation of inner-liver DIR performance

Assessment of the accuracy of DIR algorithms in low contrast areas like the liver in 4D CT data is, compared to high contrast areas like the lungs, hardly feasible due to the lack of structures that could be used to visibly or quantitatively judge registration success. We, therefore, adapted a 4D MRI data with clearly visible inner liver structures to allow for indirect liver DIR accuracy evaluation, using the same evaluation criteria applied for the DIRLAB data sets (TRE, similarity measures, transformation plausibility and smoothness).

Suppl. E.1. MRI data description

Original MRI data. Based on the data provided at `http://www.vision.ethz.ch/4dmri` (see also [5]), a ten-phase 4D MRI data set (representing one breathing cycle)

$$\left(I_i^{\mathrm{MRI}}\right)_{i\in\{1,\ldots,10\}}, I^{\mathrm{MRI}} : \Omega^{\mathrm{MRI}} \subset \mathbb{R}^3 \to \mathbb{R}$$

was selected for analysis of DIR accuracy and plausibility inside the liver. The MRI phase images exhibit clearly visible inner-liver structures, as illustrated in Fig. E.1 (left). To allow for TRE computation during DIR evaluation, an expert selected 20 corresponding inner-liver landmarks (prominent anatomical points like vasculature bifurcations) in the individual 4D MRI phase image.

(Partially) Masked MRI data. To mimic the challenges of liver registration in 4D CT images, we generated a series of 4D MRI with synthetically reduced structure-to-background contrast inside the liver. First, an expert manually segmented the liver in all phase images in the original 10-phase MRI data set. Let these images be denoted by

$$\left(M_i^{\mathrm{MRI}}\right)_{i\in\{1,\ldots,10\}}, M^{\mathrm{MRI}} : \Omega^{\mathrm{MRI}} \to \mathbb{R}$$

with $M_i^{\mathrm{MRI}}(x)$ being the median value of the liver voxels in I_i^{MRI} and zero elsewhere.
Mask and original MRI data were combined to form test data sets

$$\left(I_i^{\mathrm{MRI},\alpha}\right)_{i\in\{1,\ldots,10\}}, I^{\mathrm{MRI},\alpha} : \Omega^{\mathrm{MRI}} \to \mathbb{R} \tag{E.1}$$

with

$$I_i^{\mathrm{MRI},\alpha}(x) = \begin{cases} (1-\alpha) I_i^{\mathrm{MRI}}(x) + \alpha\left[M_i^{\mathrm{MRI}}(x) + n(x)\right] & \text{if } M_i^{\mathrm{MRI}}(x) \neq 0 \\ I_i^{\mathrm{MRI}}(x) & \text{else} \end{cases}$$

The additive term $n(x)$ represents Gaussian noise with $\mu = 0$ and σ^2 derived by analysis of the liver voxel intensity value distribution in the 4D CT data, but rescaled to the intensity dynamic range of the MRI data to yield CT-like noise characteristics inside the liver of the $I_i^{\mathrm{MRI},\alpha}$ image data.

Suppl. E.2. 4D MRI-based experiments and DIR evaluation

In total, image sequences $\left(I_i^{\mathrm{MRI},\alpha}\right)_{i\in\{1,\dots,10\}}$ with α-levels of 0%, 80%, 85%, 90%, 95%, and 100% (with $I_i^{\mathrm{MRI},0\%} = I_i^{\mathrm{MRI}}$) were computed. For each α-level, all DIR algorithms were applied for extreme phase registration ($00 \leftrightarrow 50$) of the ten-phase 4D MRI data. DIR parameters were the same as applied for 4D CT registration. DIR accuracy for the different α-levels was evaluated using the landmarks identified in the original (i.e. the $\alpha = 0\%$) images.

It can be assumed that the transformation properties and DIR accuracy in low contrast areas strongly depend on the DIR regularization weight. To better understand especially the relationship between regularization weight and DIR accuracy for different contrast levels, we repeated the said evaluation using *VarReg* [2].

Aforementioned experiments were finally complemented using the entire set of evaluation metrics described and applied for DIR of the DIRLAB data sets. For reasons of clarity and comprehensibility, we restricted the evaluation to the images of an α-level of 90%, which visibly appeared to be most similar to the clinical 4D CT data in terms of inner liver structure visibility (cf. Fig. E.1).

Suppl. E.3. 4D MRI-based DIR evaluation: Results

Suppl. E.3.1. TRE for different α-levels

The TRE values for DIR in the 4D MRI data with different α-levels are summarized in Table E.5. Mean landmark distance before registration was 14.1 mm. High TRE values for *DIRART* are noticeable and therefore visualized and compared to VarReg results in Figure E.2. Liver borders are well-aligned after *DIRART* registration; however, missing contrast inside the liver obviously led to less accurate results compared to *VarReg* for all α-levels $\geq 80\%$. It should, however, be mentioned that even for DIR of the original MRI data, *DIRART* results in by far highest TRE values, indicating that the chosen *DIRART* parameters are not ideally suited for 4D MRI registration. For the other DIR algorithms, the $\alpha = 0\%$ TRE values are at least in a similar order compared to the DIRLAB data and taking into account the relatively large landmark distances between registration. Focusing on the selected reference noise level of 90%, *VarReg*, *NiftyReg* and *Elastix* achieve best TRE results. *ANTS* and *Plastimatch* registration yield up to 1.5 mm higher values.

Table E.6 further illustrates the influence of varying regularization strength on *VarReg* DIR accuracy. A regularization of around $\sigma^2 = 2$ mm results in lowest TRE for almost all noise levels (chosen σ^2 during DIRLAB registration: 2 mm). However, the influence of the regularization strength (except for $\sigma^2 \leq 1$) on the registration accuracy is small even for larger α-values and lower inner-liver contrast, respectively.

Suppl. E.3.2. Additional DIR evaluation criteria

Results for the transformation plausibility and smoothness evaluation by means of the transformation Jacobian determinant are summarized in Table E.7 and Table E.8, respectively. The values of the different similarity measures evaluated by comparison of warped and reference image (mean squared differences, normalized cross correlation and normalized mutual information) are shown in Table E.9.

Table E.5. DIR accuracy for the different DIR algorithms for extreme phase registration ($00 \leftrightarrow 50$), evaluated by means of the MRI data sets with different α-levels. TRE without registration is 14.1 mm

α-level	**Mean TRE of $5 \mapsto 0$ registration of MRI data set (mm)**					
	ANTS	VarReg	DIRART	NiftyReg	Elastix	Plastimatch
100%	6.96 ± 3.14	7.68 ± 2.90	8.97 ± 2.22	3.20 ± 1.57	3.61 ± 1.81	5.70 ± 2.29
95%	4.86 ± 3.74	5.94 ± 3.80	9.14 ± 2.25	3.10 ± 1.48	3.16 ± 1.92	4.27 ± 1.96
90%	4.41 ± 3.74	2.93 ± 2.02	9.17 ± 2.25	2.93 ± 1.38	2.68 ± 1.84	3.79 ± 1.84
85%	4.22 ± 3.62	2.47 ± 1.52	9.19 ± 2.26	2.78 ± 1.36	2.33 ± 1.61	3.44 ± 1.59
80%	3.97 ± 3.46	1.90 ± 1.23	9.07 ± 2.24	2.56 ± 1.37	1.93 ± 1.35	2.97 ± 1.68
0%	3.58 ± 3.33	2.02 ± 1.31	5.99 ± 2.39	2.31 ± 1.39	2.58 ± 2.02	2.26 ± 1.34

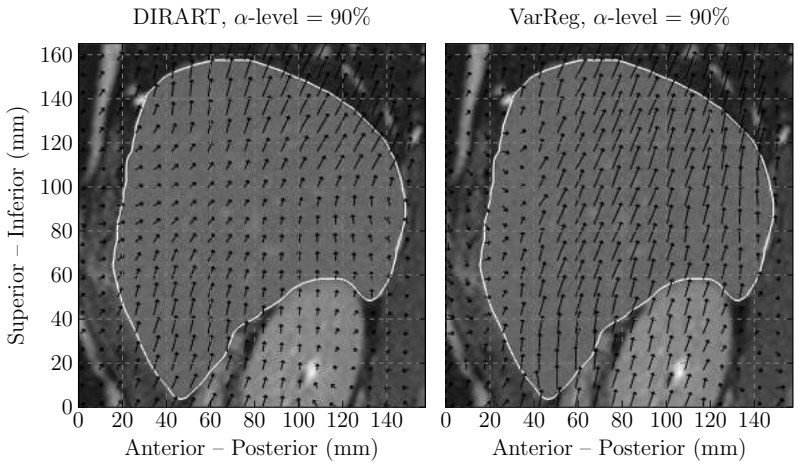


Figure E.2. Visualization of vector fields inside the liver for *DIRART* (left) and *VarReg* (right) resulting of phase 5 ↦ 0 registration.

Table E.6. DIR accuracy for *VarReg* for extreme phase registration (00 ↔ 50), evaluated by means of the MRI data sets with different α-levels added. The influence of the regularization strength is investigated by varying σ^2 (Gaussian regularization) of the algorithm. TRE without registration is 14.1 mm

α-level	**Mean TRE of 5 ↦ 0 registration of MRI data set (mm)**							
σ^2	0.5	1	1.5	2	2.5	3	3.5	4
100%	12.60 ± 4.31	8.74 ± 3.21	7.73 ± 2.87	7.68 ± 2.90	**7.60 ± 2.83**	7.74 ± 2.77	7.63 ± 2.72	7.61 ± 2.65
95%	10.23 ± 5.21	7.21 ± 4.12	6.66 ± 4.08	5.94 ± 3.80	5.80 ± 3.69	5.67 ± 3.64	5.57 ± 3.36	**5.46 ± 3.29**
90%	8.97 ± 5.86	5.50 ± 4.27	3.10 ± 2.20	**2.93 ± 2.02**	3.17 ± 2.29	3.21 ± 2.27	3.40 ± 2.39	3.39 ± 2.39
85%	4.65 ± 4.16	2.52 ± 1.85	2.60 ± 1.87	**2.47 ± 1.52**	2.61 ± 1.89	2.66 ± 1.91	2.71 ± 1.96	2.76 ± 2.00
80%	1.90 ± 1.23	1.89 ± 1.09	**1.83 ± 1.13**	1.90 ± 1.23	2.07 ± 1.35	2.16 ± 1.40	2.16 ± 1.40	2.29 ± 1.53
0%	2.46 ± 1.73	2.08 ± 1.34	2.04 ± 1.29	**2.02 ± 1.31**	2.06 ± 1.39	2.07 ± 1.37	2.08 ± 1.36	2.08 ± 1.35

Table E.7. Accuracy of the different DIR algorithms, evaluated by means of the MRI data set with α-level of 90%: Mean transformation Jacobian determinant inside the liver. The respective liver volume ratio as based on the manual liver segmentation data of the phase images refers as ground truth.

Noise	**Mean Jacobian**					
	00 ↔ 50	20 ↦ 00	20 ↦ 10	20 ↦ 30	20 ↦ 40	20 ↦ 50
ANTS	1.02	0.99	0.99	1.01	1.01	1.01
VarReg	1.00	0.98	0.98	0.99	1.00	1.00
DIRART	1.03	0.98	0.98	1.01	1.02	1.02
NiftyReg	0.99	0.97	0.99	1.00	1.00	1.00
Elastix	0.99	0.98	0.98	0.99	1.00	0.99
Plastimatch	1.03	0.98	0.99	1.00	1.01	1.02
Mask volume ratio	1.01	0.99	0.99	1.01	1.02	1.02

5.3. Dose simulation in the presence of image artifacts

Typical image artifacts in 4D CT data sets, as introduced in Section 3.1.2, can, as briefly evaluated in the last section, have an unfavourable impact on subsequent steps like image registration or 4D dose accumulation. In this section, this impact is further analyzed by manipulating artifact-free rated 4D CT data to contain typical artifact types and, subsequently, use the artifact-containing data for mentioned processing steps. The described methods and results are based on the publication below:

T. Sothmann, T. Gauer, R. Werner. Influence of 4D CT motion artifacts on correspondence model-based 4D dose accumulation. In: *Proc SPIE*, Vol. 10576, 105760F. International Society for Optics and Photonics, SPIE, 2018.

Out of the in-house patient cohort (cf. Section 4.4.4), six 4D CT data sets visually rated as artifact-free were selected for this study. To examine the influence of artifacts in 4D CT image data on correspondence modeling and 4D dose simulation, the corresponding 4D CT raw data of the selected six data sets were retrospectively manipulated to induce typical image artifacts near the target volume. Each z-slice in an already reconstructed 4D CT phase corresponds to a specific time point in the acquisition process, that is, the position of the GTV, delineated in the artifact-free 4D CT data, can be related to an interval between two time points in the simultaneously recorded external patient breathing signal (see upper part of Figure 1 [SPIE 2018]). As the patient respiration during CT imaging is utilized for reconstruction, i. e. binning of acquired 4D CT projections as described in detail in Section 3.1.1, manipulation of the breathing cycle in pre-defined temporal interval impacts the reconstruction result. Steep breathing gradients lead to faulty bin assignments and thus to a duplication of structures (double structure artifact). Removing the breathing cycle leads to missing information and thus an interpolation between slices (interpolation artifact, see lower part of Figure 1 [SPIE 2018] and Fig. 3.3 for both types of artifacts). These two artifact types were investigated for all six patient data sets and compared to the artifact-free data set by evaluating the subsequently built correspondence model using mean motion magnitudes inside the ITV as well as motion vectors in (x, y, z)-direction. Further, 4D-simulated dose distributions were compared.

Results for evaluation of correspondence models and dose simulations are summarized in Table 1 [SPIE 2018]. Correspondence models built using 4D CT data affected with double structure artifacts seem to underestimate the motion inside the ITV, but the influence on subsequent dose simulation is only minor. For the interpolation artifact-affected data, however, predicted motion vector fields are for some cases even flipped (e. g. z-motion for case 5) and thus resulting 4D-simulated dose distribution show noticeable deviations compared to the reference dose distribution. Those findings suggest that the

in reality considerably more often occurring interpolation artifact has to be considered as relevant source of uncertainty in subsequent dose simulation processing steps. Further, the influence of interpolation artifacts on e. g. radiotherapy-relevant structure delineation[4] should be investigated. For dose simulation, an estimation of registration uncertainties in artifact-affected image data is desired to allow for error propagation and thus computing, for instance, a confidence interval for resulting dose distributions. This is, however, not trivial for standard DIR approaches. In the following section, therefore, a new deep learning-based DIR framework for CT data is proposed that is able to estimate corresponding registration uncertainties.

[4] As delineation in radiotherapy is conducted in axial planes, the interpolation artifacts, which are mainly visible in coronal and sagittal planes are not considered.

Influence of 4D CT Motion Artifacts on Correspondence Model-based 4D Dose Accumulation

Thilo Sothmann[a,b], Tobias Gauer[b], and René Werner[a]

[a]Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[b]Department of Radiotherapy and Radio-Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

ABSTRACT

In radiotherapy (RT) of moving targets, motion artifacts in 4D CT planning data can be hypothesized to influence accuracy of RT treatment planning steps. Especially results of deformable image registration (DIR) of 4D CT phase images and DIR-based dose accumulation/4D dose simulation can be assumed to be directly affected. In this study, the influence of typical 4D CT "double structure" and "interpolation" artifacts on correspondence model-based 4D dose simulation is investigated. The correspondence model correlates patient-specific DIR-based internal motion information and external breathing signals, which allows for integration of respiratory variability into 4D dose simulation. Artifact-free 4D CT data of 6 lung and liver cancer patients were manipulated to contain mentioned artifacts. Correspondence model-based dose accumulation was performed in both artifact-free and artifact-affected data sets. Overall, the effect of "double structure" artifacts was negligible, whereas "interpolation" artifacts noticeably influenced dose accumulation accuracy.

Keywords: 4D CT, motion artifacts, deformable image registration, correspondence modeling, 4D dose accumulation

1. INTRODUCTION

Stereotactic body radiation therapy (SBRT) is an effective method to treat lung and liver tumors.[1] However, one of the biggest challenges is still the consideration of tumor motion and deformation while treatment planning.[2] In current clinical practice, a time resolved CT (3D+t, 4D CT) is used to define the tumor motion space, the so called internal target volume (ITV). Subsequently, the ITV is the basis of treatment planning and dose application. Patient-specific intra- and inter-fractional respiratory variability, can nevertheless lead to deviations between actual delivered and planned dose distributions. Thus, the representation of target volume dynamics by the planning 4D CT is further incorporated into dose distribution computation and optimization. This allows for precise 4D (3D+t) dose simulation and accumulation in RT to offer an understanding of motion-induced differences between planned and actual delivered dose and its potential interrelation with clinical outcome. The respective combination of patient-specific motion data and information about the dynamic dose delivery process often relies on motion information extracted by DIR from patient-specific 4D CT RT treatment planning image data.[3,4] However, clinical 4D CT data often suffer from motion artifacts (cf. Figure 1 for typical artifacts), which in turn can be hypothesized to affect the accuracy of subsequent DIR and dose accumulation.
This hypothesis motivates investigating the actual influence of typical 4D CT artifact types, e. g. "double structures" and "interpolation" artifacts, on following processing steps. Here, we retrospectively manipulated breathing curves of 6 lung and liver patients that were acquired for 4D CT reconstruction purposes at predetermined time points to induce incorrect 4D CT reconstruction (Figure 1). Based on a 4D dose accumulation approach for volumetric arc therapy (VMAT)[5,6] in combination with a well-evaluated correspondence modeling[7] and an open-source DIR framework,[8] we analyzed the artifact influence on DIR, correspondence modeling and correspondence model-based 4D dose accumulation results.

Send correspondence to Thilo Sothmann; E-mail: t.sothmann@uke.de

Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by Baowei Fei, Robert J. Webster III, Proc. of SPIE Vol. 10576, 105760F · © 2018 SPIE
CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2291481

2. METHOD(S)

This section first introduces the concept of correspondence model-based 4D dose accumulation. Second, the patient collective and the performed simulations are described.

2.1 Concept of Correspondence Model-based 4D Dose Accumulation

A correspondence model represents a patient-specific functional relationship between external breathing signal measurements and respiratory motion of internal structures (target volumes, organs at risk). Here, multivariate regression-based correspondence modeling is applied that, briefly speaking, consists of the following steps:[7]

Input 4D treatment planning CT patient data sets consist of a series of 3D CT images, i. e.

$$(I_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}},\ I_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R} \tag{1}$$

with $i \in \{1,\ldots,n_{\mathrm{ph}}\}$ as the breathing phase of a 3D CT image and n_{ph} the total number of breathing phases (here: $n_{\mathrm{ph}} = 10$). Analogously, the normally one-dimensional breathing signal measurements acquired during 4D CT imaging that are assigned to the images read

$$(\zeta_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}},\ \zeta_i = (z_i, \partial_t z_i)^T \in \mathbb{R}^2. \tag{2}$$

Here, z_i and ∂z_i are the anterior-posterior (AP) component of the Varian real-time position management (RPM)-system and its time derivative. The functional relationship between external breathing signal measurement $(\zeta_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}}$ and internal breathing phase-specific images I_i is represented by the sought correspondence model. Representation of internal motion is achieved by displacement vector fields (DVF) estimated by non-linear registration.

Estimate motion fields Let reference phase (here: $i_0 = 3$, mid expiration phase) and the corresponding CT image I_{i_0} as fixed image during DIR[8] be given. The registration process results in a series of transformations

$$(\varphi_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}},\ \varphi_i : \Omega \to \Omega \tag{3}$$

with $\varphi_{i_0} = \mathrm{id}$ and corresponding motion fields

$$(u_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}},\ u_i : \Omega \to \mathbb{R}^3 \tag{4}$$

with $u_i = \varphi_i - id$ (i. e. $u_{i_0} = 0$). The breathing signal measurements $(\zeta_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}}$ and motion fields $(u_i)_{i\in\{1,\ldots,n_{\mathrm{ph}}\}}$ are the basis for and correlated during correspondence model training using a multivariate regression approach.

Multivariate regression Breathing signal measurements and motion fields are interpreted as random variables $\mathbf{Z}_i$ ($\equiv \zeta_i$) and $\mathbf{U}_i \in \mathbb{R}^{3m}$ with m denoting the number of voxels of the reference phase image I_{i_0}. The correspondence model is finally defined by

$$\hat{\mathbf{U}} = \overline{\mathbf{U}} + \mathbf{B}\left(\hat{\mathbf{Z}} - \overline{\mathbf{Z}}\right). \tag{5}$$

At this, $\hat{\mathbf{Z}} \in \mathbb{R}^2$ represents a breathing signal observation and $\hat{\mathbf{U}} \in \mathbb{R}^{3m}$ the corresponding and sought internal motion field. The coefficient matrix $\mathbf{B} \in \mathbb{R}^{3m\times 2}$ is, based on the above mentioned tuples $(\mathbf{U}_i, \mathbf{Z}_i)$, computed in an ordinary least-squares regression approach.

Using the patient-specific functional relationship between external and internal motion data, correspondence model-based 4D dose simulation is conceptually straightforward:

4D dose simulation With the time-dependent dose rate during a single treatment fraction, $\dot{D} : \Omega \times \mathcal{T} \subset \mathbb{R} \to \mathbb{R}_+$, the dynamical dose delivery process, its interplay with patient motion and the resulting dose distribution $D_{4\mathrm{D}} : \Omega \to \mathbb{R}_+$ can be expressed as

$$D_{4\mathrm{D}}(x) = \sum_{t \in \tilde{\mathcal{T}}} D_t(\varphi(x,t))$$

with $\varphi : \Omega \times \mathcal{T} \to \mathbb{R}^3$ as the position $\varphi(x,t)$ of voxel $x \in \Omega$ of the reference phase CT I_{i_0} at time $t \in \mathcal{T} = [0;T) \subset \mathbb{R}$. With $\hat{\varphi} = id + \hat{u}$ and $\hat{\zeta} : \mathcal{T} \to \mathbb{R}^2$ as $\hat{\zeta}$ as the patient breathing signal acquired during dose delivery, correspondence model-based 4D dose simulation for a single fraction reads

$$D_{4\mathrm{D}}(x) = \sum_{t \in \tilde{\mathcal{T}}} D_t\left(x + \hat{u}\left(x, \hat{\zeta}_t\right)\right) \tag{6}$$

where $\hat{\zeta}_t = \hat{\zeta}(t)$.

In the current study, all patients were treated with VMAT (here: RapidArc, Varian). In contrast to standard static field intensity modulated RT techniques, VMAT techniques utilize dose rate and gantry speed variation as well as leaf modulation to optimize the planned dose distribution characteristics. By replacing the temporal variable for each planned arc by the gantry rotation angle, Equation (6) can be re-parametrized to

$$D_{4\mathrm{D}}(x) \approx \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} D_\alpha\left(x + \hat{u}\left(x, \hat{\zeta}_\alpha\right)\right) \tag{7}$$

with $\tilde{\mathcal{A}}_{\mathrm{ax}} = \{1/2\,\Delta\alpha; 3/2\,\Delta\alpha; \dots\}$ as a discretized version of the gantry angle range $\mathcal{A}_{\mathrm{ax}} \subset [0°; 360°)$ of the respective VMAT arc.[5]

Using the treatment planning system Eclipse 13 (Varian Medical Systems), the minimum achievable angle segment size was $\Delta\alpha = 2.3°$, which, depending on the respective gantry speed, corresponds to Δt_{Dose}-values between 1 s and 2.5 s. In contrast, the breathing signal acquisition (using the Varian RPM system) had a frequency of 25 Hz, i. e. $\Delta t_{\mathrm{RPM}} = 0.04\,s$. Taking into account this imbalance and aiming at dose simulation accuracy as high as possible, the finally implemented single fraction accumulation scheme was

$$\begin{aligned} D_{4\mathrm{D}}(x) &\approx \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} \sum_{t \in \tilde{\mathcal{T}}_\alpha} D_\alpha\left(x + \hat{u}\left(x, \hat{\zeta}_t\right)\right) \\ &= \sum_{\mathrm{ax}} \sum_{\alpha \in \tilde{\mathcal{A}}_{\mathrm{ax}}} \sum_{t \in \tilde{\mathcal{T}}_\alpha} \left(D_\alpha \circ \hat{\varphi}\right)\left(x, \hat{\zeta}_t\right) \end{aligned} \tag{8}$$

with $\tilde{\mathcal{T}}_\alpha$ denoting the RPM measurements during dose delivery at gantry angle α and the respective dose segment.

Effects arising from the limited VMAT arc discretization due to the treatment planning system capabilities were subject to extensive motion phantom studies performed as a pre-study to the current work and (for the chosen minimum angle range of 2.3°) shown to be small. Respective results and further discussion of additional technical factors that could mitigate the accuracy of the applied 4D VMAT dose simulation can be found in Sothmann *et al* 2017.[5]

2.2 Patient Collective and Simulations

In total, six visually rated as artifact-free 4D CT data sets and corresponding treatment plans of previously at our hospital treated cancer patients (three liver and three lung metastases) were selected for this study. The tumor volume was delineated in each phase of the ten-phase 4D CT yielding 10 gross tumor volume (GTV) per patient. The union of all GTV results in the motion encompassing internal target volume (ITV), and the ITV plus predefined safety margin in the planning target volume (PTV), which is the basis of patient treatment

planning. All data sets were evaluated using the workflow described in Section 2.1, yielding a reference 4D dose distribution.
Further, retrospective manipulation of the actual reconstruction CT breathing curves at predetermined time points, corresponding to z-positions of the GTV delineated in reference phase 3 (cf. Figure 1), was performed to induce 4D CT artifacts ("double structure" and "interpolation") during reconstruction. The artifact-containing image data sets were again processed using the workflow described in Section 2.1, resulting in one "double structure" and one "interpolation" artifact-affected correspondence model and 4D dose distribution per patient. The effects of the artifacts on correspondence modeling and 4D dose accumulation were analyzed. Effects on the correspondence models were evaluated by predicted motion field statistics, i. e. mean motion magnitudes inside the ITV as well as motion vectors in x, y, z-direction. Comparison of artifact-free and artifact-containing 4D dose distributions were performed by ΔD_{98}, i. e. the difference between D_{98} of the artifact-free and D_{98} of the artifact-containing dose distribution with D_{98} as the dose to 98% of the GTV.

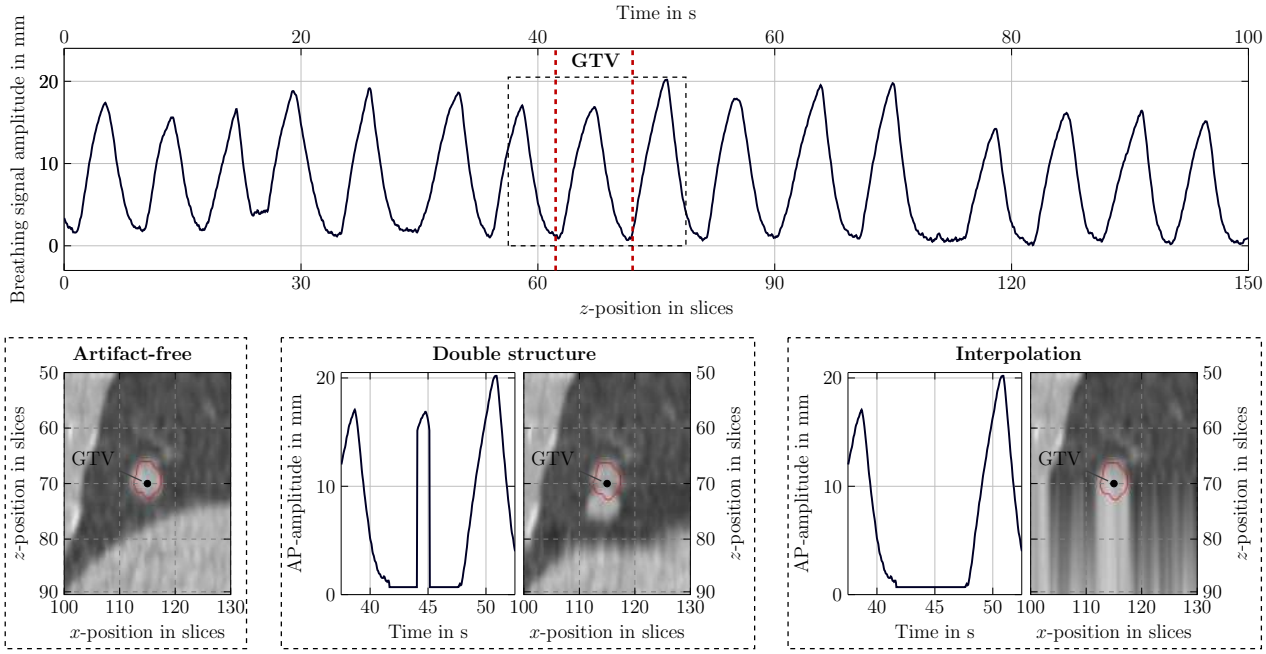


Figure 1. Manipulation of reconstruction CT breathing curve and resulting 4D CT motion artifacts. The top figure shows an actual patient breathing curve acquired during 4D CT scanning for subsequent 4D CT reconstruction. The retrospectively calculated z-position of the GTV is illustrated. Without breathing curve manipulation, an artifact-free CT image is generated, see bottom left figure. "Sorting artifacts (double structure)" (bottom middle) and "interpolation (missing data) artifacts" (bottom right) occur when the reconstruction curve is manipulated as shown.

3. RESULTS

Artifact-induced differences in D_{98} are summarized in Table 1. Further, mean magnitudes inside the ITV of DVF resulting from registering 4D CT phase 6 to 3 are presented. Detailed motion information (x, y, z-direction) is given in brackets beneath. Results for "double structure" artifacts show only minor differences between artifact-free and artifact-contained dose distribution with a range of ΔD_{98} from $-0.23\,\mathrm{Gy}$ to $+0.26\,\mathrm{Gy}$. Apparently, computed deviations of mean motion magnitudes inside the ITV have no considerable influence on resulting dose distributions, although it has to be considered that motion magnitudes are only evaluated for one phase registration ($6 \mapsto 3$).
Dose simulations for induced "interpolation" artifacts show for four out of six patients similar to "double structure" artifacts no considerable deviations (ΔD_{98} range of $-0.12\,\mathrm{Gy}$ to $-0.01\,\mathrm{Gy}$). Dose distributions of patient 1 and 5, however, are subject to dose differences of $-2.52\,\mathrm{Gy}$ and $-1.87\,\mathrm{Gy}$ compared to the artifact-free dose,

respectively. Motion magnitudes can again not necessarily be correlated to mentioned deviations but a large negative z-motion component compared to a large positive z-motion component of the artifact-free DVF for patient 5 motivates further investigation. Thus, a visualization of the results of patient 5 is given in Figure 2. The artifact influence on the DVF is clearly visible for both artifact cases. However, the "interpolation" artifact actually flips the motion field in z-direction and has therefore a bigger impact on subsequent 4D dose accumulation, as seen in the bottom row of Figure 2.

Table 1. GTV dose coverage of artifact-free and artifact-contained dose distributions. Coverage is quantified by differences of $D_{98,\text{Artifact}}$ and $D_{98,\text{Artifact-free}}$. Further, motion magnitudes (mean ± std) inside the ITV for artifact-free/artifact-contained 4D CT phase registration (phase 6 to 3) as well as mean motion vectors $(x, y, z$-direction) are shown.

	Pat.	$D_{98,\text{Artifact}} - D_{98,\text{Artifact-free}}$ (Gy)		Mean 6↦3 motion magnitude in ITV (mm)		
		Double structure	Interpolation	Artifact-free	Double structure	Interpolation
Liver	1	+ 0.26	− 2.52	9.33 ± 2.20 (0.42, 2.92, 8.76)	4.27 ± 2.16 (0.25, 1.25, 3.94)	8.83 ± 2.53 (0.66, 0.62, 8.59)
	2	+ 0.06	− 0.12	9.99 ± 2.36 (−0.36, 2.29, 9.51)	5.23 ± 1.68 (−0.42, 0.72, 5.05)	4.61 ± 1.71 (−0.83, 2.48, 3.58)
	3	+ 0.09	− 0.04	7.02 ± 1.46 (2.09, 2.02, 6.24)	4.95 ± 1.83 (1.11, 1.40, 4.53)	5.75 ± 1.42 (1.48, 1.47, 5.28)
Lung	4	+ 0.04	− 0.11	3.51 ± 0.41 (−0.54, −0.48, 3.37)	2.35 ± 0.57 (−0.43, −0.21, 2.25)	3.33 ± 1.15 (2.77, −1.40, 0.67)
	5	− 0.23	− 1.87	5.93 ± 1.42 (0.67, −1.82, 4.90)	0.98 ± 0.63 (0.45, 0.17, −0.24)	6.26 ± 1.08 (−1.30, 0.22, −6.08)
	6	+ 0.01	− 0.01	3.32 ± 0.43 (−0.41, 1.39, 2.94)	2.67 ± 0.33 (−0.52, 1.41, 2.16)	2.82 ± 0.62 (−0.27, 1.39, 2.14)

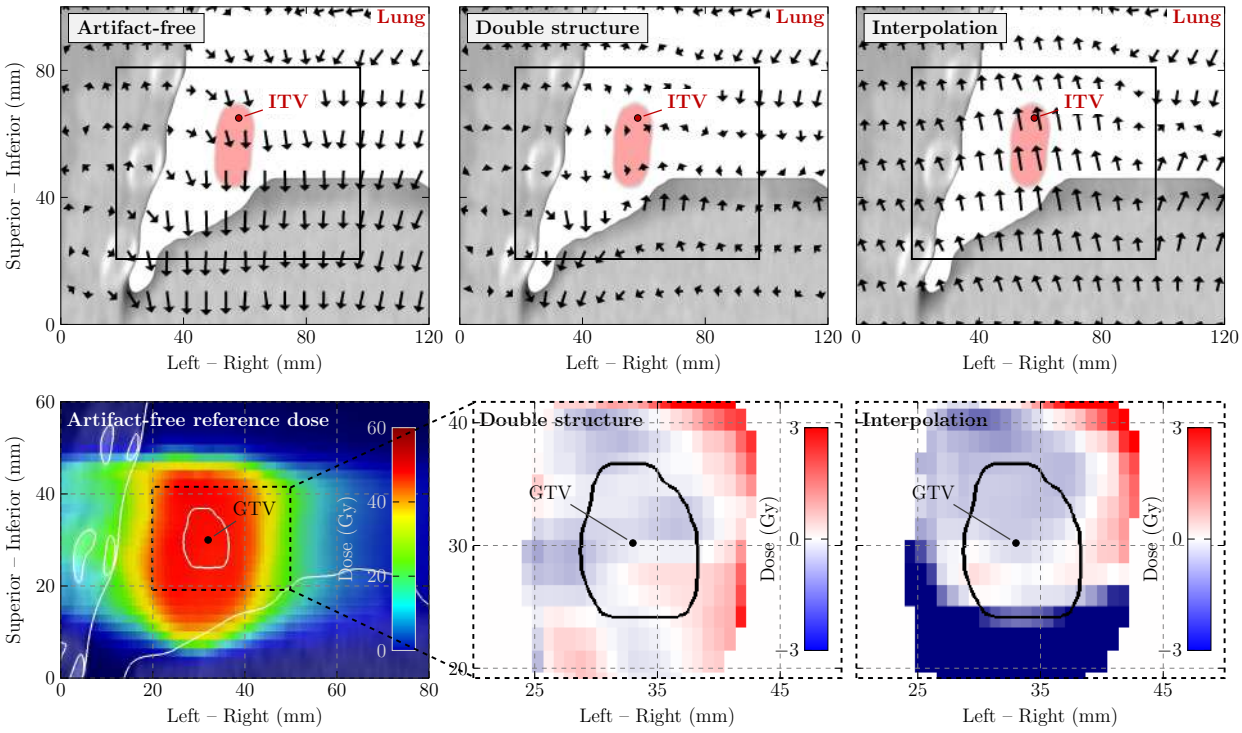


Figure 2. Illustration of results for patient 5. Top row: Comparison of DVF resulting from artifact-free/artifact-contained 4D CT phase registration (phase 6 to 3). Bottom row: Artifact-free reference dose (left) and dose difference between retrospectively simulated and reference dose distributions for "double structure" (middle) and "interpolation" (right) artifacts.

A following test of the correspondence model inside the ITV reveals a by factor 2.5 higher mean Euclidean distance between "interpolation" artifact and artifact-free correspondence model prediction for a ROI around the GTV compared to the mean Euclidean distance between "double structure" artifact and artifact-free prediction, as shown in Figure 3 for patient 5.

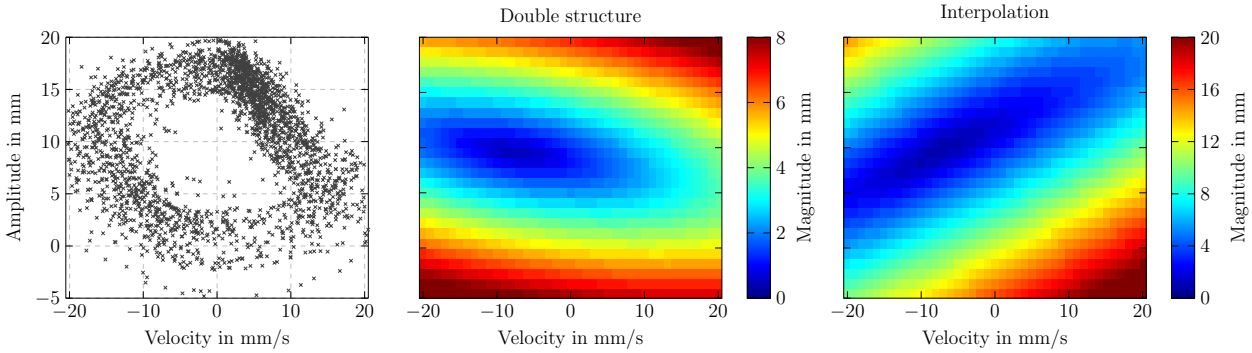


Figure 3. Test of correspondence model. Left: Joint histogram of motion and velocity of CT reconstruction curve (patient 5) spans the motion space in which the correspondence model prediction is evaluated. Middle and right: Mean Euclidean distance between artifact-containing and artifact-free correspondence model prediction inside the defined motion space. Please note that shown color maps have different maximum values.

4. CONCLUSIONS

As expected and illustrated in Section 3, the results of our study show that 4D CT motion artifacts have an impact on correspondence modeling and 4D dose accumulation. However, "double structure" artifacts in liver cancer data sets are shown to have only a minor influence on subsequent dose accumulation, which was somewhat expected. The same applies for lung patients, as the artifact often only occur in two or three 4D CT phases and, thus, are mainly averaged out during the correspondence model training process. In our case, "interpolation" artifacts occur in all 4D CT phases and, therefore, can have a bigger influence, as seen in Table 1 for patient 1 and 5. Evaluated motion magnitudes of one phase registration (phase $6 \mapsto 3$) can not necessarily be correlated to computed dose deviations between artifact and artifact-free D_{98} values.

ACKNOWLEDGMENTS

This work was in parts funded by the University of Hamburg (scholarship to TS).

REFERENCES

[1] Wulf, J., Hdinger, U., Oppitz, U., Thiele, W., Ness-Dourdoumas, R., and Flentje, M., "Stereotactic radiotherapy of targets in the lung and liver," *Strahlenther Onkol* **177**, 645–655 (2001).

[2] Keall, P. J., Mageras, G. S., Balter, J. M., Emery, R. S., Forster, K. M., Jiang, S. B., Kapatoes, J. M., Low, D. A., Murphy, M. J., Murray, B. R., Ramsey, C. R., Van Herk, M. B., Vedam, S. S., Wong, J. W., and Yorke, E., "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med Phys* **33**, 3874–3900 (2006).

[3] Velec, M., Moseley, J. L., Craig, T., Dawson, L. A., and Brock, K. K., "Accumulated dose in liver stereotactic body radiotherapy: positioning, breathing, and deformation effects," *Int J Radiat Oncol Biol Phys* **83**, 1132–1140 (2012).

[4] Samavati, N., Velec, M., and Brock, K. K., "Effect of deformable registration uncertainty on lung SBRT dose accumulation," *Med Phys* **43**, 233 (2016).

[5] Sothmann, T., Gauer, T., and Werner, R., "4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters," *PLoS One* **12**, e0172810 (2017).

[6] Sothmann, T., Gauer, T., Wilms, M., and Werner, R., "Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT," *Phys Med Biol* **62**, 9001–9017 (2017).

[7] Wilms, M., Werner, R., Ehrhardt, J., Schmidt-Richberg, A., Schlemmer, H.-P., and Handels, H., "Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy," *Phys Med Biol* **59**, 1147–1164 (2014).

[8] Werner, R., Schmidt-Richberg, A., Handels, H., and Ehrhardt, J., "Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study," *Phys Med Biol* **59**, 4247–4260 (2014).

5.4. Deep learning-based deformable image registration

Directly motivated by findings of the previous two sections, implementation of a DIR framework that is able to estimate registration uncertainties seemed to be necessary for further evaluation of the proposed 4D dose simulation algorithm and its uncertainties and limitations. Traditional DIR algorithms, as analyzed regarding their performance in lung and liver registration in Section 5.2, are commonly used and optimized for specific registration tasks but have the risk of getting stuck in local minima during optimization and are time consuming. Therefore, the idea was to develop a deep learning-based DIR framework for fast registration of clinical thoracic and abdominal 4D CT data, which is presented in the following publication:

T. Sentker[†], F. Madesta[†] and R. Werner. GDL-FIRE^{4D}: Deep Learning-based Fast 4D CT Image Registration. In: *Lect Notes Comput Sc*, 765–773. Springer, 2018.

The aim of this study was to 1) propose a general and efficient CNN-based framework, 2) evaluate the proposed method by means of 4D CT benchmark data bases, i. e. DIRLAB and CREATIS (cf. Section 4.4.1 and Section 4.4.2, respectively, as well as Table 4.1), and 3) illustrate and analyze first dropout-generated registration uncertainty maps. To understand the basic principle of deep learning-based image registration, an introduction has been given in Section 4.3. The utilized CNN architecture for the presented approach is shown in Fig. 1 [MICCAI 2018]. The application of pre-trained autoencoders allowed for a deeper network structure. Dropouts in deeper layers extended the network to be probabilistic and further aimed at an intrinsic and generalized DIR representation. Utilizing dropouts during prediction and repeated motion estimations enabled computing the sought motion field as the mean of the sampled predicted fields. Additionally, voxel-wise variances can be interpreted as local registration uncertainty estimates. Training and testing of proposed CNN was done by in-house acquired 4D CT data sets of 69 ten-phase 4D CT data sets (cf. Section 4.4.4), split into 85% and 15% train and test, respectively. Pseudo ground truth data was generated using traditional open source DIR frameworks (NiftyReg, Plastimatch, VarReg) in a plug-and-play manner to register phase images of 4D CT data sets, i. e. the CNN was trained to learn the relationship between the moving image I_M and fixed image I_F with given registration transformation φ. For each DIR algorithm, respective probabilistic CNN variants were built and cascaded up to four times. Evaluation of the output was again conducted using the TRE computed by means of the landmarks publicly available for the DIRLAB and CREATIS data. Note that 4D CT data of those image data bases was solely used for verification purposes, not for model training or testing. Further, registration output, i. e. motion vector fields, were analyzed

[†]Shared first authorship.

regarding their smoothness in terms of the standard deviation of the transformation Jacobian determinant values of the lung voxels.

Resulting TRE and transformation smoothness values are summarized in Table 1 [MICCAI 2018]. Interestingly, the mean registration accuracy for the DIRLAB data is higher for the specific CNN variants compared to the accuracy achieved by application of the original DIR algorithms. For the CREATIS data, TRE values of CNN-based registrations are similar to standard DIR. The transformation smoothness, however, is for all CNN variants lower in contrast to the smoothness of vector fields computed by original DIR frameworks. Estimated uncertainty maps, as illustrated in Fig. 2 [MICCAI 2018], differ in magnitude between individual CNN variants, but highlight that uncertainties occur in the same spatial region. The overall registration run-time for DIR of two 3D phase images was reduced from approximately 900 s of traditional DIR frameworks to 15 s of CNN-based DIR (60-fold speed-up).

In conclusion, all three aims of the proposed study were successfully implemented in a new deep learning-based registration framework. Using the registration and uncertainty output in subsequent dose accumulation processing steps directly allows for an error propagation implementation. However, beforehand the general functionality of the simulation approach has to be validated. Therefore, the next section focuses on verifying the general dose simulation approach by comparing measured dose distributions and dose simulation results. The planned uncertainty propagation is afterwards implemented in the Monte Carlo-based 4D dose accumulation scheme (cf. Sections 5.6 and 5.7).

GDL-FIRE^{4D}: Deep Learning-based Fast 4D CT Image Registration

Thilo Sentker[1,2⋆], Frederic Madesta[1,2⋆], and René Werner[1]

[1] Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
[2] Department of Radiotherapy and Radiation Oncology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
t.sentker@uke.de, f.madesta@uke.de, r.werner@uke.de

Abstract. Deformable image registration (DIR) in thoracic 4D CT image data is integral for, e.g., radiotherapy treatment planning, but time consuming. Deep learning (DL)-based DIR promises speed-up, but present solutions are limited to small image sizes. In this paper, we propose a General Deep Learning-based Fast Image Registration framework suitable for application to clinical 4D CT data (GDL-FIRE^{4D}). Open source DIR frameworks are selected to build GDL-FIRE^{4D} variants. In-house-acquired 4D CT images serve as training and open 4D CT data repositories as external evaluation cohorts. Taking up current attempts to DIR uncertainty estimation, dropout-based uncertainty maps for GDL-FIRE^{4D} variants are analyzed. We show that (1) registration accuracy of GDL-FIRE^{4D} and standard DIR are in the same order; (2) computation time is reduced to a few seconds (here: 60-fold speed-up); and (3) dropout-based uncertainty maps do not correlate to across-DIR vector field differences, raising doubts about applicability in the given context.

Keywords: Non-linear Image Registration · Registration Uncertainty · 4D CT · Deep Learning

1 Introduction

Acquisition of 4D image data (3D+t images, respiration-correlated data) is an integral part of current radiation therapy (RT) workflows for RT planning and treatment of thoracic and abdominal tumors. Especially 4D CT imaging is meanwhile widespread and currently estimated to be routinely applied in approximately 70% of the RT facilities in the United States [1]. Typical clinical use cases of 4D CT data are (semi-)automated target volume and organ at risk contour propagation; assessment of motion effects on dose distributions (4D RT quality assurance, dose warping) [2]; and 4D CT-based lung ventilation estimation and its incorporation into RT treatment planning [1].

At this, a key step is the application of deformable image registration (DIR) to the phase images of the 4D CT data. Traditional DIR approaches tackle the

⋆ equal contribution

underlying task of finding an optimal transformation mapping two phase images by minimization of a dissimilarity measure that controls local correspondences of voxel intensities [3]. Yet, the algorithms are time consuming and there exists the risk of getting stuck in local minima during optimization.

Motivated by the exceptional success of deep learning (DL) and especially convolutional neural networks (CNNs) for image segmentation and classification tasks, meanwhile a number of approaches has been proposed to also solve image registration tasks by CNNs – first in the context of optical flow estimation in computer vision [4], and later similarly for medical image registration [3, 5–7]. Yang *et al.* further extended a CNN-based DIR architecture to a probabilistic framework using dropouts [5], resulting in DIR uncertainty maps that could be of great value for RT treatment planning [8].

However, Uzunova *et al.* noted that "dense 3D registration with CNNs is currently computationally infeasible" [6], and focused on 2D (brain and cardiac) DIR only. To overcome this issue, patch-based approaches have been proposed for, e.g., 3D brain DIR [5], with the side effect that global information about the transformation to learn might be missing [3].In turn, Rohé *et al.* indeed proposed using a fully convolutional architecture; with a size of $64 \times 64 \times 16$ voxel, their cardiac MR images were, however, not even close to typical sizes of 4D CT images (in the order of $512 \times 512 \times 150$ voxel per phase image).

This paper is therefore dedicated to CNN-based registration suitable for application to fast DIR in clinical thoracic 4D CT data. Taking up the aforementioned challenges and trends in current DL-based DIR,

C1 we propose a general and efficient CNN-based framework for deep learning of dense motion fields in clinical thoracic 4D CT, called GDL-FIRE^{4D},
C2 build variants of GDL-FIRE^{4D} using common open source DIR frameworks,
C3 perform a first comprehensive evaluation thereof using publicly available 4D CT data repositories (thereby presenting first respective benchmark baseline results for DL-based DIR in 4D CT data), and
C4 compare and discuss dropout-generated registration uncertainty maps for the different GDL-FIRE^{4D} variants.

To the best of our knowledge, all aspects C1-C4 are novel contributions in the given application context.

The remainder of the paper is structured as follows: In Sec. 2, the problem formulation and the concept of GDL-FIRE^{4D} are detailed. Applied data sets and performed experiments are described in Sec. 3 and respective results given and discussed in Sec. 4. The paper closes with concluding remarks in Sec. 5.

2 Methods: DL-based Deformable Image Registration

A 4D CT image is a series $(I_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$ of 3D CT images $I_i : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^3$, representing the patient geometry at different breathing phases i with n_{ph} as number of available images and breathing phases, respectively. The phases i sample the patient's breathing cycle in time and are usually denoted by cycle

fractions, i.e. $\{1,\dots,n_{\text{ph}}\} \equiv \{0\%,\dots,50\%,\dots\}$ with 0% as end inspiration and 50% as end expiration phase. Deformable registration in 4D CT data then aims to estimate a corresponding series of transformations $(\varphi_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$ between the I_i and a reference image I_{ref}, with $\varphi_i : \Omega \to \Omega$. For the applications outlined in Sec. 1, I_{ref} usually represents one of the phase images I_i and the transformation φ_i and vector fields $u_i : \Omega \to \mathbb{R}^3$, $u_i = \varphi_i - \text{id}$ (id: identity map) the respiration-induced motion of the image structures between phase i and the reference phase.

2.1 Traditional Deformable Image Registration (DIR) Formulation

In a traditional 4D CT DIR setting, the reference image is considered the fixed image, $I_{\text{ref}} \equiv I_{\text{F}}$, and the phase images as moving images, $I_i \equiv I_{\text{M}}$, which are sequentially registered to I_{F} by $\varphi_i = \arg\min_{\varphi_i^* \in \mathcal{C}^2[\Omega]} \mathcal{J}\left[I_{\text{F}}, I_{\text{M}}; \varphi_i^*\right]$ to compute the sought transformations $(\varphi_i)_{i\in\{1,\dots,n_{\text{ph}}\}}$. The exact functional $\mathcal{J}$, i.e. dissimilarity measure, applied regularization approach and considered transformation model, and the optimization strategy vary in the community; see [9] for details.

2.2 Convolutional Neural Networks (CNNs) for DIR

Different to traditional DIR, we now assume a database of n_{pat} training tuples $\left(I_i^p, I_j^p, \varphi_{ij}^p\right)$, $i,j \in \{1,\dots,n_{\text{ph}}\}$, $p \in \{1,\dots,n_{\text{pat}}\}$ to be given; $\varphi_{ij}^p = id + u_{ij}^p$ represents a DIR result of the phase images $I_i \equiv I_{\text{F}}$ and I_j of patient p. The goal is to learn the relationship between the input data $\left(I_i^p, I_j^p\right)$ and u_{ij}^p by a convolutional neural network.

As noted by Uzunova *et al.* [6], it is currently computationally not feasible to directly feed the entire images and vector fields into a CNN or GPU memory. Instead, we propose a slab-based approach: Let $I|_{\hat{x}} := I|_{\Omega_{\hat{x}}}$ the restriction of image I to $\Omega_{\hat{x}} = \{(x,y,z) \in \Omega \mid x = \hat{x}\}$, i.e. the sagittal slice of I at x-position $\hat{x}$. Similarly, let $I|_{[\hat{x}_1,\hat{x}_2]}$ the restriction of I to $\Omega_{[\hat{x}_1,\hat{x}_2]} = \{(x,y,z) \in \Omega \mid \hat{x}_1 \le x \le \hat{x}_2\}$, i.e. an image slab comprising the sagittal slices $\hat{x}_1,\dots,\hat{x}_2$ of I. Using this notation, the aforementioned training tuples were converted to slab-based training samples $\left(I_i^p|_{[x-2,x+2]}, I_j^p|_{[x-2,x+2]}, u_{ij}^p|_x\right)$ with $x \in \{x_{\min},\dots,x_{\max}\}$ covering all sagittal slices of I. The rationale was to represent maximum information along main motion directions *inferior-superior* and *anterior-posterior* for each training sample, but also to provide some anatomical context in lateral direction.

Furthermore, the image dynamics were rescaled to $[0,1]$, the slabs resampled to isotropic resolution of 2 mm and cropped/zero-padded to identical size, and the non-patient background intensity set to zero. Similar pre-processing was applied to the displacement fields (resampling and -sizing of sagittal slices, background set to zero). In addition, x-, y- and z-displacement components were z-transformed on a voxel-level to avoid unintended suppression of small displacements during CNN training. Thus, the CNN aimed to learn normalized 3D-vectors for the individual voxels of sagittal slices, which are back-transformed to actual motion fields during final reconstruction of the fields. The pre-processed slab-based samples $(\tilde{I}_i^p|_{[x-2,x+2]}, \tilde{I}_j^p|_{[x-2,x+2]}, \tilde{u}_{ij}^p|_x)$ with $x \in \{x_{\min},\dots,x_{\max}\}$ of the n_{pat} patients were finally shuffled and used for CNN training.

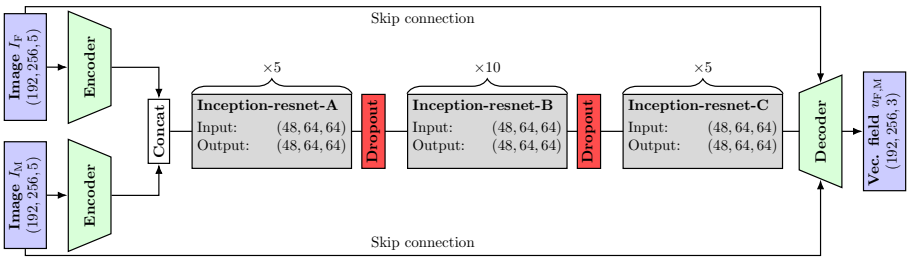


Fig. 1. CNN architecture implemented for DL-based DIR.

We tested different CNN architectures, including the classical U-Net [10]. Due to an observed increased robustness for DL-based DIR compared to the U-Net, we finally used an iterative CNN architecture with an Inception-ResNet-v2 [11] embedded in the encoder part of a pre-trained CT autoencoder, see Fig. 1, with MSE (mean squared error) loss function and NADAM optimizer (implemented in Tensorflow). *Iterative* means that we cascaded copies of the trained networks for improved coverage of large motion patterns.

2.3 Probabilistic CNN-based DIR

As detailed by Yang *et al.* [5] and references therein, deterministic CNN architectures can be extended to probabilistic using dropouts [12]. Briefly speaking, the dropout layers incorporated into the CNN architecture to prevent overfitting during model training remain enabled during motion prediction. Repeated motion prediction with respectively sampled connections to be dropped eventually enable computing the sought motion field as the mean of the sampled predicted fields; further, corresponding voxel-wise variances can be interpreted as local registration uncertainty estimates [5].

3 Materials and Study Design

All experiments were run on a desktop computer with Intel Xeon CPU E5-1620 and Nvidia Titan Xp GPU. Models and scripts required can be found at github.com/IPMI-ICNS-UKE/gdl-fire-4d.

3.1 Training and Testing 4D CT Data Cohorts

For CNN training and model optimization, a cohort of 69 in-house acquired RT treatment planning ten-phase 4D CT data sets of patients with small lung and liver tumors was used (image size: $512 \times 512 \times 159$ voxel) and a 85%/15% split into training and testing data performed. The 4D CT images of the open data repositories DIRLAB [13] and CREATIS [14] (see also www.creatis.insa-lyon.fr/rio/popi-model) served as *external* evaluation cohort of the trained CNNs (i.e. no model optimization performed by means of the external 4D CT cohorts).

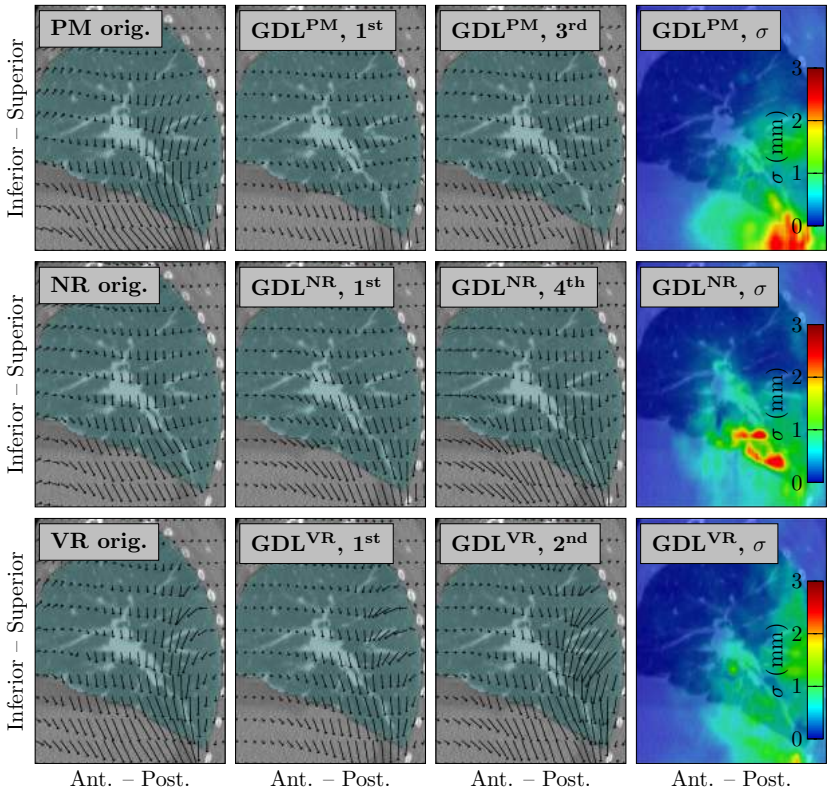


Fig. 2. Motion fields estimated by the original DIR algorithms (left column); GDL-FIRE^{4D} with only a single iteration (2nd column); GDL-FIRE^{4D} n iterations (3rd column); and GDL-FIRE^{4D} variant-specific registration uncertainty maps (right column). Data set: DIRLAB case 08, DIR of 0% and 50% phase images.

3.2 Applied DIR Frameworks and Algorithms

To provide motion field training data, the in-house 4D CT data were registered using three common open source DIR frameworks: PlastiMatch [15], NiftyReg [16], and VarReg [17]. All approaches have been proven suitable for 4D CT registration [9]; the applied parameters were similar to respective EMPIRE10 parameters [9]. However, the algorithms are applied in a plug-and-play manner (no data pre-processing or pre-registration, no masks used). For each DIR algorithm, motion fields were provided between the 20% phase image (served as I_F) and all other phase images.

3.3 Experiments and Evaluation Measures

For each DIR algorithm, a respective probabilistic GDL-FIRE^{4D} variant was built (up to 4 cascaded CNNs, 20% dropouts). DIR accuracy was evaluated by

the target registration error (TRE), computed by means of the landmarks publicly available for the DIRLAB and CREATIS data. In addition, the smoothness of transformations of the different DIR approaches and GDL-FIRE^{4D} variants was analyzed in terms of the standard deviation of transformation Jacobian determinant values of the lung voxels of the evaluation data.

4 Results and Discussion

Motion fields estimated by the original DIR algorithms and respective GDL-FIRE^{4D} variants as well as corresponding registration uncertainty maps are shown in Fig. 2 for DIRLAB case 08 (DIRLAB case with maximum motion amplitude) and phase 50% to phase 0% DIR. The similarity of the original and the GDL-FIRE^{4D} predicted fields is striking, i.e. the CNN obviously learned the DIR-specific transformation properties. This includes that the NiftyReg GDL-FIRE^{4D} variant has (similar to the original DIR) problems to directly cover larger motion amplitudes – and thereby motivates cascading several trained models for *iterative* CNN-based DIR. The success can be seen in Table 1, where the NiftyReg GDL-FIRE^{4D} outperforms the original NiftyReg DIR in terms of accuracy especially for cases with larger motion.

Still, GDL-FIRE^{4D} DIR accuracy as well as transformation properties for the other DIR approaches also resemble respective values of the traditional registration algorithm – but GDL-FIRE^{4D} offers a reduction of the runtime from approx. 15 min to a few seconds (speedup of approx. 60-fold).

Finally, it can be seen that the computed DIR uncertainty maps differ strongly between the GDL-FIRE^{4D} variants. In Fig. 3, a dataset of our internal testing cohort is shown that exhibits an artifact in the liver. This artifact led to very different motion patterns estimated by the NiftyReg and the VarReg GDL-FIRE^{4D} variant, but almost no measurable uncertainty for both DIR approaches. Being a direct consequence of the concept of probabilistic CNN-based DIR, this does, however, not match our understanding of DIR uncertainty and raises doubts regarding its applicability for RT planning and estimation of uncertainties therein.

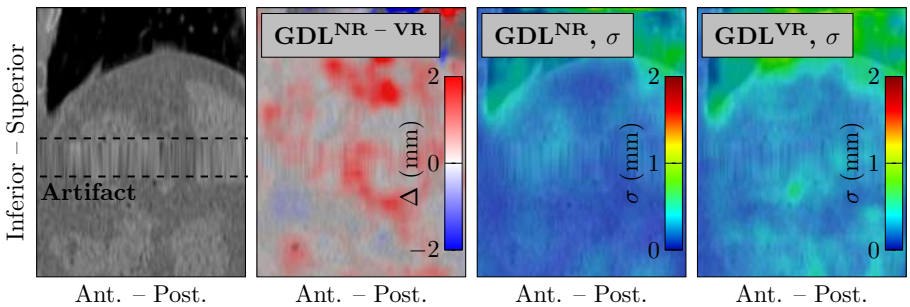


Fig. 3. From left to right: CT image serving as reference image with artifact in liver; difference of motion amplitudes estimated by the NiftyReg and the VarReg GDL-FIRE^{4D} variants, illustrating large across-DIR approach differences; NiftyReg and VarReg GDL-FIRE^{4D} uncertainty maps, showing negligible uncertainties for both variants.

Table 1. TRE values (in mm) and transformation smoothness (measured by standard deviation of lung voxel Jacobian determinant values), listed for the DIRLAB and CREATIS data, the individual DIR algorithms, and respective GDL-FIRE^{4D} variants (PM: PlastiMatch; NR: NiftyReg; VR: VarReg). Landmark distance before registration: (8.46 ± 6.58) mm for the DIRLAB and (8.11 ± 4.76) mm for the CREATIS data.

		Original DIR algorithms			**GDL-FIRE^{4D}**		
		PM [16]	**NR** [15]	**VR** [17]	**PM**	**NR**	**VR**
DIRLAB 4DCT	01	1.54 ± 0.98	1.46 ± 0.92	1.13 ± 0.54	1.69 ± 0.92	1.58 ± 0.80	1.20 ± 0.60
	02	1.74 ± 1.76	1.55 ± 1.06	1.17 ± 0.83	1.58 ± 1.07	1.65 ± 1.15	1.19 ± 0.63
	03	2.78 ± 2.20	2.53 ± 2.41	1.33 ± 0.69	2.39 ± 1.76	2.68 ± 1.78	1.67 ± 0.90
	04	2.70 ± 2.27	3.01 ± 2.45	3.08 ± 3.83	2.72 ± 1.97	2.48 ± 1.68	2.53 ± 2.01
	05	3.30 ± 3.06	3.21 ± 2.77	1.57 ± 1.33	2.83 ± 2.21	3.09 ± 2.50	2.06 ± 1.56
	06	3.80 ± 3.03	5.40 ± 3.94	5.23 ± 4.67	3.01 ± 1.97	2.73 ± 1.63	2.90 ± 1.70
	07	5.62 ± 5.32	8.36 ± 6.59	4.64 ± 3.91	4.48 ± 4.83	4.12 ± 4.21	3.60 ± 2.99
	08	7.65 ± 7.45	11.45 ± 9.08	4.58 ± 5.95	7.44 ± 6.87	8.26 ± 6.47	5.29 ± 5.52
	09	3.74 ± 2.60	5.66 ± 3.24	2.66 ± 2.46	3.56 ± 2.35	3.26 ± 1.90	2.38 ± 1.46
	10	3.15 ± 2.99	4.39 ± 4.21	2.14 ± 2.42	2.48 ± 1.99	2.55 ± 2.01	2.13 ± 1.88
∅ TRE		3.60 ± 1.83	4.70 ± 3.17	2.75 ± 1.57	3.22 ± 1.71	3.24 ± 1.81	2.50 ± 1.16
∅ $\sigma_{\lvert\nabla\varphi\rvert}$		0.10 ± 0.02	0.11 ± 0.03	0.39 ± 0.08	0.30 ± 0.13	0.24 ± 0.09	0.39 ± 0.14
CREATIS	01	1.13 ± 0.78	1.79 ± 1.26	0.90 ± 0.39	1.49 ± 0.83	1.73 ± 0.97	1.34 ± 0.74
	02	3.29 ± 3.10	4.29 ± 4.33	1.95 ± 2.87	3.59 ± 2.92	4.25 ± 3.47	2.98 ± 2.38
	03	1.95 ± 2.14	2.39 ± 2.60	1.14 ± 1.37	1.83 ± 1.42	2.05 ± 1.26	1.57 ± 1.01
	04	2.32 ± 2.95	2.51 ± 2.87	1.28 ± 2.13	1.79 ± 1.79	1.92 ± 1.73	1.64 ± 1.62
	05	1.88 ± 1.84	2.51 ± 2.73	1.17 ± 1.17	2.10 ± 1.78	2.18 ± 1.67	1.62 ± 1.09
	06	1.13 ± 0.78	1.52 ± 1.38	0.97 ± 0.72	1.60 ± 1.07	1.63 ± 1.11	1.26 ± 0.73
∅ TRE		2.01 ± 0.68	2.50 ± 0.88	1.24 ± 0.34	2.07 ± 0.78	2.29 ± 0.89	1.74 ± 0.57
∅ $\sigma_{\lvert\nabla\varphi\rvert}$		0.09 ± 0.02	0.11 ± 0.05	0.28 ± 0.05	0.31 ± 0.12	0.26 ± 0.08	0.30 ± 0.10

5 Conclusions

The presented GDL-FIRE^{4D} framework illustrates feasibility and potential of deep learning of dense vector fields for motion estimation in clinical thoracic 4D CT image data (TRE values of CNN-based DIR were in the same order than for the underlying DIR algorithms, accompanied by a speed-up factor of approximately 60), and thereby motivates continuing optimization of the framework.

Acknowledgments

We thank NVIDIA Corporation donating the applied Titan Xp GPU.

References

1. Yamamoto, T., Kabus, S., Bal, M., Keall, P., Benedict, S., Daly, M.: The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. Radiother Oncol **118** (2016) 227–31
2. Rosu, M., Hugo, G.D.: Advances in 4D radiation therapy for managing respiration: part II - 4D treatment planning. Z Med Phys **22** (2012) 272–80
3. Rohe, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-net: Learning deformable image registration using shape matching. MICCAI 2017 **10433** (2017) 266–73
4. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Smagt, P.V.D., Cremers, D., Brox, T.: FlowNet: Learning Optical Flow with Convolutional Networks. In: 2015 IEEE International Conference on Computer Vision (ICCV). Volume 2015 Inter., IEEE (2015) 2758–2766
5. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration - a deep learning approach. Neuroimage **158** (2017) 378–396
6. Uzunova, H., Wilms, M., Handels, H., B, J.E.: Training CNNs for image registration from few samples with model-based data augmentation. MICCAI 2017 **10433** (2017) 223–31
7. Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F.C., Miao, S., Maier, A.K., Ayache, N., Liao, R., Kamen, A.: Robust non-rigid registration through agent-based action learning. MICCAI 2017 **10433** (2017) 344–352
8. Amir-Khalili, A., Hamarneh, G., Zakariaee, R., Spadinger, I., Abugharbieh, R.: Propagation of registration uncertainty during multi-fraction cervical cancer brachytherapy. Phys Med Biol **62** (2017) 8116–8135
9. Murphy, K., van Ginneken, B., Reinhardt, J.M., et al.: Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Trans Med Imaging **30** (2011) 1901–20
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015 **9351** (2015) 234–41
11. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, Inception-Resnet and the impact of residual connections on learning. Proc. Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) (2017) 4278–4284
12. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res **15** (2014) 1929–58
13. Castillo, R., Castillo, E., Guerra, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys Med Biol **54** (2009) 1849–70
14. Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. Med Phys **38** (2011) 166–78
15. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed **98** (2010) 278–84
16. Shackleford, J.A., Kandasamy, N., Sharp, G.C.: On developing B-spline registration algorithms for multi-core processors. Phys Med Biol **55** (2010) 6329–51
17. Werner, R., Schmidt-Richberg, A., Handels, H., et al.: Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study. Phys Med Biol **59** (2014) 4247–4260

5.5. Accuracy of 4D dose simulation

High simulation accuracy of 4D dose accumulation in the context of VMAT quality assurance is mandatory to allow for interpretable results and further usage of estimated dose distributions for e. g. treatment plan adaption. As the dose simulation scheme requires instantiation of various input parameters that influence estimated dose distributions, the reliability of the simulation results is mitigated. Thus, the here presented study aims at analyzing the impact of 4D dose simulation parameters on the simulation accuracy. Note that uncertainties introduced by target deformation and non-linear registration were explicitly refrained from analyzing. Identification of the most influencing factors and assessment of the overall appropriateness of 4D VMAT dose simulation is conducted by comparing VMAT-based SBRT treatment simulations and motion phantom-based dose measurements, as presented in following publication:

T. Sothmann[†], T. Gauer[†] and R. Werner. 4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters. *PLoS One*, 12(2):e0172810, 2017.

Dose measurements were executed using an ionization chamber-based detector array with high spatial resolution mounted on a dedicated motion platform to allow for simulating patient and synthetic motion patterns. Further, a lung phantom add-on consisting of solid water, tumor, lung and bone inserts was placed on top of the detector, as illustrated in Fig. 1 (left) [PLOS 2017]. A VMAT treatment plan comprising of two arcs was planned on the corresponding 3D CT image of the experimental setup, with the achieved dose distribution shown in Fig. 1 (right) [PLOS 2017]. In Table 1 [PLOS 2017], the simulated motion patterns consisting of synthetic sine motion and regular, as well as irregular, patient motion trajectories are listed. By applying those trajectories, the treatment plan was delivered and the dose measured by the detector individually for each VMAT arc resulting in a motion-affected dose distribution. Similarly, reference dose distributions were acquired by irradiating the detector while no motion trajectory was applied to the phantom (static measurement). For the dose simulation, the described approach was used, with the difference that, in this study, no internal motion prediction, i. e. no application of correspondence models, was necessary as the phantom motion patterns were known. Considered and investigated as most influencing parameters on the accuracy of the 4D dose simulation were 1) accuracy of target structure and organs at risk motion representation, 2) degree of temporal discretization of technical dose delivery process and 3) accuracy of static dose calculation. To investigate these parameters, different experiments were performed. For 1), as no correspondence modeling and therefore no uncertainty of the predicted motion was included, 4D dose simulation had been repeated

[†]Shared first authorship.

with systematically varied starting phase. Using four different discretization levels, i. e. discretization of the continuous VMAT dose delivery process by splitting the total arc into sub-arcs of 2.3°, 5°, 10° and 150°, where 2.3° is the finest possible discretization and 150° represents the total arc (cf. Section 2.4.4 for VMAT discretization details), allowed to investigate parameter 2). Analysis of parameter 3) was done by computing the original treatment plan on the AvCT, on a single phase of the 4D CT and finally by exploiting the entire set of breathing phases represented by the 4D CT. Simulated and measured dose distributions were primarily evaluated and compared by a γ-evaluation (cf. Section 4.2.2) for the total dose and clinical relevant target structures (CTV, ITV). To do so, corresponding coronal 2D plane out of the simulated 3D volume dose was chosen for comparison to the measured 2D dose plane. The principle evaluation strategy is visualized in Fig. 3 [PLOS 2017]. As the focus of the simulation scheme was to illustrate possible motion-induced deviations between static and motion-affected dose distributions, the comparison was done by analyzing γ-maps representing the dose differences of static and motion affected measurements and simulations, i. e. as shown in the bottom row of Fig. 3 [PLOS 2017]. In Fig. 4 [PLOS 2017], achieved 2D γ-maps for different motion patterns and discretization level are illustrated. Repeated dose measurements (identical dose measurement series at a different day) were conducted to consider potential dose measurement uncertainties and are additionally shown; with the visual agreement of repeat measurements being high. For the dose simulation of irregular motion patterns the 2.3° level performed noticeably better, i. e. agreement between simulated and measured maps is higher, compared to the 150° level, indicating a more reliable computational simulation of the VMAT-delivered motion-affected dose. To further analyze this, in Tables 2-5 [PLOS 2017], corresponding values are given for all conducted measurements and simulations. Especially for smaller structures, i. e. CTV and ITV, highest simulation accuracy was achieved by the 2.3° level simulations. Results for varying starting phases are shown in Fig. 5 [PLOS 2017]. Here, both the finest and lowest discretization level were investigated for two motion patterns (synthetic sine and regular patient motion) by systematically changing the temporal starting point $t = 0\,\mathrm{s}$ of respective curves by adding offsets $\Delta t \in [0\,\mathrm{s}, 10\,\mathrm{s}]$. As expected and visible on the left part of the figure, varying the starting phase for a periodic motion returns a periodic evaluation metric value with $T = T_{\text{breathing cycle}}$. For the regular patient motion (cf. right part of figure) the variation of the starting phase for the simulation level with beforehand proven highest simulation accuracy, i. e. the 2.3° level, highly impacts computed metric values. On the contrary, numbers for the 150° level simulation suggest only a minor impact of the starting phase.

In conclusion, the proposed 4D dose simulation approach achieved high accuracy by applying lowest possible discretization and can therefore be used to accurately predict

motion effects on dose distributions. However, as correspondence modeling was not needed, the influence of estimated patient motion and its accuracy remains to be investigated. Limitations of simulation accuracy influencing parameters 2) and 3), meaning the maximum temporal resolution of the treatment planning system defined by the angular dose segment size and the dependency on pre-calculated static dose distributions, were consequences of the utilized dose simulation algorithm. In the next section, therefore, the re-implementation of the current framework in a Monte Carlo-based simulation approach to allow for even higher simulation accuracy is presented.

PLOS ONE

RESEARCH ARTICLE

4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters

Thilo Sothmann[1☯]*, **Tobias Gauer**[2☯], **René Werner**[1]

1 Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Hamburg, Germany, **2** Department of Radiotherapy and Radio-Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Hamburg, Germany

☯ These authors contributed equally to this work.
* t.sothmann@uke.de

OPEN ACCESS

Citation: Sothmann T, Gauer T, Werner R (2017) 4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters. PLoS ONE 12 (2): e0172810. doi:10.1371/journal.pone.0172810

Editor: Qinghui Zhang, North Shore Long Island Jewish Health System, UNITED STATES

Received: November 18, 2016

Accepted: February 9, 2017

Published: February 23, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Forschungsförderungsfond of the Medical Faculty of the University of Hamburg (to RW; URL: https://www.uke.de/english/research/funding-opportunities/faculty-of-medicine/index.html).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Radiotherapy of lung and liver lesions has changed from normofractioned 3D-CRT to stereotactic treatment in a single or few fractions, often employing volumetric arc therapy (VMAT)-based techniques. Potential unintended interference of respiratory target motion and dynamically changing beam parameters during VMAT dose delivery motivates establishing 4D quality assurance (4D QA) procedures to assess appropriateness of generated VMAT treatment plans when taking into account patient-specific motion characteristics. Current approaches are motion phantom-based 4D QA and image-based 4D VMAT dose simulation. Whereas phantom-based 4D QA is usually restricted to a small number of measurements, the computational approaches allow simulating many motion scenarios. However, 4D VMAT dose simulation depends on various input parameters, influencing estimated doses along with mitigating simulation reliability. Thus, aiming at routine use of simulation-based 4D VMAT QA, the impact of such parameters as well as the overall accuracy of the 4D VMAT dose simulation has to be studied in detail–which is the topic of the present work. In detail, we introduce the principles of 4D VMAT dose simulation, identify influencing parameters and assess their impact on 4D dose simulation accuracy by comparison of simulated motion-affected dose distributions to corresponding dosimetric motion phantom measurements. Exploiting an ITV-based treatment planning approach, VMAT treatment plans were generated for a motion phantom and different motion scenarios (sinusoidal motion of different period/direction; regular/irregular motion). 4D VMAT dose simulation results and dose measurements were compared by local 3% / 3 mm γ-evaluation, with the measured dose distributions serving as ground truth. Overall γ-passing rates of simulations and dynamic measurements ranged from 97% to 100% (mean across all motion scenarios: 98% ± 1%); corresponding values for comparison of different day repeat measurements were between 98% and 100%. Parameters of major influence on 4D VMAT dose simulation accuracy were the degree of temporal discretization of the dose delivery process (the higher, the better) and correct alignment of the assumed breathing phases at the beginning of the dose measurements and simulations. Given the high γ-passing rates between simulated motion-affected doses and dynamic measurements, we consider the simulations to provide a reliable basis for assessment of VMAT motion effects that–in the sense of 4D QA of VMAT treatment plans–allows to verify target coverage in hypofractioned VMAT-based

radiotherapy of moving targets. Remaining differences between measurements and simulations motivate, however, further detailed studies.

Introduction

Radiation therapy of lung and liver lesions has fundamentally changed from conventional 3D conformal radiation therapy (3D-CRT) to hypofractioned and even ablative-type treatment schemes such as stereotactic body radiation therapy (SBRT) or stereotactic ablative radiotherapy (SABR) [1, 2]. Treatment plans are often delivered by intensity modulated radiation therapy (IMRT)-type techniques like volumetric modulated arc therapy (VMAT) [3]. Lung and liver lesions are, however, subject to respiratory motion with well-studied motion amplitudes of up to several centimetres [4, 5]. For IMRT-type dose delivery, the target motion may lead to a risk of so-called interplay effects, i. e. the unintended interference of target motion and dynamically changing beam parameters such as gantry position and MLC segments' shape [6]. Some years ago, this has been an oft-reported reason to avoid IMRT techniques for lung and liver treatment even for conventional fractionation schemes [7]–although planning studies highlighted superior dose distribution characteristics compared to 3D-CRT [8, 9] and related interplay effects were shown to average out over the course of treatment [10].

Nowadays, lung and liver patients *are* treated in a few or even a single fraction [11], with high target doses delivered by VMAT techniques (sometimes even by means of only a single arc [12]) and employing treatment units with high dose rates, e. g. using flattening filter free beams [13, 14]. Thus, the risk associated to single fraction interplay effects appears to be increased. This, on the one hand, highlights the importance of studies on interplay effects in the context of VMAT and hypofractionation [6, 14–21]. On the other hand and from a clinical perspective, it also motivates establishing 4D quality assurances (4D QA) to assess the potential risk of motion and interplay effects for generated treatment plans when accounting for patient-specific breathing patterns before dose delivery. Related studies, including the aforementioned references, can be divided into computational 4D dose simulation-based and motion phantom, i. e. measurement-based approaches. A drawback of motion phantom-based 4D QA is that it is restricted to a small(er) number of measurements and variations of the patient-specific motion patterns. In contrast, computational approaches allow simulating almost every conceivable scenario and have the potential to provide a more comprehensive picture of motion effects and their impact on, e. g., target coverage. They, however, require instantiation of various input parameters that influence estimated dose distributions–and which therefore mitigate reliability of the simulation results.

Being placed in the in the context of the discussion about appropriate VMAT 4D QA, the present study aims to analyze the impact of 4D VMAT dose simulation parameters on the simulation accuracy, to identify the most influencing factors, and finally to assess the overall appropriateness of 4D VMAT dose simulation to assess motion-induced dose alterations for VMAT-based SBRT treatment when compared to motion phantom-based measurements. To do so, we built on a computational 4D dose simulation scheme originally introduced for the analysis of interplay effects in step-and-shoot IMRT [7], extended it to VMAT dose delivery, and studied different physiological as well as technical parameters that likely affect 4D VMAT dose simulation accuracy.

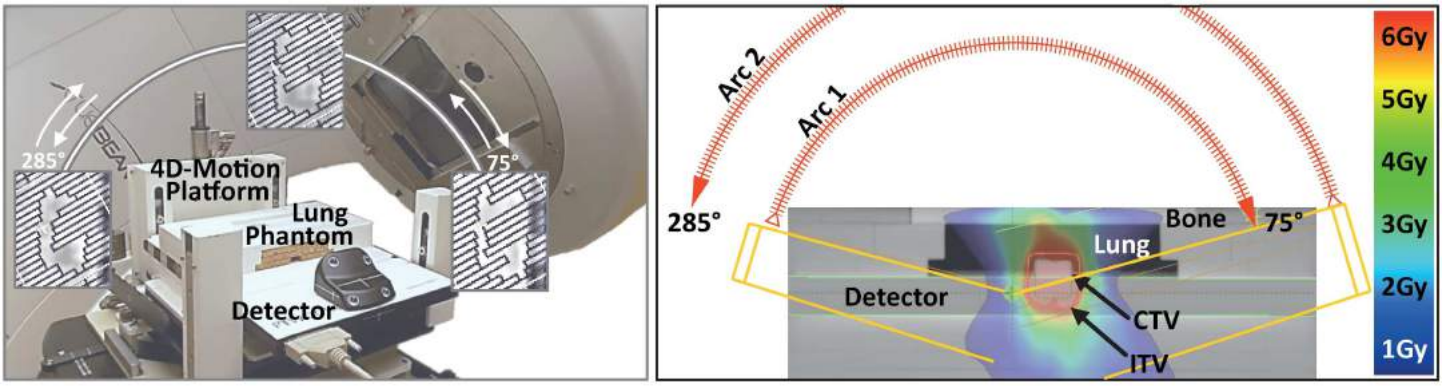


Fig 1. Experimental setup. Left: Measurement setup: 4D motion platform with detector array and lung phantom, consisting of bone, lung and tissue equivalent materials. Right: Average CT of setup, planned VMAT dose distribution, and target structures/organs at risk.

doi:10.1371/journal.pone.0172810.g001

Materials and methods

This section is structured as follows: First, the concept of VMAT dose delivery is introduced, the fundamentals of the applied 4D dose simulation scheme are described and parameters are identified that potentially influence 4D VMAT dose simulation accuracy. Finally, the study design, i. e. strategies to address these parameters, and the performed experiments are detailed.

Principles of and uncertainties in 4D VMAT dose simulation

Volumetric arc therapy exploits dose rate and gantry speed variation as well as leaf modulation to maximize benefits of classic (e. g. static field) IMRT techniques [22]. Briefly speaking, the dose is delivered during continuous gantry rotation around the patient, usually covering a wide range of gantry angles and often using a (small) number of so-called arcs as indicated in Fig 1.

Due to leaf modulation, each gantry angle is further associated with a multileaf collimator (MLC)-formed treatment field of specific shape and dose; cf. Fig 1 (left).

Computer-based simulation of VMAT dose delivery requires discretization of the continuous process. Let $\dot{D} : \Omega \subset \mathbb{R}^3 \times \mathcal{T} \subset \mathbb{R} \rightarrow \mathbb{R}_+$ be the time-dependent dose rate of the dose delivery process and $D : \Omega \rightarrow \mathbb{R}_+$ the resulting dose distribution; then, the discretization mathematically translates into

$$D(x) = \int_{\mathcal{T}} \dot{D}(x,t)\,dt \approx \sum_{t\in\tilde{\mathcal{T}}} \dot{D}(x,t)\Delta t = \sum_{t\in\tilde{\mathcal{T}}} D_t(x) \tag{1}$$

with $\mathcal{T} = (0, T]$ being the period of dose delivery and $\tilde{\mathcal{T}} = \{\Delta t, 2\Delta t, \ldots, T\}$ a sampled version of $\mathcal{T}$. In other words: $D_t(x) := \dot{D}(x,t)\Delta t$ represents an approximation of the dose delivered during the interval $(t - \Delta t, t] \subset \mathcal{T}$, and the sum of all $D_t(x)$–the *accumulated* or simulated (4D) dose [7]–approximates the entire dose D to the voxel at spatial position $x \in \Omega$.

Focusing on single arc VMAT dose delivery, the time variable t can be replaced by the gantry rotation angle α, and Eq (1) reads

$$D(x) \approx \sum_{\alpha\in\tilde{\mathcal{A}}} \dot{D}(x,\alpha)\Delta\alpha = \sum_{\alpha\in\tilde{\mathcal{A}}} D_\alpha(x) \tag{2}$$

with $\tilde{\mathcal{A}} \subset \mathcal{A}$ being a discretized version of the gantry angle range $\mathcal{A}$ of the considered arc. As

before, $D_\alpha(x) := \dot{D}(x,\alpha)\Delta\alpha$ is an approximation of the dose delivered to x during the gantry interval $(\alpha - \Delta\alpha, \alpha]$. Thus, for $\Delta\alpha \to 0$, the distributions D_α can be interpreted as the dose delivered for the individual MLC-formed fields illustrated in Fig 1 (left).

So far, Eqs (1) and (2) represent a discretization of single arc VMAT dose delivery to a *static* geometry. Including (here: breathing-induced) motion means to additionally account for a time dependence of the spatial voxel position during dose delivery. This leads to a common modification of Eq (2) [23]:

$$D_{\text{dyn}}(x) \approx \sum_{\alpha\in\tilde{\mathcal{A}}}\sum_{t\in\tilde{\mathcal{T}}_\alpha} D_\alpha\big(\varphi_t(x)\big) = \sum_{\alpha\in\tilde{\mathcal{A}}}\sum_{t\in\tilde{\mathcal{T}}_\alpha} \big(D_\alpha \circ \varphi_t\big)(x). \tag{3}$$

Here, $\tilde{\mathcal{T}}_\alpha$ represents a sampled version of the time interval corresponding to dose delivery for the gantry angle interval $(\alpha - \Delta\alpha, \alpha]$ and $\varphi_t(x) \in \Omega$ the position of the voxel originally positioned at x at time point $t \in \tilde{\mathcal{T}}_\alpha$. The other way around, $\varphi_t : \Omega \to \mathbb{R}^3$ can also be read as a transformation that maps the dose delivered to the correspondingly moved geometry $\varphi_t(\Omega)$ to the original (= reference) coordinate space and geometry Ω, with the latter being usually defined by a pre-selected phase of a planning 4D CT or a derived image like a temporal average CT. Consequently, the resulting dose distribution $D_{\text{dyn}} : \Omega \to \mathbb{R}^3$ represents an estimation of the dose delivered during the considered VMAT arc and accounting for the geometry (= patient, phantom) motion represented by the set of transformations φ_t representing all $t \in \tilde{\mathcal{T}}_\alpha$ and $\alpha \in \tilde{\mathcal{A}}$. Note that the nested summation in Eq (3) is only necessary if the temporal resolution of the target motion information is higher than the resolution of the dose delivery process; otherwise, the 4D dose simulation can be formulated by a single summation.

Extending single arc 4D dose simulation, i. e. Eq (3), to an entire treatment course would finally lead to an extended nested summation,

$$D_{\text{dyn}}^{\text{total}}(x) = \sum_{i=1}^{n_{\text{fx}}} \underbrace{\sum_{j=1}^{n_{\text{arc}}} \underbrace{\sum_{\alpha\in\tilde{\mathcal{A}}_j} \underbrace{\sum_{t\in\tilde{\mathcal{T}}_\alpha} \underbrace{(D_{i,j,\alpha} \circ \varphi_{i,j,t})(x)}_{\text{angle interval dose}}}_{\text{single arc dose}}}_{\text{treatment fraction dose}}}_{\text{total treatment dose}} . \tag{4}$$

Assuming more or less periodic motion patterns, the fractionation effect (outermost summation) on the delivered dose distribution can be simulated by repeating treatment fraction dose computation with randomly varied breathing phases associated to the beginning of the dose delivery of the individual fractions (i. e. randomly shifting the voxel trajectories in time) and summing up the dose distributions [7]. This averages out single fraction motion and interplay effects, depending on the number of fractions n_{fx}. Principle and dosimetric consequences have already been well-explained for classic IMRT techniques [10], and related conclusions can be transferred to VMAT. Similar observations can also be made for the summation over different arcs [21].

As our study aims to analyze factors that systematically (i. e. in a deterministic way) influence 4D dose simulation accuracy, we refrain from considering the randomness and averaging effects induced by fractionation and multi-arc dose delivery. Instead, we focus on single arc 4D dose simulation as given by the innermost summation and Eq (3), respectively.

Further neglecting implementation details such as the exact type of dose interpolation in Eq (3) and potential effects of, e. g., ignoring mass effects by not applying energy transformation models (previously proven to be small for small voxel sizes [23]), the following parameters are considered as potentially most affecting 4D VMAT dose simulation accuracy and are studied:

P1 *Accuracy of target structure and organs at risk motion representation* Correct representation of target structure and organs at risk motion, i. e. the estimation of the transformations $\varphi_{i,j,\alpha}$, is a prerequisite for accurate 4D dose simulation and assessment of motion/interplay effects. In real patient-scenarios, the $\varphi_{i,j,\alpha}$ are usually computed by non-linear registration in planning 4D CT data–with the problem of being confined by a limited temporal resolution of the data and uncertainties associated to the applied registration approach; see Ref. [24, 25] for recent overviews on non-linear registration accuracy in 4D CT image sequences. Phantom studies, however, usually allow applying known motion patterns, which reduces these uncertainties to a minimum. Imperfect synchronization of the phantom breathing phase at measurement beginning and the assumed starting phase used for simulation purposes remains nevertheless as source of error especially for comparison of measurements and simulations.

P2 *Degree of temporal discretization of technical dose delivery process* Discretization of the continuous VMAT dose delivery process, i. e. Eqs (2) and (3), means to ignore potential interplay effects that are due to gantry or MLC leaf movements during the considered intervals of size Δt and $\Delta\alpha$. Thus, interplay effects on a times scale of Δt ($\Delta\alpha$) will not be represented by the simulated 4D dose.

P3 *Accuracy of static dose calculation* Like for any RT treatment planning, 4D dose simulation accuracy also depends on the dose calculation algorithm and dose grid size (has to be small for interpolation-based 4D dose simulation schemes) applied for computation of the individual $D_{i,j,\alpha}$. In addition and especially for lung SBRT, the choice of appropriate CT images and density distributions considered for $D_{i,j,\alpha}$ computation further remains an open issue and a potentially influencing parameter when comparing 4D dose simulation results to measurements.

Study design and experiments

Following the previous section, our study aimed at analyzing the influence of the above-mentioned aspects on single fraction single arc 4D dose simulation accuracy by comparison of motion phantom dose measurements and corresponding simulated dose distributions; the measurements were considered as ground truth to be resembled by the simulations as closely as possible. The motion phantom setup has already been shown in Fig 1 and is–together with the applied dose measurement equipment, treatment planning aspects, motion scenarios, the performed experiments and our evaluation strategy–detailed in the following.

Motion phantom and dose measurement setup. The motion phantom consisted of a lung phantom add-on compiled by solid-water, lung, tumor and bone inserts, and the Octavius 1000 SRS detector array (PTW Freiburg, Germany). The detector provided high spatial resolution dose measurements by means of 977 liquid filled ionization chambers, distributed over an area of 11×11 cm, with a 2.5 mm chamber spacing in the inner 5.5×5.5 cm and 5 mm spacing in the outer detector area [26]. The add-on was mounted on a computer-controlled motion platform with three spatial degrees of freedom (Euromechanics, Germany). According to Ref. [27], the phantom-based regular breathing pattern simulation accuracy is higher than 0.5 mm.

Motion scenarios and treatment planning. The characteristics of the motion patterns programmed to the phantom are listed in Table 1. Five sinusoidal curves with varying motion

Table 1. Motion characteristics: maximum and mean peak-to-peak amplitudes, mean breathing cycle lengths.

Case		Variability	max. Amplitude (mm)			ø Amplitude (mm)			ø Period (s)
			SI	AP	LR	SI	AP	LR	
1	a	none (sine)	20	10	0	20	10	0	3.1
	b		20	10	0	20	10	0	4.5
	c		20	10	0	20	10	0	5.8
	d		20	0	0	20	0	0	4.5
	e		20	10	10	20	10	10	4.5
2	a	regular	16.3	11.8	N/A	12.8 ± 3.3	9.4 ± 3.3	N/A	4.8 ± 0.8
	b	irregular	22.5	23.9	N/A	13.0 ± 5.2	13.6 ± 5.5	N/A	4.4 ± 1.0

SI: superior-inferior; AP: anterior-posterior; LR: left-right.

doi:10.1371/journal.pone.0172810.t001

period and directions as well as two real-patient tumor motion paths extracted from CyberKnife lung SBRT treatment logfiles were considered. The (ir)regularity of the real-patient curves differed significantly (Fig 2); the two scenarios are subsequently denoted as 'regular' and 'irregular'.

With the tumor insert of the lung phantom defining an 2×2×2 cm clinical target volume (CTV), the internal target volume (ITV) being chosen sufficiently large to compensate for simulated motion patterns, and the lungs surrounding the ITV as organ at risk, dual-arc VMAT (RapidArc, Varian Medical Systems, USA) plans with standard MLC modulation were created using Eclipse 13 (Varian Medical Systems). Taking into account an angular dependency of the Octavius 1000 SRS detector array as reported in [28] (larger differences between measured dose and the dose calculated by the treatment planning system for dose delivery to the rear of the detector), we restricted the utilized gantry angle interval to 285˚ to 75˚ and vice versa, cf. Fig 1. Thus, beam incidence is always on the front of the detector (the angle perpendicular to the detector front plane is 360˚/0˚). ITV dose coverage was optimized based on the average CT corresponding to a 10-phase 4D CT of the moving phantom [motion pattern 1b of Table 1; CT scanner: Siemens Definition AS+ (Siemens Healthcare, Germany) with Real-Time Position

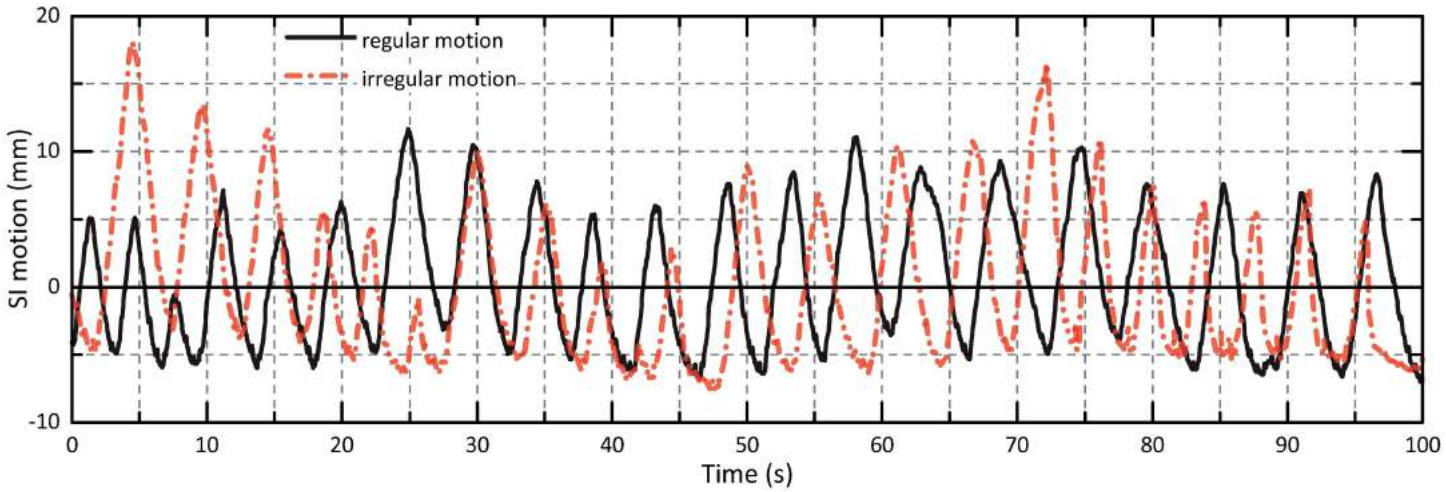


Fig 2. Patient motion scenarios. SI motion amplitudes of applied regular and irregular tumor trajectories.

doi:10.1371/journal.pone.0172810.g002

Management system (Varian Medical Systems)]. For 6 Gy prescribed dose, a nominal 600 MU/min dose rate and an energy of 6 MV, the 'beam on' time per arc was 50 s, corresponding to a mean arc velocity of 3˚/s.

Experiments I: dose measurements. The treatment plans were delivered without (static measurement) and with phantom motion (dynamic measurement) by a TrueBeam linear accelerator (Varian Medical Systems, USA). The delivered dose was separately measured for the individual arcs, and two measurement series were acquired at different days for each motion scenario to estimate related dose measurement uncertainties. The motion phantom position for static measurements corresponded to mid-respiration states of the individual motion scenarios. This position also represented the starting position and breathing phase for the dynamic measurements (synchronized with 'beam on' signal of the treatment unit).

Experiments II: 4D VMAT dose simulation. As the motion phantom was limited to rigid movements and to be able to separately analyze the impact of parameters P2 and P3, we explicitly refrained from using image-based obtained motion trajectories $\varphi_{i,j,t}$. Instead, the trajectories programmed to the phantom and used throughout the measurements were applied as $\varphi_{i,j,t}$ in Eq (4). With a resolution of 0.04 s, uncertainties arising from temporal motion trajectory discretization were assumed to be negligible. To further illustrate the influence of inaccuracies regarding the synchronization of the breathing phases at motion phantom measurement and simulation beginning (P1), 4D dose simulation has been repeated with systematically varied starting phase. This part of the experiments was also intended to demonstrate potential advantages of a simulation-based approach in comparison to measurement studies.

With the motion trajectories being assumed to agree between dose measurements and simulation, the influence of the temporal discretization of the dose delivery process (P2) was studied by varying the angle interval size $\Delta\alpha$. Four discretization levels were applied: 2.3˚ (corresponds to 65 segments per arc), 5˚ (30 segments), 10˚ (15 segments), and 150˚ (1 segment). The corresponding angle interval dose segments D_α were determined using the arc splitting option implemented in Eclipse for verification purposes. The smallest possible angle interval size in Eclipse was 2.3˚, which explains the aforementioned choice for the finest discretization level. The last discretization choice actually means that the planned arc is not split into segments. This, in turn, represents the situation that effects of dynamically changing beam parameters are not accounted for during 4D dose simulation; the geometry is moved inside the originally planned 'dose cloud' and interplay effects are neglected. The hypothesis was that, if interplay effects influenced the measurements, the agreement between measurement and simulation should increase with finer discretization.

For studying the influence of differences with respect to calculation of the static dose distributions $D_{i,j,\alpha}$, 4D dose simulation was repeated with the arc segment dose distributions being computed based on the intensity distribution of the average CT, based on a single phase CT of the acquired 4D CTs, and exploiting the entire set of breathing phases represented by the 4D CTs. In the latter case, a dose distribution $D_{i,j,\alpha}$ was assumed to correspond to the dose distribution computed by means of the CT with the breathing phase closest to the actual phase represented by $\varphi_{i,j,t}$. In addition, the outputs of the dose calculation algorithms applied to lung SBRT in our facility–the analytical anisotropic algorithm (AAA) and Acuros XB–were exemplarily compared.

Evaluation and comparison of measured and simulated dose distributions. Measured and simulated dose distributions were primarily compared by 2D-γ-evaluation [29], with γ-value maps and γ-passing rates analyzed for the CTV, ITV and the entire measurement area determined by the flat panel detector [30]. The coronal slice of the simulated 3D dose distributions that was selected for comparison to the measured dose distributions corresponded to the

slice of the planning average CT that, in terms of visual inspection, most closely corresponded to the central plane of the detector (cf. Fig 1). During plan generation, the center of the detector plane visible in the coronal CT slice was further marked by a reference point. Centered in this point, a 2D-region of interest (ROI) of similar size as the detector area was extracted from the coronal dose slice, resampled to the spatial resolution of the measurement dose distribution, and remaining small spatial shifts caused by imperfect phantom setup were manually corrected. In agreement with standard QA parameters, a local γ-criterion of 3%/3 mm was applied and only pixels with dose values $\geq$ 20% of the maximum dose value were considered. Software packages used for evaluation were Verisoft 6.0 (PTW Freiburg, Germany) and Matlab R2015a (MathWorks, USA). All measured and simulated 2D dose ROIs underlying subsequent result tables and figures are provided as supporting information (S1 File) to allow interested readers to reproduce the data.

Results

Fig 3 illustrates the concept underlying our study design and evaluation strategy: Focusing on a single VMAT arc of a dual-arc treatment plan, the left column of the figure shows the planned dose distribution ('simulated dose without motion'; top), the accumulated dose representing the simulated motion effects ('simulated dose with motion'; middle), and the γ-map demonstrating differences between the two distributions, i. e. the simulated motion effects (γ-criterion 3%/3 mm; bottom). The middle column represents the same information for the measurements (from top to bottom: static measurement, measurement with motion, γ-map for comparison of static and moved measurement).

Focussing on the question of the appropriateness of simulation-only based 4D VMAT QA and therefore being primarily interested in an assessment of the 4D VMAT dose simulation accuracy and parameters affecting it, the motion effects represented in the left and middle γ-maps were only of secondary interest; instead, it was the agreement of the two γ-maps as well as the similarity of underlying simulated and measured motion affected dose distributions that was to appraise. The motivation for choosing the given arc and its inhomogeneous dose distribution within the ITV as a showcase was in line with this argument: the inhomogeneous dose distribution simplified visual assessment of motion effects and respective differences in comparison to homogeneous high ITV doses that would result for single arc treatment planning.

Direct comparison of simulated and measured motion-affected dose distributions by means of the corresponding γ-map is finally illustrated in the middle of the right column. For ideal agreement, the γ-map would correspond to an 100% γ-passing rate. This ideal scenario is, however, not realistic due to unavoidable uncertainties influencing the comparison. The figure represents two such sources. In the right top corner, the planned dose and the static measurement were compared. The γ-map corresponds to a total γ-passing rate of 97%; related uncertainties, of course, also affect a comparison of dynamic simulated and measured doses. In addition, the right bottom corner provides differences between repeat dynamic measurements; the γ-passing rate was 98%. These reference values have to be taken into account for discussion of γ-passing rates between measured and simulated motion-affected dose distributions. Nevertheless, for the shown case, the γ-passing rate between dynamic measurement and motion-affected simulation was still 98%.

Influence of VMAT arc discretization on 4D VMAT dose simulation accuracy

The influence of the degree of temporal discretization of the technical dose delivery process on the 4D VMAT dose simulation results is illustrated in Fig 4, again using the first arc of the

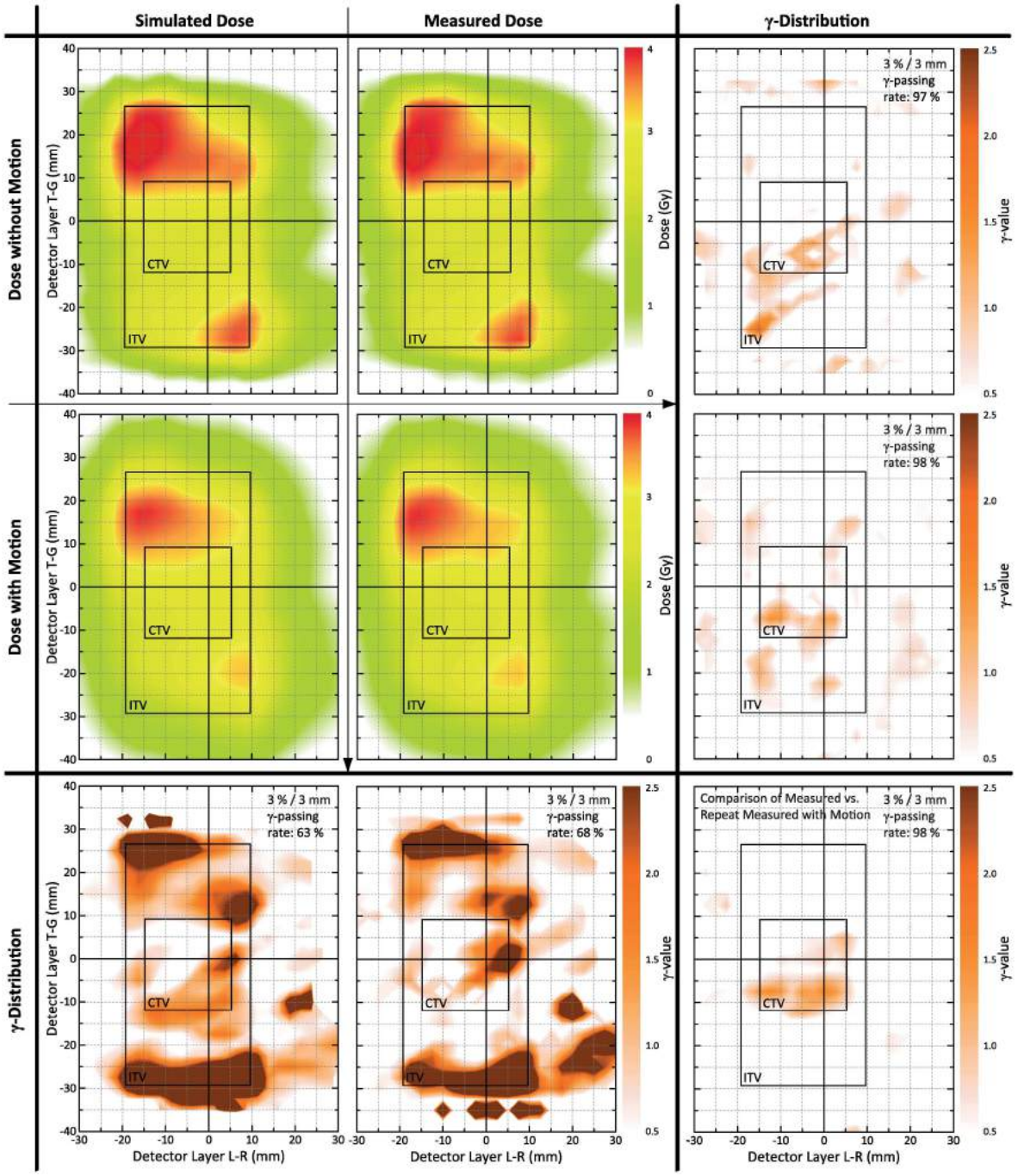


Fig 3. Study design and evaluation strategy. Illustration of performed experiments for the SI-only sinusoidal motion with 4.5 s period (i. e. case 1d); for details see text. Left column: planned dose distribution (top), simulated motion-affected dose (middle; arc discretization of 2.3˚), *γ*-map for comparison of the two (bottom). Middle column: measured static dose (top), measured dynamic dose (middle), *γ*-comparison (bottom). Right column: *γ*-comparison of planned and measured static dose (top), *γ*-comparison of simulated motion-affected and corresponding measured dose (middle), *γ*-comparison of repeat dynamic measurements (bottom).

doi:10.1371/journal.pone.0172810.g003

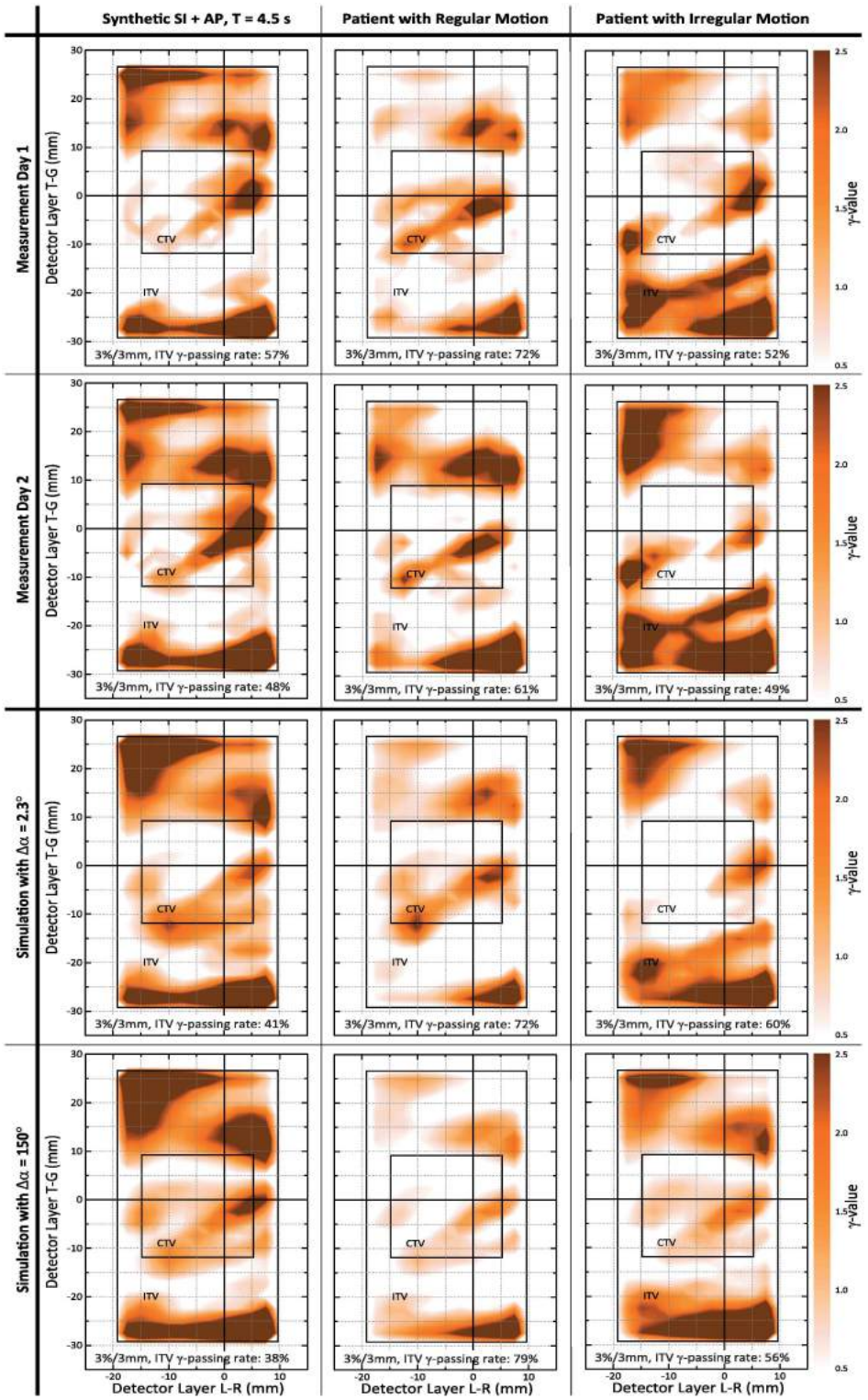


Fig 4. Influence of arc discretization. Illustration of the influence of arc discretization on simulated motion effects. 3rd row: γ-comparison to planned dose for finest possible arc discretization; 4th row: no discretization. Results have to be compared to γ-maps between static and motion-affected measurements in 1st and 2nd row. Differences between the simulation γ-maps and the measurement γ-maps should be as small as possible.

doi:10.1371/journal.pone.0172810.g004

Table 2. Total γ-passing rates for comparison of dynamic dose measurements of day 1 ($D_{dyn,day1}$) and dynamic day 2 dose measurements and simulated 4D dose distributions ($D_{dyn,\cdot}$) with respect to the impact of the arc discretization on the simulation accuracy.

	$D_{dyn,day1}$ vs. $D_{dyn,\cdot}$ [%]							$\bar{\gamma}$ [%]
	1a	1b	1c	1d	1e	2a	2b	
Day 2 measurement	99	98	98	98	99	99	100	**99 ± 1**
4D VMAT simulation: 65 seg. (2.3˚)	99	98	98	98	98	100	97	**98 ± 1**
4D VMAT simulation: 30 seg. (5˚)	98	97	96	97	96	99	94	**97 ± 1**
4D VMAT simulation: 15 seg. (10˚)	99	94	91	94	95	96	91	**94 ± 3**
4D VMAT simulation: 1 seg. (150˚)	99	96	93	95	95	97	82	**94 ± 5**

$\bar{\gamma}$ represents the mean gamma passing rates, averaged over all motion scenarios. The closer the numbers to 100%, the better. Ideally, γ-passing rates between day 1 measurements and 4D VMAT simulations are in the same order than those for comparison of day 1 and day 2 measurements.

doi:10.1371/journal.pone.0172810.t002

respective dual-arc VMAT plans. The figure shows γ-maps obtained by comparison of static and dynamic measurements (first two rows) and γ-maps for comparison of the statically planned and simulated motion-affected dose distributions (last two rows); the columns represent different motion scenarios. The γ-maps therefore represent measured and simulated motion effects. Consequently, a simulation is superior to another when its γ-map more closely resembles the γ-map of the measurement.

For the regular motion patterns, the γ-maps obtained for both $\Delta\alpha = 2.3^\circ$ and $\Delta\alpha = 150^\circ$ simulations appear to well agree with the γ-maps for the measurements–especially, when considering between-measurement differences (comparison of rows 1 and 2). Deviations between measurement and simulation increased, however, for the irregular motion pattern (see right column), with the agreement between the $\Delta\alpha = 150^\circ$ γ-map and the measurements being noticeably smaller than for $\Delta\alpha = 2.3^\circ$. This already indicated that taking into account the interplay between the dynamic dose delivery process and target structure motion on a fine-scale temporal resolution allows for a more reliable computational simulation of the VMAT-delivered motion-affected dose.

The aforementioned impression was further supported by the quantitative evaluation of the γ-maps summarized in Tables 2–5. The visual agreement between measurement- and

Table 3. Total γ-passing rates for comparison of static dose measurements to dynamic measurements (lines 'Day 1' and 'Day 2') and γ-passing rates for comparison of the statically planned dose and the dose distributions containing simulated motion effects.

	$D_{sta,\cdot}$ vs. $D_{dyn,\cdot}$ [%]							$\bar{\Delta}_\gamma$ [%]
	1a	1b	1c	1d	1e	2a	2b	
Day 1 measurement	65	67	64	68	62	86	68	–
Day 2 measurement	59	62	61	62	60	79	68	**4.1 ± 2.4**
4D VMAT simulation: 65 seg. (2.3˚)	61	60	57	63	60	86	69	**3.7 ± 2.6**
4D VMAT simulation: 30 seg. (5˚)	61	60	56	62	61	85	68	**3.9 ± 3.0**
4D VMAT simulation: 15 seg. (10˚)	61	58	54	59	59	87	67	**5.3 ± 3.7**
4D VMAT simulation: 1 seg. (150˚)	60	59	60	60	60	89	74	**5.1 ± 2.2**

Cf. Table 2 for symbols and indices. Comparing static to dynamic measurements / simulations, the simulations should now (different to Table 2) resemble the numbers of the measurements as closely as possible. Consequently, the absolute difference $\bar{\Delta}_\gamma$ between the γ-passing rates obtained by the simulations and the γ-passing rates of the day 1 measurements should be (averaged over all motion scenarios) as low as possible and in the same order as the differences between day 1 and day 2 measurements.

doi:10.1371/journal.pone.0172810.t003

Table 4. CTV γ-passing rates for comparison of static dose distributions and dynamic dose measurements/simulations.

	CTV: $D_{sta,\cdot}$ vs. $D_{dyn,\cdot}$ [%]							$\bar{\Delta}_\gamma$ [%]
	1a	1b	1c	1d	1e	2a	2b	
Day 1 measurement	45	73	59	73	73	59	80	–
Day 2	32	59	41	55	71	71	82	**11.3 ± 6.3**
4D VMAT simulation: 65 seg. (2.3˚)	66	68	57	68	82	71	93	**9.6 ± 6.0**
4D VMAT simulation: 30 seg. (5˚)	70	66	45	64	77	68	86	**10.6 ± 6.6**
4D VMAT simulation: 15 seg. (10˚)	63	54	27	41	68	88	79	**19.4 ± 11.7**
4D VMAT simulation: 1 seg. (150˚)	59	57	59	48	86	93	82	**14.9 ± 11.1**

Cf. Table 2 for symbols and indices. Similar to Table 3, a simulation result is considered superior to another if its γ-passing rate values are closer to corresponding measurement values.

doi:10.1371/journal.pone.0172810.t004

simulation-based γ-maps for regular motion, independent of the degree of arc discretization, is mirrored by high γ-passing rates when directly comparing the simulated motion-affected dose distributions and the dynamic measurements; cf. Table 2. In turn, the visual differences for scenario 2b (3rd column of Fig 4) translated into a drop of the γ-passing rate from 97% ($\Delta\alpha = 2.3^\circ$) to 82% ($\Delta\alpha = 150^\circ$).

A similar tendency could be observed when directly studying the γ-passing rates obtained by comparison of, on the one hand, static and dynamic measurements, and, on the other hand, γ-passing rates for comparison of statically planned and simulated motion-affected dose distributions (i. e. γ-passing rates corresponding to, for instance, the γ-maps of Fig 4). As these γ-passing rates directly quantify measured and simulated motion effects, the simulation values should be as close as possible to the measurement values. In turn, related differences were assumed to indicate simulation uncertainties. Table 3 already reflects that the average differences of simulation to measurement γ-rates are smallest for the finest discretization level $\Delta\alpha = 2.3^\circ$. The differences become, however, more pronounced when switching from total to structure-based γ-rates (i. e. when focusing on higher dose areas) shown in Tables 4 and 5. In this case, $\Delta\alpha = 2.3^\circ$ not only resulted in lowest $\bar{\Delta}_\gamma$ values but also (more or less on a par with $\Delta\alpha = 5^\circ$) γ-passing rates differences to the day 1 measurement that were in the order of the day 2-to-day 1 measurement differences for, e. g., the CTV.

Observations for the other arcs of the treatment plans were similar.

Table 5. ITV γ-passing rates for comparison of static dose distributions and dynamic dose measurements/simulations.

	ITV: $D_{sta,\cdot}$ vs. $D_{dyn,\cdot}$ [%]							$\bar{\Delta}_\gamma$ [%]
	1a	1b	1c	1d	1e	2a	2b	
Day 1 measurement	48	57	47	58	50	72	52	–
Day 2 measurement	38	48	41	46	48	61	49	**7.6 ± 3.7**
4D VMAT simulation: 65 seg. (2.3˚)	43	41	33	42	44	72	60	**9.3 ± 5.7**
4D VMAT simulation: 30 seg. (5˚)	43	41	29	40	46	71	56	**9.4 ± 7.0**
4D VMAT simulation: 15 seg. (10˚)	42	35	26	32	41	76	53	**12.7 ± 9.3**
4D VMAT simulation: 1 seg. (150˚)	39	38	39	35	48	79	56	**10.3 ± 7.2**

Cf. Table 2 for symbols and indices and Table 4 for further explanations.

doi:10.1371/journal.pone.0172810.t005

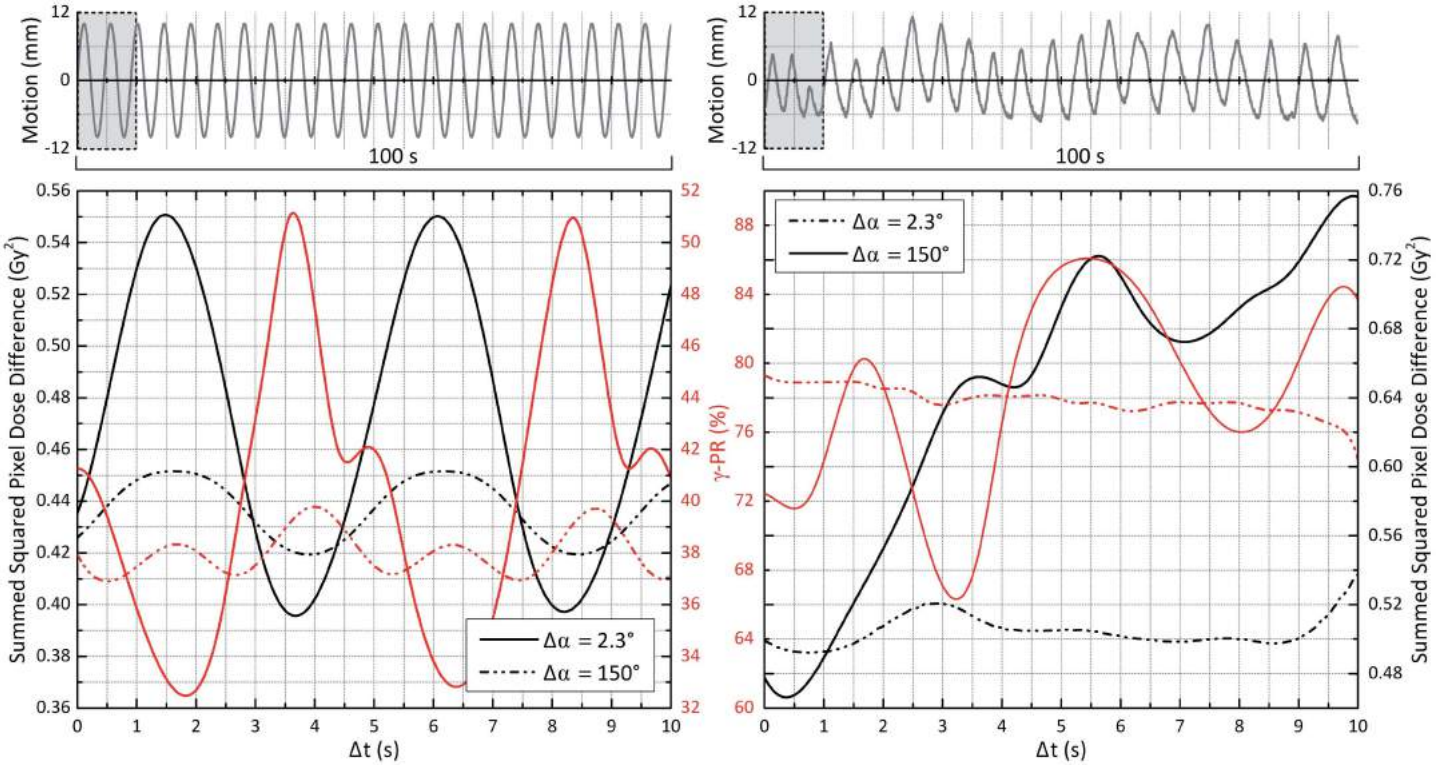


Fig 5. Starting phase influence. Influence of breathing phase at dose delivery beginning. Left, top: In accordance with the measurements, all previous results were computed with the simulations starting at the breathing phase at $t = 0$ s of the curve (here: case 1b). Now, this starting phase was systematically varied by adding offsets $\Delta t \in [0\,\mathrm{s}; 10\,\mathrm{s}]$. Left, bottom: The ITV γ-passing rates for comparison of planned static and motion-affected simulated dose distributions are shown as red lines (solid lines: $\Delta\alpha = 2.3^\circ$; dashed: $\Delta\alpha = 150^\circ$); the black lines visualize the dependence of the difference between dynamic measurement and simulated motion-affected dose on the starting phase. Right: similar information but for the regular real tumor trajectory (case 2a).

doi:10.1371/journal.pone.0172810.g005

Influence of starting phase/synchronization uncertainties

So far, perfect agreement between the initial breathing phase of the motion phantom at dose delivery beginning and the breathing phase applied to $t = 0$ s of the simulations was assumed. As signal latencies potentially led to phase shifts between measurement and simulation, in the next step, the breathing phase at $t = 0$ s of the simulations was systematically varied. The effect on the ITV γ-passing rates between statically planned and simulated motion-affected dose on the one hand and absolute pixel-wise squared dose differences (SDD) between dynamic measurement and simulation on the other hand are shown in Fig 5 for synthetic (case 1b) and real (case 2a) motion scenarios. For both $\Delta\alpha = 2.3^\circ$ and $\Delta\alpha = 150^\circ$, it can be seen that the SDD values were close to minimum for $\Delta t = 0$ s, with the actual SDD minima, however, being slightly shifted in time (in negative direction for case 1b, in positive direction for case 2a).

In addition to the illustration of uncertainties due to mismatches of measurement and simulation starting phases, the figure also demonstrates the dependence of predicted VMAT interplay effects on the assumed starting phase (in terms of γ-passing rates between static planned and simulated motion-affected dose distributions) by varying Δt over a longer period (here: 10 s). It becomes obvious that the $\Delta\alpha = 2.3^\circ$ simulations (beforehand proven to be the most accurate ones) predicted a substantial impact of the starting phase on the γ-passing rate and amount of interplay effects, respectively–whereas such details were not reflected by the $\Delta\alpha = 150^\circ$ simulations.

Static dose calculation uncertainties

In agreement with Ref. [23, 31], uncertainties related to static dose calculation approaches proved subordinate to the temporal discretization and the assumed breathing phase at the beginning of dose delivery. 3%/3 mm γ-comparison of simulated motion-affected dose distributions with the doses $D_{i,j,\alpha}$ computed based on the closest neighbored phase CT, a fixed single phase CT or the average CT resulted in total γ-passing rates between 98% and 100% for all motion scenarios. Uncertainties due to AAA or Acuros XB dose calculation were in the order of ±3%.

Discussion and conclusions

Being placed in the context of the current discussion of appropriate VMAT 4D quality assurance approaches, the present study aimed at assessing 4D VMAT dose simulation accuracy and associated influencing parameters by comparing simulated motion-affected dose distributions to corresponding ground truth dynamic measurements–and so study potential limitations of computational simulation-only 4D VMAT QA to assess appropriateness of a VMAT treatment plan when taking into account patient-specific breathing and motion characteristics.

We identified and illustrated the temporal discretization of the dose delivery process as major technical factor and the breathing phase at dose delivery beginning as most relevant physiology/breathing curve-related parameter. Applying the highest VMAT arc discretization of 2.3˚ achievable by the treatment planning system, high total γ-passing rates of on average 98% between simulations and dynamic measurements [cf. Table 2] rise, from our perspective, the question whether measurement-based assessment of VMAT motion effects still remains necessary or if it can be replaced by 4D VMAT dose simulations–with our opinion being the latter.

As 4D VMAT dose simulation accuracy has been shown to decrease for arc discretization levels larger than 2.3˚, remaining small differences to the measurements may be in parts due to the mentioned software limitations (i. e., $<$ 2.3˚ discretization not possible). However, uncertainties such as existence of a potential shift between breathing phases at dose measurement and simulation beginning, uncertainties of the applied dose calculation algorithm (although illustrated to be small), and, e. g., the angular and dose rate dependence of the detector array (although also reported to be small for the gantry angles and dose rates exploited in our study, cf. [28] for respective details) are likely to also affect the numbers. This superposition of uncertainties motivates further detailed studies.

In terms of limitations of our study, we would again emphasize that we explicitly refrained from analyzing uncertainties introduced by registration errors associated to, e. g., motion estimation in clinical 4D CT data. This has been in parts due to the motion phantom design (only rigidly moving phantom); we, however, also believe that uncertainty estimation and quantification of deformable image registration represents an issue that has yet not been solved in its entirety. We consider this topic to be beyond the scope of this study but to represent an important aspect of our future work.

As a consequence of the aforementioned limitation, the impact of potential breathing-induced deformation on, e. g., the interplay effect, target coverage and the numbers presented in Tables 3 to 5 remains unclear. This aspect could be of interest in terms of future work as (although not being the primary focus of our QA-oriented study) our results indeed demonstrate interplay effects and clinically relevant low CTV γ-passing rates for at least our single-arc scenarios; and even for the full dual-arc plans and simulations, low CTV γ-passing rates $<$ 60% could be observed for irregular and longer breathing period motion, despite the well-

reported averaging effect for multiple arcs [10, 23]. However, these results are obtained using only rigid motion patterns. In addition, the interplay effects are shown to depend on the breathing phase at dose delivery beginning, and we would also like to note that only a standard dose rate of 600 MU/min was used in the current study. Higher dose rates, e. g. provided by flattening filter-free (FFF) dose delivery, are likely to result in a further increased risk of VMAT interplay effects [32]. A detailed discussion of such aspects and related questions (How to account for potentially different breathing phases at the beginning of VMAT dose delivery during 4D QA? How to account for breathing variability during 4D QA? How to stabilize 4D VMAT treatment planning in terms of a robust target dose coverage even in the presence of motion variability?) would again be beyond the scope of this paper but motivates additional studies.

Supporting information

S1 File. File containing dose distributions underlying the manuscript figures and numbers. The zip-file contains the simulated 4D VMAT dose distributions and the the dose measurements (serving as ground truth data) that are underlying the figures and numbers presented in the manuscript. For further details see the Readme.md file contained in the zip-file.
(ZIP)

Author Contributions

Conceptualization: TS TG RW.

Data curation: TS TG RW.

Formal analysis: TS RW.

Funding acquisition: RW.

Investigation: TS TG.

Methodology: TS RW.

Project administration: TS TG RW.

Resources: TG RW.

Software: TS RW.

Supervision: RW TG.

Validation: TS TG RW.

Visualization: TS.

Writing – original draft: RW.

Writing – review & editing: TS TG RW.

References

1. Katz AW, Carey-Sampson M, Muhs AG, Milano MT, Schell MC, Okunieff P. Hypofractionated stereotactic body radiation therapy (SBRT) for limited hepatic metastases. Int J Radiat Oncol Biol Phys. 2007; 67(3):793–798. doi: 10.1016/j.ijrobp.2006.10.025 PMID: 17197128
2. Simeonova AO, Fleckenstein K, Wertz H, Frauenfeld A, Boda-Heggemann J, Lohr F, et al. Are three doses of stereotactic ablative radiotherapy (SABR) more effective than 30 doses of conventional

radiotherapy? Transl Lung Cancer Res. 2012; 1(1):45–53. doi: 10.3978/j.issn.2218-6751.10.01 PMID: 25806154

3. McGrath SD, Matuszak MM, Yan D, Kestin LL, Martinez AA, Grills IS. Volumetric modulated arc therapy for delivery of hypofractionated stereotactic lung radiotherapy: A dosimetric and treatment efficiency analysis. Radiother Oncol. 2010; 95(2):153–157. doi: 10.1016/j.radonc.2009.12.039 PMID: 20116115
4. Keall PJ, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med Phys. 2006; 33(10):3874–3900. doi: 10.1118/1.2349696 PMID: 17089851
5. Korreman SS. Motion in radiotherapy: photon therapy. Phys Med Biol. 2012; 57(23):R161–91. doi: 10.1088/0031-9155/57/23/R161 PMID: 23165229
6. Stambaugh C, Nelms BE, Dilling T, Stevens C, Latifi K, Zhang G, et al. Experimentally studied dynamic dose interplay does not meaningfully affect target dose in VMAT SBRT lung treatments. Med Phys. 2013; 40(9):091710. doi: 10.1118/1.4818255 PMID: 24007143
7. Werner R, Ehrhardt J, Schmidt-Richberg A, Albers D, Frenzel T, Petersen C, et al. Towards accurate dose accumulation for Step-&-Shoot IMRT: Impact of weighting schemes and temporal image resolution on the estimation of dosimetric motion effects. Z Med Phys. 2012; 22(2):109–22. doi: 10.1016/j.zemedi.2011.08.001 PMID: 21924880
8. Dogan N, King S, Emami B, Mohideen N, Mirkovic N, Leybovich LB, et al. Assessment of different IMRT boost delivery methods on target coverage and normal-tissue sparing. Int J Radiat Oncol Biol Phys. 2003; 57(5):1480–91. doi: 10.1016/S0360-3016(03)01569-4 PMID: 14630288
9. Palma DA, Verbakel WFAR, Otto K, Senan S. New developments in arc radiation therapy: a review. Cancer Treat Rev. 2010; 36(5):393–399. doi: 10.1016/j.ctrv.2010.01.004 PMID: 20181430
10. Bortfeld T, Jokivarsi K, Goitein M, Kung J, Jiang SB. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation. Phys Med Biol. 2002; 47(13):2203–2220. doi: 10.1088/0031-9155/47/13/302 PMID: 12164582
11. Goodman KA, Wiegner EA, Maturen KE, Zhang Z, Mo Q, Yang G, et al. Dose-escalation study of single-fraction stereotactic body radiotherapy for liver malignancies. Int J Radiat Oncol Biol Phys. 2010; 78(2):486–493. doi: 10.1016/j.ijrobp.2009.08.020 PMID: 20350791
12. Guckenberger M, Richter A, Krieger T, Wilbert J, Baier K, Flentje M. Is a single arc sufficient in volumetric-modulated arc therapy (VMAT) for complex-shaped target volumes? Radiother Oncol. 2009; 93(2):259–265. doi: 10.1016/j.radonc.2009.08.015
13. Zhang GG, Ku L, Dilling TJ, Stevens CW, Zhang RR, Li W, et al. Volumetric modulated arc planning for lung stereotactic body radiotherapy using conventional and unflattened photon beams: a dosimetric comparison with 3D technique. Radiat Oncol. 2011; 6(1):152. doi: 10.1186/1748-717X-6-152 PMID: 22070866
14. Ong CL, Dahele M, Slotman BJ, Verbakel WFAR. Dosimetric impact of the interplay effect during stereotactic lung radiation therapy delivery using flattening filter-free beams and volumetric modulated arc therapy. Int J Radiat Oncol Biol Phys. 2013; 86(4):743–8. doi: 10.1016/j.ijrobp.2013.03.038 PMID: 23773394
15. Ehrbar S, Lang S, Stieb S, Riesterer O, Stark LS, Guckenberger M, et al. Three-dimensional versus four-dimensional dose calculation for volumetric modulated arc therapy of hypofractionated treatments.
16. Li X, Yang Y, Li T, Fallon K, Heron DE, Huq MS. Dosimetric effect of respiratory motion on volumetric-modulated arc therapy-based lung SBRT treatment delivered by TrueBeam machine with flattening filter-free beam. J Appl Clin Med Phys. 2013; 14:4370. PMID: 24257279
17. Poulsen PR, Schmidt ML, Keall P, Worm ES, Fledelius W, Hoffmann L. A method of dose reconstruction for moving targets compatible with dynamic treatments. Med Phys. 2012; 39(10):6237–6246. doi: 10.1118/1.4754297 PMID: 23039659
18. Rao M, Wu J, Cao D, Wong T, Mehta V, Shepard D, et al. Dosimetric impact of breathing motion in lung stereotactic body radiotherapy treatment using image-modulated radiotherapy and volumetric modulated arc therapy. Int J Radiat Oncol Biol Phys. 2012; 83(2):e251–e256. doi: 10.1016/j.ijrobp.2011.12.001 PMID: 22365622
19. Riley C, Yang Y, Li T, Zhang Y, Heron DE, Huq MS. Dosimetric evaluation of the interplay effect in respiratory-gated RapidArc radiation therapy. Med Phys. 2014; 41:011715. doi: 10.1118/1.4855956 PMID: 24387507
20. Zou W, et al. Dynamic simulation of motion effects in IMAT lung SBRT. Radiat Oncol. 2014; 9:225. doi: 10.1186/s13014-014-0225-3 PMID: 25365935
21. Ong C, Verbakel WF, Cuijpers JP, Slotman BJ, Senan S. Dosimetric impact of interplay effect on RapidArc lung stereotactic treatment delivery. International Journal of Radiation Oncology* Biology* Physics. 2011; 79(1):305–311. doi: 10.1016/j.ijrobp.2010.02.059

22. Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. Br J Radiol. 2011; 84(1007):967–96. doi: 10.1259/bjr/22373346 PMID: 22011829

23. Milz S, Wilkens JJ, Ullrich W. A dose error evaluation study for 4D dose calculations. Phys Med Biol. 2014; 59(21):6401–15. doi: 10.1088/0022-3727/59/21/6401 PMID: 25295741

24. Murphy K, et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Trans Med Imaging. 2011; 30(11):1901–20. doi: 10.1109/TMI.2011.2158349 PMID: 21632295

25. Werner R, Schmidt-Richberg A, Handels H, Ehrhardt J. Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study. Phys Med Biol. 2014; 59(15):4247–60. doi: 10.1088/0031-9155/59/15/4247 PMID: 25017631

26. Poppe B, Stelljes TS, Looe HK, Chofor N, Harder D, Willborn K. Performance parameters of a liquid filled ionization chamber array. Med Phys. 2013; 40(8):082106. doi: 10.1118/1.4816298 PMID: 23927343

27. Grohmann C, Frenzel T, Werner R, Cremers F. Design, performance characteristics and application examples of a new 4D motion platform. Z Med Phys. 2015; 25:156–67. doi: 10.1016/j.zemedi.2014.09.003 PMID: 25270980

28. Markovic M, Stathakis S, Mavroidis P, Jurkovic IA, Papanikolaou N. Characterization of a two-dimensional liquid-filled ion chamber detector array used for verification of the treatments in radiotherapy. Med Phys. 2014; 41(5):051704. doi: 10.1118/1.4870439 PMID: 24784370

29. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys. 1998; 25(5):656–661. doi: 10.1118/1.598248 PMID: 9608475

30. Sothmann T, Blanck O, Poels K, Werner R, Gauer T. Real time tracking in liver SBRT: comparison of CyberKnife and Vero by planning structure-based γ-evaluation and dose-area-histograms. Phys Med Biol. 2016; 61(4):1677–91. doi: 10.1088/0031-9155/61/4/1677 PMID: 26836488

31. Tsuruta Y, Nakata M, Nakamura M, Matsuo Y, Higashimura K, Monzen H, et al. Dosimetric comparison of Acuros XB, AAA, and XVMC in stereotactic body radiotherapy for lung cancer. Med Phys. 2014; 41(8):081715. doi: 10.1118/1.4890592 PMID: 25086525

32. Gauer T, Sothmann T, Werner R. EP-1763: Experimental analysis of interplay effects in flattening filter free VMAT treatment techniques. Radiotherapy and Oncology. 2016; 119:S826–S827. doi: 10.1016/S0167-8140(16)33014-6

5.6. Monte Carlo-based 4D dose simulation

In the previous section, the accuracy and affecting parameters of the currently implemented dose simulation scheme were evaluated. Corresponding findings suggest that the simulation accuracy can be further improved by decreasing dose segment angular spacing, which is, however, limited by the TPS. Moreover, as the method introduced before is based on pre-calculated dose distributions extracted out of the original treatment plan, density effects in dose distributions due to internal patient motion are not considered. Therefore, the idea was to re-implement the current approach into a Monte Carlo-based method that enables highest possible discretization level and, at the same time, takes density changes into account by computing the dose for each time dependent patient geometry. The proposed scheme is based on following publication:
T. Sentker, F. Madesta and R. Werner. Patient-specific 4D Monte Carlo dose accumulation using correspondence model-based motion prediction. In: *Proc SPIE*, Vol. 10951, 1095109. International Society for Optics and Photonics, SPIE, 2019.

In line with previous sections, prediction of internal patient motion was conducted by a patient-specific correspondence model trained on the 4D CT data set of the patient. The dose simulation, however, was completely re-implemented to allow for dose calculation utilizing a Monte Carlo-based algorithm. Basics of Monte Carlo-based dose calculation and accumulation are detailed in Section 4.2.1. Necessary information for the 4D Monte Carlo dose simulation was directly extracted from the original patient treatment plan. Here, the smallest possible segments are control points that define for each time point gantry angle/position, MLC leaf positions defining the field opening and MU to be delivered. The temporal resolution of control points is depending on the gantry rotation speed and approximately 0.1 s to 1 s. At each time point, the inverse of the predicted motion vector field was used to transform the reference 3D CT image to match the patient geometry according to the corresponding breathing state. On this virtual moving image, the Monte Carlo-estimated dose was computed using corresponding beam parameters. After dose calculation, the predicted motion vector field was applied to transform the dose image back to the reference frame and yield the sought motion-affected dose distribution. Summation over all simulated motion-affected doses yielded the total motion affected patient dose. Afterwards, an absolute dose calibration was necessary for the Monte Carlo-based dose calculation to allow for an absolute dose comparison between 4D-simulated dose and planned dose distribution. Calibration was accomplished by correlating dose measurements under reference conditions and corresponding simulation. Applying the proposed simulation scheme to the in-house acquired patient data sets (cf. Section 4.4.4) allowed to compare the new and initial 4D dose simulation approach. Not yet published results for dose simulation accuracy of the new approach, evaluated using the dose

Table 5.1.: Additional γ-passing rates for analysis of 4D dose accumulation accuracy gain achieved by using new Monte Carlo-based dose simulation scheme. The mean absolute difference $\langle \Delta_\gamma \rangle$ between the γ-passing rates obtained by the simulations and the γ-passing rates of the day 1 measurements should be (averaged over all motion scenarios) as low as possible.

		γ, $D_{\text{sta,.}}$ vs. $D_{\text{dyn,.}}$ (%)							$\langle \Delta_\gamma \rangle$ (%)
		1a	1b	1c	1d	1e	2a	2b	
Total	Day 1 measurement	65	67	64	68	62	86	68	–
	4D MC sim.: 194 seg. ($\approx 0.8°$)	69	65	62	64	57	88	67	**2.9 ± 1.5**
	4D VMAT sim.: 65 seg. (2.3°)	61	60	57	63	60	86	69	**3.7 ± 2.6**
CTV	Day 1 measurement	45	73	59	73	73	59	80	–
	4D MC sim.: 194 seg. ($\approx 0.8°$)	44	66	59	63	72	68	90	**5.4 ± 4.2**
	4D VMAT sim.: 65 seg. (2.3°)	66	68	57	68	82	71	93	**9.6 ± 6.0**
ITV	Day 1 measurement	48	57	47	58	50	72	52	–
	4D MC sim.: 194 seg. ($\approx 0.8°$)	51	49	49	48	45	74	61	**5.6 ± 3.2**
	4D VMAT sim.: 65 seg. (2.3°)	43	41	33	42	44	72	60	**9.3 ± 5.7**

measurements presented in the previous section, are additionally described and put into context to the initial dose simulation accuracy.

Results of the comparison of initial and new dose simulation approach are summarized in Table 1 [SPIE 2019]. For the liver cases, predicted dose differences are similar but underdosages for patients with local metastasis recurrence are more pronounced for the new Monte Carlo-based simulation approach. Deviations for the lung cases are less prominent for the investigated metastases. However, for metastases with local recurrence, estimated underdosages are still more pronounced when using the Monte Carlo-based simulation scheme. Further and similar to the liver cases, some metastases without local recurrence exhibit noticeable overdosages. The assumption that the higher temporal resolution in combination with consideration of density changes of the internal patient geometry during dose simulation improves the general simulation accuracy is, however, not yet demonstrated. Thus, the accuracy of the finest temporal resolution of the initial approach ($\approx$ 0.5 s to 2 s) is directly compared to the proposed simulation scheme (temporal resolution of $\approx$ 0.1 s to 1 s) by ground truth dose measurements as conducted in Section 5.5. And indeed, due to higher temporal resolution and an actual dose recalculation, and thus consideration of density changes in the patient geometry, the new approach achieves a higher dose simulation accuracy, as numbers in Table 5.1 indicate. In particular values for smaller and highly radiotherapy-relevant structures (CTV, ITV) are improved and in better agreement to measurements. Consequently, the proposed 4D Monte Carlo dose simulation scheme for quality assurance in 4D radiotherapy has the potential to replace measurement-based assessment of VMAT motion effects.

Patient-specific 4D Monte Carlo Dose Accumulation using Correspondence Model Based Motion Prediction

Thilo Sentker[a,b], Frederic Madesta[a,b], and René Werner[a]

[a]Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[b]Department of Radiotherapy and Radio-Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

ABSTRACT

Quality assurance in current 4D radiotherapy workflows is of great importance to assure a positive treatment outcome, i.e. total tumor eradication. Especially for the treatment of lung and liver tumors, which are subject to high motion magnitudes due to patient breathing, it is crucial to verify the applied dose to the target volume. In this study, we present a new 4D Monte Carlo dose accumulation approach that accounts for internal patient motion during treatment and is therefore able to predict the actual 3D dose distribution delivered to the patient for quality assurance purposes. Monte Carlo simulations are conducted using the EGSnrc software toolkit, which models the propagation of photons, electrons and positrons. However, to consider dynamic beam parameters and the movement of internal patient geometry, we developed a method to compute the dose for each control point of the actual VMAT patient treatment plan to account for breathing induced internal patient motion. The internal motion during treatment is predicted using correspondence modeling, which correlates patient-specific DIR-based internal motion information and external breathing signals and is trained on 4D CT data of the patient. For each VMAT control point, a corresponding motion vector field is predicted and applied to the original patient CT to allow for dose computation on the patient geometry as it was irradiated during treatment. Thus, density changes while treatment due to patient breathing motion are taken into account during computation of the resulting dose distribution.

Keywords: Monte Carlo simulation, 4D dose accumulation, correspondence modeling, deformable image registration

1. INTRODUCTION

In extracranial stereotactic radiation therapy (SBRT), a high radiation dose is delivered in a few fractions to the target volume while simultaneously sparing organs at risk. Thus, for SBRT of lung and liver tumors, which are subject to high motion magnitudes due to patient breathing,[1] the consideration of tumor motion and deformation while treatment is crucial. To do so, a time resolved CT (3D+t, 4D CT) is acquired pre-treatment and used for defining the tumor motion space (so called internal target volume, ITV) in the treatment planning process. By using an ITV-based treatment planning it is usually assumed that the target volume is sufficiently irradiated during dose application even if the patient is breathing freely. However, the combination of dynamically changing beam parameters (gantry angle, field form, dose rate and fluence) and patient-specific intra- and inter-fractional respiratory variability can nevertheless lead to deviations between actual delivered and planned dose distributions.[2,3] Therefore, retrospective quality assurance in current 4D radiotherapy workflows is of great importance to assure a positive treatment outcome, i.e. the total tumor eradication. This directly motivates the development of an algorithm that is capable of computing the actual to the patient delivered dose. Our existing correspondence model-based 4D dose accumulation approach[4–6] provides a first retrospective estimation of this applied dose. However, in its current form, the algorithm lacks some important features: 1) density changes of the patient geometry due to internal motion are not considered during 4D dose accumulation, as the dose is not actually re-calculated, and 2) the temporal resolution is with 2.3 s relatively low.[3] Therefore, we decided to enhance our algorithm by implementing an actual dose simulation algorithm using the EGSnrc/DOSXYZnrc user code.[7]

Send correspondence to Thilo Sentker né Sothmann; E-mail: t.sentker@uke.de

Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling,
edited by Baowei Fei, Cristian A. Linte, Proc. of SPIE Vol. 10951, 1095109 · © 2019 SPIE
CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2512423

In line with our previous work, the internal patient motion during treatment is predicted using correspondence modeling, which correlates patient-specific deformable image registration (DIR)-based internal motion information and external breathing signals and is trained on 4D CT data. For each VMAT control point, a corresponding motion vector field is predicted and applied to the original patient CT to allow for dose computation on the patient geometry as it was irradiated during treatment. This allows for consideration of density changes during treatment due to patient breathing motion and increases the temporal resolution to 0.2 s to 0.8 s, which is defined by by control point sampling.

In this manuscript, we present our new approach and show first preliminary results for lung and liver tumor patients treated at our hospital.

2. METHOD(S)

In this section, first, the basic correspondence model-based dose accumulation approach is briefly explained before the integration of this method into a 4D Monte Carlo dose simulation scheme is described. Further, absolute dose calibration and treatment plans of our patient collective used for retrospective dose simulation are introduced.

2.1 Concept of Correspondence Model-based 4D Dose Accumulation

The mathematical background of correspondence model-based 4D dose accumulation is detailed in our previous publications.[4–6] For comprehensibility, however, a brief summary of the approach is given in this section.

The main idea of correspondence modeling is to train a model that is able to predict the internal patient motion during radiation therapy treatment using only an external breathing signal measurement. Training of the model is conducted using a pre-treatment acquired 4D CT that consists of a series of, in our case, 10 3D CT phase images. During CT imaging, the usually one-dimensional external breathing signal of the patient is recorded and subsequently used for CT reconstruction as well as correspondence modeling. The functional relationship between external breathing signal measurements and internal breathing phase-specific images is represented by the sought linear correspondence model. Distinction between patient in- and exhalation is accomplished by using not only the amplitude signal of recorded patient breathing but additionally the computed velocity information. Representation of internal motion is achieved by displacement vector fields (DVF) estimated by non-linear registration.

2.2 4D Monte Carlo Dose Simulation

For each patient, the corresponding treatment plan

$$RT_{\text{plan}}(t_{\text{cp}}) : t_{\text{cp}} \mapsto \{\text{beam parameters}\} \tag{1}$$

is divided into control points (cp) that define the beam parameters (e.g. gantry angle, leaf positions defining the field, beam energy etc.) for a specific time point t_{cp}. Here, the time resolution of following dose accumulation is directly determined by the temporal spacing of consecutive t_{cp}, ranging from 0.2 s to 0.8 s. At every t_{cp}, a transformation $T^{I_{\text{ref}}}_{\hat{I}(t_{\text{cp}})} \in \mathcal{T}$ can be predicted using the patient specific correspondence model M_{C} and respective external patient motion information $z_i(t_{\text{cp}})$ and $\partial_t z_i(t_{\text{cp}})$

$$M_{\text{C}}(z_i, \partial_t z_i) : \mathbb{R}^2 \to \mathcal{T}, (z_i, \partial_t z_i)^\top \mapsto T^{I_{\text{ref}}}_{\hat{I}(t_{\text{cp}})}, \quad I_{\text{ref}}, \hat{I} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}. \tag{2}$$

Here, $T^{I_{\text{ref}}}_{\hat{I}(t_{\text{cp}})}$ is the transformation of the hypothetical moving image $\hat{I}(t_{\text{cp}})$ to the pre-defined reference image I_{ref}.

Correspondence model-based 4D Monte Carlo dose simulation is then conceptually straightforward:

$$d_{t_{\text{cp}}} = D_{\text{MC}} \left(\underbrace{I_{\text{ref}} \circ \left(T^{I_{\text{ref}}}_{\hat{I}(t_{\text{cp}})} \right)^{-1}}_{\hat{I}(t_{\text{cp}})}, \{\text{beam parameters}\} \right) \circ T^{I_{\text{ref}}}_{\hat{I}(t_{\text{cp}})}. \tag{3}$$

The Monte Carlo dose simulation $D_{\mathrm{MC}} : \Omega \to \mathbb{R}^+$ is executed on the hypothetical moving image $\hat{I}$ resulting by warping the reference image I_{ref} with the predicted inverse transformation $\left(T^{I_{\mathrm{ref}}}_{\hat{I}(t_{\mathrm{cp}})}\right)^{-1}$ and applying corresponding beam parameters. Warping the Monte Carlo simulated dose back using $T^{I_{\mathrm{ref}}}_{\hat{I}(t_{\mathrm{cp}})}$ yields sought motion affected dose distribution $d_{t_{\mathrm{cp}}}$. Finally, summation over all fractions fx and t_{cp} results in the total 4D dose distribution

$$D_{4\mathrm{D}} = \sum_{\mathrm{fx}} \sum_{t_{\mathrm{cp}}} d_{t_{\mathrm{cp}}}. \tag{4}$$

2.2.1 Absolute Dose Calibration

Absolute dosimetric comparison between applied 4D simulated dose and planned 3D dose distribution requires an absolute dose calibration of Monte Carlo simulated doses. Typically, a dose measurement inside a water phantom using an ionization chamber in different depths along the central axis (z-direction) under standard conditions is necessary to determine sought calibration factor. To calibrate the simulated dose, the same setup used for dose measurement is simulated utilizing the implemented Monte Carlo method.

General output of the Monte Carlo simulation is a dose that is normalized by an estimate of the number of particles incident from the original, non-phase space source. This means the simulated dose at a specific z-position, $D_{z,\mathrm{MC}}$, can be correlated to the measured absolute dose at the same position, $D_{z,\mathrm{chamber}}$, generated by a defined number of delivered monitor units (MU). For calibration measurements, $100\,\mathrm{MU} \equiv 1\,\mathrm{Gy}$ at dose maximum applies. Finally, for any simulation result at any x, y, z-position, D_{xyz}, absolute dose calibration can be performed using following equation

$$D_{xyz} = \frac{D_{z,\mathrm{MC}}}{D_{z,\mathrm{chamber}}} \times \frac{1\,\mathrm{cGY}}{1\,\mathrm{MU}} \times \mathrm{MU}_{\mathrm{tot}}, \tag{5}$$

with $\mathrm{MU}_{\mathrm{tot}}$ being the total number of applied MU during treatment.

Note that calibration has to be performed for each beam energy independently.

2.3 Patient Collective and Simulations

In total, 10 4D CT data sets and corresponding SBRT treatment plans of cancer patients previously treated at our hospital (five liver patients with in total nine metastases and five lung patients with in total six metastases) were selected for 4D dose accumulation. Treatment plans were directly used for the workflow described in Section 2.2. Number of control points per treatment plan were in the range of 400 to 600 and the Monte Carlo simulation of one control point dose took approximately 1000 s to 3000 s when simulating 5×10^5 histories. Therefore, total number of particles ranged between 2×10^8 and 3×10^8 per treatment fraction. As a computation cluster of about 100 CPU cores were available, the total 4D dose accumulation for one patient treatment fraction took about 1.5 h to 3 h. The phase space file provided by the linear accelerator manufacturer was used for initialization of the treatment beam above the dynamically changeable accelerator geometry (jaws, multi leaf collimator). Beam parameters for each control point simulation input file were directly extracted from the specific patient treatment plan. Corresponding 4D CT data was utilized to build the required correspondence model for predicting internal patient motion during treatment by applying the provided external patient breathing information. Possible simulated underdosages (3D vs. 4D dose simulation) were correlated to known clinical endpoints (local recurrence yes/no) for each metastases.

3. RESULTS

Evaluation of results are conducted by comparison of static and motion affected dose distributions and correlation of underdosages to known clinical endpoints. Dose differences are expressed by deviations of D_{98} (dose to 98% of the GTV volume) of 4D and 3D simulated dose distributions. In Table 1, the results of all investigated metastases are summarized for each individual treatment fraction and in total. Further, 4D total dose differences simulated based on our previous dose accumulation approach (cf. Sothmann et al.[4]) are shown for comparison.

Table 1. GTV dose coverage of treatment plan and retrospectively 4D Monte Carlo simulated dose distribution. GTV dose coverage is quantified by differences ΔD_{98} of $D_{98,4D\text{-}MC}$ and $D_{98,3D\text{-}MC}$. For the proposed correspondence model-based 4D dose simulation, the data is shown separately for each patient treatment plan fraction and for the accumulated fraction dose distributions (= total dose). 4D dose accumulation based on our previous approach refers to the estimated total dose (last column). Additional information about dose difference values in each fraction can be found in corresponding publication.[4] For metastasis 1, i. e. the only treatment plan with 8 treatment fractions, results for fraction 6, 7 and 8 are shown in a separate line (indicated by the notation 'Frac. 1/6' etc.). Gray row = local metastasis recurrence.

	Met.	**$\Delta D_{98} = D_{98,4D\text{-}MC} - D_{98,3D\text{-}MC}$ (Gy)**						
		Frac. 1/6	**Frac. 2/7**	**Frac. 3/8**	**Frac. 4**	**Frac. 5**	**Total**	***Old***
Liver	1	−0.38	−0.50	−0.40	−0.94	−0.40		
		−0.42	−0.37	−0.46			−3.52	*−1.49*
	2.1	+0.03	+0.04	−0.22	−0.02	−0.08	−0.14	*−0.25*
	2.2	+0.09	+0.03	+0.01	+0.04	+0.05	+0.35	*+0.29*
	2.3	−0.11	−0.05	+0.10	−0.04	−0.13	−0.04	*+0.33*
	3.1	+0.16	+0.14	+0.19	+0.17	+0.17	+0.86	*+0.10*
	3.2	+0.12	+0.16	+0.17	+0.13	+0.17	+0.76	*+0.10*
	3.3	−3.56	−5.26	−5.07	−5.04	−5.20	−24.13	*−13.28*
	4	+0.02	−0.06	−0.09	−0.12	−0.03	+0.33	*−0.18*
	5	−0.29	−0.03	−0.29	−0.15	−0.16	−0.68	*−0.71*
Lung	6	+0.02	+0.06	+0.10	+0.28	+0.12	+0.79	*+0.50*
	7.1	+0.12	−0.22	+0.10	−0.13	−0.07	−0.20	*−0.08*
	7.2	−0.07	−0.37	−0.11	−0.30	−0.09	−0.88	*−0.21*
	8	+0.29	+0.11	+0.48	+0.41	+0.52	+2.22	*+0.44*
	9	−0.51	−0.20	+0.01	+0.07	−0.30	−0.71	*−0.47*
	10	+0.11	−0.06	−0.05	+0.02	−0.06	+0.16	*−0.03*

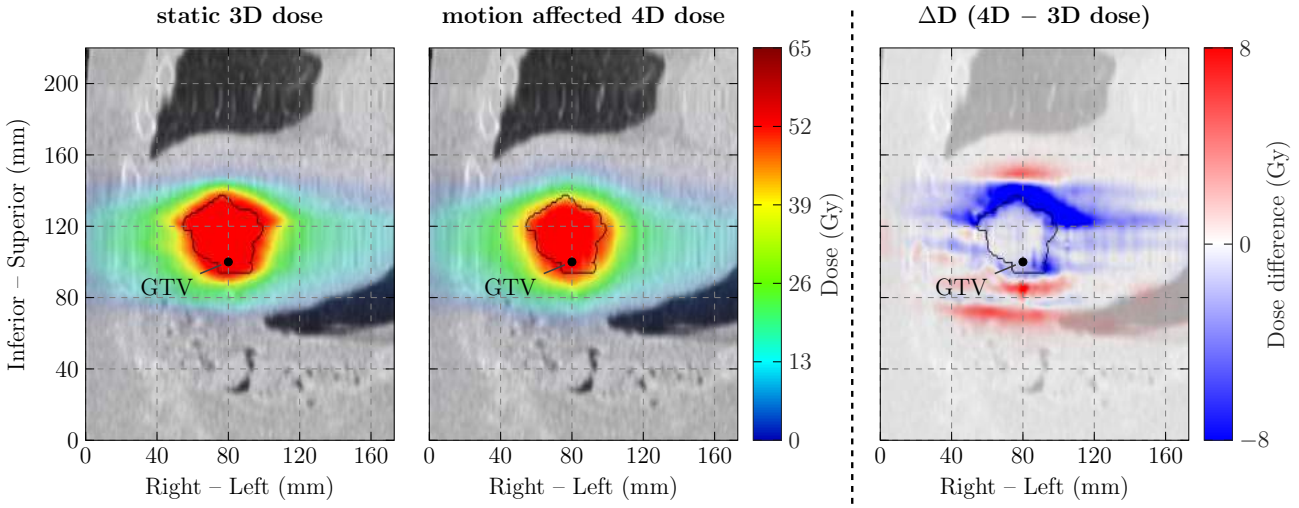


Figure 1. Comparison of static 3D (left) and motion affected 4D (middle) dose distribution for patient data set 1 (liver), simulated using our new Monte Carlo dose accumulation scheme. The effect of breathing induced internal motion is clearly visible on the resulting dose difference (right), especially at the upper border of the superimposed shape of the gross target volume (GTV).

4D Monte Carlo dose accumulation inside the liver shows larger negative dose differences for metastases with local recurrence (i.e. case 1 and 3.3) compared to our old approach. Simulated 3D and 4D Monte Carlo doses and corresponding dose differences for case 1 are shown in Figure 1. Computed underdosages mainly occurred at the upper border of the target volume. For, e.g., cases 3.1 and 4, the Monte Carlo based simulation predicts higher positive differences, as also illustrated for case 4 in Figure 2. High dose differences of ±8 Gy are visible, as also noticeable in Figure 1, but because of an apparently sufficiently dimensioned ITV located outside the target volume.

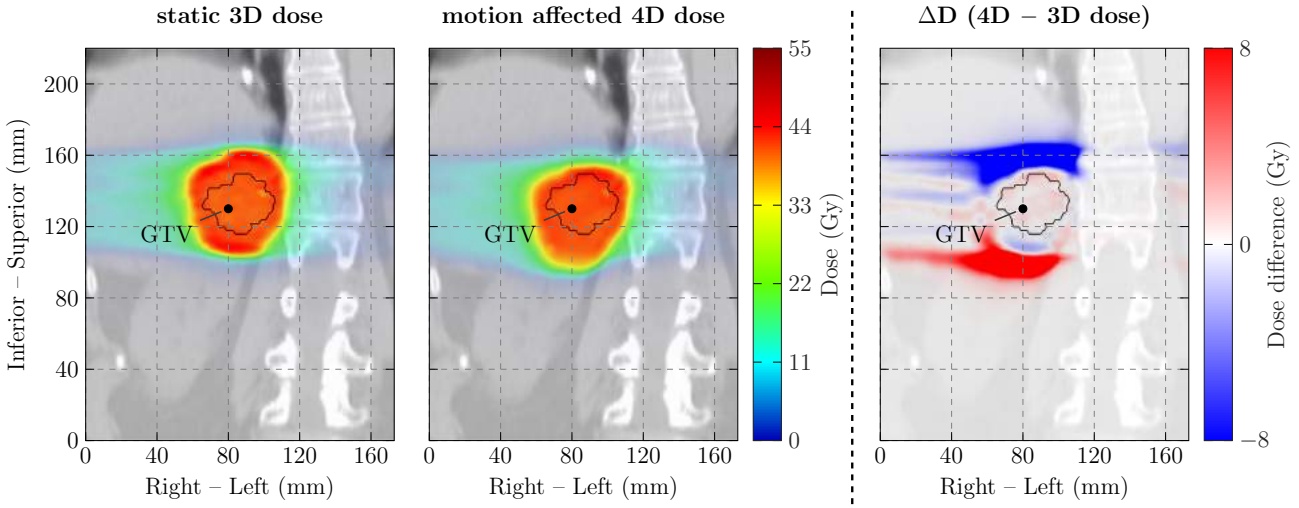


Figure 2. Comparison of static 3D (left) and motion affected 4D (middle) dose distribution for patient data set 4 (liver), simulated using our new Monte Carlo dose accumulation scheme. Again, breathing induced internal motion has a clearly visible effect on the resulting dose difference (right). However, an apparently sufficiently dimensioned ITV prevents occurrence of underdosages inside the GTV.

The comparison between new and old 4D dose accumulation results show lesser distinctions for the investigated lung metastases. However, underdosages for metastases with local recurrence are still more pronounced when using the Monte Carlo based approach (cf. case 7.2 and 9). For illustration, simulated dose distributions and difference for case 9 are shown in Figure 3. The effect of breathing induced internal motion is visible but not as distinctive as for the liver cases.

Similar to the liver cases, some metastases without local recurrence exhibit high positive dose differences (cf. case 6 and 8) as illustrated in Figure 4 for case 8. Again, breathing induced internal motion is only slightly visible in direct comparison of 3D and 4D dose distributions. The dose difference, however, shows overdosages inside and around the target volume of up to 5 Gy.

4. CONCLUSIONS

The submitted work presents, to our knowledge, the first approach that incorporates internal patient motion information while radiotherapy treatment into a 4D Monte Carlo dose simulation scheme. Thus, density changes of internal patient geometry structures as well as displacements due to breathing motion are considered in the course of the 4D dose accumulation. Further, high temporal resolution is achieved by splitting the patient treatment plan into its smallest possible segments. These advantages over our previously utilized dose accumulation method offer optimal features for a retrospective quality assurance tool in 4D radiotherapy of moving tumors.

ACKNOWLEDGMENTS

This work was in parts funded by the University of Hamburg (scholarship to TS) and DFG research grant WE 6197/2 (to RW).

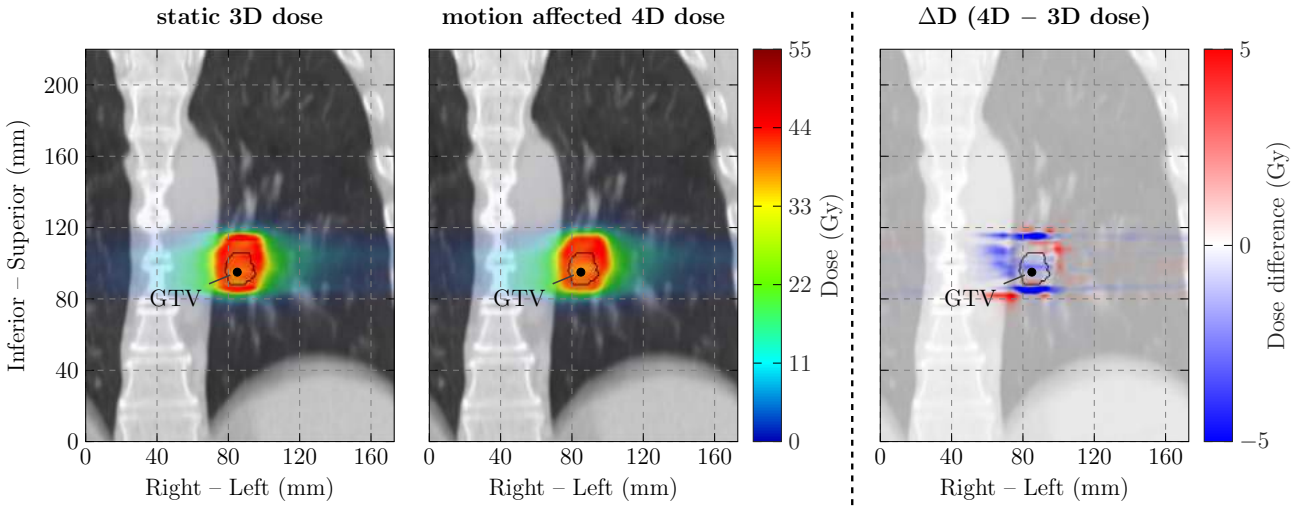


Figure 3. Comparison of static 3D (left) and motion affected 4D (middle) dose distribution for patient data set 9 (lung), simulated using our new Monte Carlo dose accumulation scheme. The effect of breathing induced internal motion is slightly visible on resulting dose difference (right), but inside the GTV underdosages are small.

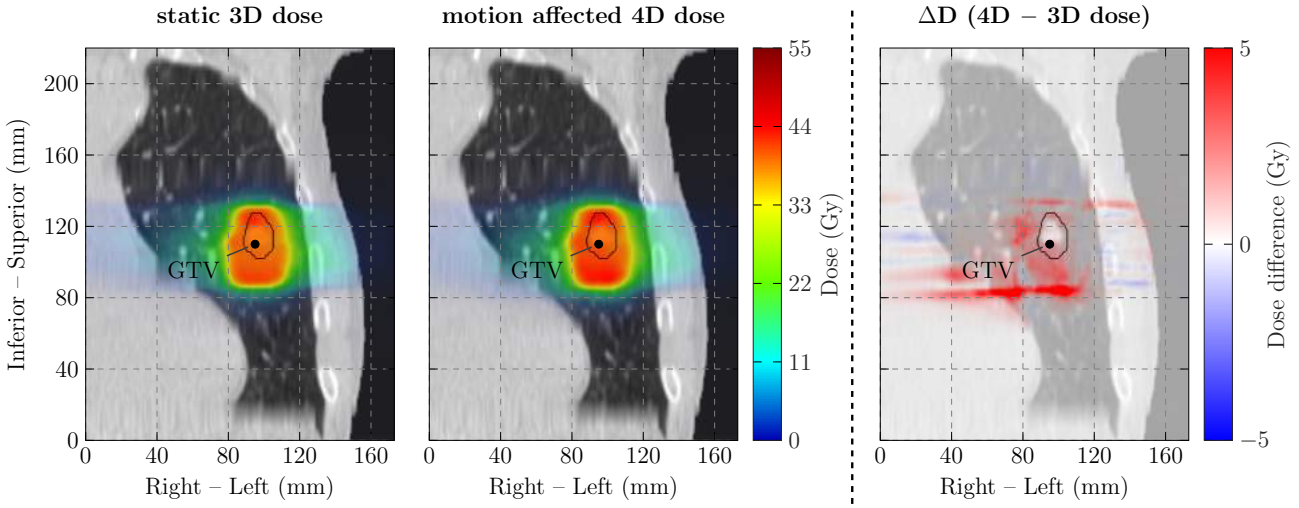


Figure 4. Comparison of static 3D (left) and motion affected 4D (middle) dose distribution for patient data set 6 (lung), simulated using our new Monte Carlo dose accumulation scheme. The effect of breathing induced internal motion is slightly visible on resulting dose difference (right).

REFERENCES

[1] Keall, P. J., Mageras, G. S., Balter, J. M., Emery, R. S., Forster, K. M., Jiang, S. B., Kapatoes, J. M., Low, D. A., Murphy, M. J., Murray, B. R., Ramsey, C. R., Van Herk, M. B., Vedam, S. S., Wong, J. W., and Yorke, E., "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med Phys* **33**, 3874–3900 (2006).

[2] Gauer, T., Sothmann, T., Blanck, O., Petersen, C., and Werner, R., "Under-reported dosimetry errors due to interplay effects during VMAT dose delivery in extreme hypofractionated stereotactic radiotherapy.," *Strahlenther Onkol* **194**, 570–579 (June 2018).

[3] Sothmann, T., Gauer, T., and Werner, R., "4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters," *PLoS One* **12**, e0172810 (2017).

[4] Sothmann, T., Gauer, T., Wilms, M., and Werner, R., "Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT," *Phys Med Biol* **62**, 9001–9017 (2017).

[5] Sothmann, T., Gauer, T., and Werner, R., "Influence of 4D CT motion artifacts on correspondence model-based 4D dose accumulation," in [*Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*], **10576**, 105760F, International Society for Optics and Photonics (2018).

[6] Mogadas, N., Sothmann, T., Knopp, T., Gauer, T., Petersen, C., and Werner, R., "Influence of deformable image registration on 4D dose simulation for extracranial SBRT: A multi-registration framework study," *Radiother Oncol* **127**(2), 225–232 (2018).

[7] Kawrakow, I., "The egsnrc code system, monte carlo simulation of electron and photon transport," *NRCC Report Pirs-701* (2001).

5.7. Error propagation in 4D dose simulation

The results presented in this section are based on a first implementation of uncertainty propagation into the previously described 4D Monte Carlo dose accumulation scheme. Uncertainties regarding the registration process, as estimated by the registration framework proposed in Section 5.4, are considered. The application of the probabilistic, i. e. dropout-based, registration approach allows comparing multiple motion vector fields for one reference/template image pair. Thus, registration between all phase images of the patient-specific 4D data sets can be repeated. For each repetition, i. e. new registration result, an individual correspondence model can be built. The application of each model for motion prediction during 4D dose accumulation propagates the registration uncertainties through motion estimation by correspondence modeling into resulting 4D-simulated dose distributions. That is, (voxel-wise) confidence intervals to estimate the reliability of calculated dose distributions can be computed. Preliminary results are obtained by building and applying ten independent motion models with above described proceeding for each in-house patient data set (cf. Table 4.2). 4D Monte Carlo dose simulation is subsequently performed for each correspondence model and the first treatment fraction of all patients (ten 4D dose simulations per patient/fraction, i. e. in total 100 simulations). Obtained results are again evaluated by means of ΔD_{98} and compared to the in previous section presented dose simulation results (i. e. results of 4D Monte Carlo dose simulation using the by traditional DIR algorithm estimated motion vector fields and thus solely one unique motion model per patient data set). Corresponding results are summarized in Table 5.2.

For the liver cases, ΔD_{98} values of the initial simulation and the mean over the individual motion model-based 4D-simulated doses $\langle \Delta D_{98} \rangle$ values of the uncertainty propagation simulation are similar. In particular for the two metastasis with local recurrence, i. e. case 1 and 3.3, the confidence intervals, i. e. the uncertainty, is in comparison to the other cases with an absolute interval size of 0.18 Gy and 0.36 Gy, respectively, relatively high. Similar but not as pronounced are the findings for the uncertainty estimations for the lung cases. Case 7.2 and 9, i. e. cases with local metastasis recurrence, show highest dose uncertainty interval sizes of 0.16 Gy and 0.08 Gy, respectively. Values of ΔD_{98} for case 7.2 noticeable differ between initial and uncertainty propagation scheme. The new simulation scheme estimates higher underdosages. For illustration purposes, the achieved results for metastasis 1 and 7.2 are visualized in Fig. 5.2, where dose volume histograms for the Monte Carlo initial 4D dose and the Monte Carlo uncertainty propagation mean 4D-simulated dose with corresponding 95%-confidence interval ($CI_{95\%}$)-values are compared to the static, i. e. reference, dose. The initially 4D-simulated dose volume histogram for

Table 5.2.: Comparison of ΔD_{98} numbers obtained by 4D Monte Carlo simulation using the initial (i. e. traditional DIR for motion modeling, no uncertainty propagation) and the uncertainty propagation scheme (i. e. probabilistic DIR for motion modeling, ten models per patient data set). Simulation is performed for the first fraction of the in-house patient data set. For the uncertainty propagation scheme, mean of ΔD_{98} and corresponding 95%-confidence interval ($CI_{95\%}$) are given. * is denoting local metastasis recurrence.

	Met.	MC initial	MC uncertainty prop.	
		ΔD_{98} (Gy)	$\langle \Delta D_{98} \rangle$ (Gy)	$CI_{95\%}$ (Gy)
Liver	1*	−0.38	−0.34	[−0.43, −0.25]
	2.1	+0.03	−0.12	[−0.20, −0.04]
	2.2	+0.09	−0.06	[−0.08, −0.04]
	2.3	−0.11	−0.14	[−0.15, −0.13]
	3.1	+0.16	+0.12	[+0.11, +0.13]
	3.2	+0.12	+0.16	[+0.15, +0.17]
	3.3*	−3.56	−2.97	[−3.15, −2.79]
	4	+0.02	−0.05	[−0.07, −0.03]
	5	−0.29	−0.18	[−0.20, −0.16]
Lung	6	+0.02	−0.23	[−0.25, −0.21]
	7.1	+0.12	−0.13	[−0.16, −0.10]
	7.2*	−0.07	−0.43	[−0.51, −0.35]
	8	+0.29	+0.38	[+0.37, +0.39]
	9*	−0.51	−0.61	[−0.65, −0.57]
	10	+0.11	+0.11	[+0.10, +0.12]

metastasis 1 lies inside the 95%-confidence interval of the estimated dose simulated by the Monte Carlo uncertainty propagation scheme. Small differences to the mean 4D dose volume histogram are, however, present. The consideration of patient motion during 4D dose simulation leads to a dose blurring and thus deviations to the static dose volume histogram occur. Regarding dose volume histograms for metastasis 7.2, the initially Monte Carlo simulated dose shows only minor deviations to the static dose. The new Monte Carlo scheme, however, estimates higher underdosages. A potential reason for this is the application of the developed and applied deep learning-based DIR framework that obviously predicted large motion vector fields inside the lung. For metastasis 1 the total 4D-simulated dose, i. e. the dose summed up over all treatment fractions, is additionally computed for two selected and individual motion models. A comparison between resulting dose distributions is illustrated in Fig. 5.3. The impact of the selected motion model onto the simulation results is clearly visible, in particular in the dose difference distribution where ΔD-values of ±3 Gy occur.

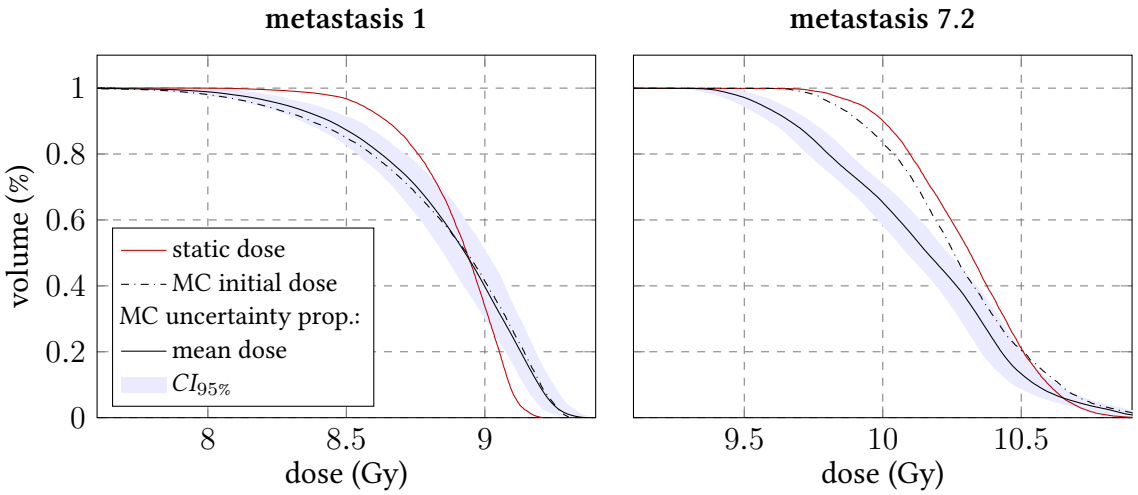


Figure 5.2.: Exemplary visualization of ΔD_{98} numbers obtained by 4D Monte Carlo simulation for metastasis 1 and 7.2. Dose volume histograms of the CTV for the Monte Carlo initially simulated doses and the Monte Carlo uncertainty propagation mean 4D-simulated doses with corresponding $CI_{95\%}$-numbers are shown and compared to the static, i. e. reference, doses. Left: Dose volume histograms of initial and uncertainty Monte Carlo dose simulation scheme for metastasis 1 show good agreement. Motion-induced deviations compared to the static dose volume histogram are visible. Right: The initial Monte Carlo scheme estimates only small underdosages for metastasis 7.2, thus the dose volume histogram is similar to the static dose volume histogram. The uncertainty propagation simulation scheme estimated underdosages are more pronounced.

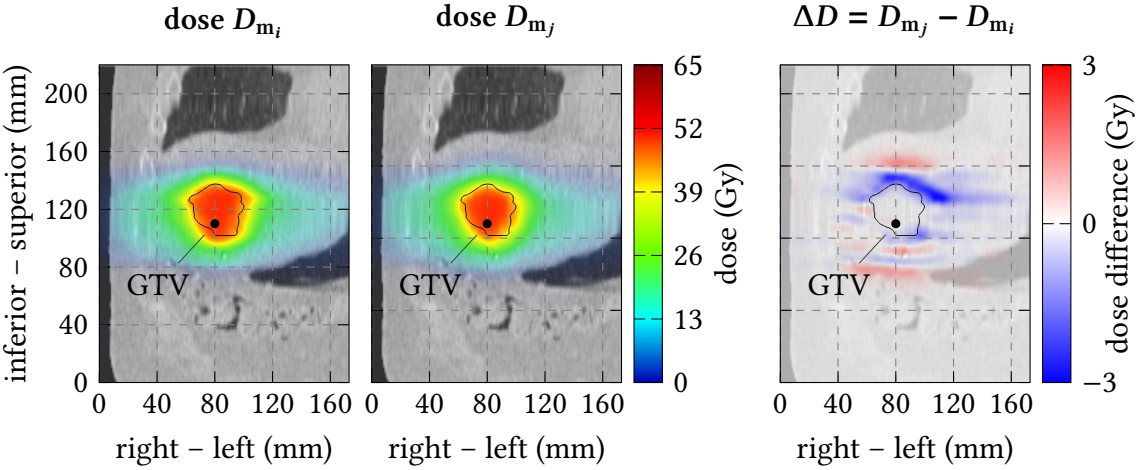


Figure 5.3.: Comparison of total 4D-simulated dose distributions for metastasis 1. Different motion models (model m_i and m_j) were applied for dose simulation of doses D_{m_i} (left) and D_{m_j} (middle). The dose difference distribution ($\Delta D = D_{m_j} - D_{m_i}$, right) illustrates the impact of the motion model onto the estimated dose distribution.

Chapter 6

Discussion

In this chapter, results presented in this thesis and the proposed methodological approaches to retrospective quality assurance in 4D radiotherapy are discussed and set into context to corresponding literature. First, a detailed interpretation of the correspondence model-based 4D dose simulation algorithm and related uncertainties, i. e. the influence of 4D CT image artifacts and choice of registration algorithm onto proposed dose accumulation scheme, is given. The potential of uncertainty estimation during the registration process is further discussed, followed by an analysis of parameters that impact the accuracy of the proposed dose accumulation scheme. An evaluation of in this process encountered limitations, their consideration in a Monte Carlo-based reimplementation of the proposed dose accumulation scheme and eventually the possibility of an uncertainty propagation to yield dose confidence intervals, follows. Furthermore, remaining questions and challenges as well as future research direction in the field of quality assurance in 4D radiotherapy are identified.

6.1. Interpretation of results

Correspondence model-based 4D dose simulation

One of the main goals of this thesis was to develop a 4D dose simulation approach that takes internal patient motion during treatment into account. Hence, a retrospective quality assurance of the dose delivery during patient treatment could be performed. The general approach to achieve this goal was to integrate an external, and thus commonly clinically acquired, breathing signal information during dose delivery into a retrospective 4D dose simulation scheme for VMAT by correspondence modeling, i. e. the correlation of internal and external motion signal. Similar to 4D dose accumulation approaches proposed by Velec et al. [33], Werner et al. [31], Ehrbar et al. [32], Samavati et al. [34] or Freislederer et al. [93][1], patient 4D CT image data is the basis of the simulation approach. In contrast to these approaches, however, the patient-specific breathing variability during

[1] Freislederer et al. [93] actually state that in their workflow the patient breathing trajectory acquired during treatment is incorporated for 4D Monte Carlo dose calculations. However, as an acquired external breathing signal measurement is solely correlated to the corresponding 4D CT phase image, an explicit consideration of patient motion variability during treatment is not given.

treatment is explicitly taken into account and thus a more realistic dose estimate is yielded. Correlating the results, achieved by applying the developed dose accumulation algorithm to actual patient data sets, to corresponding clinical endpoints, i. e. whether the patient treatment was successful or not by means of local metastasis recurrence, indicated a possible linkage between estimated underdosages and metastasis recurrence. Following these results, the interplay effect, i. e. the unfavourable interaction between patient motion, its variability and highly dynamic dose delivery technique, seems to impact the treatment at least for some patient cases. This is further supported by phantom-based dose measurements and corresponding results as presented in Section 5.5. Yet, the influence of the interplay effect is still controversial discussed in the literature, where e. g. Ong et al. [94] state that interplay effects are unlikely to be clinically significant (at least for more than two treatment fractions). Gauer et al. [19], on the contrary, find proof of considerable influence of interplay effects. Nevertheless, the results obtained by applying the proposed correspondence model-based 4D dose simulation potentially allow to explain patient motion-related dose uncertainties. The information about the actually delivered dose can further be used to modify the patient treatment plan between fractions, for instance by safety margin reduction or expansion as, to some extent, already performed in MR guided radiotherapy treatments [95, 96]. Therefore, aiming for a clinical application of the proposed scheme directly motivated to investigate the dose simulation accuracy and, more importantly, remaining uncertainties in the utilized pipeline and their general impact on the estimated dose distributions.

Uncertainties of 4D dose simulation

The basis of the developed scheme is the individual patient 4D CT data set, which is utilized to extract internal patient motion information to allow for correspondence model training and subsequent prediction. The motion extraction by the mostly image intensity-based DIR algorithms is prone to uncertainties, especially in image areas like the abdomen with low-to-no image contrast information available [97]. Further, the 4D CT is based on retrospectively sorted projection data, i. e. solely one mean respiratory cycle is represented and inter-cycle motion variability neglected. More specifically, there exists no 1-to-1 matching between the reconstructed 3D images and breathing signal values, and thus artifacts as introduced in Section 3.1.2 can occur due to patient motion variability. In 2013, Yamamoto et al. [98] concluded that such artifacts impact not only the image quality but consequently also the accuracy of DIR-estimated motion vector fields. That is, using motion vector fields estimated by DIR of artifact-affected 4D CT data sets is assumed to highly influence the accuracy of correspondence modeling and in consequence 4D dose accumulation.

A first step to examine the general influence of registration algorithms onto resulting motion vector fields was to utilize different open source available DIR frameworks for the task of motion extraction out of 4D CT data sets. The impact of estimated motion vector fields on subsequent steps of the proposed dose simulation pipeline were additionally analyzed. Interestingly, results of this evaluation showed that for the lung cases indeed a correlation between DIR and correspondence modeling accuracy exists, which is in line with the hypothesis of Liu et al. [99] that motion model accuracy is comparable to DIR accuracy. The impact on simulated dose distributions was however rather low. Consequently, predicted dose deviations for radiotherapy treatments in the lung seemed to be partly trustworthy. In contrast, predicted dose deviations for the liver tumor patients strongly differed. Even the MRI test case scenario (cf. Section 4.4.3) did not lead to a correlation between DIR characteristics and registration accuracy. However, some of the investigated frameworks showed, despite the low contrast in the abdomen, good registration performance, which consequently leads to an accurate motion model, and to some extent, to a more reliable dose estimation. The performed extensive analysis of DIR algorithms, with a strong focus on their accuracy and thus on estimated motion vector fields, further showed that especially image artifacts in 4D CT data strongly impact the registration result and correspondence model formation, as previously assumed. Therefore, a systematic analysis of typical 4D CT artifacts and their impact on the proposed dose simulation seemed necessary, especially since Yamamoto et al. [100] already stated in 2008 that 90% of abdominal/thoracic 4D CT data sets are artifact-affected. And this has not significantly changed, as proven in recent publications [63, 101].

In this thesis, the novel and most optimal idea was to choose (mostly) artifact-free 4D CT patient data sets and retrospectively manipulate the corresponding CT reconstruction breathing curve to induce typical artifacts after reconstruction to allow for a comparable analysis of artifact-free and artifact-containing image data and their impact on the dose simulation pipeline. Corresponding results showed that especially interpolation artifacts, i. e. missing projection data for some CT image slices, strongly impact the registration process. Additionally, subsequent steps like correspondence modeling, motion prediction and dose simulation are affected. This strongly motivates efforts to reduce or even prevent respective artifact type. Especially technological solutions are desired and investigated by related research community. Promising approaches are utilizing training periods to learn features of the breathing signal before the actual 4D CT acquisition and subsequently use this information to adapt to the patient breathing during CT scanning. Martin et al. [101], for instance, prospectively gate the CT scans based on the detection of irregular patient breathing. An even more advanced approach is the recently proposed intelligent 4D CT scanning technique by Werner et al. [65]. Here, a sequence scanning mode is applied,

guided by the online acquired respiratory signal. Additionally, if projection data is missing, a rapid breathing record analysis will allow to trigger potential local re-scannings. Unfortunately, these methods were clinically not available when the patient image data sets employed in this thesis were acquired. The other typical 4D CT artifact type, i. e. double structure artifacts, is in general more challenging regarding its complete prevention, but in novel CT reconstruction techniques the usage of advanced protocols can to some extent reduce their occurrence [66]. The impact of this artifact type on proposed pipeline, however, tends to be much smaller, as illustrated in Fig. 3 [SPIE 2018]. Nevertheless, obtained results showed that a consideration of artifacts is not only mandatory for treatment planning but also for the proposed 4D dose simulation scheme. Furthermore, in the specific task of 4D CT motion extraction using DIR, image quality and imaged anatomy (e. g. thoracic/abdominal with high/low image contrast, respectively) highly influence the registration result. Following this argumentation, and the fact that DIR is the basis of the proposed dose simulation scheme, directly motivated to develop a registration approach that allows estimating registration uncertainty maps. Inspired by the approach of Yang et al. [75], a probabilistic registration framework based on a CNN with integrated dropout layers was developed for registration of 4D CT image data. Training the network with an in-house acquired 4D CT data base, with corresponding pseudo ground truth vector fields estimated by standard DIR algorithms, allowed to analyze the DIR accuracy by means of additional external data bases. That is, no bias towards the external image data regarding the scanner type, reconstruction method or image dimensionality was introduced during training. Strikingly, the registration accuracy of proposed model variants (depending on the DIR algorithm used for ground truth generation) were similar or even higher compared to standard DIR accuracy. Further, a 60-fold run time reduction was achieved. However, extending the deterministic network to be probabilistic using dropouts, and in consequence allowing for an uncertainty estimation, was only partly successful. More specifically, computation of registration uncertainty maps was possible, yet, the applicability of such information for e. g. radiotherapy treatment planning and estimation of uncertainties therein, as proposed by Amir-Khalili et al. [102], remains unclear. Nevertheless, a consideration of predicted uncertainties in the proposed dose simulation pipeline was investigated. Beforehand, however, the general accuracy of the dose simulation scheme was metrologically determined to identify influencing parameters and potential improvements.

Accuracy assessment and simulation improvements

Similar to the approach of Sothmann et al. [16] in the context of a comparison of real-time tracking systems, dynamically dose measurements were acquired using a measurement

setup consisting of a detector array mounted on a programmable motion platform that allows to simulate arbitrary motion patterns with high accuracy (accuracy of the simulated motion paths are specified to be within $\approx 0.5\,\mathrm{mm}$) [103]. Comparing simulated motion-affected and dynamically measured dose distributions made the assessment of parameters influencing the 4D dose simulation possible. Corresponding results demonstrated a major impact of the arc disrectization on simulation accuracy, which was at least to some extent anticipated. As the highest possible level of disrectization also yields highest similarity between measurement and simulation, one could expect an even higher similarity when the angular resolution is further increased. Yet, this was not possible as the arc disrectization is TPS-dependent and limited to 2.3°. Overall, however, the interplay effect-induced dose differences between static and motion-affected dose distributions were reproducible with high similarity (at least by means of utilized metric, i. e. the Gamma index, with a mean over all motion scenarios $\langle \Delta\gamma \rangle$ of 3.7 ± 2.6 when comparing simulation to initial measurement[2]). That is, a purely simulation-based evaluation of influencing parameters was due to a reliable simulation approach feasible. Here, the impact of the intrafractional patient motion variability on estimated dose distributions was of particular interest. Repeated simulations for equidistantly shifted motion starting phases were therefore performed. And indeed, a substantial impact on obtained motion-affected dose distributions was observed, suggesting a high correlation between individual patient breathing and amount of occurring interplay effects. However, especially the limited arc discretization, not taking into account the TPS-based "black-box" dose computation and density changes in the patient geometry during dose simulation (i. e. the dose is computed on the AvCT of the 4D CT data set), motivated a re-implementation of proposed dose simulation scheme.

Monte Carlo simulations for dose calculations in radiotherapy are considered to provide reference dose distributions, i. e. achieve highest dose calculation accuracy as stated by the American Association of Physicists in Medicine task group report 105 [104]. It was, therefore, the obvious choice with regard to simulation accuracy to utilize a Monte Carlo dose calculation approach for the proposed dose simulation scheme. In line with the initially implemented dose simulation approach, correspondence modeling was applied to predict internal patient motion during treatment. The Monte Carlo simulation was however not performed on the AvCT but on the inversely warped reference phase of the 4D CT data set, as described in Section 4.2.1. This directly made the consideration of breathing-induced density changes in the patient geometry feasible. Further, smallest segments (here: control points; beforehand: arc segments of size 2.3° consisting of 2 to 5 control points) defining the dynamical dose delivery process were employed to achieve

[2]Comparing repeated measurements to initial measurements resulted in a $\langle \Delta\gamma \rangle$ of 4.1 ± 2.4, i. e. the simulation lies in the range of the measurement uncertainties.

highest possible temporal resolution. Application of the new Monte Carlo based simulation scheme to the ten already investigated patient cases showed, compared to the initial approach, larger negative dose differences for metastases with local recurrence. This results did not imply a gain in simulation accuracy as ground truth dose information was missing; yet, the feasibility of the new approach was illustrated. However, the assumption that simulation accuracy improves when the angular resolution is further increased was examined by means of beforehand already utilized dose measurements. And indeed, the Monte Carlo simulated dose distributions showed higher similarity to measurements compared to the initially simulated distributions with finest arc discretization. Therefore, it was assumed that remaining uncertainties of the dose simulation approach are primarily caused by DIR-based 4D CT motion extraction, subsequent correspondence model formation and related motion prediction. Dose calculation uncertainties are estimated to be small as Monte Carlo simulations are generally considered to be extremely accurate regarding the simulation of photon beams generated by electron accelerators [105].

A first step to consider the DIR uncertainty was achieved by utilizing the aforementioned CNN-based DIR approach for 4D CT registration. The concept of repeatedly predicting vector fields with enabled dropout was previously exploited to compute the sought motion field as the mean of the sampled predicted fields and the registration uncertainty as the voxel-wise variances. Now, instead of computing the mean prediction, each individual prediction was used to build individual correspondene models and simulate corresponding dose distributions by the 4D Monte Carlo approach. In doing so, a number of dose distributions was estimated, allowing for an uncertainty propagation by evaluating voxel-wise dose variances. The results were interpreted as confidence intervals, i. e. how certain is the 4D dose simulation about dose delivered to a voxel, solely affected by the CNN-based DIR uncertainty and its propagation through the simulation pipeline. In general, for all patient cases the mean over the simulated dose distributions is similar to the initial dose accumulation approach. Results for the lung cases illustrated that corresponding uncertainties are small, mainly because the performance of intensity-based DIR inside the lung is due to the available contrast information good. Dose uncertainties for the liver cases are higher, at least for the two cases with local metastasis recurrence. However, the CNN-based registration lead to relatively small motion vector fields[3]. Consequently the uncertainty i. e. the range of estimated dose distributions, is potentially underestimated. That is, the simulated dose distributions are only sparsely impacted.

[3]The cause for that remains unclear. A reason, however, could be the not explicitly performed regularization during CNN-based registration.

6.2. Challenges, limitations and outlook

In the last section, the main results of this thesis were briefly summarized, discussed and, if possible, contextualized. Remaining challenges and limitations and how to potentially tackle them in future research were, however, not directly referred to, which is therefore conducted in the following.

4D CT imaging

Not only for the discussed dose simulation but also for actual 4D radiotherapy patient treatments, the quality of pre-treatment acquired 4D CT image data is one of the most impacting parameters. Especially the previously discussed and investigated 4D CT artifact types greatly reduce the reliability of contained image information. In consequence, all following processing steps, for instance the DIR-based motion extraction or the more clinically relevant delineation of radiotherapy important structures (GTV and/or OAR), are affected. As briefly mentioned before, this is the reason why a strong effort exists in the related research community to develop and implement approaches that counter the influence of unfavourable patient motion during CT acquisition. Performance-wise, the recently proposed intelligent 4D CT protocol seems to be the most promising approach [65]. Combination of initial learning of patient motion information before imaging and online evaluation of the motion trajectory during CT acquisition allows for sequence scanning, where imaging is solely conducted in corresponding z-position when an acceptable breathing cycle is detected. That is, missing projection data due to too long patient breathing cycles with regard to the employed z-table feed (cf. Section 3.1.2) can no longer occur. As missing projection data is the reason for interpolation artifacts, this type of artifact is completely eliminated. Furthermore, exclusion criteria for an acceptable breathing cycle is its amplitude irregularity. Thus, the occurrence of double structure artifacts is reduced. However, some motion artifacts still remain, especially since cardiac or bowel motion is not considered during CT imaging. Further, presented results are only simulation-based, i. e. phantom and real patient measurements have to be performed and analyzed before a clinical application is possible. Until then, a more advanced reconstruction [66], compared to the standard phase- and amplitude-based reconstruction, is a first step to improve 4D CT image quality.

Deformable image registration

The accuracy of intensity-based deformable registration in medical image data still remains one of the biggest challenge in 4D dose simulation, especially for image areas with

low-to-no contrast information available. Registration approaches based on biophysical and finite element modeling (FEM), as for instance analyzed by Velec et al. [106], appear promising. An evaluation of DIR accuracy in the liver performed in the MIDRAS study [107] showed that such approaches can to some extent outperform purely intensity-based DIR in the liver. However, FEM-based DIR frameworks are usually not available as open source and similar to standard DIR equally dependent on a sufficient image quality. The CNN-based DIR algorithm developed in this thesis can therefore be specifically utilized to not only estimate the registration result but also a corresponding uncertainty distribution. For the lung registration, feasibility was illustrated. However, some limitations of the proposed framework prevented even higher registration accuracy, especially in low contrast image areas:

Supervised learning: For supervised training of a network, ground truth information has to be available. For the proposed network, which is trained to solve a registration task, traditional DIR algorithms were applied for pseudo ground truth generation that is subsequently used in the training process. More specifically, solely an estimate, i. e. pseudo ground truth information, and no real ground truth information, was available for training.

Slab-based approach: Due to the large image size of 4D CT data sets, a slab-based model was implemented, where main motion directions in the patient anatomy were considered. However, the model never "sees" the whole image while training; hence, important context information is to some extent not available during training and prediction. Further, to allow for a deeper network structure, the image input size was reduced by a pre-trained autoencoder so that the network was trainable with the available GPU memory. That is, in-plane image information was to some extent lost.

MSE loss between vector fields: Traditional DIR algorithms are typically optimized by a similarity measure between reference and template image that is maximized during the registration process. For the proposed CNN-based approach, similarity between (pseudo) ground truth and predicted vector field was evaluated by means of a MSE loss function. Thus, the model does not learn the direct relationship between template and reference image.

Cascaded network: To achieve higher registration accuracy, the network was applied iteratively. This was possible as registration duration was greatly reduced. However, the stopping condition, i. e. which iteration yields highest registration accuracy, was chosen arbitrarily.

Uncertainty estimation: Available image contrast information in thoracic 4D CT data leads to reliable image registration results. Thus, the model estimates solely minor uncertainties while registration. For abdominal image data, however, the registration result shows potentially unrealistically small motion predictions in low contrast areas, mainly because the regularization functional as utilized in traditional DIR is missing. Consequently, the voxel-wise variance, i. e. registration uncertainty, is estimated as being low. Thus, the obtained uncertainty estimation remains questionable.

For future work, especially the approach of reinforcement learning in combination with deep learning seems promising to highly improve the CNN-based image registration accuracy of thoracic and abdominal CT image data sets. Similar to traditional DIR, this allows for a loss function that is directly computed between reference and warped template image. That is, the model will "see" whole images, maximize similarity between both and be independent of a (pseudo) ground truth motion vector field. Further, a specific smoothness of resulting motion vector field, i. e. some sort of regularization, as well as stopping criteria while registration can be defined.

Correspondence modeling

In its current form, the utilized correspondence model is solely based on computed vector fields between one reference and all remaining phases of one pre-treatment acquired patient 4D CT image data set and corresponding external breathing motion information. Thus, no inter-cycle variability as defined by McClelland et al. [108] is considered during model training[4]. Further, variations of external surrogate signal and internal motion before first treatment fraction as well as between treatment fractions, which are apparent according to McClelland et al. [109], are not accounted for. Minimizing these uncertainties is potentially possible by using for instance pre-treatment acquired CBCT data. In clinical practice, 3D CBCT imaging is performed to verify, and if necessary adapt, the patient positioning before dose delivery. However, the 3D CBCT raw-data in combination with a simultaneously recorded external breathing signal can be utilized to reconstruct a (sparse-view) time-resolved CBCT, i. e. a 4D CBCT, as recently exploited by Madesta et al. [110]. The reconstructed 4D CBCT image data is, due to the only sparsely available projection data, extremely affected by sparse-view streaking artifacts. Yet, the implementation of sophisticated solutions, e. g. cyclic registration approaches as applied by Brehm et

[4]The inter-cycle variability is defined as the motion between breathing cycles. That is, the motion during one breathing cycle is different to that during another breathing cycle. Inter-cycle variability is in the applied approach of correspondence modeling not considered as the 4D CT only represents one average patient breathing cycle.

al. [111] or even deep learning based boosting as shown by Madesta et al. [110], achieve high image quality improvements. Therefore, an update or simply a verification of 4D CT-based correspondence model could be possible by application of corresponding patient 4D CBCT data sets.

4D dose simulation

The actual 4D dose simulation concept itself, i. e. not considering uncertainties/limitations introduced by previous motion estimation techniques, is in general an elaborated framework with in principle no limitations or shortcomings. Especially after re-implementation of the proposed scheme as a Monte Carlo-based dose accumulation framework, previously identified limitations, like e. g. not considering density changes in patient geometry or the dependency on pre-treatment computed dose distributions, are for the most parts eliminated or at least their impact strongly reduced. However, some challenges and limitations remain, as described in the following:

Measurement-based dose simulation verification: The simulation approach was verified by artificial motion patterns and dose measurements conducted by a 2D detector array mounted on a programmable motion platform. This was possible as the dose measurements were assumed to contain ground truth information. However, as already shown in corresponding publication (cf. Section 5.5), dose measurements are for numerous reasons (e. g. statistical fluctuations, accelerator performance, detector resolution, angular dependence of detector, motion platform accuracy) not completely reproducible, i. e. to some extent day 1 and day 2 measurements differ. However, the simulation accuracy can solely reliably assessed to the point where it is in the range of the measurement uncertainties.

Patient cohort-based dose simulation verification: The verification of retrospective 4D-simulated dose distributions for real patient treatments is not as trivial as the above described phantom and dose measurement-based approach, as ground truth information is missing. In this thesis, additional information about the treatment outcome (local metastasis recurrence yes/no) was available and used to allow for a first dose simulation feasibility estimation. That is, a linkage between predicted underdosages and metastasis recurrence was found. However, this correlation is rather a plausibility consideration for individual cases than a real validation for the simulated dose distribution. An idea to improve this proceeding is to use refining information about the location of metastasis recurrence. For instance, Van den Begin et al. [112] used CT and PET follow-up images to delineate the contour of the recurrent tumor volume to examine the correlation between initially

planned dose distribution and local recurrence. A similar approach could be used to correlate the location of the metastasis recurrence on a voxel level, registered onto the planning CT image, and predicted underdosages by 4D dose simulation. However, such follow-up image data with contoured recurrence volumes is at the moment not routinely available. Bypassing this issue is potentially possible by utilizing the routinely, after radiotherapy treatment, acquired MRI data. As shown in a recent study, DIR between planning CT image and post-treatment acquired MRI scan allows to assess the dose delivery accuracy by means of conformity measures regarding MRI dose imprint and treatment-relevant structures [113]. The shorter the time span between treatment end and MRI acquisition the deeper is corresponding dose imprint. Applying this approach in the context of 4D dose simulation, i. e. correlating the dose imprint to an iso-dose line estimated by the simulation approach, could be used as verification tool. That is, the patient cohort could be extensively increased and at the same time the estimated iso-doses verified by the measurable dose imprints.

Patient cohort The number of patient eligible for 4D dose simulation and subsequent verification, i. e. treated by VMAT and necessary data for dose simulation available, was (ten cases) relatively small and only sufficient to allow for a proof of principle investigation. Further verification of achieved results will be only possible if the patient cohort size is significantly increased. However, obtaining information about the clinical outcome is a long and time consuming process and thus the information is often not available in patient data sets.

Uncertainty propagation In its current form, the proposed registration uncertainty propagation through the dose simulation pipeline is a first proof of concept of its applicability. In a next step, remaining technical uncertainties, as for instance introduced by extra- and interpolation of the correspondence model-based motion estimation, has to be included. Further, and aiming at an application of the proposed scheme as a clinical quality assurance tool, is the consideration of human errors. An example of a highly influencing error introduced by humans is the contouring process of radiotherapy-relevant structures, performed by radiologists. Louie et al. [114] analyzed the intra- and inter-observer variability for lung cancer target volume delineation in 4D CT data and found that the variability in especially the target volume delineation is a major source of error in lung 4D CT treatment planning. The consideration of such errors in the proposed dose accumulation pipeline is, however, complex and needs beforehand an extensive evaluation of the intra- and inter-observer variability.

Interpretation and consideration of estimated underdosages Even if the 4D dose simulation approach would be extensively validated and found sufficiently accurate/reliable, the question how to employ predicted dose distributions remains. This means, the application of the algorithm with corresponding delivered dose information as a patient specific and retrospective quality assurance tool is desirable, but how to use, for instance, estimated underdosages to improve the patient treatment and thus the treatment success is unclear. Daily treatment plan adaption, as for example proposed by Palacios et al. [95], is possible, however, it is currently solely performed to adapt treatment plan margins to inter-fractional organ changes. Reacting to early identified underdosages in the target area by safety margin adaptation and, for instance, introduction of local dose boosts can in this form not yet be found in the literature. Nevertheless, the proposed dose accumulation framework provides valuable information for adaptive SBRT schemes of extracranial metastases.

Chapter 7

Conclusion

The aim of this thesis was to develop and analyze a 4D dose accumulation approach applicable for retrospective quality assurance in 4D radiotherapy. It was shown that correspondence modeling to consider individual patient motion variability during dose delivery in combination with VMAT dose accumulation is feasible. A potential linkage of estimated tumor underdosages to the treatment success was found. Based on the fundamental correspondence model-based 4D dose accumulation implementation, influencing parameters regarding the dose simulation accuracy were identified and further analyzed. In this process, it became clear that especially 4D CT image data, typical 4D CT artifacts and the DIR-based motion extraction for subsequent correspondence model training and prediction were most influencing parameters. Further, the comparison of dose simulation and corresponding phantom-based dose measurements illustrated potential limitations of the 4D dose accumulation algorithm. Aiming at a highest possible simulation accuracy to enable an application of the framework as a quality assurance tool in the clinical workflow of 4D radiotherapy, improvements and thus modifications of the proposed pipeline were investigated and established.

Based on the main findings of this thesis, it can be concluded that the proposed 4D dose accumulation scheme, modified to actually re-compute dose distributions by Monte Carlo simulations, is in principle able to be integrated in the clinical 4D radiotherapy workflow for quality assurance purposes. Remaining uncertainties in the pipeline can to some extent be propagated and thus considered in the final 4D-simulated dose, as exemplary shown for registration uncertainties. Application of the 4D dose accumulation scheme to actual patient data sets highlighted its feasibility. The prospect of further insight into respiratory motion and motion variability-related error sources in VMAT-based SBRT of extracranial metastases is provided. The hypofractionated VMAT treatment that is due to its high complexity and less probable averaging out effect mainly affected by motion variability-related error sources and thus by the interplay effect should further be investigated regarding its general accuracy and limitations.

However, before a clinical implementation and utilization of the 4D dose accumulation framework is possible, remaining uncertainties have to be analyzed and ruled out. This does not only improve the dose accumulation accuracy but also the general radiotherapy treatment process. The basis of both, treatment planning and dose accumulation, is a

patient specific 4D CT data set. That is, a reduction or complete elimination of typical 4D CT image artifacts by novel scanning techniques will highly increase image quality and consequently benefit all subsequent and dependent processing steps like e. g. delineation of OAR/GTV (treatment planning) and internal motion extraction using DIR (dose accumulation). In a recent study (2019) by Sentker et al. [115], the correlation of artifact severeness and treatment success was statistically analyzed. The findings showed a significant association between presence and severity of image artifacts in 4D CT data and local metastasis control (i. e. negative treatment outcome) after SBRT treatment. In consequence, using less artifact-affected 4D CT data for treatment planning has a high potential to improve the general treatment success rate.

Even if image artifacts in 4D CT data can be highly reduced, the for correspondence model building required internal motion extraction by DIR would still be prone to uncertainties in low-to-no contrast areas. Investigations performed in this thesis of current open source available DIR frameworks illustrated their shortcomings. Such frameworks cannot be considered ready for "plug-and-play", and the application of DIR by inexperienced users is questionable. In addition, the application of DIR for target delineation in the liver, commonly applied for CT and MRI data set registration to simplify the delineation process by additional MRI information inside the liver, has a direct impact on the radiotherapy treatment. The introduction of a deep learning-based registration framework with the potential to estimate registration uncertainties is a first and important step to consider such DIR-based shortcomings in subsequent processing steps. Improving the existing framework in future work is therefore highly encouraged. For instance, the application of deep reinforcement learning, i. e. having a neural network as the policy of an agent, is an approach that should be further prosecuted.

The dose measurement-based verification of the general simulation accuracy allows to conclude that a reliable basis for assessment of VMAT motion effects is now given. Nevertheless, more advanced and patient realistic measurement setups are desirable. That is, an acquisition of 3D dose distributions (in this thesis: measurement of 2D dose distributions) would be beneficial. Even more advanced non-rigid, i. e. deformable, phantom setups for measuring a motion-affected dose are unfortunately not yet available. For future work, however, the development of appropriate measurement-based methods to verify the general 4D dose accumulation scheme is important when aiming at a clinical implementation of proposed quality assurance framework.

In summary, the development, application and evaluation of a novel framework for post-treatment 4D dose simulation of highly complex treatments techniques that combines current developments in the fields of dose accumulation, correspondence modeling and DIR was successfully accomplished.

Bibliography

[1] G. R. Dagenais, D. P. Leong, S. Rangarajan, et al. Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. *The Lancet*, 2019.

[2] S. Yusuf, P. Joseph, S. Rangarajan, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *The Lancet*, 2019.

[3] L. A. Torre, F. Bray, R. L. Siegel, et al. Global cancer statistics, 2012. *CA Cancer J Clin*, 65(2):87–108, 2015.

[4] F. Bray, J. Ferlay, I. Soerjomataram, et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68(6):394–424, 2018.

[5] R. L. Siegel, K. D. Miller and A. Jemal. Cancer statistics, 2016. *CA Cancer J Clin*, 66(1):7–30, 2016.

[6] K. D. Miller, R. L. Siegel, C. C. Lin, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin*, 66(4):271–289, 2016.

[7] T.-S. Su, P. Liang, J. Liang, et al. Long-Term survival analysis of stereotactic ablative radiotherapy versus liver resection for small hepatocellular carcinoma. *Int J Radiat Oncol*, 98(3):639–646, 2017.

[8] S.-H. Lu, J. C.-H. Cheng, S.-H. Kuo, et al. Volumetric modulated arc therapy for nasopharyngeal carcinoma: A dosimetric comparison with TomoTherapy and step-and-shoot IMRT. *Radiother Oncol*, 104(3):324–330, 2012.

[9] A. Choudhury, G. Budgell, R. MacKay, et al. The future of image-guided radiotherapy. *Clin Oncol*, 29(10):662–666, 2017.

[10] C. Njeh. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J Med Phys*, 33(4):136–140, 2008.

[11] S. Seuntjens, E. Lartigau, S. Cora, et al. ICRU report 91. Prescribing, recording, and reporting of stereotactic treatments with small photon beams. *Journal of the ICRU*, 14:2, 2014.

[12] F.-M. S. Kong, T. Ritter, D. J. Quint, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int J Radiat Oncol*, 81(5):1442–1457, 2011.

[13] K. A. Goodman, E. A. Wiegner, K. E. Maturen, et al. Dose-escalation study of single-fraction stereotactic body radiotherapy for liver malignancies. *Int J Radiat Oncol*, 78(2):486–493, 2010.

[14] P. J. Keall, G. S. Mageras, J. M. Balter, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys*, 33(10):3874–3900, 2006.

[15] I. S. Grills, G. Hugo, L. L. Kestin, et al. Image-guided radiotherapy via daily online cone-Beam CT substantially reduces margin requirements for stereotactic lung radiotherapy. *Int J Radiat Oncol*, 70(4):1045–1056, 2008.

[16] T. Sothmann, O. Blanck, K. Poels, et al. Real time tracking in liver SBRT: Comparison of CyberKnife and Vero by planning structure-based γ-evaluation and dose-area-histograms. *Phys Med Biol*, 61(4):1677–1691, 2016.

[17] S. D. McGrath, M. M. Matuszak, D. Yan, et al. Volumetric modulated arc therapy for delivery of hypofractionated stereotactic lung radiotherapy: A dosimetric and treatment efficiency analysis. *Radiother Oncol*, 95(2):153–157, 2010.

[18] A. O. Simeonova, K. Fleckenstein, H. Wertz, et al. Are three doses of stereotactic ablative radiotherapy (SABR) more effective than 30 doses of conventional radiotherapy? *Transl Lung Cancer Res*, 1(1):45–53, 2012.

[19] T. Gauer, T. Sothmann, O. Blanck, et al. Under-reported dosimetry errors due to interplay effects during VMAT dose delivery in extreme hypofractionated stereotactic radiotherapy. *Strahlenther Onkol*, 194(6):570–579, 2018.

[20] J. F. Fowler. The linear-quadratic formula and progress in fractionated radiotherapy. *BJR*, 62(740):679–694, 1989.

[21] A. E. Nahum. The radiobiology of hypofractionation. *Clin Oncol*, 27(5):260–269, 2015.

[22] J. Boda-Heggemann, A.-C. Knopf, A. Simeonova-Chergou, et al. Deep inspiration breath hold-based radiation therapy: A clinical review. *Int J Radiat Oncol*, 94(3):478–492, 2016.

[23] C. Riley, Y. Yang, T. Li, et al. Dosimetric evaluation of the interplay effect in respiratory-gated RapidArc radiation therapy. *Med Phys*, 41(1):011715, 2014.

[24] T. Bortfeld, K. Jokivarsi, M. Goitein, et al. Effects of intra-fraction motion on IMRT dose delivery: Statistical analysis and simulation. *Phys Med Biol*, 47(13):2203–2220, 2002.

[25] A. Edvardsson, F. Nordström, C. Ceberg, et al. Motion induced interplay effects for VMAT radiotherapy. *Phys Med Biol*, 63(8):085012, 2018.

[26] B. Fraass, K. Doppke, M. Hunt, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning. *Med Phys*, 25(10):1773–1829, 1998.

[27] L. Lee, Q.-T. Le and L. Xing. Retrospective IMRT dose reconstruction based on cone-beam CT and MLC log-file. *Int J Radiat Oncol*, 70(2):634–644, 2008.

[28] S. Milz, J. J. Wilkens and W. Ullrich. A dose error evaluation study for 4D dose calculations. *Phys Med Biol*, 59(21):6401–6415, 2014.

[29] D. Sarrut. Deformable registration for image-guided radiation therapy. *Z Med Phys*, 16(4):285–297, 2006.

[30] M. Velec, J. L. Moseley, C. L. Eccles, et al. Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration. *Int J Radiat Oncol*, 80(1):265–272, 2011.

[31] R. Werner, J. Ehrhardt, A. Schmidt-Richberg, et al. Towards accurate dose accumulation for Step-&-Shoot IMRT: Impact of weighting schemes and temporal image resolution on the estimation of dosimetric motion effects. *Z Med Phys*, 22(2):109–122, 2012.

[32] S. Ehrbar, S. Lang, S. Stieb, et al. Three-dimensional versus four-dimensional dose calculation for volumetric modulated arc therapy of hypofractionated treatments. *Z Med Phys*, 26(1):45–53, 2016.

[33] M. Velec, J. L. Moseley, T. Craig, et al. Accumulated dose in liver stereotactic body radiotherapy: Positioning, breathing, and deformation effects. *Int J Radiat Oncol*, 83(4):1132–1140, 2012.

[34] N. Samavati, M. Velec and K. K. Brock. Effect of deformable registration uncertainty on lung SBRT dose accumulation. *Med Phys*, 43(1):233–240, 2015.

[35] N. Andratschke, H. Alheid, M. Allgäuer, et al. The SBRT database initiative of the German Society for Radiation Oncology (DEGRO): patterns of care and outcome analysis of stereotactic body radiotherapy (SBRT) for liver oligometastases in 474 patients with 623 metastases. *BMC Cancer*, 18(1):283, 2018.

[36] J. Rieber, N. Abbassi-Senger, S. Adebahr, et al. Influence of institutional experience and technological advances on outcome of stereotactic body radiation therapy for oligometastatic lung disease. *Int J Radiat Oncol*, 98(3):511–520, 2017.

[37] F. Alongi, S. Arcangeli, A. R. Filippi, et al. Review and uses of stereotactic body radiation therapy for oligometastases. *The Oncologist*, 17(8):1100–1107, 2012.

[38] D. T. Chang, A. Swaminath, M. Kozak, et al. Stereotactic body radiotherapy for colorectal liver metastases. *Cancer*, 117(17):4060–4069, 2011.

[39] R. Timmerman, R. Paulus, J. Galvin, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA*, 303(11):1070–1076, 2010.

[40] T. M. Buzug. Computed tomography. In: *Springer handbook of medical technology*, 311–342. Springer Berlin Heidelberg, 2011.

[41] W. Demtröder. *Experimentalphysik 4*. Springer Berlin Heidelberg, 2014.

[42] W. Demtröder. *Experimentalphysik 3*. Springer Berlin Heidelberg, 2010.

[43] H. Reich and U. Burmester. *Dosimetrie ionisierender Strahlung: Grundlagen und Anwendungen*. Teubner, 1990.

[44] O. Klein and Y. Nishina. Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. *Z Med Phys*, 52(11-12):853–868, 1929.

[45] M. J. Berger, J. H. Hubbell, S. M. Seltzer, et al. XCOM: Photon cross sections database, NIST standard reference database 8 (XGAM). *The National Institute of Standards and Technology (NIST)*, 2010.

[46] M. J. Berger, M. Inokuti, H. H. Anderson, et al. ICRU report 37. Stopping powers for electrons and positrons. *Journal of the ICRU*, 19(2), 1984.

[47] P. Mayles, A. E. Nahum and J.-C. Rosenwald. *Handbook of radiotherapy physics*. CRC Press, 2007.

[48] IAEA. *Diagnostic radiology physics: A handbook for teachers and students*. IAEA - International Atomic Energy Agency, 2014.

[49] H. H. Rossi and M. Zaider. *Microdosimetry and its applications*. Springer Berlin Heidelberg, 1996.

[50] J. Booz, L. Braby, J. Coyne, et al. ICRU report 36. Microdosimetry. *Journal of the ICRU*, 19(2), 1983.

[51] T. Han, F. Mourtada, K. Kisling, et al. Experimental validation of deterministic Acuros XB algorithm for IMRT and VMAT dose calculations with the radiological physics center's head and neck phantom. *Med Phys*, 39(4):2193–2202, 2012.

[52] A. Fogliata, F. Lobefalo, G. Reggiori, et al. Evaluation of the dose calculation accuracy for small fields defined by jaw or MLC for AAA and Acuros XB algorithms. *Med Phys*, 43(10):5685–5694, 2016.

[53] P. J. Keall, J. V. Siebers, R. Jeraj, et al. The effect of dose calculation uncertainty on the evaluation of radiotherapy plans. *Med Phys*, 27(3):478–484, 2000.

[54] I. Kawrakow and M. Fippel. Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC. *Phys Med Biol*, 45(8):2163, 2000.

[55] I. Kawrakow. The EGSnrc code system, Monte Carlo simulation of electron and photon transport. *NRCC Report Pirs-701*, 2001.

[56] M. Fippel and F. Nüsslin. Grundlagen der Monte-Carlo-Methode für die Dosisberechnung in der Strahlentherapie. *Z Med Phys*, 11(2):73–82, 2001.

[57] M. J. Berger. Monte Carlo calculation of the penetration and diffusion of fast charged particles. In: *Methods in computational physics volume 1*, 135–215. Academic Press, 1963.

[58] I. Kawrakow. Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version. *Medical Physics*, 27(3):485–498, 2000.

[59] W. Schlegel, C. P. Karger and O. Jäkel. *Medizinische Physik*. Springer Berlin Heidelberg, 2018.

[60] K. Otto. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med Phys*, 35(1):310–317, 2008.

[61] J. M. Park, H. G. Wu, J. H. Kim, et al. The effect of MLC speed and acceleration on the plan delivery accuracy of VMAT. *Br J Radiol*, 88(1049):20140698, 2015.

[62] R. Werner. *Strahlentherapie atmungsbewegter Tumoren*. Springer Fachmedien Wiesbaden, 2013.

[63] R. Werner, C. Hofmann, E. Mücke, et al. Reduction of breathing irregularity-related motion artifacts in low-pitch spiral 4D CT by optimized projection binning. *Radiat Oncol*, 12(1):100, 2017.

[64] W. Lu, P. J. Parikh, J. P. Hubenschmidt, et al. A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4D CT. *Med Phys*, 33(8):2964–2974, 2006.

[65] R. Werner, T. Sentker, F. Madesta, et al. Intelligent 4D CT sequence scanning (i4DCT): Concept and performance evaluation. *Med Phys*, 46(8):3462–3474, 2019.

[66] R. Werner, C. Hofmann and T. Gauer. Optimized projection binning for improved helical amplitude- and phase-based 4DCT reconstruction in the presence of breathing irregularity. In: *Proc SPIE*, Vol. 9783, 978313. International Society for Optics and Photonics, SPIE, 2016.

[67] H. Handels. *Medizinische Bildverarbeitung*. Vieweg+Teubner, 2007.

[68] J. Modersitzki. *FAIR: flexible algorithms for image registration*, Vol. 6. SIAM, 2009.

[69] I. Goodfellow, Y. Bengio and A. Courville. *Deep learning*. MIT Press, 2016, `http://www.deeplearningbook.org`.

[70] S. K. Zhou, H. Greenspan and D. Shen. *Deep learning for medical image analysis*. Academic Press, 2017.

[71] M. Tan and Q. V. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946, 2019.

[72] D. E. Rumelhart, G. E. Hinton and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[73] M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *J Am Stat Assoc*, 82(398):528–540, 1987.

[74] N. Srivastava, G. Hinton, A. Krizhevsky, et al. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res*, 15(1):1929–1958, 2014.

[75] X. Yang, R. Kwitt, M. Styner, et al. Quicksilver: Fast predictive image registration – A deep learning approach. *Neuroimage*, 158:378–396, 2017.

[76] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: *International conference on machine learning*, 1050–1059. 2016.

[77] M. Wilms, R. Werner, J. Ehrhardt, et al. Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy. *Phys Med Biol*, 59(5):1147–1164, 2014.

[78] T. Sothmann, T. Gauer, M. Wilms, et al. Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT. *Phys Med Biol*, 62(23):9001–9017, 2017.

[79] D. A. Low, W. B. Harms, S. Mutic, et al. A technique for the quantitative evaluation of dose distributions. *Med Phys*, 25(5):656–661, 1998.

[80] J. B. Smilowitz, I. J. Das, V. Feygelman, et al. AAPM medical physics practice guideline 5.a.: Commissioning and QA of treatment planning dose calculations—megavoltage photon and electron beams. *J Appl Clin Med Phys*, 16(5):14–34, 2015.

[81] H. Uzunova, M. Wilms, H. Handels, et al. Training CNNs for image registration from few samples with model-based data augmentation. In: *Lect Notes Comput Sc*, 223–231. Springer, Cham, 2017.

[82] N. Otsu. A threshold selection method from gray-Level histograms. *IEEE Trans Syst, Man, Cybern*, 9(1):62–66, 1979.

[83] T. Sothmann, F. Madesta and R. Werner. GDL-FIRE^{4D}: Deep learning-based fast 4D CT image registration. In: *Lect Notes Comput Sc*, 765–773. Springer, 2018.

[84] C. Szegedy, V. Vanhoucke, S. Ioffe, et al. Rethinking the inception architecture for computer vision. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, 4278–4284. IEEE, 2016.

[85] O. Ronneberger, P. Fischer and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In: *Lect Notes Comput Sc*, 234–241. Springer, 2015.

[86] R. Castillo, E. Castillo, R. Guerra, et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol*, 54(7):1849–1870, 2009.

[87] J. Vandemeulebroucke, S. Rit, J. Kybic, et al. Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. *Med Phys*, 38(1):166–178, 2010.

[88] C. Paganelli, B. Whelan, M. Peroni, et al. MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Phys Med Biol*, 63(22):22TR03, 2018.

[89] M. v. Siebenthal, G. Székely, U. Gamper, et al. 4D MR imaging of respiratory organ motion and its variability. *Phys Med Biol*, 52(6):1547–1564, 2007.

[90] N. Mogadas, T. Sothmann, T. Knopp, et al. Influence of deformable image registration on 4D dose simulation for extracranial SBRT: A multi-registration framework study. *Radiother Oncol*, 127(2):225–232, 2018.

[91] K. Murphy, B. Van Ginneken, J. M. Reinhardt, et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imag*, 30(11):1901–1920, 2011.

[92] J. Vandemeulebroucke, O. Bernard, S. Rit, et al. Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT. *Med Phys*, 39(2):1006–1015, 2012.

[93] P. Freislederer, A. von Münchow, F. Kamp, et al. Comparison of planned dose on different CT image sets to four-dimensional Monte Carlo dose recalculation using the patient's actual breathing trace for lung stereotactic body radiation therapy. *Med Phys*, 46(7):3268–3277, 2019.

[94] C. L. Ong, M. Dahele, B. J. Slotman, et al. Dosimetric impact of the interplay effect during stereotactic lung radiation therapy delivery using flattening filter-free beams and volumetric modulated arc therapy. *Int J Radiat Oncol*, 86(4):743–748, 2013.

[95] M. A. Palacios, O. Bohoudi, A. M. Bruynzeel, et al. Role of daily plan adaptation in MR-guided stereotactic ablative radiation therapy for adrenal metastases. *Int J Radiat Oncol*, 102(2):426–433, 2018.

[96] T. Finazzi, M. A. Palacios, F. O. Spoelstra, et al. Role of on-table plan adaptation in MR-guided ablative radiation therapy for central lung tumors. *Int J Radiat Oncol*, 104(4):933–941, 2019.

[97] R. Varadhan, T. Magome and S. Hui. Characterization of deformation and physical force in uniform low contrast anatomy and its impact on accuracy of deformable image registration. *Med Phys*, 43(1):52–61, 2015.

[98] T. Yamamoto, S. Kabus, C. Lorenz, et al. 4D CT lung ventilation images are affected by the 4D CT sorting method. *Med Phys*, 40(10):101907, 2013.

[99] F. Liu, Y. Hu, Q. Zhang, et al. Evaluation of deformable image registration and a motion model in CT images with limited features. *Phys Med Biol*, 57(9):2539–2554, 2012.

[100] T. Yamamoto, U. Langner, B. W. Loo, et al. Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. *Int J Radiat Oncol*, 72(4):1250–1258, 2008.

[101] S. Martin, R. O' Brien, C. Hofmann, et al. An in silico performance characterization of respiratory motion guided 4DCT for high-quality low-dose lung cancer imaging. *Phys Med Biol*, 63(15):155012, 2018.

[102] A. Amir-Khalili, G. Hamarneh, R. Zakariaee, et al. Propagation of registration uncertainty during multi-fraction cervical cancer brachytherapy. *Phys Med Biol*, 62(20):8116–8135, 2017.

[103] C. Grohmann, T. Frenzel, R. Werner, et al. Design, performance characteristics and application examples of a new 4D motion platform. *Z Med Phys*, 25(2):156–167, 2015.

[104] I. J. Chetty, B. Curran, J. E. Cygler, et al. Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning. *Med Phys*, 34(12):4818–4853, 2007.

[105] D. Sheikh-Bagheri, D. W. O. Rogers, C. K. Ross, et al. Comparison of measured and Monte Carlo calculated dose distributions from the NRC linac. *Medical Physics*, 27(10):2256–2266, 2000.

[106] M. Velec, J. L. Moseley, S. Svensson, et al. Validation of biomechanical deformable image registration in the abdomen, thorax, and pelvis in a commercial radiotherapy treatment planning system. *Med Phys*, 44(7):3407–3417, 2017.

[107] K. K. Brock. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol*, 76(2):583–596, 2010.

[108] J. McClelland, D. Hawkes, T. Schaeffter, et al. Respiratory motion models: A review. *Med Image Anal*, 17(1):19–42, 2013.

[109] J. R. McClelland, S. Hughes, M. Modat, et al. Inter-fraction variations in respiratory motion models. *Phys Med Biol*, 56(1):251–272, 2010.

[110] F. Madesta, T. Gauer, T. Sentker, et al. Self-consistent deep learning-based boosting of 4D cone-beam computed tomography reconstruction. In: *Proc SPIE*, Vol. 10949, 1094902. International Society for Optics and Photonics, SPIE, 2019.

[111] M. Brehm, P. Paysan, M. Oelhafen, et al. Artifact-resistant motion estimation with a patient-specific artifact model for motion-compensated cone-beam CT. *Med Phys*, 40(10):101913, 2013.

[112] R. Van den Begin, B. Engels, T. Gevaert, et al. Impact of inadequate respiratory motion management in SBRT for oligometastatic colorectal cancer. *Radiother Oncol*, 113(2):235–239, 2014.

[113] J. Boda-Heggemann, A. Jahnke, M. K. Chan, et al. In-vivo treatment accuracy analysis of active motion-compensated liver SBRT through registration of plan dose to post-therapeutic MRI-morphologic alterations. *Rad Onc*, 134:158–165, 2019.

[114] A. V. Louie, G. Rodrigues, J. Olsthoorn, et al. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. *Radiother Oncol*, 95(2):166–171, 2010.

[115] T. Sentker, V. Schmidt, C. Petersen, et al. 4D CT image artifacts affect local control in SBRT of lung and liver metastases. *Submitted to Int J Radiat Oncol*, 2019.

Appendix A

List of publications

- **T. Sentker**†, V. Schmidt†, C. Petersen, F. Madesta, C. Hofmann, R. Werner, T. Gauer. 4D CT image artifacts affect local control in SBRT of lung and liver metastases. *Submitted to Radiother Oncol*, 2019.
- N. Gessert, **T. Sentker**, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, A. Schlaefer. Skin Lesion Classification Using CNNs with Patch-Based Attention and Diagnosis-Guided Loss Weighting. *IEEE Trans Biomed Eng*, 2019.
- R. Werner, **T. Sentker**, F. Madesta, T. Gauer, C. Hofmann. Intelligent 4D CT sequence scanning (i4DCT): Concept and performance evaluation. *Med Phys*, 46(8):3462–3474, 2019.
- **T. Sentker**, F. Madesta and R. Werner. Patient-specific 4D Monte Carlo dose accumulation using correspondence model based motion prediction. In: *Proc SPIE*, Vol. 10951, 1095109. International Society for Optics and Photonics, SPIE, 2019.
- F. Madesta, T. Gauer, **T. Sentker**, R. Werner. Self-consistent deep learning-based boosting of 4D cone-beam computed tomography reconstruction. In: *Proc SPIE*, Vol. 10949, 1094902. International Society for Optics and Photonics, SPIE, 2019.
- **T. Sentker**†, F. Madesta† and R. Werner. GDL-FIRE^{4D}: Deep Learning-based Fast 4D CT Image Registration. In: *Lect Notes Comput Sc*, 765–773. Springer, 2018.
- D. Schetelig, J. Sedlacik, J. Fiehler, A. Frölich, T. Knopp, **T. Sothmann**, J. Waschkewitz, R. Werner. Analysis of the influence of imaging-related uncertainties on cerebral aneurysm deformation quantification using a no-deformation physical flow phantom. *Sci Rep*, 8(1):11004, 2018.
- T. Gauer, **T. Sothmann**, O. Blanck, C. Petersen, R. Werner. Under-reported dosimetry errors due to interplay effects during VMAT dose delivery in extreme hypofractionated stereotactic radiotherapy. *Strahlenther Onkol*, 194(6):570–579, 2018.
- N. Mogadas†, **T. Sothmann**†, T. Knopp, T. Gauer, C. Petersen, R. Werner. Influence of deformable image registration on 4D dose simulation for extracranial SBRT: A multi-registration framework study. *Radiother Oncol*, 127(2):225–232, 2018.

- **T. Sothmann**, T. Gauer, R. Werner. Influence of 4D CT motion artifacts on correspondence model-based 4D dose accumulation. In: *Proc SPIE*, Vol. 10576, 105760F. International Society for Optics and Photonics, SPIE, 2018.

- R. Werner, **T. Sothmann**, F. Madesta, T. Gauer, C. Hofmann. Technical considerations for automated low-pitch spiral 4D CT scanning protocol selection. In: *Proc SPIE*, Vol. 10573, 105733Y. International Society for Optics and Photonics, SPIE, 2018.

- **T. Sothmann**, T. Gauer, M. Wilms, R. Werner. Correspondence model-based 4D VMAT dose simulation for analysis of local metastasis recurrence after extracranial SBRT. *Phys Med Biol*, 62(23):9001–9017, 2017.

- **T. Sothmann**[†], T. Gauer[†] and R. Werner. 4D dose simulation in volumetric arc therapy: Accuracy and affecting parameters. *PLoS One*, 12(2):e0172810, 2017.

- **T. Sothmann**, O. Blanck, K. Poels, R. Werner, T. Gauer. Real time tracking in liver SBRT: comparison of CyberKnife and Vero by planning structure-based γ-evaluation and dose-area-histograms. *Phys Med Biol*, 61(4):1677–1691, 2016.

- E. Colvill, J. Booth, S. Nill, M. Fast, J. Bedford, U. Oelfke, M. Nakamura, P. Poulsen, E. Worm, R. Hansen, T. Ravkilde, J. S. Rydhög, T. Pommer, P. Munck Af Rosenschöld, S. Lang, M. Guckenberger, C. Groh, C. Herrmann, D. Verellen, K. Poels, L. Wang, M. Hadsell, **T. Sothmann**, O. Blanck, P. Keall. 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*, 119(1):159–165, 2016.

  [†] shared first authorship