## Title

Development, validation, and translation of a respiratory motion model-based 4DCT technique for use as a clinical protocol for radiation therapy treatment planning



## Author

O'Connell, Dylan

## Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Development, validation, and translation of a respiratory motion model-based 4DCT technique for use as a clinical protocol for radiation therapy treatment planning

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Biomedical Physics

by

Dylan Patrick O'Connell

2018

ABSTRACT OF THE DISSERTATION

Development, validation, and translation of a respiratory motion model-based 4DCT technique for use as a clinical protocol for radiation therapy treatment planning

by

Dylan Patrick O'Connell
Doctor of Philosophy in Biomedical Physics
University of California, Los Angeles, 2018
Professor Daniel Abraham Low, Chair

Breathing motion in radiotherapy is commonly managed with four-dimensional computed tomography (4DCT). 4DCT datasets consist of multiple breathing-gated images that display motion of the subject's anatomy over a breathing period. Commercial 4DCT protocols are susceptible to image artifacts when the subject breathes irregularly. Unlike commercial protocols, model-based 4DCT techniques describe a correspondence between tissue motion and an external signal used as a surrogate for respiratory phase. One such technique, termed '5DCT,' uses free-breathing fast helical acquisition and characterizes lung tissue motion as function of five degrees of freedom: x,y, and z position in a reference geometry, breathing amplitude, and breathing rate. The overarching goal of this dissertation is to develop the 5DCT technique for clinical use.

A validation study was conducted involving comparing 5DCT images to commercial 4DCT using an animal model. Reproducible and periodic breathing patterns were achieved through mechanical ventilation and image similarity was quantified using landmark displacement.

Differences in measured tumor motion between 5DCT and commercial 4DCT were examined in a cohort of 20 lung cancer patients. Solutions for the unique challenges of using model model-based techniques clinically, such as appropriate amplitude interval selection and presentation of a quantitative error map, were developed.

A quality management program was developed to ensure the safety of the protocol using risk analysis methods such as process mapping and failure modes and effects analysis. In addition, safety of the in-house software required to implement the technique was managed through use cases and quantitative, testable safety requirements.

The physiological significance of the 5D model parameters was investigated, particularly their dependence on breathing rate during acquisition. It was shown that the model parameter relating tissue motion to breathing amplitude was largely invariant with breathing rate during acquisition.

A prospective scanning method was developed to reduce the number of fast helical scans, and associated imaging dose, necessary to perform 5DCT while maintaining motion modeling accuracy. A simulation study was conducted using patient breathing traces, and the results demonstrated that adequate sampling of the respiratory cycle could be achieved using six scans, compared to the previously-published protocol that employs twenty-five.

The dissertation of Dylan Patrick O'Connell is approved.

Michael McNitt-Gray

James Michael Lamb

Jonathan Goldin

Nzhde Agazaryan

Daniel Abraham Low, Committee Chair

University of California, Los Angeles

2018

For Rebecca and Richard

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

# ACKNOWLEDGMENTS

Chapter 2 is a version of O'Connell, D., Thomas, D. H., Dou, T. H., Lamb, J. M., Feingold, F., Low, D. A., ... and Hofmann, C. (2015). Comparison of breathing gated CT images generated using a 5DCT technique and a commercial clinical protocol in a porcine model. Medical physics, 42(7), 4033-4042.

Chapter 3 is a version of O'Connell, D., Shaverdian, N., Kishan, A. U., Thomas, D. H., Dou, T. H., Lewis, J. H., ... & Low, D. A. (2017). Comparison of lung tumor motion measured using a model-based 4DCT technique and a commercial protocol. Practical Radiation Oncology.

Chapter 4 is a version of O'Connell, D., Thomas, D. H., Dou, T. H., Aliotta, E., Lewis, J. H., Lamb, J. M., ... & Low, D. A. (2017). Adaptive weighted median filtering for reduced blurring when fusing co-registered fast helical CT images. Biomedical Physics & Engineering Express.

Chapter 6 is a version of O'Connell, D., Thomas, D. H., Lamb, J.M., Lewis, J.H., Dou, T. H., Sieren, J., Saylor, M., Hofmann, C., Hoffman, E. A., Lee, P., Low, D. A. Dependence of subject-specific parameters for a fast helical CT respiratory motion model on breathing rate: an animal study. Physics in Medicine and Biology.

Chapter 7 is a version of O'Connell, D., Ruan, D., Thomas, D.H., Dou, T.H., Lewis, J.H., Santhanam, A.P., Lee, P.P. and Low, D.A., 2018. A prospective gating method to acquire a diverse set of free-breathing CT images for model-based 4DCT. Physics in Medicine and Biology.

Funding for the work described in this dissertation was provided by the UCLA Graduate Division through the Eugene V. Cota-Robles Fellowship, the National Institutes of Health Institutional Training Grant, and from Siemens Healthcare.

I would like to thank the admissions committee of the Physics and Biology in Medicine (PBM) graduate program for granting me the opportunity to study at UCLA. In particular, I

am profoundly grateful to Dr. McNitt-Gray, our program director, for his advocacy, support and mentorship throughout my graduate studies. Dr. McNitt-Gray, through his dedication, kindness and approachability, helped me and many other PBM students find success in our graduate studies.

Dr. Lamb was the first faculty member that I worked with. I would like to sincerely thank him for his patience and guidance during the beginning of my research career. Dr. Lamb has always been generous with his time and provided assistance on many occasions. Dr. Lamb's integrity and commitment to being a good scientist have left a lasting impression on me. I look forward to continuing to work with him during my time as a physics resident.

I am grateful to Dr. Goldin for taking the time to serve as a member on my dissertation committee and for his valuable insight, engagement, and many suggestions.

I would like to thank Dr. Agazaryan for his mentorship and support during my time in the Department of Radiation Oncology. Dr. Agazaryan went above and beyond to make me feel welcome in the department. Without his help, I would not have been able to participate in many of the clinically-oriented activities that added greatly to my education. Dr. Agazaryan's professionalism and devotion to excellence in research and patient care make him an exemplary medical physicist, and I am grateful to have had the opportunity to work with him early in my career.

Dr. David Thomas was tremendously helpful in my early research efforts, and I am very thankful for his guidance, frequent assistance, and friendship.

Lastly, I would like to thank my advisor Dr. Daniel Low. I am incredibly fortunate to have spent the last few years working with an advisor who was a skilled researcher, a leader in his field, and perhaps most importantly a genuinely good and caring person. I greatly admire Dr. Low's inquisitiveness, intellectual tenacity, and emphasis on looking critically at data to avoid being misled by false preconceptions. Dr. Low's enthusiasm and passion for discovery continually inspired me. All of my accomplishments during my graduate studies were made possible through his guidance and tireless support. Dr. Low's adventurous spirit and positive outlook extend beyond his work, and getting to know him over the past several

years has helped me grow not only as a researcher, but also as an individual.

# VITA

2010 Research Intern, NASA Jet Propulsion Laboratory. Developed software using C to extract positional data from images of Jupiter's satellites using optical navigation techniques in order to improve models of their orbits.

2011 Research Fellow, Tufts University Summer Scholars. Designed and executed a study on the effectiveness of Tikhonov-regularization methods on restoration of astronomical images

2012 Intern, Department of Radiation Oncology, Rhode Island Hospital. Development work on an Android application to reduce radiation treatment errors through safety checklists and information sharing.

2013 B.S. (Physics), Tufts University

2013–present Graduate Student Researcher, Department of Radiation Oncology, UCLA

# PUBLICATIONS

O'Connell, D., Thomas, D. H., Dou, T. H., Lamb, J. M., Feingold, F., Low, D. A., ... and Hofmann, C. (2015). Comparison of breathing gated CT images generated using a 5DCT technique and a commercial clinical protocol in a porcine model. Medical physics, 42(7), 4033-4042.

O'Connell, D., Thomas, D. H., Dou, T. H., Aliotta, E., Lewis, J. H., Lamb, J. M., ... & Low, D. A. (2017). Adaptive weighted median filtering for reduced blurring when fusing co-registered fast helical CT images. Biomedical Physics & Engineering Express.

O'Connell, D., Shaverdian, N., Kishan, A. U., Thomas, D. H., Dou, T. H., Lewis, J. H., ... & Low, D. A. (2017). Comparison of lung tumor motion measured using a model-based 4DCT technique and a commercial protocol. Practical Radiation Oncology.

O'Connell, D., Ruan, D., Thomas, D.H., Dou, T.H., Lewis, J.H., Santhanam, A.P., Lee, P.P. and Low, D.A., 2018. A prospective gating method to acquire a diverse set of free-breathing CT images for model-based 4DCT. Physics in Medicine and Biology.

O'Connell, D., Thomas, D. H., Lamb, J.M., Lewis, J.H., Dou, T. H., Sieren, J., Saylor, M., Hofmann, C., Hoffman, E. A., Lee, P., Low, D. A. Dependence of subject-specific parameters for a fast helical CT respiratory motion model on breathing rate: an animal study. Physics in Medicine and Biology.

# CHAPTER 1

# Introduction

Accurate characterization and management of respiratory motion is essential to delivering prescription radiation dose to lung tumors while sparing surrounding healthy tissues [1]. Stereotactic body radiation therapy (SBRT), a treatment paradigm involving higher radiation doses in fewer fractions than conventional radiotherapy, is used in many early stage primary lung cancer cases due to its superior local control rates [2][3]. Due to the conformality and steep gradients in planned dose distributions, SBRT requires a highly precise definition of the target volume. Underestimation of breathing motion can result in systematic underdosing and reduced tumor control probability (TCP) [4]. Excessively irradiating adjacent tissues, especially for centrally located lesions, can lead to pulmonary toxicity [5]. Increasing prevalence of SBRT as the primary treatment for early stage lung cancer highlights the need for continued refinement of techniques to minimize uncertainty in respiratory motion estimation.

## 1.1 Commercial 4DCT and its shortcomings

One widely used strategy to manage breathing motion is to define the internal target volume (ITV) using four dimensional computed tomography (4DCT), a set of images at different respiratory phases. The ITV is commonly defined on the maximum intensity projection (MIP) of the 4DCT dataset[6][7]. Commercial 4DCT relies on sorting acquired data according to breathing phase or amplitude to produce gated images. Commercial techniques produce accurate images when breathing rate and depth are consistent during acquisition, but the sorting process can introduce image artifacts in the presence of irregular breathing [8] and

misrepresent tumor motion [9]. Lung cancer patients tend to breathe irregularly, therefore the compromised performance of commercial 4DCT in such situations warrants concern.

Underestimation of tumor motion magnitude by 4DCT is well reported in both phantom and patient studies. [10] measured the ranges of target motions captured by 4DCT MIPs using a programmable lung phantom. Movement patterns with varying amplitudes and periods resulted in systematic underestimation of motion magnitude, in some cases by 1 cm or more. A similar phantom study by [11] also reported ITV underestimations for targets with large and varied motion amplitudes. [12] used implanted fiducials to measure lung tumor motion and simulated multiple ciné 4DCT acquisitions for each patient with different start times, then defined an ITV for each simulated acquisition on its MIP. Variations of greater than 100% volume were observed for the ITVs of 10 mm diameter tumors, and none of the ITVs completely covered the entire range of tumor motion for any of the eight patients in the study. [13] subsequently evaluated the dosimetric consequences of such errors in target volume definition with digital phantoms. In one case, over 25% of the tumor was underdosed.

Similarly, [14] found that 4DCT MIPs underestimated target volumes up to 40% for patients with irregular breathing patterns when compared to high temporal resolution dynamic MRI (dMRI). Further dMRI studies by the group consistently reported target volume underestimations by 4DCT and that the magnitude of the underestimations depended on tumor size and breathing variation [15] [16]. Again, the largest discrepancies were observed in patients with small tumors who breathed irregularly.

## 1.2 Model-based 4DCT

Respiratory motion model-based techniques were developed as an alternative approach to overcome the weaknesses associated with sorting-based methods. Model-based techniques correlate the externally measurable breathing surrogate signal, typically an abdominal belt or fiducial markers, to internal organ motion rather than only the surrogate signal for binning. [17]. '5DCT,' proposed by [18] is one such model-based technique. The method uses fast helical scanning to sample the breathing cycle and deformable image registration to measure

tissue displacement. [19] demonstrated 5DCT's improved image quality over commercial methods on a cohort of 10 lung cancer patients. Fast helical image acquisition affords 5DCT superior temporal resolution and the technique is robust in the presence of irregular breathing due to its parametrization of the respiratory cycle as a function of amplitude and rate rather than an explicit function of time.

## 1.3 Specific aims and overview

The goal of this dissertation is to develop the 5DCT protocol for use in a clinical setting, including validation, software safety and quality assurance, and to address practical challenges relating to the model parameter reproducibility and imaging dose in order to facilitate use of 5DCT for applications other than radiation therapy.

1. Independently validate 5DCT through comparison to commercial 4DCT under controlled conditions.

2. Develop a clinical workflow that manages the unique aspects of motion model-based breathing-gated imaging techniques.

3. Design and implement a quality management program to ensure the safety of the protocol and the in-house software used to generate 5DCT images.

4. Establish a foundation for application of 5DCT to areas outside of radiotherapy, namely biomechanical modeling and chronic obstructive pulmonary disease (COPD) staging, by addressing the physiological significance of the '5D' model parameters and lowering the imaging dose associated with the protocol.

Chapter 1 addresses specific aim (SA)1, an independent validation of 5DCT. Chapters 2 and 3 treat SA2, developing a clinical workflow for a motion model-based technique. Chapter 4 deals with the software safety and quality management aspects of the project, as outlined in SA3. Chapters 5-7 involve SA4, applying 5DCT and fast helical free-breathing CT acquisition to biomechanical modeling and COPD staging.

# CHAPTER 2

# Comparison of breathing gated CT images generated using a 5DCT technique and a commercial clinical protocol in a porcine model

Chapter 2 deals with a validation of the 5DCT against commercial 4DCT under controlled conditions designed to allow the commercial protocol to produce high-quality, artifact-free images. Validation was challenging because in order for a fair comparison, the respiratory motion needed to be the same across acquisitions with the two different protocols. This requirement precluded use of patients, where exactly reproducible breathing is impossible. Additionally, multiple pairs of datasets had to be acquired for a meaningful analysis, which introduced dose constraints. While a phantom study was possible, it would be of limited utility due to the deformable image registration used in 5DCT – any phantom would need to realistically model the human lung and its motion.

To accomplish the validation, we performed an animal study using an anesthetized and mechanically ventilated swine in order to achieve reproducible and perfectly regular breathing. Porcine models have been used extensively in pre-clinical imaging studies due to similarities in lung anatomy with humans. Commercial 4DCT was performed, along with 5DCT, at various respiratory rates and tidal volumes. 5DCT images were generated at the same breathing phases as the commercial reconstructions. Corresponding images were compared using automatically generated landmarks. Mean landmark displacement was found to be less than 2 millimeters.

A version of this chapter has been published in Medical Physics: O'Connell, D., Thomas, D. H., Dou, T. H., Lamb, J. M., Feingold, F., Low, D. A., ... and Hofmann, C. (2015). Com-

parison of breathing gated CT images generated using a 5DCT technique and a commercial clinical protocol in a porcine model. Medical physics, 42(7), 4033-4042

## 2.1 Abstract


### 2.1.1 Purpose

To demonstrate that a '5DCT' technique which utilizes fast helical acquisition yields the same respiratory gated images as a commercial technique for regular, mechanically produced breathing cycles.

### 2.1.2 Methods

Respiratory-gated images of an anesthetized, mechanically ventilated pig were generated using a Siemens low-pitch helical protocol and 5DCT for a range of breathing rates and amplitudes and with standard and low dose imaging protocols. 5DCT reconstructions were evaluated independently for precision by measuring the distance between tissue positions predicted by a 5D motion model and those given by deformable registration, as well as for accuracy by reconstructing the originally acquired scans. Discrepancies between the 5DCT and commercial reconstructions were measured using landmark correspondences.

### 2.1.3 Results

Mean distance between model predicted tissue positions and deformably registered tissue positions over the nine datasets was 0.65 $\pm$ 0.28 millimeters. Reconstructions of the original scans were accurate to 0.78 $\pm$ 0.57 mm on average. Mean landmark displacement between the commercial and 5D images was 1.76 $\pm$ 1.25 mm while the maximum lung tissue motion over the breathing cycle had a mean value of 27.2 $\pm$ 4.6 mm. Use of an averaged image for scans acquired with a low dose protocol resulted in single standard deviation image noise of 6 HU in the heart versus 31 HU for the commercial images.

### 2.1.4 Conclusions

A gold standard evaluation of the 5DCT technique was conducted through landmark based comparison to breathing gated images acquired with a commercial protocol under highly

regular ventilation. The techniques were found to agree to within 2 mm for most respiratory phases and most points in the lung.

## 2.2 Introduction

Breathing-induced motion can cause errors in the delivery of therapeutic radiation to lung tumors [20]. Currently, respiratory motion is managed in radiation therapy treatment planning through the acquisition of four-dimensional computed tomography (4DCT), a series of volumetric images at different breathing phases [1]. Commercial 4D-CT image datasets are typically acquired using low-pitch helical [21] or ciné [22] sequences. The acquired projections or images are retrospectively grouped according to breathing phase or amplitude using simultaneously acquired breathing surrogate measurements. Sorting-based 4DCT techniques produce artifact-free and clinically useful images when the breathing rate and depth remain consistent throughout acquisition; however, breathing irregularities cause sorting artifacts with the potential to introduce errors in breathing amplitude estimations or tumor volume segmentation [23].

The 5DCT data acquisition and analysis technique that utilizes fast helical CT scans presented by Low et al. [18] and described in detail by Thomas et al. [19] has been shown to provide quantitative 4DCT images free of sorting artifacts. In brief, 5DCT samples multiple breathing cycles with several scans and uses a voxel-based 5D respiratory motion model [24] to describe lung tissue displacement as a function of tidal volume and airflow, equivalent to breathing depth and rate, which are measured during image acquisition using an abdominal bellows.

In addition to offering artifact-free breathing gated images, 5DCT potentially allows the entirety of the imaging dose to contribute to image quality and can be used with common clinical equipment. The aim of this study is a gold standard validation of the 5D technique by demonstrating that it yields the same images as a commercial 4DCT protocol currently employed clinically for highly regular breathing cycles. This paper presents comparison of the images produced by the 5DCT technique and a Siemens 4DCT protocol on a mechan-

ically ventilated anesthetized pig. Controlled ventilation provided ideal conditions for the commercial protocol to produce high quality, motion sorting artifact free images that served as a reference against which the 5DCT reconstructions could be compared.

## 2.3 Methods

### 2.3.1 Animal Model

Procedures were conducted with approval from the Institutional Animal Care and Use Committee of University of Iowa. One farm-bred, wild-type pig (42.7 kg; male) was sedated with a pre-anesthetic mixture of Telazol, Ketamine, and Xylazine and anesthetized with 3-5% delivery of Isoflurane via nose cone inhalation. The pig was mechanically ventilated with 100% oxygen using a volume-controlled ventilator (Model 613; Harvard Apparatus Bioscience Company, Holliston, MA). A tracheostomy was performed, and the animal was intubated with a 7.5-mm diameter cuffed endotracheal tube.

### 2.3.2 Image Acquisition

A multidetector row CT scanner (Somatom Definition Flash; Siemens Healthcare, Forschheim, Germany) was used to acquire nine 4DCT sequences, with both the 5D technique and a Siemens amplitude-gated clinical protocol, of the anesthetized and mechanically ventilated pig. In order to gauge the relative performance of the two techniques over a range of periodic breathing conditions, sequences were acquired using 12.4 bpm, 17.1 bpm, and 24.1 bpm, approximately 450 mL and 650 mL tidal volumes. Experiments were done with typical clinical and low dose image acquisition protocols using 40 mAs and 10 mAs per scan, respectively. The same medium smoothing kernel was used to reconstruct all scans.

Acquisition using the 5DCT technique consisted of 30 repeated fast helical scans, performed in succession with alternating directions, and simultaneous monitoring of the breathing cycle using a surrogate. The scan speed was much greater than the rate of respiratory motion; therefore each axial slice was treated as a collection of instantaneous voxel positions

at a single point in the breathing cycle, given by the surrogate measurement at the slice acquisition time. Each scan was approximately 2.5 seconds in duration, with a 2 second system-imposed delay between scans. Images were acquired with a total collimation width of 3.84 cm, a pitch of 1.2 and a scanner rotation period of 0.285 seconds which resulted in in a table movement of 46.08 mm per rotation. The scans used 120 kVp and were reconstructed using the Siemens B30f medium smoothing kernel.

A pneumatic bellows device (Lafayette Instrument, Lafayette IN) consisting of an elastic hollow belt placed at the abdomen, was used as the breathing surrogate for the 5DCT protocol. The bellows expanded and contracted during inhalation and exhalation, reducing and increasing the relative air pressure in the bellows respectively A pressure transducer was used to generate a voltage signal, which has been previously demonstrated to be proportional to tidal volume in humans [25]. The derivative of the voltage signal with time [V/s] was used as the surrogate measurement of breathing rate. The bellows signal was synchronized to CT acquisition by simultaneously acquiring the CT-on signal from the scanner and correlating the signal to CT slice.

### 2.3.3 5D Motion Model

For each 5DCT acquisition sequence, the first scan was arbitrarily selected as the reference scan and was deformably registered to the following 29 scans. Registration was done using a multi-resolution B-spline transform [26] and with a similarity metric based on local self-similarity proposed by Heinrich et al. [27]. The deformable registration provided the position of each lung voxel from the reference image in each of the other 29 images. Deformation vectors and their corresponding breathing amplitude and rate measurements were used to independently calculate the per-voxel parameters for the 5D breathing motion model. The position of a voxel in the reference image was described by

$$\mathbf{X}(v, f, \mathbf{X}_0) = \mathbf{X}_0 + \boldsymbol{\alpha}(\mathbf{X}_0)v + \boldsymbol{\beta}(\mathbf{X}_0)f \tag{2.1}$$

where $\mathbf{X}$ was the displacement of the voxel, $\mathbf{X}_0$ its position at zero breathing amplitude.

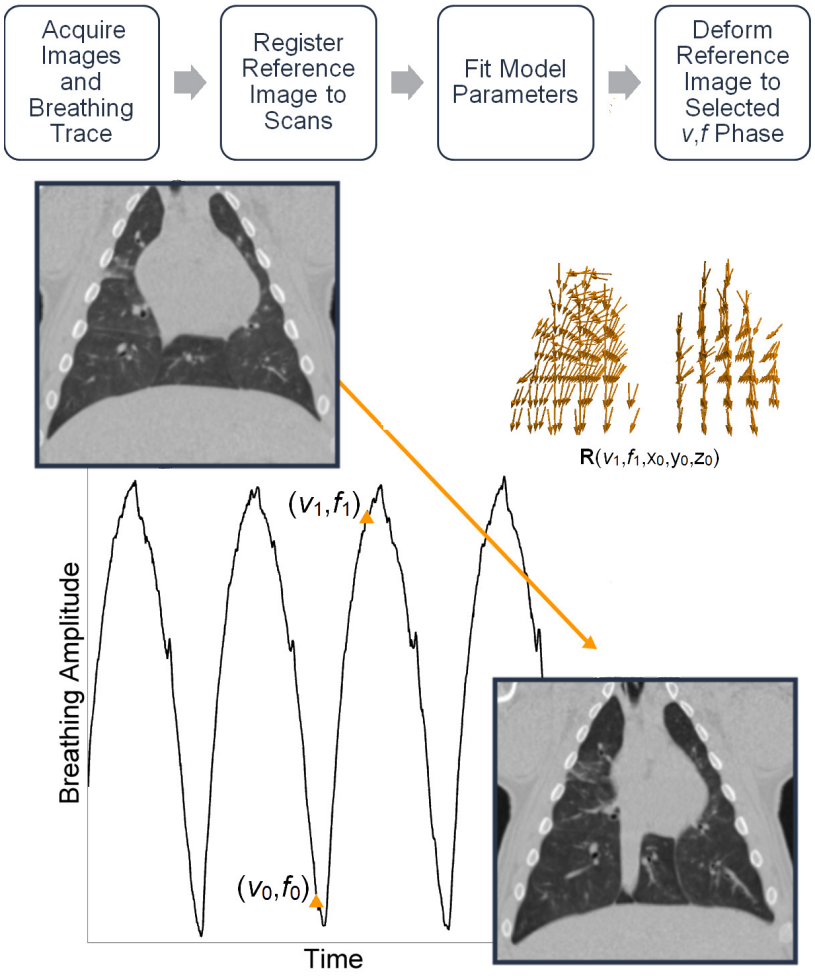


Figure 2.1: Flowchart describing 5DCT image generation. Several scans were acquired with breathing surrogate measurements. One scan was chosen as the reference image and deformably registered to the others. A motion model was used to generate a deformation vector field, R, which gave each voxel's displacement from the reference breathing phase ($v_0$, $f_0$) to a user selected phase ($v_1$,$f_1$) as a function of breathing amplitude $v$, rate $f$, and position $x_0$,$y_0$,$z_0$. $\mathbf{R}$ was then used to deform the reference, creating a new image at breathing phase ($v_1$,$f_1$).

defined as the bellows pressure transducer voltage, selected to be the 5th percentile inhalation breathing amplitude, and zero breathing rate, $v$ was the breathing amplitude and $f$ was the breathing rate. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were vector fields that related tissue motion to $v$ and $f$ respectively. The 30 measured positions for each voxel along with 30 related breathing amplitude and

rate measurements formed an over-determined system for the vector parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were solved independently for each voxel. The least squares solutions were found with QR factorization on a graphics processing unit (Tesla K40; Nvidia, Santa Clara, CA) using a cuBLAS routine. The motion model parameters were used to compute the displacement of each lung voxel, $\mathbf{R}$, from the zero amplitude and rate reference phase, $v_0$ =0 and $f_0 = 0$, to any user selected breathing phase $v_1$ and $f_1$.

$$\mathbf{R} = \boldsymbol{\alpha} v_1 + \boldsymbol{\beta} f_1 \tag{2.2}$$

Images were then generated using the displacement field $\mathbf{R}$ to deform the reference image to the selected breathing phase. Each displacement vector was the sum of two independent functions of breathing amplitude and rate. Therefore, the trajector of each voxel over the breathing cycle lay in a plane. Figure 2.1 shows a flowchart describing the image generation process. A low noise image in the reference geometry was created by averaging the deformably registered scans. For the acquisitions that employed the low dose CT protocol, the averaged image was used in place of the reference image for subsequent deformations and analysis against the Siemens commercial protocol. The 5D motion model precision was evaluated by measuring the distance between the tissue locations in the deformably registered 30 acquired images and the corresponding breathing motion model predicted tissue locations. Model accuracy was assessed by generating images at the breathing phase of each of the 30 originally acquired scans. Similarity between the model generated and original images was quantified by deformably registering the original images to their corresponding reconstructions and measuring deformation vector lengths.

### 2.3.4 Commercial Clinical Protocol

Images generated using the 5DCT motion model were compared to reconstructions produced by a Siemens amplitude-gated 4DCT protocol. The Siemens protocol employed a single, low-pitch helical scan with simultaneous monitoring of intratracheal pressure, which was used as a breathing surrogate. Table feed per rotation was 3.40 mm and the total collimation

width was 38.4 mm, resulting in a spiral pitch factor of 0.09. The scanner rotation time was 0.5 seconds and total scanning time was 42 seconds. CT data was reconstructed into 10 respiratory phases according to pressure amplitude at the time of acquisition.

### 2.3.5 Image Generation

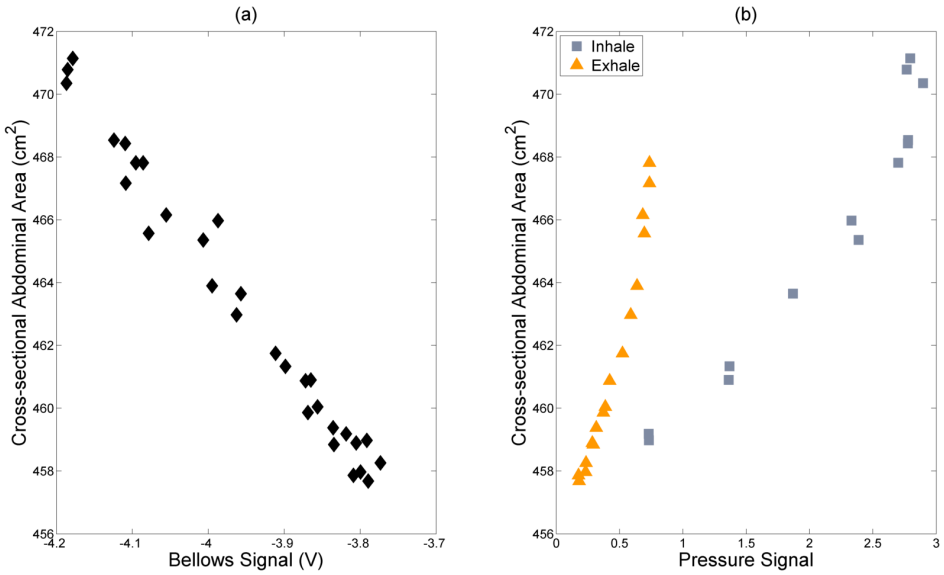


Figure 2.2: Relationship between abdominal cross-sectional area and (a) bellows signal, and (b) intratracheal pressure signal.

For each dataset, a direct comparison between the 5DCT technique and the Siemens 4DCT protocol was performed by using the 5D motion model to deform the reference image to the same breathing phases as the ten images comprising the clinical Siemens 4DCT. Each pair of in-phase images was then compared.

The intratracheal pressure and bellows-measured breathing amplitude were not measured at the same time, so a method was developed to synchronize the two different measurements of the breathing cycle. The complex relationship between intratracheal pressure, used as a breathing surrogate for the Siemens protocol, and the $v$ and $f$ definition of phase used by the 5D model was made evident by the lack of a one to one correspondence between the surrogate measurements. This was likely due to the interplay between intratracheal pressure and airflow into and out of the lungs. In order to relate the two breathing surrogate measurements

to lung volume, the cross-sectional abdominal area at the level of the L2 vertebra was used as a common feature that, in humans, had been shown to correlate to breathing amplitude $v$ [25]. Figure 2.2 shows the relationship between the cross-sectional abdominal area and the bellows voltage (Figure 2.2a) and the intratracheal pressure (Figure 2.2b). The bellows voltage exhibited the expected linear relationship to cross-sectional abdominal area, while intratracheal pressure had a cyclic relationship to abdomen cross-section with a dependence on airflow direction.

The $v$ and $f$ phases corresponding to the commercial reconstruction points were determined by generating images using the first 5DCT acquisition at relative breathing amplitudes of 0% to 100% in increments of 1%, for both inhalation and exhalation, and comparing to the commercial images. For each Siemens reconstruction in the first dataset, the $v$ and $f$ phase of the image which minimized the sum of squared intensity differences within the lung region was selected and used for all subsequent experiments.

### 2.3.6 Landmark Analysis

The discrepancy between images produced by the commercial and 5DCT techniques was determined by measuring the displacement of landmarks between image pairs. 100 landmark points in the each of the commercial phase images (1000 landmarks per 4D-CT) were automatically identified using an algorithm published by Murphy et al. [28] based on intensity gradient magnitude. The landmark points were uniformly distributed throughout the lungs and were reliably identified in the corresponding image. A semi-automatic software system, developed by Murphy et al. [28], was used to establish the correspondence of landmarks between commercial and 5DCT images. For each image pair, a subset of landmarks (30 or more) with the steepest intensity gradients was manually matched. The annotation software presented landmarks one at a time by displaying a coronal, sagittal and axial slice view of the landmark in the commercial image as well as three orthogonal views of the 5D image. Landmark correspondence was determined by matching the the three orthogonal slices or by clicking the appropriate point in any one slice. Based on the manually selected corre-

spondences, a thin-plate-spline model [29] of the deformation between the two images was created which was then used to automatically match the remaining points. The thin-plate-spline warping of the reference image was interpolated to estimate locations of the landmarks in the second image. For each estimated location, a block-matching search was conducted to select a final point which minimized the local sum of squared intensity differences. All landmark correspondences were visually inspected for accuracy. The mean, standard deviation and 95th percentile distance between landmark positions across images were calculated for each 4D-CT dataset.

## 2.4 Results

Table 2.1: Scan mAs, breathing rate, and maximum lung tissue motion for each experiment. Two approximate tidal volumes were used. The maximum respiratory motion was calculated using the motion model by taking the 99th percentile tissue displacement from zero amplitude to the 99th percentile breathing amplitude for each experiment.

| Experiment | mAs | Breathing Rate (bpm) | Maximum Respiratory Motion (mm) |
|---|---|---|---|
| 1 | 40 | 12.4 | 20.3 |
| 2 | 40 | 12.4 | 21.9 |
| 3 | 40 | 17.1 | 24.7 |
| 4 | 40 | 24.1 | 23.1 |
| 5 | 10 | 24.1 | 32.7 |
| 6 | 40 | 24.1 | 32.7 |
| 7 | 40 | 17.1 | 34.2 |
| 8 | 40 | 12.4 | 29.9 |
| 9 | 10 | 12.4 | 28.3 |
| Mean ± SD | | | 27.2 ± 4.6 |

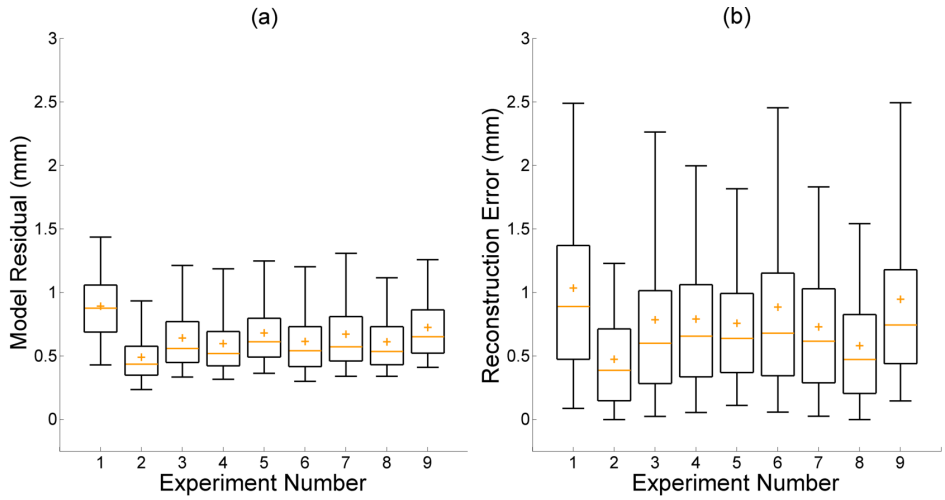


Figure 2.3: Box plots displaying model precision (a) and error in reconstruction of the original scans (b). For each experiment, interquartile range is shown by the box, median by the line and mean by the plus sign. Whiskers mark the 5th and 95th percentiles.

### 2.4.1 5DCT Error Evaluation

The 5D motion model precision, given by the difference between the model predicted and deformably registered positions of all lung voxels, had a mean value of 0.65 mm ± 0.28 mm across the nine datasets. The mean value of the 95th percentile differences was 1.22 mm ± 0.14 mm. The mean 5DCT reconstruction error, defined by the length of deformation vectors from the originally acquired scans to images generated at the same breathing phase using the motion model after registration, was 0.78 mm ± 0.57 mm. The mean 95th percentile deformation magnitude value for all nine datasets was 1.79 mm ± 0.67 mm. Descriptions of the imaging protocol and breathing conditions for each experiment are given in Table 2.1. Figure 2.3 shows box plots of the model residuals which describe precision (a), and reconstruction error of the originally acquired scans (b). The medians, means and interquartile ranges of both quantities were predominantly uniform. Neither model precision nor reconstruction error exhibited a significant dependence on respiratory rate, amplitude or scan mAs.

Calculation of voxel specific motion model parameters in parallel, using software written

in MATLAB (version R2013a; Mathworks, Natick, MA) and CUDA, took approximately 2.7 seconds per dataset. Each individual deformable registration took approximately 12 minutes. Image volumes were 270 x 270 x 314 voxels with isotropic 1 $mm^3$ resolution.

### 2.4.2 Comparison with Commercial Protocol

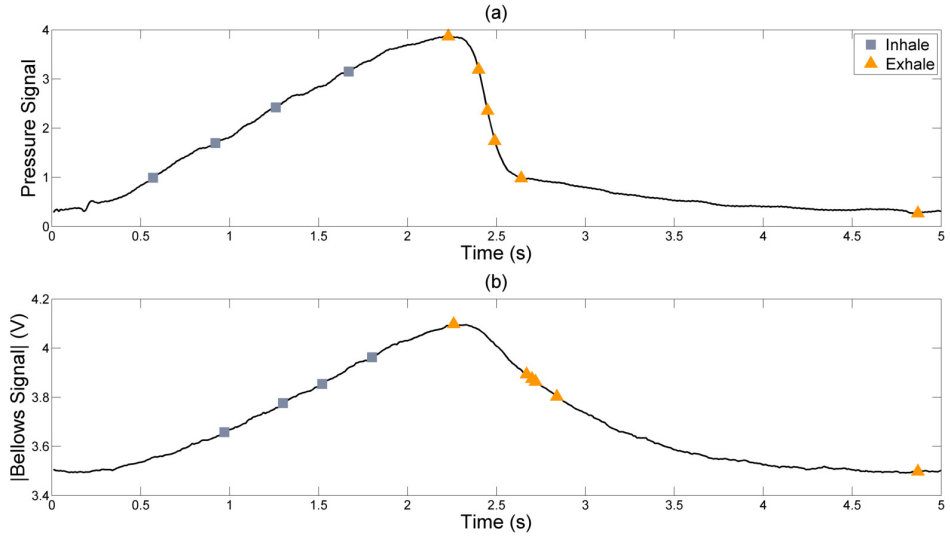


Figure 2.4: Breathing amplitude points at which the Siemens images were reconstructed (a) and their corresponding bellows voltages (b).

Figure 2.4a shows the time points where the intratracheal pressure based amplitudes were selected for reconstruction with the commercial protocol. The four inhalation phases were approximately evenly spaced on the pressure and bellows waveforms; but there were three exhalation phases within a short period of time due to the rapid falloff of the intratracheal pressure. The corresponding bellows voltages at those points are shown in Figure 2.4b. Because there were three time points in rapid succession and with little change in lung volume, only one of these (the middle point) was used in the analysis to avoid biasing the statistics to that bellows-defined phase.

A side-by-side comparison, difference images, and overlays of the commercial protocol and 5DCT images for experiment 3 are shown in Figure 2.5. Images at maximum expiration and inspiration were selected for display. Although the same reconstruction kernel was used,

the 5DCT images appear sharper because of the increased temporal resolution and reduced motion blurring afforded by fast helical acquisition. During experiment 3 the animal was ventilated at a rate of 17.1 breaths per minute, which falls within the typical rate for adult humans [30]. Additionally, the difference between scanner rotation times, and consequently the temporal resolution of the two protocols, was minimized.

Mean landmark displacement, across the eight analyzed phases of all nine experiments, was $1.76 \pm 1.25$ mm. Figure 2.6 shows the distribution of landmark locations throughout the lung for experiment 1 and an example match. The results of the landmark analysis are summarized in Figure 2.7. No significant trends were observed between landmark displacement and respiratory rate, amplitude or scan mAs. Experiments 3 and 5 had the lowest mean and median displacements due to a large number of 1 mm displacements, which indicated a single voxel difference in landmark positions between image pairs. Experiment 7, where respiratory motion was the largest, had the fewest displacements of 1 mm or less, though the mean and median values remained similar to other experiments.

The relationship between landmark displacement and maximum tissue motion during the breathing cycle at the landmark position for experiment 4 is shown in Figure 2.8. The Maximum expiration was chosen to minimize the influence of discrepancy between the Siemens and $v$ and $f$ definitions of breathing phase. Landmark displacement was predominantly uniform with respect to tissue motion, though the largest displacements occurred in high motion regions.

Experiments 5 and 9 were acquired using a low dose (10 mAs) imaging protocol. For these cases, the 5D motion model was used to deform an averaged image. Figure 2.9a and 2.9b show the commercial maximum inspiration image for experiment 5 and the averaged image deformed to the corresponding phase. Pixel-to-pixel noise in the heart was 31 HU in commercial image and 6 in the 5D reconstruction.

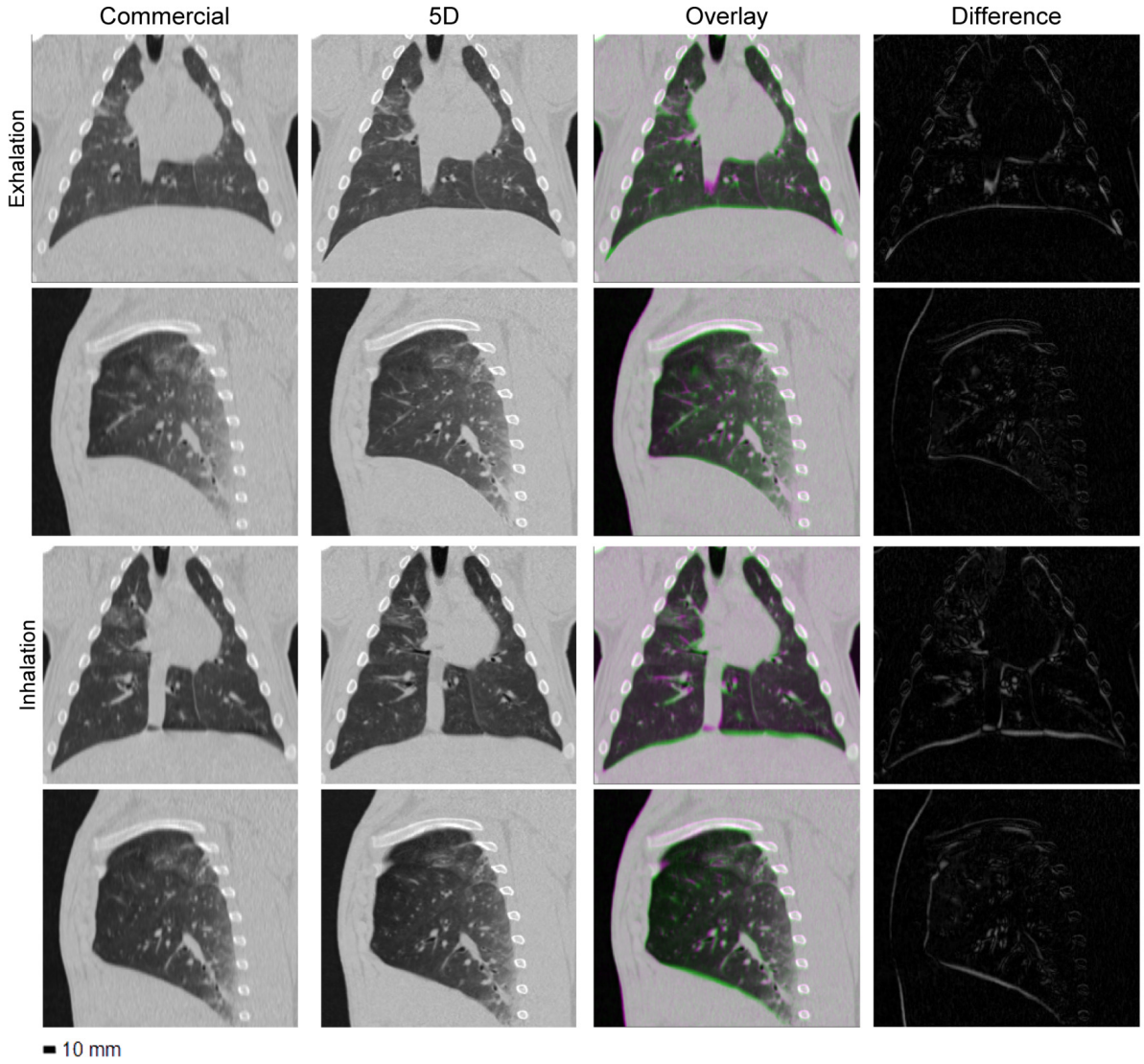


Figure 2.5: Images generated using the commercial and 5D protocols for experiment 3. Columns, from left to right: commercial image, corresponding 5D image, the pair overlaid, and a difference image. The two upper rows show full exhalation and the two lower rows show maximum inspiration. With the exception of the difference images, window and level are 1700 and -300 HU, respectively.

## 2.5 Discussion

The results presented here demonstrate that the 5DCT technique was able to generate respiratory gated images at arbitrary breathing phases accurate to 0.78 $\pm$ 0.57 mm on average for the periodic breathing of a ventilated pig during nine experiments with varied respiratory

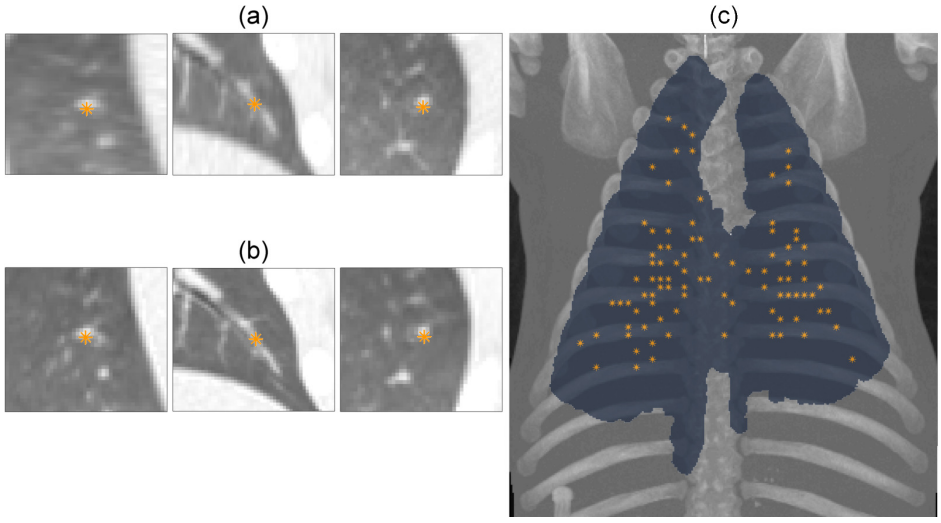


Figure 2.6: Orthogonal views of one of the 100 points selected in the commercial image (a) and the manually identified location of the landmark in the corresponding 5D image (b). Maximum intensity projection of the maximum inspiration image from experiment 1 and its lung segmentation with projected landmark locations marked (c).

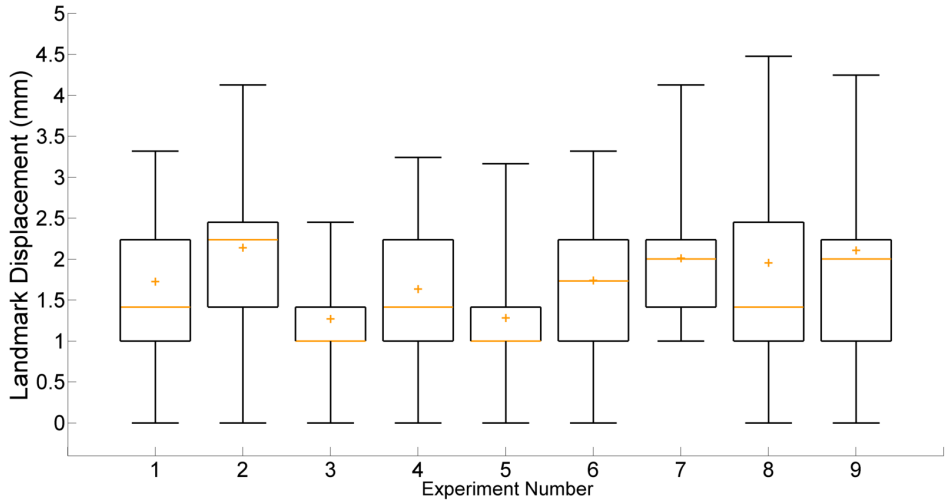


Figure 2.7: Box plot showing the displacement of landmark points between the commercial images and corresponding 5D images. Whiskers mark the 5th and 95th percentiles.

rate, amplitude and scan mAs. Images produced using the 5D technique agree with images generated by a commercial, amplitude sorting based protocol used clinically to within 1.76 $\pm$ 1.25 mm on average. Error in the 5D images and discrepancy between them and the commercial reconstructed was independent of respiratory rate, amplitude and acquisition

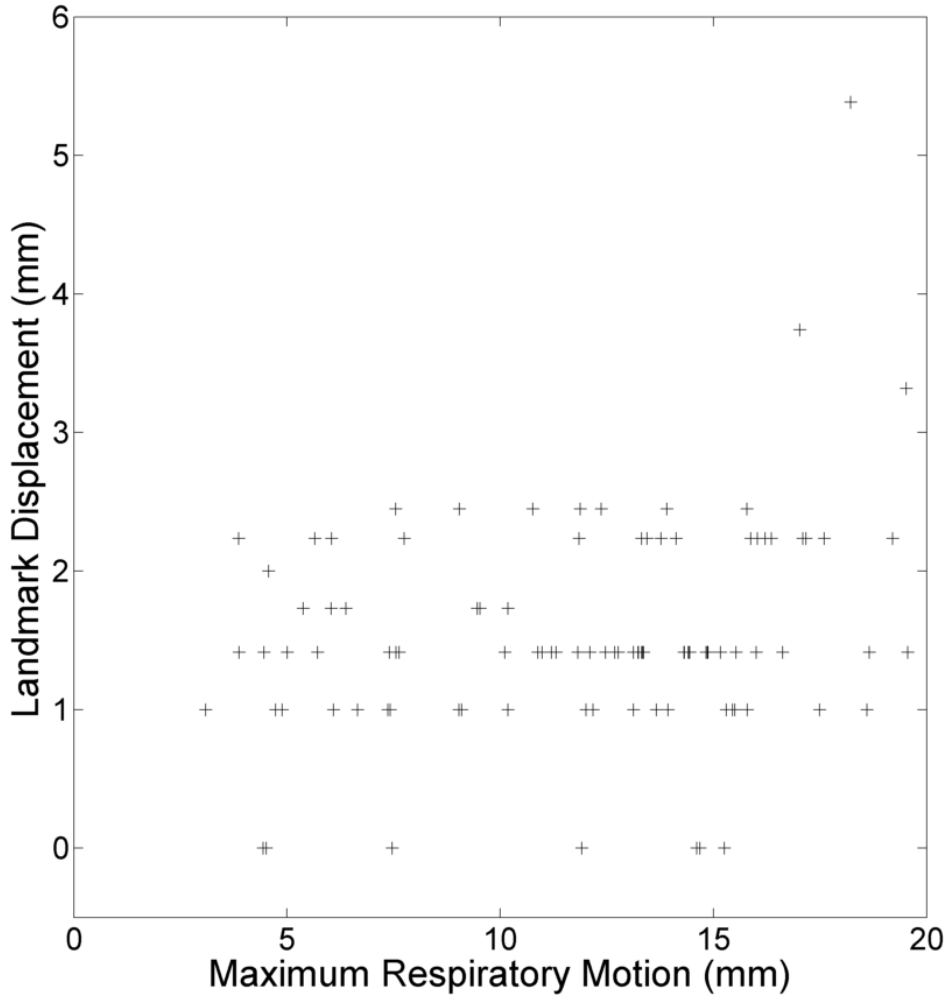


Figure 2.8: Displacement of landmarks between images versus their voxel's maximum motion during the breathing cycle for the maximum expiration phase of experiment 4.

mAs.

While previous work [18] has demonstrated advantages of 5DCT, such as its robustness to breathing irregularites and ability to generate images at arbitrary phases, this paper is the first to assess 5DCT results in terms of scan reconstruction accuracy and landmark based comparison to commercial 4DCT. Agreement between images generated by 5DCT and a commercial protocol, under conditions which allowed the commercial amplitude-based sorting technique to yield accurate and artifact free results, is important for the acceptance of 5DCT as clinically useful.

Two measures of model error were reported, difference in predicted and registered tissue

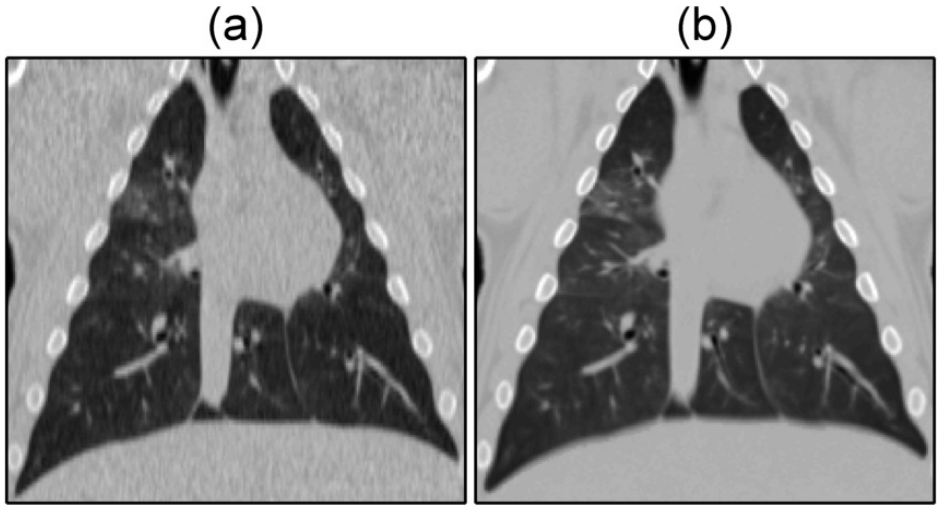


Figure 2.9: Commercial (a) and 5D reconstruction (b), created by deforming an averaged image in the reference geometry, maximum inspiration images for experiment 5. Images are shown with a window and level of 1700 and -300 HU, respectively.

locations (model residual) and deformation vector magnitude between the originally acquired scans and model-based reconstructions. The former provided a metric of the motion models precision. The original scans provided a ground truth and a comparison against their recreation using the motion model was the strongest descriptor of reconstruction accuracy. Deformation vector length was used to quantify reconstruction error in place of landmark correspondence due to the large number of image reconstructions involved and the small magnitude of differences (less than 1 mm on average) between image pairs.

An initial clinical study of the 5DCT technique, described by Thomas et al. [19] achieved a mean model error of $1.19 \pm 0.37$ mm across 10 patients. The quanity reported as model error by Thomas et al. [19] is referred to as model residual in this study because similarity between the original scans and corresponding 5DCT reconstructions was subsequently adopted as the primary error metric. The mean model residual of $0.65 \pm 0.2$ was significantly lower (two sample t-test, $p < 0.01$) in this study, likely due to the use of mechanical ventilation.

Motion model errors were nonnuniform throughout the lung region. Voxels with the largest magnitude errors were in the basal regions near the ribs. The breathing motion model was based on the assumption that tissue motion was linear with tidal volume and

flow, which did not fully account for the nonlinear motion around the ribs. Additionally, the trajectory of each voxel was approximated as being planar, which was based on the analysis of lung tumor motion presented by Seppenwoolde et al. [31]. Whether the inclusion of a third model parameter to allow for nonplanar trajectories would afford increased accuracy is a topic of future research.

Discrepancies between the 5D and Siemens reconstructions were smallest at the maximum and minimum inspiration phases which, as demonstrated in Figure 2.4, occurred at stable regions of the pressure waveform. The commercial sorting method reconstructed individual phase images based on projections occurring within a window of pressure amplitudes while the 5D reconstructions were based on single time points at what were determined to be the mean of these windows. Discrepancy between the images, therefore, was influenced by the range of the pressure values which were included in the reconstruction of each phase image.

Disagreement between image pairs in Figure 2.5 is most prominent in the region surrounding the heart, due to uncompensated cardiac motion, and at the lower lung, which underwent the largest displacement during the breathing cycle. Landmark displacement based error measurements and maximal tissue motion were largely uncorrelated, as shown in Figure 2.9, but the greatest landmark displacements occurred in the basal regions of the lung. The instantaneous positions of these tissues, upon which the analysis was based, were the most sensitive to differences in the specification of respiratory phase. Larger discrepancies occurring at their locations suggest that the inherent difference in the v, f phase and corresponding pressure window definition used by the commercial protocol, which would be most exaggerated in the regions of highest motion, contribute substantially to the discrepancies between image pairs.

## 2.6 Conclusion

This study demonstrated agreement between a commercial 4DCT acquisition technique which is used clinically and a fast helical-based 5DCT technique in a porcine model. The porcine model provided reproducible breathing conditions that allowed for direct comparison

of images. Images produced by the techniques were found to agree within 2 mm for most points in the lungs and most phases and the agreement was independent of respiratory rate, breathing amplitude, and scan mAs.

## 2.7 Acknowledgment

The authors would like to thank Dr. Keelin Murphy of the Image Sciences Insititute at the University of Utrecht (Utrecht, The Netherlands) for use of the landmark selection and annotation software. The Tesla K40 used for this research was donated by the NVIDIA Corporation. This work was supported by NIH R01 CA0096679.

# CHAPTER 3

# Comparison of lung tumor motion measured using a model-based 4DCT technique and a commercial protocol

Chapter 3 involves comparing lung tumor measured using commercial 4DCT and 5DCT on a cohort of 20 patients, interpreting and managing the difference and designing a workflow to support the clinical use of a model-based technique such as 5DCT. Specifically, one of the major challenges associated with model-based 4DCT techniques is that images can be generated at any breathing amplitude. A crucial step unique to such techniques is determining the appropriate amplitude interval to represent when producing a dataset of images for treatment planning. The amplitude selection is critical because it can directly affect target volume definition, and it is the primary focus of this chapter.

A version of this chapter has been published in Practical Radiation Oncology: O'Connell, D., Shaverdian, N., Kishan, A. U., Thomas, D. H., Dou, T. H., Lewis, J. H., ... & Low, D. A. (2017). Comparison of lung tumor motion measured using a model-based 4DCT technique and a commercial protocol. Practical Radiation Oncology.

## 3.1 Abstract


### 3.1.1 Purpose

To compare lung tumor motion measured with a model-based technique to commercial 4DCT and describe a workflow for using model-based 4DCT as a clinical simulation protocol.

### 3.1.2 Methods

20 patients were imaged using a model-based technique and commercial 4DCT. Tumor motion was measured on each commercial 4DCT dataset, and was calculated on model-based datasets for three breathing amplitude percentile intervals: 5th to 85th, 5th to 95th and 0th to 100th. Internal target volumes (ITV) were defined on the 4DCT and 5th to 85th interval datasets and compared using Dice similarity. Images were evaluated for noise and rated by two radiation oncologists for artifacts.

### 3.1.3 Results

Mean differences in tumor motion magnitude between commercial and model-based images were 0.47  3.0 mm, 1.63  3.17 mm, and 5.16  4.90 mm for the 5th to 85th, 5th to 95th, and 0th to 100th amplitude intervals, respectively. Dice coefficients between ITVs defined on commercial and 5th to 85th model-based images had a mean value of 0.77  .09. Single standard deviation image noise was 11.6  9.6 HU in the liver and 6.8  4.7 HU in the aorta for the model-based images compared to 57.7  30 and 33.7  15.4 for commercial 4DCT. Mean model error within the ITV regions was 1.71  0.81 mm. Model-based images exhibited reduced presence of artifacts at the tumor compared to commercial.

### 3.1.4 Conclusion

Tumor motion measured with the model-based technique using the 5th to 85th percentile breathing amplitude interval corresponded more closely to commercial 4DCT than the 5th to 95th or 0th to 100th intervals, which showed greater motion on average. The model-based

technique tended to display increased tumor motion when breathing amplitude intervals wider than 5th  85th were used due to the influence of unusually deep inhalations. These results suggest that care must be taken in selecting the appropriate interval during image generation when using model-based 4DCT methods.

## 3.2 Introduction

Four dimensional computed tomography (4DCT) is widely used to provide treatment planning guidance for mobile tumors [32], [1]. In general, commercial 4DCT acquisition involves acquiring multiple images or projections of the same tissues while simultaneously monitoring breathing using a surrogate device and retrospectively sorting to reconstruct gated images[33].

Sorting-based process assume that breathing periods are similar enough to ignore intercycle variations. However, when period and/or amplitude vary markedly the sorting process introduces artifacts [8] [9] [34] that can cause errors in target volume delineation and dose calculation [10][14][13].

Various methods have been proposed to mitigate sorting artifacts. Langer and Keall [35] demonstrated that using both displacement and velocity of external markers reduced artifact severity. Gianoli et al. [36] used multiple infrared markers distributed on the thorax and abdomen to increase robustness. Recent work by Werner et al [37] defined amplitude bins by analyzing the entire trace rather than defining bins each period. Castillo et al. [38] determined that oversampling led to fewer artifacts.

Respiratory motion model-based techniques, an alternative approach, correlate the surrogate to internal organ motion rather than only using it for sorting [17]. The '5D' respiratory motion model, introduced by Low et al [24] [18], termed for its description of tissue displacement as a function of five degrees of freedom, was shown to produce artifact-free images [19]. The model-based technique in this paper utilizes the 5D motion model and will be referred to as '5DCT.' Briefly, 5DCT uses fast-helical acquisition to sample the respiratory cycle, and deformable image registration to a common reference geometry to measure tissue

displacement.

Although there is a great deal of literature describing the methodology and advantages of model-based techniques, including 5DCT, little work has been done in regards to developing these approaches into viable clinical simulation protocols. Importantly, Model-based techniques are capable of generating images at user-selected breathing phases. A task unique to model-based 4DCT workflows is determining the appropriate amplitude range that images used for target volume delineation should represent. A natural approach would be to use the entire range, but inclusion of unusually deep inhalations can misrepresent the overall breathing pattern. Amplitude interval selection directly affects target delineation and is a critical step, and developing a suitable selection method is one of the first challenges faced when employing model-based 4DCT techniques clinically. In this paper, we describe a method to select an appropriate amplitude interval, and investigate the relationship between the selected amplitude interval and differences in measured tumor motion between 5DCT and commercial 4DCT, and detail the approach we used to manage the differences in order to develop the technique for clinical use at our institution.

## 3.3 Methods

### 3.3.1 Patient Population

Twenty patients were enrolled under an institutional review board-approved study to be scanned using the 5DCT protocol following their simulation with standard 4DCT.

Nine patients were male, and eleven female. Ages ranged from 26 to 87, with a median age of 70. Two patients had two lesions. 4 lesions were located in the left upper lobe, 9 in the right upper lobe, 4 in the left lower lobe and 5 in the right lower lobe.

### 3.3.2 Commercial 4DCT

Commercial 4DCT images were acquired using slow-helical (spiral pitch of 0.1), 800 mAs on a 16-slice scanner (Siemens Sensation Open) with respiratory monitoring using a pressure

belt (AZ-733V; Anzai Medical Corp, Tokyo, Japan). Retrospective sorting was performed by the scanner software using the standard clinical protocol to reconstruct eight breathing-gated images from each acquisition.

### 3.3.3 5DCT

#### 3.3.3.1 Image acquisition

Patients were imaged using 64-slice CT scanners (Biograph TruePoint, Definition AS; Siemens Healthcare, Forchhiemm, Germany). Each patient was scanned 25 times while breathing freely using a low mAs (40), fast rotation (0.33 s) and a pitch of 1.5. 25 scans are used to ensure a high probability of observing all tissues at multiple points in the breathing cycle. The low mAs of each individual scan results in a total CTDIvol comparable to commercial 4DCT.

Breathing signals were generated using an abdominal bellows (Model 76513NM12G; Lafayette Instrument, Lafayette, IN), a pressure-to-voltage transducer and LabVIEW (National Instruments, Austin, TX). Traces were smoothed and derivatives were calculated using Savitzy-Golay filtering [39] [40]. Analysis was performed using in-house MATLAB (version R2016b; MathWorks, Natick, MA) and C++ software.

Respiratory surrogate measurements, the signals from the bellows, were synchronized to the CT data using the scanner x-ray on signal. Using the fast helical acquisition protocol, the scan was rapid enough that the tissues at a single slice location were imaged for a sufficiently short time (0.23s) to treat individual slices of the free-breathing scans as static images. Slices were assigned a breathing phase based on the surrogate signal at acquisition.

#### 3.3.3.2 Image registration and motion modeling

Tissue motion was measured by deformably registering the first scan, the reference image, to the remaining 24 scans using the 'deeds' DIR algorithm [26] [41] [42]. The resulting deformation vectors for each voxel were related to breathing amplitude, v, and rate, f, measurements

using the 5D motion model:

$$\mathbf{X}(v, f, \mathbf{X}_0) = \mathbf{X}_0 + \boldsymbol{\alpha}(\mathbf{X}_0)v + \boldsymbol{\beta}(\mathbf{X}_0)f \tag{3.1}$$

where $\mathbf{X}$ was the voxel displacement, $\mathbf{X}_0$ its position in the reference geometry, defined as zero amplitude and zero rate, the parameter $\boldsymbol{\alpha}$ related breathing depth to respiratory surrogate amplitude, and the parameter $\boldsymbol{\beta}$ related hysteresis motion to surrogate rate. Parameters were solved on a graphics processing unit (Tesla K40; NVIDIA, Santa Clara, CA) using a cuBLAS batched QR factorization routine. The motion model described tissue displacement as a function of five degrees of freedom: x,y,z position in the reference geometry, v and f. The model allowed for the generation of images by providing a means to calculate a deformation vector field (DVF) to deform the reference image.

#### 3.3.3.3 Model error metric

5DCT image accuracy was assessed by comparing originally acquired scans, taken as ground truth, to simulated scans at the same breathing phases generated with the model [43]. Image similarity was quantified by registering each pair and computing the magnitude of deformation vectors. This process yielded 25 error values for every voxel, one for each scan. The approximate 75th percentile error was chosen to be representative of the error distribution and was presented as an error map. The co-registered error map was imported into MIM (version 6.27; MIM Software, Cleveland, OH) to facilitate the viewing. The error metric describes how well the motion model, given only breathing surrogate measurements, can reproduce the tissue positions observed in each of the originally acquired scans. It does is does not describe similarity between 5DCT and commercial 4DCT images.

#### 3.3.3.4 Representative breath generation

A representative breath [44] was constructed to describe the patient's average breathing dynamics over the acquisition period. Slow variation in amplitude caused by physiological drift was characterized using a technique for mean respiratory position tracking [45][46]. A

non-interpolant smoothing spline was fit to the signal [47].

#### 3.3.3.5 Amplitude interval selection and image generation

For each patient, the representative breath was scaled to three amplitude intervals: (1) taking the 5th percentile breathing amplitude as end exhalation, intended to reject abnormal exhalations, and the 85th percentile as maximum inspiration, similarly to mitigate the effect of abnormally deep inhalations, (2) using the 5th to 95th percentiles and (3) the true minimum and maximum. Patient-specific breathing dynamics were quantitatively described with a histogram of the time spent at each amplitude as a percentage of the total acquisition time, shown in Figure 3.1.

### 3.3.4 Tumor motion comparison

Tumor motion was measured on the 4DCT datasets by deformably registering the end-exhalation phase image to the maximum inspiration image and determining the center-of-mass displacement of the gross tumor volume (GTV), which was contoured on the end-exhalation phase. Tumor motion on the 5DCT datasets was found using the model to compute center-of-mass GTV displacement for each amplitude interval.

### 3.3.5 ITV comparison

ITVs were delineated on 5DCT and commercial 4DCT datasets. Target volumes were defined using the maximum intensity projection (MIP) as well as a free-breathing scan [48]. 5DCT MIPs were generated using the 5th to 85th percentile breathing amplitude interval. ITVs were compared using Dice similarity coefficient [49], [50] after a rigid center-of-mass alignment.

### 3.3.6 Image quality evaluation

Image quality differences were quantified to determine their contribution, if any, to discrepancies in tumor motion or target volumes. For each patient, both datasets were evaluated by two radiation oncologists in a blinded manner for presence and severity of artifacts at the tumor location(s). The observer was instructed to view both datasets and assign each a score from 1 to 4. Scoring was according to the following rubric: (1) no apparent artifacts; (2) minimal, target volume delineation is not compromised; (3) moderate, contouring would be more difficult due to the presence of artifacts; (4) severe, target volume could not be reliably delineated due to artifacts.

Irregular breathing during 5DCT acquisition was quantified by calculating the coefficients of variation, defined as the ratios of the standard deviation to the mean, of amplitude and period. Irregularity was correlated to artifact severity. Single standard deviation image noise within a 20-mm diameter region of interest (ROI) at the liver and aorta were measured on the mid-exhalation phase image of both datasets.

## 3.4 Results

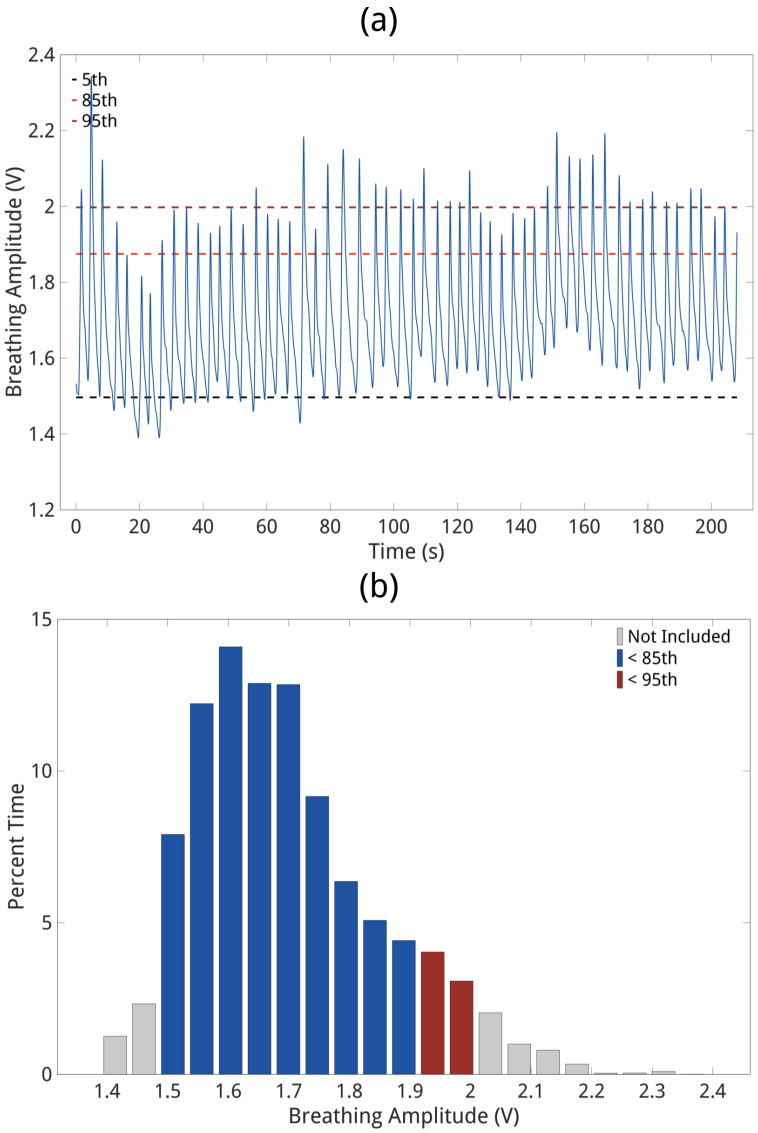


Figure 3.1: Example patient breathing trace (A) and histogram (B) displaying percent of total image acquisition time measured at breathing amplitudes for patient 16. The 5th to 85th percentile amplitude range included 82.2% of all measurements. Blue, amplitudes included in the 5th to 85th percentile interval; red, amplitudes between the 86th and 95th percentiles; gray, amplitudes less than the 5th and greater than the 95th.

Table 3.1: Tumor location, volume, artifact score on 4DCT and 5DCT datasets, and mean 5DCT model error within the GTV region.

| Patient number | Lesion location | Tumor Volume (mL) | 4DCT score | 5DCT sore | Mean error (mm) |
|---|---|---|---|---|---|
| 1 | LUL | 9.15 | 2.5 | 1.0 | 0.9 |
| 2 | RUL | 13.30 | 2.0 | 1.5 | 0.9 |
| 3 | RUL | 9.15 | 2.5 | 1.0 | 2.2 |
| 4 | RUL | 28.16 | 3.5 | 1.5 | 2.7 |
| 5 | RUL | 2.36 | 1.5 | 1.5 | 2.7 |
| 6 | LUL | 3.91 | 3.5 | 1.5 | 2.7 |
| 7a | LLL | 1.28 | 2.0 | 1.0 | 1.8 |
| 7b | RUL | 2.65 | 2.0 | 1.0 | 2.0 |
| 8 | RUL | 2.81 | 2.0 | 1.5 | 0.6 |
| 9 | RUL | 8.97 | 2.0 | 1.5 | 0.8 |
| 10 | RLL | 10.23 | 3.0 | 2.0 | 2.1 |
| 11 | RLL | 1.93 | 3.0 | 1.0 | 2.0 |
| 12 | RUL | 5.10 | 3.0 | 1.0 | 2.3 |
| 13 | RLL | 2.40 | 2.0 | 1.0 | 1.4 |
| 14a | RLL | 54.75 | 1.5 | 1.0 | 3.0 |
| 14b | LLL | 5.22 | 2.0 | 1.0 | 1.8 |
| 15 | LLL | 2.20 | 2.5 | 1.0 | 1.4 |
| 16 | LUL | 1.29 | 1.0 | 1.0 | 1.0 |
| 17 | RLL | 17.36 | 2.0 | 1.0 | 0.9 |
| 18 | LLL | 23.74 | 2.5 | 1.0 | 0.9 |
| 19 | LUL | 11.83 | 2.0 | 1.0 | 3.0 |
| 20 | LUL | 5.12 | 2.5 | 1.5 | 0.6 |

A summary of tumor measurements and statistics is shown in Table 2. Mean model, as displayed in the error map for each patient and defined as the mean of the 75th percentile error values, error within the GTV was 1.71 ± 0.81 mm. No significant correlation was found between model error within the GTV region and tumor displacement or tumor volume (Spearman correlation).

Mean differences in tumor motion magnitude between the 4DCT and 5DCT were 0.47 ± 3.0 mm, 3.17 ± 3.17 mm, and 5.16 ± 4.90 mm for the 5th to 85th, 5th to 95th, and 0th to 100th breathing amplitude intervals, respectively. As demonstrated in Table 3.2 and Figure 3.2a, taking the true maximum and minimum as full inspiration and exhalation results in greater tumor center-of-mass displacement than was observed on the 4DCT in 20/22 tumors. During 5DCT acquisition, images and surrogate signals are recorded over 23 minutes. Consequently, it is more likely that unusually deep breaths will be measured using the technique and a deep breath occuring at any point in the acquisition will contribute to the measured tumor motion if the 0th to 100th interval is used. Conversely, with commercial 4DCT acquisition, abnormally deep breaths need to occur as the tumor is imaged to affect measured motion.

Excluding surrogate measurements in the upper and lower 5 percentiles, on average, markedly reduced the motion magnitude difference with 4DCT. With the 5th to 95th percentile amplitude interval, greater motion was observed in 15/22 tumors. Figure 3.2b shows absolute differences in superior/inferior tumor motion between the 5th to 85th 5DCT and 4DCT plotted against 4DCT artifact score. For differences greater than 1 mm, the discrepancy correlates well with artifact severity score (Spearman's R = .79).

Average Dice similarity coefficient between ITV as defined on commercial and 5DCT datasets was 0.77 ± 0.09. Figure 3.3b shows a bar chart with Dice scores for each tumor.

Mean single standard deviation image noise was 11.6 ± 9.6 HU in the liver and 6.8 ± 4.7 HU in the aorta for the 5D images. The commercial images had mean values of 57.7 ± 30.0 and 33.8 ± 15.4 for the liver and aorta, respectively. Box plots of image noise measurements are shown in Figure 3.3a. $CTDI_{vol}$ for the 5DCT acquisitions was 42 mGy on average and

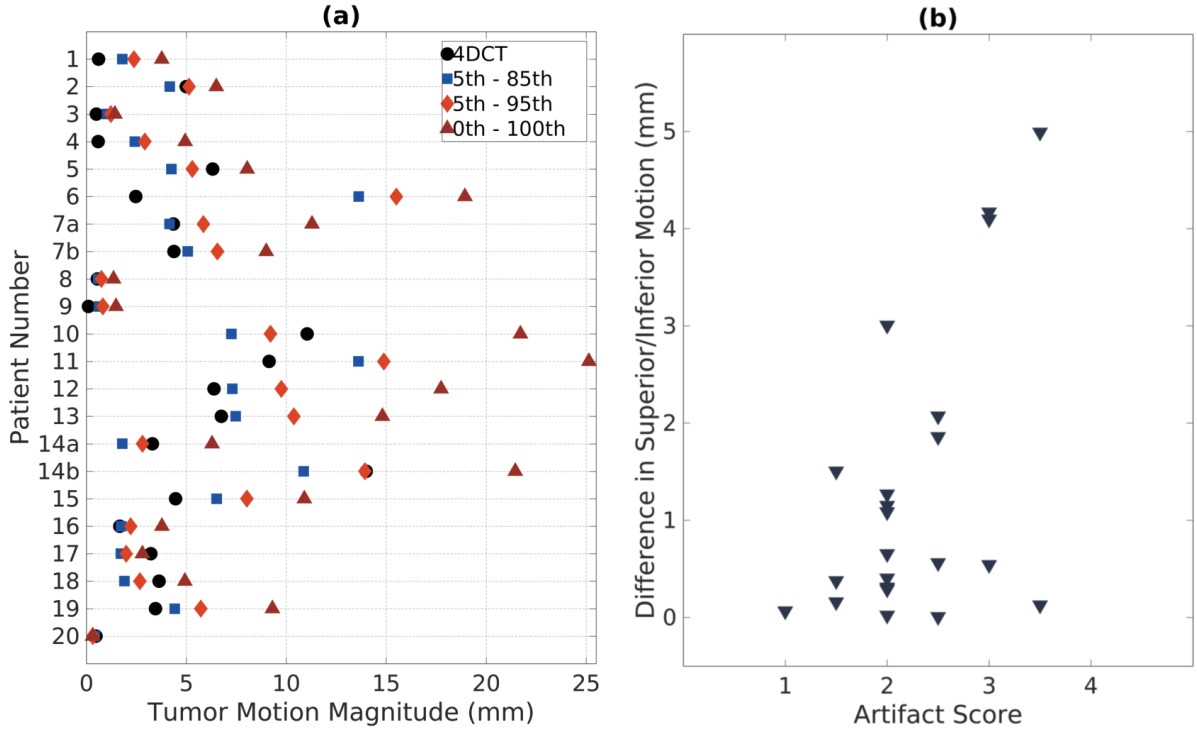


Figure 3.2: (A) Tumor center-of-mass displacement magnitude as measured on 4DCT (circle) and 5DCT datasets for 3 amplitude intervals: 5th to 85th (square), 5th to 95th (diamond), and 0th to 100th (triangle). (B) Absolute difference in superoinferior tumor motion between the 4DCT and 5th to 85th 5DCT plotted against 4DCT artifact score. 4DCT, 4-dimensional computed tomography; 5D, 5 degrees of freedom computed tomography model.

69 mGy for the commercial acquisitions.

Mean artifact severity rating for was 1.2 for 5DCT and 2.2 for commercial 4DCT. No significant correlation between breathing amplitude or period irregularity and artifact severity of the 5DCT images was found (Spearman correlation).

## 3.5 Discussion

5DCT allows for the generation of images at any breathing phase, introducing a choice as to which amplitude interval to represent. Amplitude interval selection can directly influence target volume definition. This study described the differences in measured tumor motion

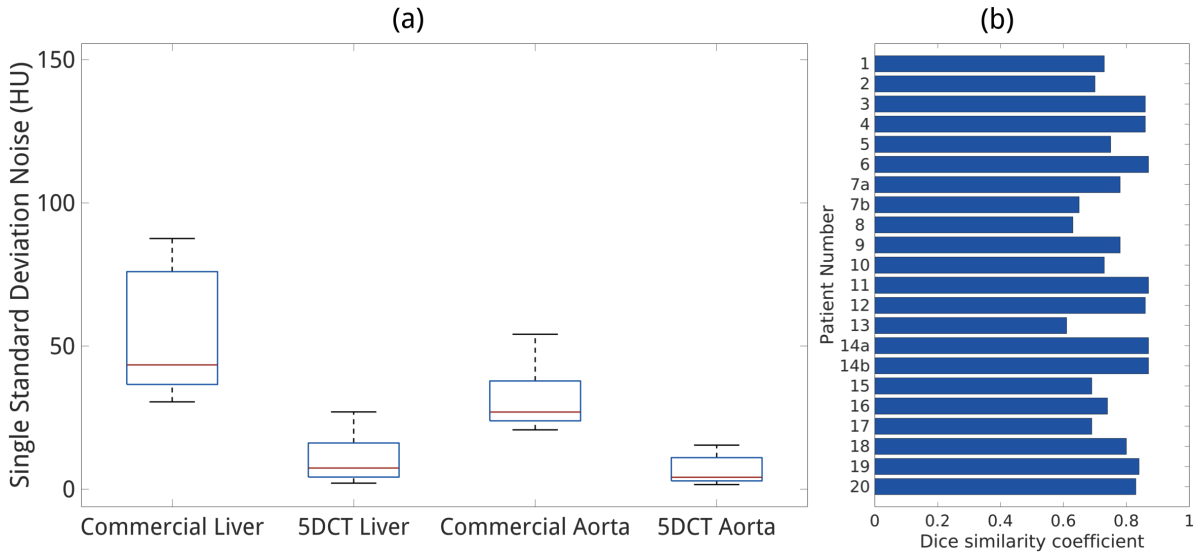


Figure 3.3: (A) Box plot shows image noise measured using a 20 mm ROI on the mid-exhalation images of both datasets. ROIs were taken on axial slices in homogeneous liver and the ascending aorta. (B) Dice similarity coefficient between ITV contours defined on commercial and 5DCT datasets. ITV, internal target volume; ROI, region of interest. Other abbreviation as in Fig 3.2.

Table 3.2: Difference (5DCT  4DCT) in measured tumor centroid motion on each axis from maximum inspiration to maximum exhalation between 4DCT and three 5DCT breathing amplitude intervals.

| Motion axis | Anterior/posterior (mm) | Left / right (mm) | Superior/inferior (mm) |
|---|---|---|---|
| 5th - 85th | 0.64 | 0.05 | 0.20 |
| 5th - 95th | 1.03 | 0.28 | 1.15 |
| 0th - 100th | 2.41 | 1.03 | 4.05 |

between 5DCT images generated with three different amplitude intervals and commercial 4DCT. It was also observed that 5DCT overall yielded images with fewer artifacts and reduced noise, along with quantitative descriptions of image accuracy and the range of amplitudes represented by the resulting 4D dataset.

The large increase in tumor motion measured using the 0th to 100th amplitude interval 5DCT images compared to commercial 4DCT in almost all patients (4 mm in the SI direction on average) suggests that this interval is not indicative of ordinary breathing and therefore not appropriate to use clinically. Furthermore, the 5D motion model was designed to describe quiet respiration and approximates the relationship of tissue motion to breathing amplitude and rate as linear. The accuracy of this approximation decreases as tissue motion becomes nonlinear in amplitude during deep inspiration. Figure 3.4b-d shows 5DCT images reconstructed at the 85th and 95th percentile and the maximum breathing amplitude observed for one patient. Some distortion is visible at the boundary of the diaphragm and the liver in 3.4d, but not in 3.4b or 3.4c.

Out of the three 5DCT amplitude intervals used in this study, the 5th  85th most closely agreed with commercial 4DCT. Using the 5th- 85th amplitude interval to generate images is a heuristic solution that can be used to implement 5DCT clinically, but the results of this study demonstrate the impact of choosing the 5th-85th as opposed to the 0th to 100th interval on tumor motion magnitude is substantial and therefore the specific amplitude interval of any 5DCT images should be made apparent to clinicians. To this end, we have developed an automated tool to report the patient breathing trace with relevant amplitude percentiles overlaid, as well as a histogram of the amount of time each amplitude measurement was observed as a percentage of the total image acquisition time. Example plots for patient 16 are shown in Figure 3.1a and 3.1b. The report is provided to clinicians in PDF format to provide context for the corresponding 5DCT image dataset. Exclusion of the upper 15 amplitude percentiles by definition results in a reduction tumor motion in reconstructed images. The primary utility of the automatically generated reports lies in the visualization of how the upper 15% of amplitudes are distributed. The included plots depict whether these datapoints are evenly distributed over all breaths and remain close to the upper limit of the selected interval, as would be the case in a regularly breathing patient, or if they are concentrated in a few deep inhalations, as in Figure 3.1a. For such cases, where the deep inhalations are repeated and potentially a characteristic of that patients breathing pattern, rather than an aberration, it may be more beneficial to reconstruct images over a wider amplitude interval

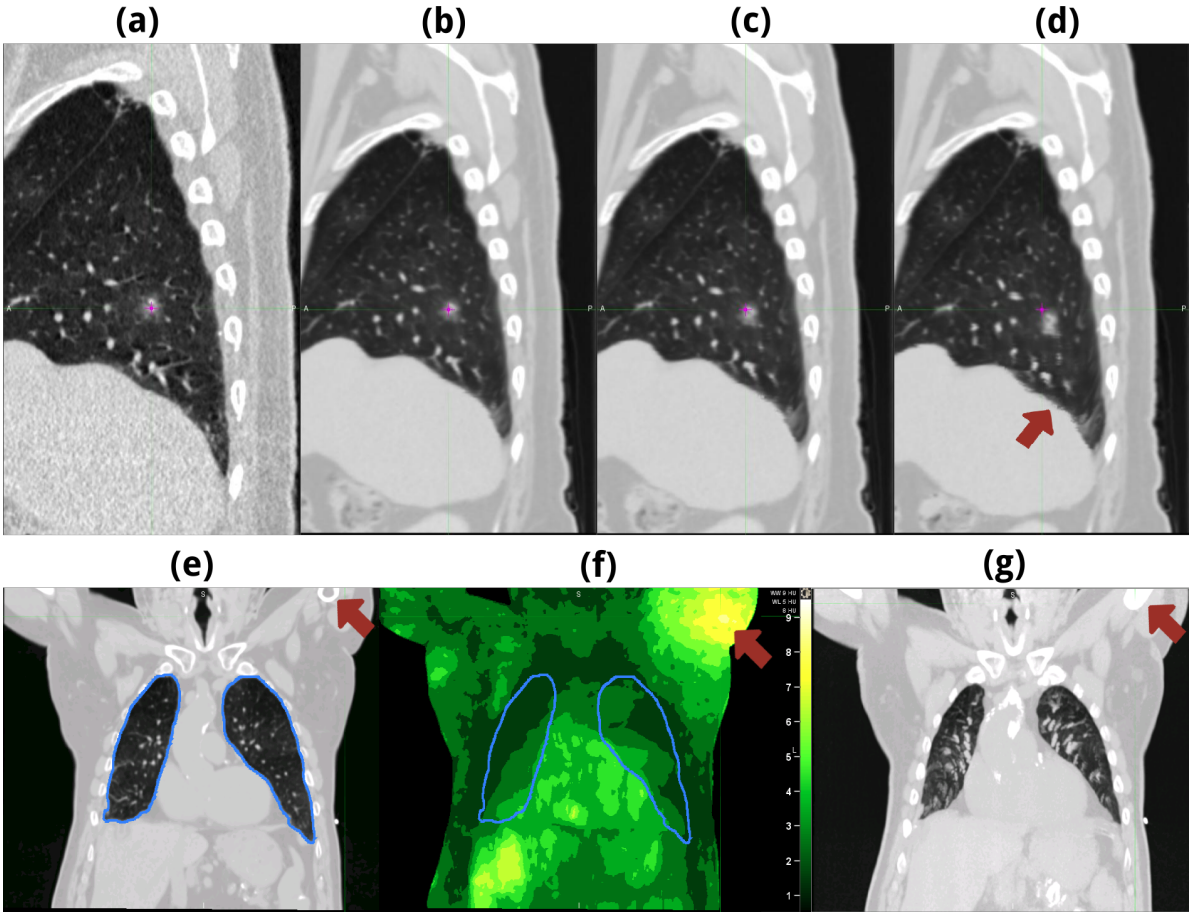


Figure 3.4: (A) Maximum inspiration 4DCT image for patient 6. Gross tumor volume centroid marked with crosshair. (B-D) Corresponding 5DCT images for the 85th, 95th and 100th percentile breathing amplitudes. (E) The 85th percentile maximum inspiration 5DCT image for patient 20. (F) Corresponding error map. (G) Maximum intensity projection of the 25 original fast helical scans. Arrow marks extended appearance of the humerus compared with (E), indicating that the arm was moved during acquisition, resulting in the high local error.

in order to capture these deep breaths in planning. This would be particularly important for free-breathing lung SBRT treatments where there are fewer fractions to average out motion-induced differences between the planned and delivered dose distributions. A major advantage of model-based 4DCT techniques is that once the correspondence model is fit, images can be generated with minimal computational burden, therefore it would be feasible for a physician to view the report and request images representing a different amplitude

interval with little disruption to the fast-paced clinical workflow. Decision metrics have been developed to categorize patient breathing patterns by their level of irregularity, as demonstrated by Blanco Kiely et al in [51]. Such a classifier could potentially be used to determine a suitable initial reconstruction interval on a patient-specific basis, but further study is needed.

The mean Dice similarity coefficient of 0.77 indicated reasonable agreement, on average, between ITV contours defined on commercial 4DCT and 5DCT images generated using the 5th to 85th amplitude. In the absence of a ground-truth, there was no straightforward interpretation of the differences between the two sets of contours. Furthermore, the images were acquired at different times, ranging from several minutes to hours apart, and breathing patterns can vary with time. In order to truly assess the accuracy of the motion measured using both techniques, an additional time-resolved imaging modality such as fluoroscopy or dynamic MRI would be necessary to provide a benchmark against which both 4DCT and 5DCT could be compared. Inter-observer variation could potentially contribute to target volume discrepancies. For example, a study on a contour propagation tool by van Dam et. al [52] described an inter-observer variance across three clinicians in lung tumor definition which resulted in a mean Dice coefficient of 0.88 for a 6 patient study.

Additionally, one of the limitations of this study is that only the tumor motion range was examined, not the absolute positions of the ITV in the 4DCT and 5DCT images. The first ten patients in the cohort were scanned while the 5DCT protocol was still under development and not were immobilized consistently with their 4DCT simulation positions. Furthermore, because different scanners were used for the two protocols, a comparison of absolute ITV position would be influenced by error in reproducing the patient set-up. An earlier animal study performed by our group to validate the technique has shown, using a landmark-based analysis, that 5DCT images correspond to 4DCT images under controlled breathing conditions to within 2 mm [53], suggesting that offsets induced by systematic error are unlikely. Moreover, the error map generated along with 5DCT datasets provides a measure of the techniques spatial accuracy by presenting a voxel-by-voxel comparison of model-generated images at the breathing phases of the 25 originally acquired helical scans with the actual

scans themselves.

The error map described in section 3.2.C.3 is a unique advantage of model-based 4DCT . It offers a quantitative description of uncertainty in the image. The error map is provided to clinicians along with the 5DCT images and amplitude interval report. It is a volumetric image in the same geometry of a MIP of the 5DCT images and is intended to convey regional modeling uncertainty. In Figure 3.4f, a high regional error (greater than 9mm) is apparent at the upper left arm. A maximum intensity projection of the originally acquired free-breathing scans, shown in 3.4g, reveals that the patient's arm moved during acquisition, explaining the discrepancy. While this motion did not affect the lung model accuracy, it is an excellent example of the feedback provided to the clinician in case discrepancies did exist within the relevant lung regions. In this case, it can be concluded that the images are still useful for target volume definition because the issue was localized to the arm region. The main function of the error map is to provide a visual indicator of whether or not 5DCT images are sufficiently accurate for planning purposes. While sorting artifacts in commercial 4DCT images are easily identifiable, a large amount of error in the motion model does not necessarily cause obvious artifacts in reconstructed 5DCT images, as shown in Figure 3.4e where there is no visual indication of the poor modeling. Importantly, the error map is not intended to influence decisions on margin sizes when contouring, only to serve as a warning mechanism to alert clinicians if major errors in the imaging or modeling process occur, much as a severe sorting artifact in a commercial 4DCT would signal that corrective action needs to be taken. Given the lack of data on the accuracy of 4DCT, it is difficult to provide context for the error values and further study needs to be done to establish thresholds for acceptable modeling error.

The mean artifact evaluation score for the 5DCT images of 1.1 was close to the scale's minimum and the most severe artifact was localized blurring. A minor loss of sharpness is introduced when the 25 co-registered scans are fused to create the single low-noise reference image.

Our institutions initial clinical implementation of the 5DCT protocol was designed to mimic commercial 4DCT in order to minimize disruption to the existing workflow: eight im-

age phases were generated to match the existing protocol, but with improved image quality and no sorting artifacts. However, the motion model provides tools for alternate workflows that could reduce clinical workload, such as rapid propagation of contours to all phases, and also has the potential to provide increased quantitation in plan quality and delivery assessment through 4D dose calculation and retrospective dose accumulation. Despite their promise, model-based 4DCT techniques have seen little adoption in the clinic. The focus of this work was to overcome one of the earliest implementation challenges, appropriate amplitude interval selection. Future projects will explore the utility and validation of alternate workflows using the motion model.

## 3.6 Conclusion

We have shown that 5DCT reconstructions using the 5th and 85th percentile breathing amplitudes as minimum and maximum inspiration result in tumor motion magnitude measurements which more closely correspond with commercial 4DCT than using the 5th to 95th or true minimum and maximum. Appropriate breathing amplitude selection is a crucial step in using a model-based technique such as 5DCT clinically. A tool was developed to provide clinicians with a report describing the breathing amplitude interval used to generate 5DCT images as a percentage of time to provide context and determine if it is appropriate on a case-by-case basis.

# CHAPTER 4

# Adaptive weighted median filtering for reduced blurring when fusing co-registered fast helical CT images

Chapter 4 addresses a practical challenge of 5DCT: during free-breathing acquisition, multiple low dose scans are performed. The reconstructed images are then registered to a common reference frame. Due to the low effective mAs technique, each image is relatively noisy. Given a set of co-registered CT images, how should the images be combined to produce a fused image with improved noise characteristics without compromising image sharpness? Typically, one would take the mean or median intensity at each voxel location. Although using the mean or median results in good noise reduction, blurriness is introduced into the fused image due to residual misalignment after deformable registration. This chapter describes a method, adaptive weighted median filtering (AWM), which results in sharper fused images with similar noise reduction as using the mean or median. AWM modulates the amount of smoothing according to local contrast in order to preserve sharpness at image boundaries, such as edges of blood vessels and other small structures in the lung.

A version of this chapter has been published in Biomedical Physics and Engineering Express: O'Connell, D., Thomas, D. H., Dou, T. H., Aliotta, E., Lewis, J. H., Lamb, J. M., ... & Low, D. A. (2017). Adaptive weighted median filtering for reduced blurring when fusing co-registered fast helical CT images. Biomedical Physics & Engineering Express.

## 4.1 Abstract


### 4.1.1 Purpsose

Four dimensional computed tomography (4DCT) images can be deformably registered to a common geometry and combined through mean or median averaging to create a single low-noise fusion. Registration error causes blurring in the resulting image. In this study, we present a method to combine multiple 4DCT images using adaptive weighted median filtering (AWM) which achieves a similar noise reduction but induces less blurring.

### 4.1.2 Methods

Ten lung cancer patients were imaged 25 times using a low mA, fast helical protocol. For each patient, the first scan was selected as a reference and deformably registered to the following 25 scans. The co-registered images were combined using mean and median averaging and the proposed technique. Image noise and sharpness were assessed for the three techniques.

### 4.1.3 Results

Mean pixel-to-pixel image noise in the liver was 7.6 HU for mean averaging, 9.3 HU for median averaging and 9.8 HU for AWM. AWM images were found to be significantly sharper than both mean and median averaged images ($p = .002$) using the Wilcoxon signed rank test.

### 4.1.4 Conclusion

AWM filtering is capable of producing sharper fused images with similar noise reduction to mean and median averaging. The proposed technique also allows for a user-adjustable balance between smoothing and sharpness preservation.

## 4.2 Introduction

In radiation therapy treatment planning for lung cancer, four dimensional computed tomography (4DCT) is commonly employed to determine the extent of respiratory-driven tumor motion so patient-specific planning margins can be used which minimize irradiation of healthy tissue while achieving target coverage [32] [1]. A 4DCT dataset is made up of a sequence of breathing-gated CT images at points distributed throughout the respiratory cycle. Together, the images depict patient respiratory motion over one breath.

Delineation of target volumes and normal structures on each phase of the 4DCT dataset, typically 8 or 10 images, is labor intensive. Several approaches have been adopted to produce single derived images which represent the motion information of the 4DCT and can be used for contouring, such as the maximum intensity projection (MIP) [6], average intensity projection [33], [54], and mid-ventilation [55], [56]. The mid-ventilation technique involves reconstruction of an image in which the tumor and anatomy occupy their time-weighted average positions. Due to hysteresis [31], lung tissue motion differs between inhalation and exhalation and consequently the mean position may not lie on the tumor trajectory. The mid-position concept was developed by Wolthaus et al [57] to cope with hysteresis motion and improve image quality. To generate the mid-position image, deformable image registration is performed between each image of a 4DCT and a common reference, such as the maximum exhalation image, and the resulting deformation vector fields (DVF) are averaged. The mean DVF is then numerically inverted and subtracted from each DVF to yield a set of DVFs which transform each phase of the 4DCT to the mean position. The final mid-position image is then generated by combining the set of transformed images using mean or median averaging. The resulting mid-position image benefits from reduced noise compared to the original images. Noise reduction is advantageous because the individual breathing-gated image of a 4DCT are noisier than a conventional CT image due to the reduced exposure per frame.

Another application in which multiple co-registered thoracic CT images must be fused to create a single low-noise image arises in 5DCT, a motion model-based technique for generating breathing-gated images using fast helical scanning [18]. 25 low mA scans are

acquired under free-breathing conditions. A reference scan is deformably registered to the remaining 24 scans and the resulting DVFs are used to solve for voxel-specific parameters in a breathing motion model [24]. Each individual low mA scan is noisier than a conventional 4DCT image, but a high quality image in the reference geometry is obtained by mean averaging of the 25 co-registered scans. The motion model is then used to deform the fused image to any user-selected breathing phase, resulting in a low-noise and artifact free 4DCT [19].

In both mid-position 4DCT and 5DCT, mean or median averaging of co-registered scans is performed. In both studies cases, mean averaging results in slight blurring of small structures due to registration error. The blurring is more pronounced in 5DCT due to the increased number of image pairs. In this work, we propose a method that uses adaptive weighted median filtering (AWM) to combine multiple co-registered thoracic CT images rather than mean or median averaging to achieve similar noise reduction in homogeneous regions while reducing blurring of edges.

## 4.3 Methods

### 4.3.1 CT image data

Ten patients were imaged using one of two 64-slice scanners (Biograph TruePoint, Definition AS; Siemens Healthcare, Forchhiemm, Germany). Each patient was scanned 25 times in alternating craniocaudaul and caudocranial directions with a low mAs (40) technique, fast rotation time (0.33 s) and a pitch of 1.2 or 1.5. Patients were breathing freely during image acquisition.

### 4.3.2 Image registration

Tissue motion was measured by deformably registering the first scan, chosen as the reference image, to the remaining 24 scans. The 'deeds' DIR algorithm [26] [27] [42], which used a multilevel B-spline transform, a discrete optimization framework and an image similarity

metric based on self-similarity context.

### 4.3.3 Adaptive weighted median filter

For a set of N co-registered CT images, there are N intensity values at all voxel locations. In our method of creating a single low-noise fused image, 1-D weighted median filtering is performed over the N intensity values for every voxel and the weights are independently determined based on local registration error and image gradient.

Conceptually, the weighted median of a sequence $\{S_i\}$ and associated positive integer weights $\{w_i\}$ is found by forming a new sequence $\{S\prime_j\}$ where each element $S_i$ is repeated $w_i$ times. The median of $\{S\prime_j\}$ is equal to the weighted median of $\{S_i\}$ and $\{w_i\}$. Formally, the weighted median of n ordered elements $s_1$, ..., $s_n$ with weights $w_1$, ..., $w_n$ is defined as the smallest element $s_k$ which satisfies:

$$\sum_{i=0}^{k-1} w_i \leq \frac{(\sum w_i)}{2} \tag{4.1}$$

Adjustment of the weights allows for tuning of the balance between noise reduction and edge preservation. In homogeneous regions such as the liver where the N intensity values for a given voxel are likely to differ only due to image noise, it is beneficial to assign approximately uniform weights for maximum noise suppression. Conversely, for regions where the N values may differ prominently due to registration error, such as at the boundary between a blood vessel and parenchymal tissue, it is desirable to weight the central values more strongly in order to increase edge preservation.

We use a modified version of the method proposed by Loupas et al. [58] to determine the set of weights $\{w_i\}$ for a given voxel $\vec{x}$.

$$\begin{aligned} &w_i = w_{central} - cr_i|\nabla(\vec{x})| \text{ if } w_{central} - cr_i|\nabla(\vec{x})| > 0 \\ &w_i = 0 \text{ if } w_{central} - cr_i|\nabla(\vec{x})| \leq 0 \end{aligned} \tag{4.2}$$

$w_i$ is the weight corresponding to the intensity value of image $i$. $w_{central}$ is a fixed central

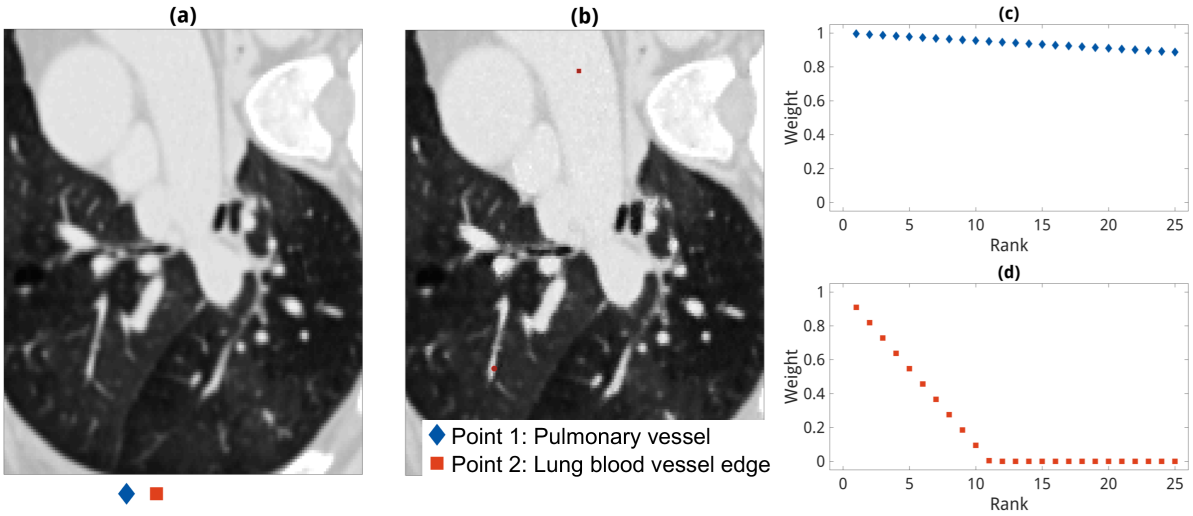


Figure 4.1: (a) Axial slice of image obtained after averaging 25 co-registered scans. (b) AWM filtered image. Two points are marked, the first in a homogeneous region of a pulmonary vessel and the second at the boundary of a blood vessel and parenchymal tissue in the lung. (c) and (d) show the weights for the aorta and vessel, respectively, for each of the 25 intensities at their voxel locations. Intensities are ranked according to local sum of squared differences. A window and level of 1700 and -600 are shown.

weight which is the same for all voxels, $c$ is a tuning constant, $r_i$ is the rank of the image $i$ according to the local difference with the reference image. The $r_i$ term is intended to capture the local registration error and is computed by taking the sum of squared differences over all immediately adjacent voxels to the given voxel, referred to as the 27-neighborhood, between each image and the reference and then ranking the images accordingly. $|\nabla(\vec{x})|$ is the gradient magnitude at location $\vec{x}$ in the reference image. The gradient term allows the filter's smoothing to vary with image content; weights decline more rapidly at boundaries than in homogenous regions. $w_{central}$ and $c$ are set empirically. Figure 4.1 displays the distribution of weights for two such voxels.

### 4.3.4 AWM algorithm

For each voxel $\vec{x}$:

1. Read the 27-neighborhood in the reference image $i = 1$.
2. Compute the gradient magnitude at $\vec{x}$.
3. For each image $i = 2, ..., N$:
   (a) Read the 27-neighborhood in image $i$
   (b) Compute the sum of squared differences (SSD) with the reference 27-neighborhood.
4. Rank images $i = 2, ..., N$ in ascending order according to their local SSD with the reference image.
5. Determine a weight for each of the N intensity values using equation 4.2.
6. Find the weighted median intensity value and write it at $\vec{x}$ in the output image.

### 4.3.5 Parallel implementation

Filtering is performed independently for each voxel therefore the algorithm is straightforward to parallelize. The filter was implemented in CUDA using a Telsa K40 graphics processing unit (NVIDIA, Santa Clara, CA). A MATLAB (MathWorks, Natick, MA) wrapper was written using the MEX functionality.

### 4.3.6 Image quality assessment

Pixel-to-pixel noise of images generated using AWM filtering, median filtering, and averaging was measured by taking the standard deviation of a 20-mm radius circular region of interest (ROI) within a homogenous region of the liver on a single axial slice. Image bluriness was assessed using a technique based on comparing intensity variations between neighboring pixels before and after low-pass filtering [59]. The blur metric described in [59] relies on the assumption that low-pass filtering will degrade edges more in a sharp image than a blurry image. A normalized blur metric value was computed for all axial slices of AWM filtered, median filtered and averaged images. The mean value of all slices was taken to obtain a

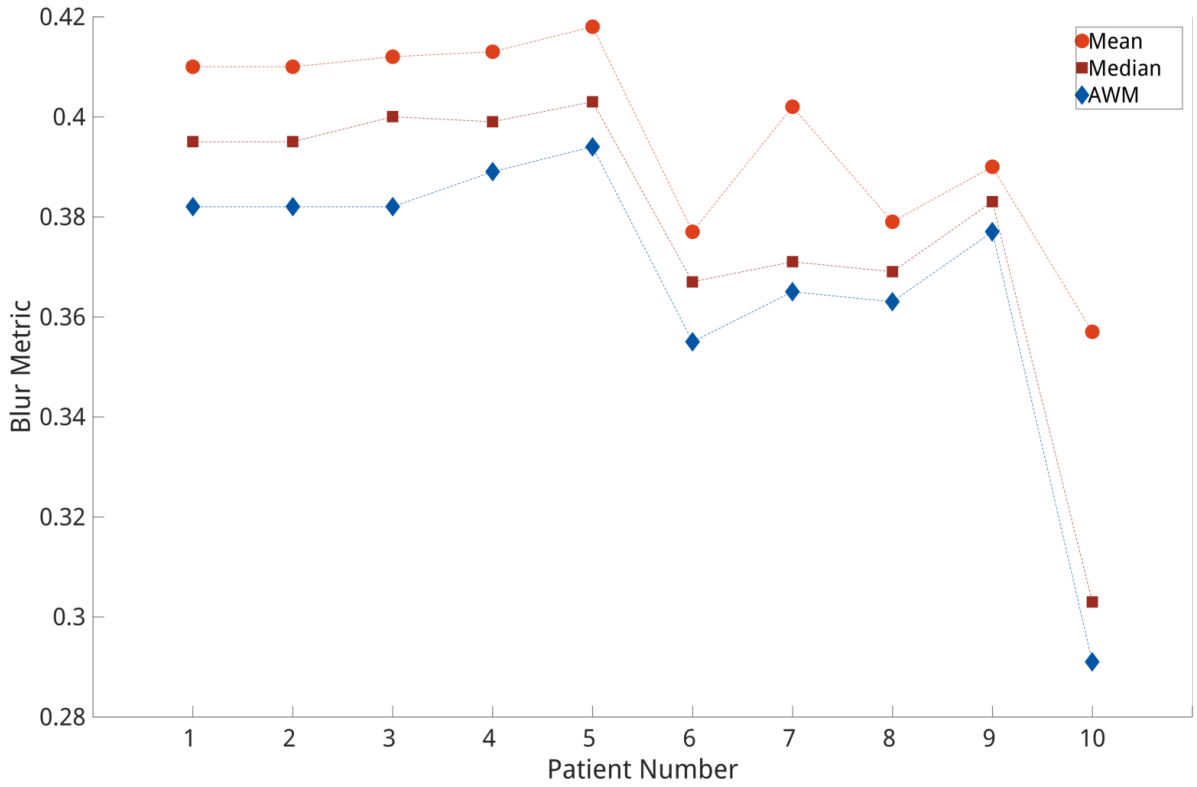


Figure 4.2: Blur metric values after combining 25 co-registered images using mean (circle), median (square) and AWM (diamond) for each patient.

single score per technique per patient. The metric ranged from [0, 1], where 0 indicated the least blur.

## 4.4 Results

Computation time to filter a typical dataset on the GPU, 25 co-registered images with isotropic 1 $mm^3$ voxels and dimensions of 500 x 500 x 425, was 8.4 seconds. Values of 1000 and 0.1 for $w_{central}$ and $c$, respectively, were found to produce a desirable compromise between smoothing and edge preservation in all cases.

Image blurriness was quantified using a technique based on different degrees of blurring perceptible on the single image which was demonstrated in [59] to agree with subjective assessments by human observers. Blur metric values for AWM filtered images were found to be significantly less than both the average ($p = .002$) and median ($p = .002$) images

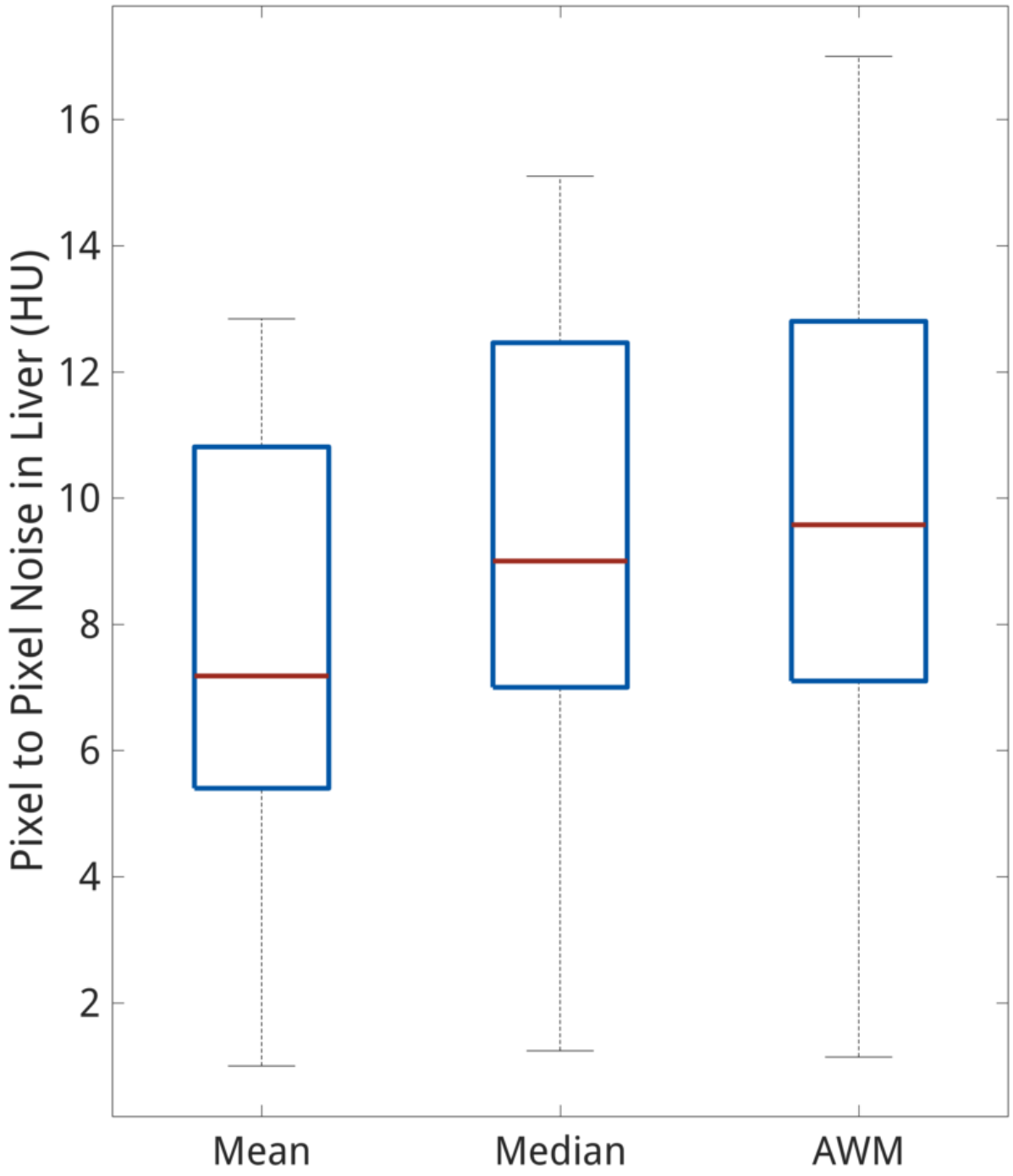


Figure 4.3: Box plot showing the distribution of pixel-to-pixel noise measured in the liver using a 20 mm ROI after image fusion with the three techniques.

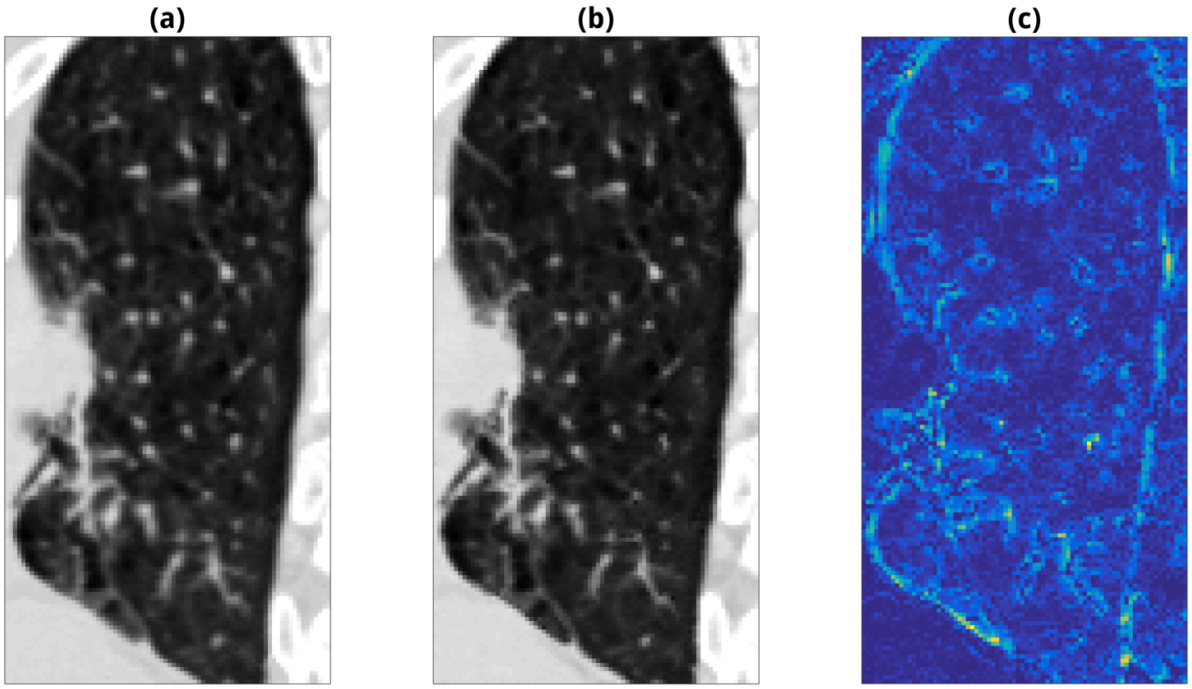


Figure 4.4: Comparison of median averaging and AWM on subsection of a coronal lung slice. (a) shows the median averaged image, (b) the image obtained using AWM and (c) the difference image. In (a) and (b) the window is 1700 and the level is -600.

using the Wilcoxon signed rank test. Values for all patients are given in Figure 4.2. Mean pixel-to-pixel noise in the homogeneous liver was 7.6 HU for the averaged images, 9.3 HU for the median and 9.9 HU for the AWM filtered images. A summary of the image noise measurements is shown in Figure 4.3. For all 10 patients, AWM filtering was able to produce sharper images than mean and median averaging with mean pixel-to-pixel noise in the liver within 3 HU and 1 HU, respectively.

Mean pixel-to-pixel noise in the homogeneous liver was 7.6 HU for the averaged images, 9.3 HU for the median and 9.9 HU for the AWM filtered images. Blur metric values for AWM filtered images were found to be significantly less than both the average ($p = .002$) and median ($p = .002$) images using the Wilcoxon signed rank test.

## 4.5 Discussion and conclusions

We have described a method that uses adaptive weighted median filtering to combine multiple co-registered thoracic CT scans to produce a single low-noise fused image which results in the introduction of less blur than taking the average or median. Our method can be applied to the generation of mid-position images for lung cancer treatment planning and the 5DCT technique. Computation time using the GPU to process 25 images is approximately 8 seconds, which is sufficiently short to employ the method clinically.

AWM filtering allows for a user-selectable compromise between noise reduction and sharpness in the resulting fused image through adjustment of the $w_{central}$ and $c$ parameters. The parameters are sufficiently robust to variance in HU that the same values of $w_{central}$ and $c$ resulted in desirable fused images for the ten patients in this study. Although the AWM filter uses the local image gradient and registration error to determine weights, filtering is only performed over the set of intensities at a given voxel location in each image. Extending the filter over the spatial dimensions of the co-registered images may allow for greater noise reduction, but at higher computational cost and with the potential of introducing image artifacts that are more clinically detrimental than the noise.

Blurriness for images produced by mean and median averaging and AWM filtering was evaluated on individual axial slices, as that is how they would most commonly be viewed while delineating targets and normal structures, and a final score was obtained by averaging the score for each slice. Using AWM, a modest but statistically significant reduction in blurring metric values was observed (7.3% less than using the mean and 2.6% less than the median). The actual metric values were sensitive to the choice of low-pass filter but the relative performance of the techniques was not. Consequently, the blur metric was primarily treated in this study as an indication of which image was blurrier rather than by what degree. Ordering according to blur metric values agreed qualitatively with image data, as shown in the comparisons between AWM and mean averaging in Figure 4.1 and median averaging in Figure 4.4.

The results presented here demonstrate the favorable performance of AWM filtering in

generating a fused image compared to mean or median averaging in terms of blur reduction. One potential issue with AWM filtering is that the non-uniform noise in the resulting fused image could lead to the appearance of spurious edges. While the effect was not observed in this study, it warrants further investigation. Additionally, residual misalignment in poorly registered image regions could lead to the introduction of doubling artifacts at structure boundaries. The $r_i$ term in the weight calculation, which ranks images according to local SSD with the reference image is intended to mitigate this effect. Further study is needed to quantify the impact of the improvement in edge preservation afforded by AWM on target volume and normal structure delineation and to determine the extent, if any, that artifacts which could have a deleterious effect on contouring accuracy arise in the fused image.

AWM filtering offers a straightforward approach to creating a single low-noise fused image dataset from multiple co-registered FHCT images. Similar levels of noise-reduction to mean and median averaging can be obtained with better edge preservation. The amount of smoothing can be controlled on a case-by-case basis, and processing as many as 25 images can be performed sufficiently quickly to be integrated into clinical treatment planning workflows.

## 4.6 Acknowledgement

The Tesla K40 used for this research was donated by the NVIDIA Corporation. This work was supported in part by NIH R01 CA0096679 and funding from Siemens Healthcare.

# CHAPTER 5

# Improving quality and safety of in-house software for translating techniques from research to clinical practice: a case study on a novel model-based 4DCT protocol

Chapter 5 details the safety and quality management aspect of using 5DCT clinically. The chapter focuses on two main areas: software safety and process-oriented quality interventions. The chapter is intended to serve as a case study on how to promote safety in in-house clinical software, and to develop quality management programs for new techniques using established risk analysis methods. Two software engineering techniques are introduced for the software aspect of the project: use cases, and the easy approach to requirements syntax (EARS). The process level quality management is addressed using techniques promoted in the American Association of Physicists in Medicine (AAPM) task group (TG) 100 report, specifically process mapping and failure modes and effects analysis (FMEA).

This chapter is currently being prepared for submission.

## 5.1 Abstract


### 5.1.1 Purpose

In-house software is commonly employed to implement new imaging and therapy techniques before commercial solutions are available. Risk analysis methods, as detailed in the TG-100 report of the AAPM, provide a framework for quality management of processes but offer little guidance on software design. In this work, we examine a novel model-based 4DCT protocol using the TG-100 approach and describe two additional methods for promoting safety of the associated in-house software.

### 5.1.2 Methods

To implement a previously published model-based 4DCT protocol, in-house software was necessary for tasks such as synchronizing a respiratory signal to CT images, deformable image registration (DIR), model parameter fitting, and interfacing with a treatment planning system. A process map was generated detailing the workflow. Failure modes and effects analysis (FMEA) and fault tree analysis (FTA) were performed to identify critical steps and guide quality interventions. Software system safety was addressed through writing use cases, narratives that characterize the behavior of the software, for all major operations to elicit safety requirements. Safety requirements were codified using the easy approach to requirements syntax (EARS) to ensure testability and eliminate ambiguity.

### 5.1.3 Results

61 failure modes were identified and assigned risk priority numbers using FMEA. Resulting quality management interventions include integration a comprehensive reporting and logging system into the software, mandating daily and monthly equipment quality assurance procedures, and a checklist to be completed at image acquisition. Use cases and resulting safety requirements informed the design of needed in-house software as well as a suite of tests performed during the image generation process.

### 5.1.4 Conclusions

TG-100 methods were used to construct a process-level quality management program for a 4DCT imaging protocol. Two supplemental tools from the field of requirements engineering facilitated elicitation and codification of safety requirements that informed the design and testing of inhouse software necessary to implement the protocol. These general tools can be applied to promote safety when in-house software is needed to bring new techniques to the clinic.

## 5.2 Introduction

In-house software is commonly employed at academic cancer centers to implement new imaging and radiation therapy techniques before commercial solutions become available. Unlike the software embedded in medical devices, which must adhere to standards such as International Electrotechnical Commission (IEC) 62304 [60], applicable regulation and formal, professional society-driven guidelines for in-house software remain open questions [61] [62].

Medical physics specific literature on clinical software quality and safety is sparse. Rosen, in [63], provided one of the earliest contributions by presenting several general programming and design guidelines. In particular, Rosen described the importance of determining and adequately documenting specifications to reduce the likelihood of accidents due to unintended behavior, and conducting a hazard analysis during the design phase of development to identify and mitigate potential errors. Similarly, Gersten and Langer [64] published a set of recommendations aimed at the solo developer, or small team with limited resources, typical of in-house software written for radiation oncology clinics. The authors presented a simplified example of a linear accelerator data capture system as a case study.

Salomons and Kelly wrote about the state of software safety standards and practices in the field in [65], pointing out the lack of a dedicated American Association of Physicists in Medicine (AAPM) task group (TG) report on the topic. The authors called for the development of a common set of guidelines, arguing that the safety goals and priorities

inherent to clinical medical physics are different enough from other safety-critical industries to prevent the unconditional adoption of practices from other safety-critical fields such as aviation from being an effective solution.

In [66], Salomons and Kelly detail responses to a survey about in-house software usage and quality assurance practices sent to medical physicists at various Canadian cancer centers. 78% of respondents reported using in-house software for dose calculation quality assurance, and 18% for treatment planning or delivery. All 46 respondents stated that in-house software at their institutions was used by people other than just the developer, illustrating the importance of communicating how the system will behave, particularly in edge cases that may be encountered by other users. Additionally, when asked about the relative importance of software quality assurance activities, testing methodology was found to be of the greatest concern, suggesting that additional tools to guide the testing process would be of interest to the medical physics community.

The recent AAPM TG-100 report [67] presented a comprehensive framework for using risk analysis methodology, namely process mapping, failure modes and effects analysis (FMEA), and fault tree analysis (FTA) for quality management in a general radiotherapy context. These three techniques have been mainstays of safety-critical industries for decades, and multiple studies have detailed the application of TG-100 methods to radiotherapy processes as high dose-rate brachytherapy [68], stereotactic body radiation therapy [69], and linear accelerator quality assurance [70]. A major benefit of the TG-100 paradigm is that it can be used to proactively develop quality management programs for new planning, delivery, and imaging techniques.

In this work, we applied TG-100 methods to a novel model-based 4DCT protocol that was previously used in research studies at our institution as we developed it into a clinical simulation protocol. Similar to many other advances that derive from research, extensive in-house software was necessary to perform our 4DCT technique. TG-100 generally focuses on processes and does not contain specific provisions for software design. To address this shortcoming, we employed two methods used in the software engineering field [71]: use cases [72] and the Easy Approach to Requirements Syntax (EARS) [73]. Use cases are a narrative

formalism for describing how a piece of software should behave at a greater level of detail than commonly found in other forms of documentation such as method summaries or user guides, but are much more acceessible than a line-by-line examination of code. Use cases were beneficial for communicating implementation details between the physicist writing the software and other physicists involved in development in order to reach a consensus on how various conditions, including faults, should be handled, rather than leaving the decision solely up to person writing that section of the software. EARS consists of a ruleset for writing software requirements that is designed to enforce testability and minimize ambiguity, overcomplexity and vagueness. Once elicited through FMEA or use cases, all of our projects safety requirements were specified using EARS templates and formed the basis for test design. The intent of this work is not to prescribe a formula for in-house software development, or to claim that our chosen methods are more effective than alternatives. Rather, we aim to share our institutions experience applying TG-100 methodology to the clinical implementation a novel imaging protocol and in particular how we managed the in-house software quality aspect of the project in the hopes of working towards establishing community-driven standards and best practices.

## 5.3 Methods

### 5.3.1 Case study: 5DCT

In order to avoid the sorting artifacts associated with commercial 4DCT reconstruction [9], [38] and their potential deleterious effects [74] [75] [76] on target volume delineation, our group previously proposed a novel model-based 4DCT protocol, termed 5DCT for its parameterization of lung tissue motion as a function of 5 degrees of freedom: x,y,z position in a reference geometry, breathing amplitude, and breathing rate [24]. The technique has been detailed in previous publications [19], [43]. In brief, multiple free-breathing CT scans with simultaneous breathing surrogate monitoring using an abdominal bellows are acquired to sample the respiratory cycle. Images are registered to a common geometry using deformable image registration (DIR). The deformation vector fields (DVF), along with synchronized

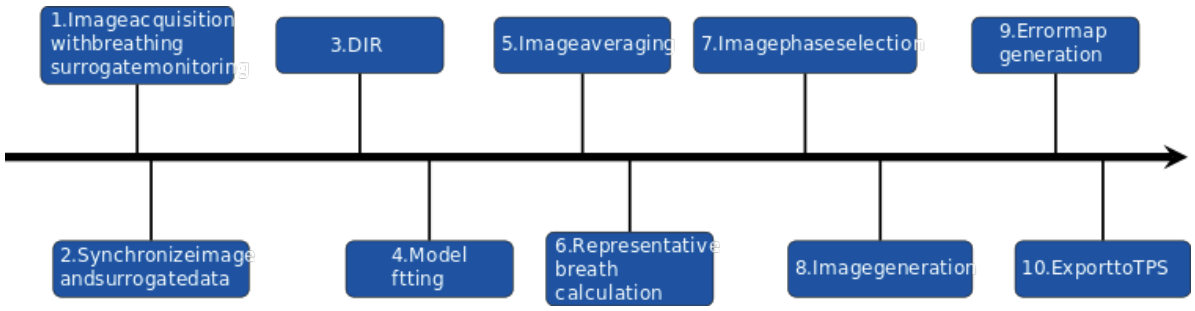


Figure 5.1: Simplified process map showing the primary steps involved in 5DCT image generation.

breathing amplitude and rate measurements, are used to fit patient-specific model parameters that relate amplitude and rate to tissue motion. After parameters are fit, images can be generated at any user-defined breathing state by using the model to calculate a DVF from the reference geometry to the desired state, and applying it to the reference image. Figure 5.1 shows an overview of the workflow necessary to generate a set of breathing gated images using the 5DCT technique for treatment planning. In-house software is employed in some capacity at all stages, from respiratory surrogate monitoring to exporting the images to the treatment planning system (TPS.)

Accuracy of the 5DCT technique was validated against commercial 4DCT in an animal model under controlled breathing conditions [53]. Twenty-two patients were enrolled under an institutional review board-approved study to be scanned using the 5DCT protocol following their simulation with commercial 4DCT. Data from the initial research patient cohort allowed for comparison of tumor motion measured using 5DCT against commercial 4DCT [77], and to refine the workflow and necessary in-house software before clinical use.

### 5.3.2 Process mapping

Firstly, a process map [78], a graphical depiction of the steps required to perform 5DCT, was created in order to summarize the workflow. The process map serves two primary purposes: to concisely describe the process to team members who are not familiar with it, and to clearly define the scope of the project and the quality management program that will be developed.

A simplified version of our process map is shown in figure 5.1. The process is divided into ten major steps, shown as rounded rectangles. The steps proceed in sequence from the left, where free-breathing scans are acquired, to right, where the dataset of gated images are ready for contouring. The map was constructed using LaTeX (LaTeX Project) with the TikZ package. Similar diagrams can be created using PowerPoint (Microsoft, Redmond, WA) or the free and open-source LibreOffice Impress (The Document Foundation, Berlin, Germany).

### 5.3.3 Failure modes and effects analysis

Failure modes and effects analysis (FMEA) is a technique to evaluate the likelihood of failures in a process and their impact on the outcome of that process. For each potential failure mode, numerical values are estimated for the following three parameters: occurrence (O), the likelihood that the failure will happen; severity (S), the degree to which the failure will affect the outcome; lack of detectability (D), the likelihood that the failure will not be detected prior to it having a deleterious impact [67]. Typically, these values range from 1 to 10, with 10 being the most severe. The three parameters are multiplied together to produce a risk priority number (RPN) for each failure mode. Resources can be directed to prevent failure modes according to their rank order of RPN, as well as for the most severe failure modes. The primary benefit of FMEA is as a means of allocating limited resources in the most effective manner.

Our group began by writing a standardized list of effects to categorize the outcomes of all discovered failure modes. The list served as a reference when assigning S values, and reinforced the scope of the analysis. Effects and their associated S values are given in table 5.1. For our case, the scope began at image acquisition and ended after exporting a 4DCT dataset. No distinction in effect was made between errors that would likely be caught in the treatment planning phase and which were more likely to impact delivery; both were treated with the same severity. The rationale behind this decision was to avoid unnecessary assumptions about treatment planning practices and other separate process

in the radiotherapy workflow which may change after the analysis. The reference values suggested for O and D in TG-100 were used, which also ranged from 1-10.

Table 5.1: List of effects and associated severity values used in the FMEA for 5DCT.

| Effect | S value range | Description |
|---|---|---|
| Inaccurate images | 5-8 | Images do not accurately represent patients anatomy, positioning, or respiratory motion. e.g. a 1 mm shift from correct position; breathing amplitude over or under-represented |
| Very inaccurate images | 9-10 | Relevant anatomy not imaged; position very wrong (1 cm shift); anatomy distorted by artifacts; non-physical respiratory motion |
| Non-radiation related physical injury | 5-10 | Patient injured; e.g. fall from CT couch |
| Unnecessary imaging dose | 3-5 | CT scan(s) performed but no viable images produced. |
| Patient inconvenience | 2-3 | Delay or annoyance which impacts patient, e.g. rescheduled scan |
| Staff inconvenience or increased cost | 1-2 | Delay or annoyance/cost increase. e.g. operator must repeat manual portion(s) of running reconstruction software |

### 5.3.4 Use cases

Three basic concepts are essential to the notion of use cases: systems, stakeholders, and actors. Systems are pieces of software at various granularity. Depending on the scope of the use case, the system under discussion can be an entire program, a class, or a subroutine. Stakeholders are individuals or other systems that are directly affected by the behavior of the system under discussion. For example, some stakeholders of a treatment planning system

would be the planner, the patient, and the delivery system. Actors are stakeholders that interact with the system to achieve a goal.

A use case is a collection of actions and events that define an interaction between an actor and a system to achieve a goal. The use case details how the system responds to requests from an actor under various conditions. It characterizes the systems behavior with respect to the interests of the stakeholders. Use cases can be expressed in a text format with natural language, typically using a structured template. The primary purposes of use cases are to provide a framework for discussion about how a system should be designed, and to aid in deciding and documenting what that system must do.

Use cases were initially proposed in [79] and have since seen widespread adoption in the software engineering industry as a tool for the requirements specification process. The basic series of actions that a software system goes through to accomplish a goal is described in the main success scenario section, and alternative flows due to erroneous or uncommon conditions are described in the extensions section. An excerpt from one of our projects use cases is shown in table 5.5. One of the major benefits of use cases is that their simple narrative structure makes them easily understandable to unfamiliar readers and they can be incorporated into team discussions without the need for prior training. Despite their simplicity, use cases are a powerful tool for eliciting safety-related software requirements because writing the extensions encourages what if style thinking about potentially uncommon failure scenarios, which if left unplanned for can combine with hazards to create unsafe conditions [80]. After establishing the extension scenarios and failure handling, the necessary functionality of the system to reduce the likelihood or mitigate the severity of these failures can be codified into a set of safety requirements.

### 5.3.5 Easy approach to requirements syntax (EARS)

Software requirements describe the needs of stakeholders that the system should address, and how the system should generally behave. Frequently, software-related accidents stem from incorrect, or incompletely defined behavior under uncommon conditions rather than

programming errors [81] [82] [83] [84]. Requirements engineering is a broad discipline, and there is extensive literature on how requirements specification relates to system safety [85]. In this work, we are limiting our focus to the projects safety requirements, defined in [86] as the subset of requirements that specify a minimum amount of safety regarding an aspect of the system and a minimum level of an associated metric. Verifiability is key to this definition any safety requirement must be testable in order to demonstrate that the software complies with it.

The easy approach to requirements syntax (EARS) was proposed to address limitations associated with natural language without the training burden of formal methods [73]. The EARS rule set consists of a few simple sentence structures that provide a practical and efficient means to state requirements. Five types of requirements are described: ubiquitous; event-driven; unwanted behavior; state-driven, and optional. Example requirements using these templates are given in table 5.2.

Table 5.2: EARS templates for various requirement types.

| Requirement type | Syntax |
|---|---|
| Ubiquitous | The system shall [response] |
| Event-driven | When [precondition (optional)] [trigger] the system shall [response] |
| Unwanted behavior | If [precondition (optional)] [trigger], then the system shall [response] |
| State-driven | While [specific state] the system shall [response] |
| Optional | Where [feature] the system shall [response] |

The EARS syntax allows for accurate, concise presentation of safety requirements. The requirements are less ambiguous than unstructured language, and most importantly can be verified through testing.

### 5.3.6 Testing safety requirements

In order to demonstrate that the software is safe, tests need to be designed that prove the system will operate correctly in its intended context. The safety requirements elicited

through use cases and codified using EARS form the foundation for testing. The suite of tests should show that the software appropriately detects and handles abnormal operating conditions, and failures [87]. Software testing is a broad topic and an adequate treatment of the discipline is beyond the scope of this manuscript. The interested reader is referred to the seminal text in [88]. In our project, we found it useful to design tests around each safety requirement.

### 5.3.7 Traceability

Traceability refers to the degree to which system requirements can be related to the hazards and faults they are designed to mitigate [89]. Standards for traceability are called for in documents governing standards for multiple industries, such as the Federal Aviation Administrations DO-178C [90], the automotive industry's safety standard ISO 26262:2011 [91]. The degree of complexity necessary to provide adequate traceability varies with the scale of the project, and any relevant certification processes. Tools such as IBMs Rational Dynamic Object Oriented Requirements System, or DOORS (IBM, Armonk, NY) or JAMA (Jama Software, Portland, OR) offer industrial grade solutions for larger projects. In our case, a simple traceability information model was sufficient: all failure modes, use cases, safety requirements, and software tests were assigned unique identifiers. In the safety requirements document, all use cases, failure modes and tests associated with a given requirement were listed so that its rationale and test coverage could quickly and clearly be determined. An example is shown in Table 5.3.

## 5.4 Results

A full process map detailing the 5DCT technique, complete with sub-steps, was constructed. The workflow was divided into primary 10 steps, each consisting of about 5 sub-steps on average.

A total of 61 failure modes were identified through FMEA. Out of the 10 process steps, image acquisition involved the largest number of failure modes at 24. This step is detailed

Table 5.3: Excerpt from requirements specification document. Unique identifiers for requirements, failure modes, use cases and safety tests establish basic traceability. Requirements are given the prefix R followed by the process step they are associated with (R02 for a requirement involving the 2nd step), then enumerated sequentially (e.g. -4). Failure modes, use cases and tests are similarly denoted by the prefixes FM, UC, and T respectively. For all requirements, their motivations and associated tests can be determined at a glance.

| Requirement | ID | Failure mode(s) | Use case(s) | Tested by |
|---|---|---|---|---|
| When importing scans, the study object shall only tolerate a maximum shift of +/- 5 mm from the reference geometry | R02-4 | FM02-8: Error induced by incorrectly managed shift | UC02: Synchronize CT and surrogate data | T02-4: study.validate_scan_alignment |

in figure 5.2. The FMEA was conducted after the cohort of 22 research patients had been imaged using the technique and associated research software. At this point, the clinical version of the in-house software was in the design phase. Although the FMEA informed the design of the software, the majority of interventions resulting from it involved modifications to the workflow or insertion of quality assurance steps.

Example failure modes, and the quality management interventions devised to mitigate them are given in table 5.4. FM01-09, FM01-11, and FM01-14 relate to image acquisition. Two interventions designed to mitigate these and similar failure modes were the establishment of daily and monthly quality assurance for the 5DCT-specific equipment in addition to the CT simulator itself, and a check-list to be completed at each acquisition. FM02-4, error due to associating the wrong surrogate data with a set of images, is mitigated by a safety requirement dictating a minimum level of agreement between abdominal height measured in images and surrogate signal (figure 5.3). FM03-5 involves deformable image registration, and resulted in a software design change: a document displaying the quality of all registrations

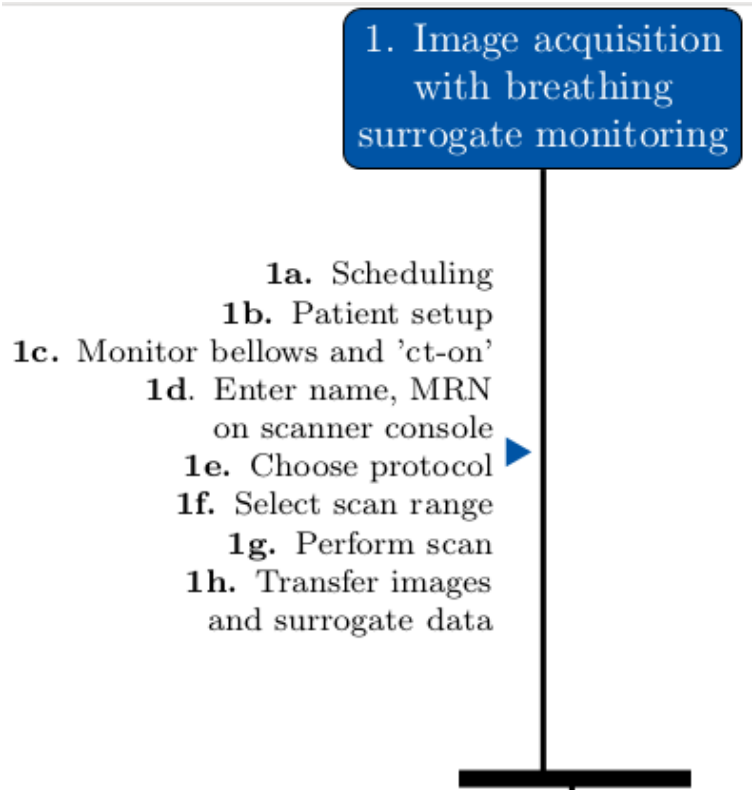


Figure 5.2: Sub-steps involved in image acquisition and breathing surrogate monitoring.

performed is automatically generated and sent out for physicist approval.

Table 5.4: Example failure modes and corrective actions.

| ID | Description | Effect | Corrective action |
|---|---|---|---|
| FM01-09 | No CT on signal due to malfunction in scanner or data acquisition system. | Unnecessary imaging dose | Verify that scanner outputs correct signal and acquisition system is functioning properly in monthly QA. For all patients, verify that CT on signal is recorded when performing initial topogram and mark on acquisition checklist. |
| FM01-11 | Erroneous bellows signal due to leak in vacuum tubing or malfunction in pressure transducer. | Very inaccurate images | Check the transducer response and integrity of tubing using a vacuum gauge during daily QA. |
| FM01-14 | Incorrect patient name or MRN entered on scanner console | Very inaccurate images | Checklist completed at acquisition must require a check of patient name and MRN against CT sim request form. |
| FM02-4 | Mismatch between surrogate data and CT images (i.e. wrong patient error) | Very inaccurate images | Safety requirement R02-3: The study object shall measure a Pearson correlation coefficient of at least 0.85 between measured abdominal height and bellows amplitude. |
| FM03-5 | DIR produces non-physical motion due to incorrect parameters or excessively large displacement between images. | Very inaccurate images | For each registration, capture coronal and sagittal views of an overlay showing the fixed and deformed moving images and compile to a PDF. Physicist must approve document before 5DCT images are used for treatment planning. |

Table 5.5: Main success scenario and list of extensions for the use case describing how the software synchronizes the signal from the abdominal bellows to the CT image data.

| **Main success scenario** |
|---|
| 1. Study object imports the signal data, plots it and prompts user to identify which channels correspond to the bellows, x-ray on and ekg signals. Input is validated. Study object plots the x-ray on signal and prompts user to select a data valid range beginning before the first scans beam on and last scans beam off. |
| 2. Study object detects rising and falling edges of x-ray on signal to identify the start and stop times of each scan. The total number of scans detected is validated against the number of scans input by the user. |
| 3. Study imports DICOM header data and parses it, then validates the number of scans found using the SeriesInstanceUID tag against the number of scans entered by the user. |
| 4. Study object verifies that all slices are contiguous and any scans with missing slices are discarded. The slice locations of each scan are plotted over an AP projection image of the reference scan. Scan range, and whether or not the slice positions are coincident with the reference scan are indicated. The user is prompted to exclude any scans with insufficient coverage. For all remaining scans, a breathing amplitude and rate measurement are assigned to each slice using the DICOM AcquisitionTime tag. |
| 5. Study object prompts user to select a slice location in the abdominal area, then draw a line profile on that axial slice across the edge of the upper abdomen. The abdomen height is detected in all scans by fitting a sigmoid to the intensity values along the line profile and computing the midpoints. The height measurements are correlated to breathing amplitude measurements taken when that slice was scanned. The heights and amplitudes are plotted and the correlation coefficient is displayed. |

6. A Nelder-Mead simplex search is performed to identify a linear drift that when subtracted from the breathing trace yields a maximal correlation coefficient between breathing amplitude and abdominal height. Amplitude and rate measurements associated with each slice are updated using the drift correction.

7. Study object updates its state to synchronized and saves results.

---

**Extensions**

---

1a. Missing or incorrectly formatted bellows file (file I/O error, not enough channels).

1a1. Request new file.

1b. Wrong channels selected.

1b1. Correlation coefficient between abdominal height and breathing amplitude will be low; incorrect number of scan start/stop times will be found.

2a. Number of detected scans does not match number entered by user.

2a1. Synchronization terminates.

3a. Part or all of the DlICOM headers for one or more scans cannot be imported because they are missing or corrupted.

3a1. Flag scans with corrupted files for exclusion from all future process steps. If the minimum number of scans specified in R04-2 is not met, abort synchronization.

3b. Number of detected scans does not match the number entered by user.

3a1. Synchronization terminates.

4a. One or more scans with missing slices are detected.

4a1. Flag scans with missing slices for exclusion from all future process steps. If the minimum number of scans specified in R04-2 is not met, abort synchronization.

4b. Maximum shared scan range does not include all anatomy necessary for treatment planning (ie one or more scans do not have sufficient sup/inf coverage and restrict the maximum shared range)

4b1. User can exclude short scans. If minimum number of remaining scans is below that specified in R04-2, abort.

5a. User selects an undesirable axial slice  bellows, immobilization devices, or other objects prevent accurate measurement of abdominal height.

5a1. User is prompted to confirm their slice selection and may opt to choose a different slice after viewing it.

5b. Abdominal height not accurately determined using sigmoid fitting method.

5b1. Maximum residual for all sigmoid fits must be lower than the value specified in R02-6. If any sigmoid fit exceeds that value, user is prompted to choose a new line profile.

6a. Calculated drift results in a lower correlation coefficient.

6a1. Per R02-7, the drift correction must improve the correlation coefficient to be valid. User must select a new line profile.

6b. Magnitude of drift correction is unrealistically large.

6b1. Calculated drift correction is invalid if its magnitude exceeds the value specified in R02- 8. User must select a new line profile.

6c. Drift correction does not have the expected sign.

6c1. Calculated drift correction is invalid if it does not have the sign specified in R02-9. User must select a new line profile.

5c. Calculated correlation coefficient between the measured abdominal heights and breathing amplitudes is low.

5c1. Correlation coefficient must exceed the value specified in R02-3. User is prompted to select a new line profile and/or axial slice. Synchronization will not proceed until satisfactory correlation is achieved.

7a. Unable to save study object due to I/O error.

7a1. Study object returned to its original state. Synchronization aborted.

---

Table 5.6: Sample safety requirements for synchronization elicited through use cases.

| ID | Requirement |
|---|---|
| R02-1 | When measuring abdominal height in images, the study object shall use slices at the same couch position as the reference scan. If any scans are shifted, the study object shall correct for shifts up to 5 mm using interpolation. |
| R02-3 | The study object shall measure a Pearson correlation coefficient of at least 0.85 between measured abdominal height and bellows amplitude. |
| R02-7 | When calculating a drift correction, the study object shall not allow the drift-adjusted correlation to be lower than the initial correlation. |
| R02-8 | When calculating a drift correction, the study object shall only allow a maximum magnitude of 5e-3 V/s or less. |

## 5.5 Discussion

In this work, we have applied two risk analysis techniques from AAPMs TG-100, process mapping and FMEA, in order to develop a quality management program for the clinical implementation of a novel model-based 4DCT image acquisition and analysis technique. Extensive in-house was necessary to perform the technique. To promote safety of the required in-house software before putting the technique into our clinic, two methods from the software engineering field were employed: use cases to document and reach consensus about how the software should behave under normal and abnormal conditions, and EARS to specify a set of safety requirements in a way that enforces testability and unambiguity. Failure modes, use cases, safety requirements and software tests were assigned unique identifiers in order to facilitate requirements traceability.

It should be noted that we opted not to employ fault tree analysis (FTA) for this project, an additional risk management method described in TG-100. FTA is commonly used in many safety-critical industries and can provide valuable insight by illustrating how component failures lead to system failures in order to reveal the dependability of a system [92]. Two of the main shortcomings of working with fault trees is that they require extensive manual effort

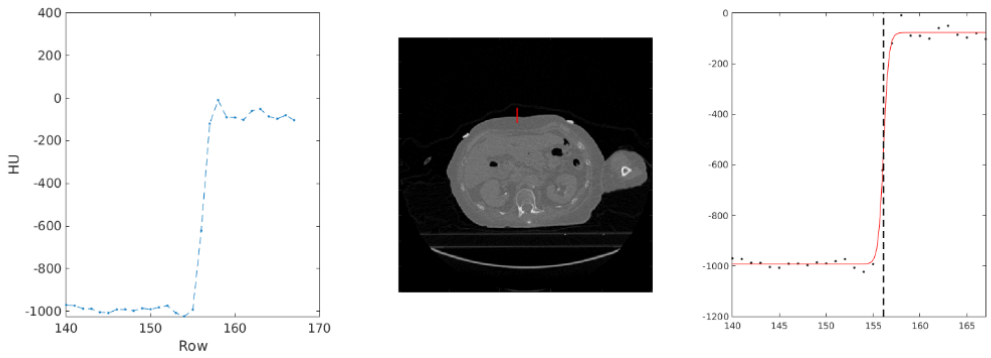


Figure 5.3: Example quality intervention. Abdominal height is measured in each free-breathing scan by taking a profile, fitting a sigmoid to the Hounsfield unit (HU) values and computing the midpoint. A safety requirement dictates that a minimum correlation be measured between the bellows signal and abdominal height to ensure the surrogate measurements are sufficiently accurate for motion modeling.

to generate and maintain, and using graphical models is not necessary preferable to working with text for all team members [93], [94]. Furthermore, it is difficult to create complex fault trees using commonly available office software such as Power Point so it becomes necessary to find and utilize dedicated software. Several open source software packages are freely available [95], but all require a non-trivial amount of initial time to install and use. In contrast, all of the tools used in our analysis aside from the process map were text-based. This facilitated easy collaboration among team members by storing copies on a commonly accessible network drive. Cloud-based services such as Google Docs and Sheets (Google LLC, Mountain View, CA) could also be used for more convenient access. Use cases and EARS were selected for their effectiveness as well as their low barrier to entry.

Assigning unique identifiers to all components of the analysis was straightforward, but updating them as design changes occurred was manual. An extension application to automatically manage the associations between various components either stored locally as Microsoft Word documents or remotely as Google Docs would be valuable. Developing such an extension is a subject of future work.

A weakness of our approach is the neglect of software requirements other than those per-

taining to safety, such as functional requirements that clearly document what the software must do. Our approach is not intended to serve as a replacement for established requirements specifications methodology but rather as a safety-focused supplement to small-scale software projects, likely developed by physicists with no software engineering training, for which formal requirements would otherwise not be written. Similarly, although EARS forces requirements to be written such that they can be tested, the resulting suite of tests should be supplementary and not replace other forms of testing.

There was considerable overlap between the type of thinking encouraged by the FMEA and extension conditions of use cases. However, we found that the FMEA was well suited to handle process-level safety, such as interactions between the user and the software while use cases were most valuable when dealing with workflows within the software, or interactions between subsystems. Employing the two tools together allow for exploiting the unique situation of writing clinical software as a physicist  unlike in the development of commercial software, the physicist can take a holistic view of the workflow surrounding the software and make changes as necessary to promote safety rather than only modifying the design of the software itself.

## 5.6 Conclusion

Two risk analysis methods described in TG-100, process mapping and FMEA, were used to construct a process-level quality management program for the clinical implementation of a novel model-based 4DCT image acquisition and analysis technique. In order to promote safety in the in-house software required to perform the technique, two supplemental tools from the field of requirements engineering, use cases and EARS, were employed to elicit and codify a set of safety requirements that informed the design and testing. These general tools can be applied to promote safety when in-house software is needed to bring new techniques to the clinic.

# CHAPTER 6

# Dependence of subject-specific parameters for a fast helical CT respiratory motion model on breathing rate: an animal study

Chapter 6 explores the dependence of 5D model parameters on breathing rate during acquisition in a porcine model. The 5D model is physiologically motivated – it was formulated based on the hypothesis that lung tissue motion during quiet respiration can be decomposed into two independent functions of breathing amplitude and rate. In contrast to data-driven motion models without physiological underpinnings, where it is only important whether or not the model is accurate given a new set of surrogate measurements, understanding the behavior of parameters in the 5D model with respect to specific changes in the breathing pattern is crucial for developing methods to update model parameters determined at CT simulation for use with surrogate measurements collected at treatment. Specifically, we sought to determine: how stable are parameters in this model from one acquisition to the next? How well can model parameters determined from data acquired at one respiratory rate describe the tissue motion measured at a different rate? We found that the model parameters are largely invariant with respect to changes in the breathing rate. Furthermore, our results support the fundamental hypothesis of the 5D model that tissue motion due to amplitude and rate can be treated independently.

A version of this chapter has been published in Physics in Medicine and Biology: O'Connell, D., et al. Dependence of subject-specific parameters for a fast helical CT respiratory motion model on breathing rate: an animal study. Physics in Medicine and Biology.

## 6.1 Abstract


### 6.1.1 Purpose

To determine if the parameters relating lung tissue displacement to a breathing surrogate signal in a previously published respiratory motion model vary with the rate of breathing during image acquisition.

### 6.1.2 Methods

An anesthetized pig was imaged using multiple fast helical scans to sample the breathing cycle with simultaneous surrogate monitoring. Three datasets were collected while the animal was mechanically ventilated with different respiratory rates: 12 bpm (breaths per minute), 17 bpm, and 24 bpm. Three sets of motion model parameters describing the correspondences between surrogate signals and tissue displacements were determined. Model error was calculated individually for each dataset, as well as for pairs of parameters and surrogate signals from different experiments. The values of one model parameter, a vector field denoted which related tissue displacement to surrogate amplitude, determined for each experiment were compared.

### 6.1.3 Results

The mean model error of the three datasets was 1.00 $\pm$ 0.36 mm with a 95th percentile value of 1.69 mm. Mean error computed from all combinations of parameters and surrogate signals from different datasets was 1.14 $\pm$ 0.42 mm with a 95th percentile of 1.95 mm. Mean difference in over all pairs of experiments was 4.7% $\pm$ 5.4%, and the 95th percentile was 16.8%. Mean angle between pairs of was 5.0 $\pm$ 4.0 degrees, with a 95th percentile of 13.2 mm.

### 6.1.4 Conclusions

Motion model parameters were largely unaffected by changes in the breathing rate during image acquisition. The mean error associated with mismatched sets of parameters and surrogate signals was 0.14 mm greater than the error achieved when using parameters and surrogate signals acquired with the same breathing rate, while maximum respiratory motion was 23.23 mm on average.

## 6.2 Introduction

If left uncompensated, respiratory motion can cause discrepancies between the planned and delivered dose in radiotherapy of lung and abdominal tumors [32] [20]. In order to better ensure that the prescription dose is delivered to the target, breathing motion is commonly managed using four dimensional computed tomography (4DCT) [1]. Commercial 4DCT protocols generally operate by acquiring multiple projections of the same tissues at different points in the breathing cycle and binning the projections according to the amplitude or phase of a simultaneously monitored breathing surrogate signal. Abdominal pressure belts, fiducial markers placed on the skin, and pneumatic bellows are commonly used respiratory surrogate devices [33]. Breathing-gated images are then reconstructed separately from the projections in each bin. While sorting-based methods function well when breathing is reproducible, irregularity in amplitude or period can result in image artifacts [9] [38].

Model-based 4DCT techniques present an alternative approach to generating breathing-gated images by modeling the correspondence between the measure surrogate signal and internal tissue motion [17]. Typically, internal tissue motions are observed using repeated imaging throughout the breathing cycle and deformable image registration to a reference image. Once the parameters describing the underlying correspondence are solved for, images at any breathing phase, specified by a surrogate value, can be generated by calculating the appropriate deformation vector field and using it to warp a reference image. Resulting images are free of the sorting artifacts commonly found in commercial reconstructions.

Accuracy of the correspondence depends on the specific motion model used. Furthermore if the model is to be used for purposes other than target volume delineation, such as tracking or dose accumulation using a surrogate signal measured at treatment, it is important to determine if the parameters accurately reflect the correspondence between external and internal motion if and when the breathing pattern deviates from that during image acquisition [96]. One approach is to continually assess and update the correspondence model during treatment, such as the Synchrony Respiratory Tracking System on the CyberKnife (Accuray Inc., SunnyVale CA) does using implanted markers and orthogonal kV imagers [97]. The correlation model employed by Synchrony is based on data analysis  different types of functions, such as linear and quadratic, are used to relate the marker positions to surrogate signals and the best fit is selected. Similarly, the principal component analysis (PCA) model proposed by Zhang and Li et al [98] [99] uses a set of principal components derived from deformation vector fields to the relate lung tissue motion to the diaphragm position. In both the Synchrony and PCA correspondence models, the parameters (polynomials or principal components, respectively) lack a straightforward physical interpretation. In contrast, the 5D motion model previously published by Low et al [18] is physiological  tissue motion is described a function of position the reference geometry and breathing amplitude and rate. For each voxel, the two free parameters are intended to independently characterize the amplitude and hysteresis components of motion.

It was hypothesized that treating tissue motion due to respiratory amplitude and rate independently is physiologically sound, and the accuracy of the 5D model in patient studies suggests that it effectively models quiet respiration [43] to within 2 mm. However, it order to verify that hypothesis, and to determine if, and to what extent, model parameters need to be updated to reflect specific changes in the breathing pattern, such as a different rate, a series of experiments with carefully controlled breathing dynamics was needed. To that end, we present an examination of the relationship between breathing rate and the 5D model parameters in a ventilated porcine model.

## 6.3 Methods

An overview of the workflow of this study is shown in Figure 6.1. An anesthetized, mechanically ventilated pig was imaged multiple times using the 5DCT acquisition protocol while the animals respiratory rate was systematically varied. Each 5DCT acquisition consisted of 30 consecutive fast helical scans with simultaneous respiratory monitoring using an abdominal bellows to provide a breathing surrogate signal. For clarity, each set of 30 scans will be referred to as an experiment. A different respiratory rate was used for each experiment.

After image reconstruction using the scanner software, all scans were deformably registered to an end-exhalation breath-hold image, which will be referred to as the reference image. The resulting deformation vector fields (DVF) described tissue displacement from the reference image geometry to all scans. Resulting DVFs and breathing surrogate measurements were used to fit 3 distinct sets of motion model parameters  one set for each experiment. The model parameters related tissue displacement to the breathing surrogate signal. Given a respiratory phase, specified by a breathing surrogate signal amplitude and rate, a set of model parameters could be used to calculate the voxel-specific displacements to warp the reference image to that phase.

Model error was evaluated for each set of parameters by comparing model-predicted tissue positions to positions given by deformable image registration (DIR). Variation in the parameters between experiments was investigated, as well as the accuracy of using parameters fit from one experiment to predict the tissue positions observed in a different experiment. For each experiment, three error values were computed: an intra-model error, and two inter-model error values. The intra-model error for a given experiment was found through comparison to to that experiments DIR results. The inter-model error values were calculated by using each set of parameters to predict the tissue displacements of the two other independent experiments given only the breathing surrogate signal measurements.

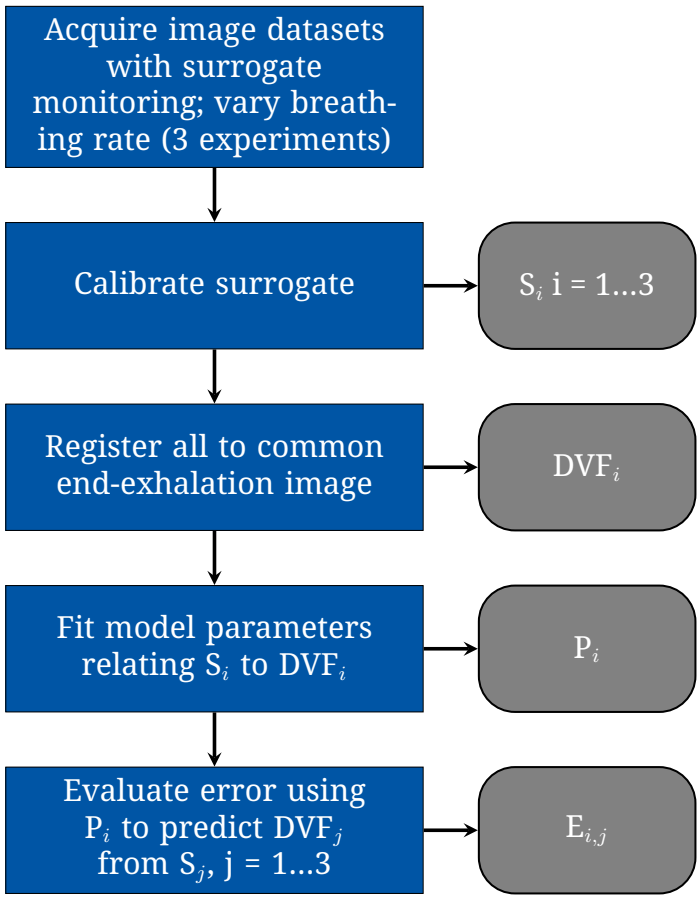


Figure 6.1: Flowchart describing the steps performed in the data acquisition and analysis for this study and the quantities computed. $S_i$: the calibrated surrogate signal data for experiment $i$. $DVF_i$: the deformation fields obtained by registering the helical scans acquired in experiment $i$ to the reference image. $P_i$: Model parameters fit to relate tissue displacement observed in $DVF_i$ to $S_i$. $E_{i,j}$: Model error, the difference in tissue displacements predicted by $P_i$ and $S_j$ and the displacements observed in $DVF_j$.

### 6.3.1 Animal model

This work involves a retrospective analysis of data acquired during a previous animal study, described in full in [53]. All imaging experiments described in this study were conducted using a single wild-type pig (42.7 kg; male). The animal was sedated using a mixture of telazol, ketamine, and xylazine before anesthetization with 3% - 5% delivery of isoflurane

through nose cone inhalation. A tracheostomy was done in order to intubate the animal with a 7.5-mm diameter cuffed endotracheal tube. A volume-controlled ventilator (Model 613; Harvard Apparatus Bioscience Company, Holliston, MA) was used to mechanically ventilate the animal with 100% oxygen. All procedures were conducted with prior approval from the Institutional Animal Care and Use Committee of University of Iowa.

### 6.3.2 Image acquisition

Three experiments, each consisting of 30 consecutive fast helical acquisitions (0.285 s rotation time, 1.2 spiral pitch factor) on a multidetector row CT scanner (Siemens Definition Flash; Siemens Healthcare, Forschhiem, Germany) were performed as the animal was ventilated with respiratory rates of: 12, 17 and 24 breaths per minute. An approximate tidal volume of 450 mL was used for all experiments. Additionally, a single end-exhalation breath-hold image was acquired. Images were reconstructed using the Siemens B30f medium smoothing kernel with a slice thickness of 1.0 mm, and an in-plane resolution of 0.53 mm.

### 6.3.3 Surrogate calibration

A pneumatic abdominal bellows (Lafayette Instrument, Lafayette, IN) and a pressure-to-voltage transducer were used to monitor the breathing cycle during all scans. The surrogate signal was synchronized to the image data using the x-ray on signal from the CT scanner, and each axial image slice was assigned a breathing amplitude and a breathing rate measurement.

It has previously been shown that the bellows was linearly proportional to tidal volume [100], and that this linear relationship remained constant over the course of a single acquisition [53] [101]. However, a time-dependent drift [100] of the bellows signal along with differences in belt tension and positioning between experiments necessitated that a calibration be performed in order to obtain consistent of surrogate values.

For each scan, an image slice at a fixed couch position was used to measure the abdomen height radiographically. The animal was not moved between experiments, therefore abdominal height at a fixed position provided a common quantity to calibrate the bellows signals

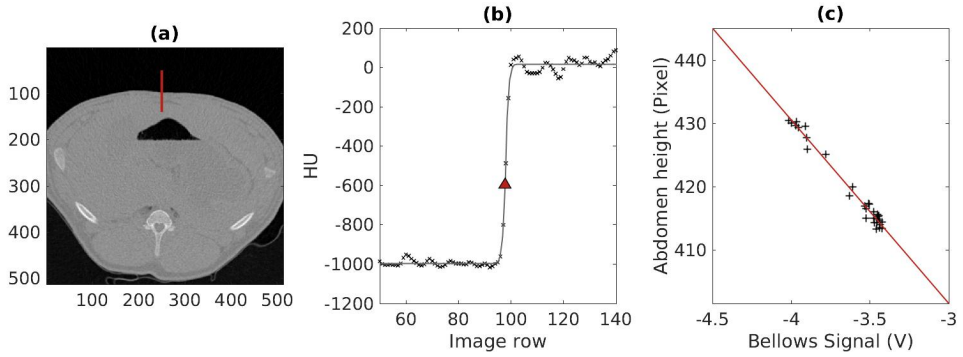


Figure 6.2: Abdominal height measurement and fitting. (a) An axial slice located approximately 1 cm superior to the bellows. Line along which HU values were sampled is shown. (b) Plot of HU values as a function of image row. Sampled values are marked by 'x,' with sigmoid fit. Triangle marks the midpoint of the sigmoid, taken as the location of the boundary. (c) The measured bellows values corresponding to the couch position shown in (a) for each of the 30 scans comprising this experiment plotted against measured abdominal heights with least-squares linear fit.

against. Figure 6.2 demonstrates the procedure used to determine the abdominal height to sub-pixel precision: a vertical line profile was taken at the abdomen/air boundary and a sigmoidal error function was fit to the Hounsfield unit values along the profile. The midpoint of the sigmoid was calculated. The fitting process was performed four additional times at adjacent image locations. The median midpoint was taken as the abdominal height.

Because each experiment consisted of 30 scans, the relevant slice location was imaged 30 times, and each image had a corresponding bellows measurement. Separate linear regressions were performed for each experiment to correlate the bellows data to abdominal height. Using the resulting fits, all bellows values were converted to abdominal height to yield a consistent surrogate.

### 6.3.4 Model fitting

The deeds (Heinrich et al. 2013) DIR algorithm was used to register all images to the end-exhalation breath-hold reference image, and resulting DVFs provided measurements of the

tissue displacement from the reference geometry to each scan. The previously published 5D motion model (Low et al. 2013) was used to describe the correspondence between tissue displacement and breathing amplitude and rate:

$$\mathbf{X}(v, f, \mathbf{X}_0) = \mathbf{X}_0 + \boldsymbol{\alpha}(\mathbf{X}_0)v + \boldsymbol{\beta}(\mathbf{X}_0)f \tag{6.1}$$

Displaced tissue location,$\mathbf{X}$, is modeled as a function of position in the reference image,$\mathbf{X}_0$, breathing amplitude $v$ and rate $f$. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and are voxel-specific parameters which relate the displacement of each voxel location to amplitude and rate, respectively. Collectively, the vector fields $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ will be referred to as model parameters. The 30 amplitude and rate measurements and corresponding displacement vectors formed an over-determined system for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and model parameters were solved for on a graphics processing unit (Tesla K40; NVIDIA, Santa Clara, CA) using a batched cuBLAS QR decomposition routine. Three sets of model parameters were computed, one set for each experiment.

### 6.3.5 Intra-model error

Individual model error for each experiment was computed by comparing the tissue displacements predicted by each set of parameters and breathing surrogate measurements to the displacements given by DIR. The mean Euclidean difference in displacement was computed for all voxels. A single figure-of-merit was obtained by averaging the mean error values of all lung voxels displaced by 5 mm or more between end-exhalation and full inspiration. The error values corresponding to less mobile tissues such as regions of the upper lung, which would intuitively be very small, were not included to avoid biasing the error evaluation.

### 6.3.6 Paramter similarity and inter-model error

Out of the three experiments, each pair of model parameters was compared: $\boldsymbol{\alpha}_i$ to $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_i$ to $\boldsymbol{\beta}_j$ , where $i, j = 1...3$ and $i \neq j$. The differences in magnitudes, as well as the angles between corresponding vectors were found.

Model error was computed using parameters from one experiment and surrogate measurements from another to predict tissue displacements and the results compared to DIR given surrogate signal, how well can parameters $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ predict the tissue displacements observed in $DVF_j$?

## 6.4 Results

Table 6.1: Error, defined as the difference between model-predicted tissue displacements and displacements measured using DIR, for combinations of parameters and surrogate measurements. Bold denotes error using the parameters and surrogate from the same experiment. Experiments are labeled by their approximate respiratory rate in breaths per minute (bpm).

| | | **Surrogate measurements** | | |
|---|---|---|---|---|
| **Model parameters** | | 12 bpm | 17 bpm | 24 bpm |
| 12 bpm | Mean ± std | **1.00 ± 0.32** | 1.04 ± 0.31 | 1.01 ± 0.34 |
| | 95th | **1.59** | 1.59 | 1.72 |
| 17 bpm | Mean ± std | 1.13 ± 0.41 | **1.00 ± 0.37** | 1.03 ± 0.39 |
| | 95th | 1.88 | **1.69** | 1.76 |
| 24 bpm | Mean ± std | 1.36 ± 0.52 | 1.17 ± 0.42 | **1.00 ± 0.39** |
| | 95th | 2.43 | 2.05 | **1.79** |

A summary of statistics on the modeling error for all combinations of parameters and surrogate measurements is given in Table 6.1. Figure 6.3a displays the distribution of all intra- and inter-model error values. Mean intra-model error was 1.00 ± 0.36 mm with a 95th percentile value of 1.69 mm, while inter-model error had a mean value of 1.14 ± 0.42 mm with a 95th percentile of 1.95 mm. The distribution of intra-model error on a coronal slice for experiment 2, which used a respiratory rate of 17.1 breaths per minute, is shown in Figure 6.4a. Figure 6.4b shows the error when using parameters fit from the experiment 2 data to predict tissue positions in experiment 3, acquired at 24.1 bpm, using only the surrogate measurements. The relative distribution of error is similar in both cases, with the

highest error at the lower lung, where motion is greatest, and near the heart, but the average magnitude is slightly larger in 6.4b.

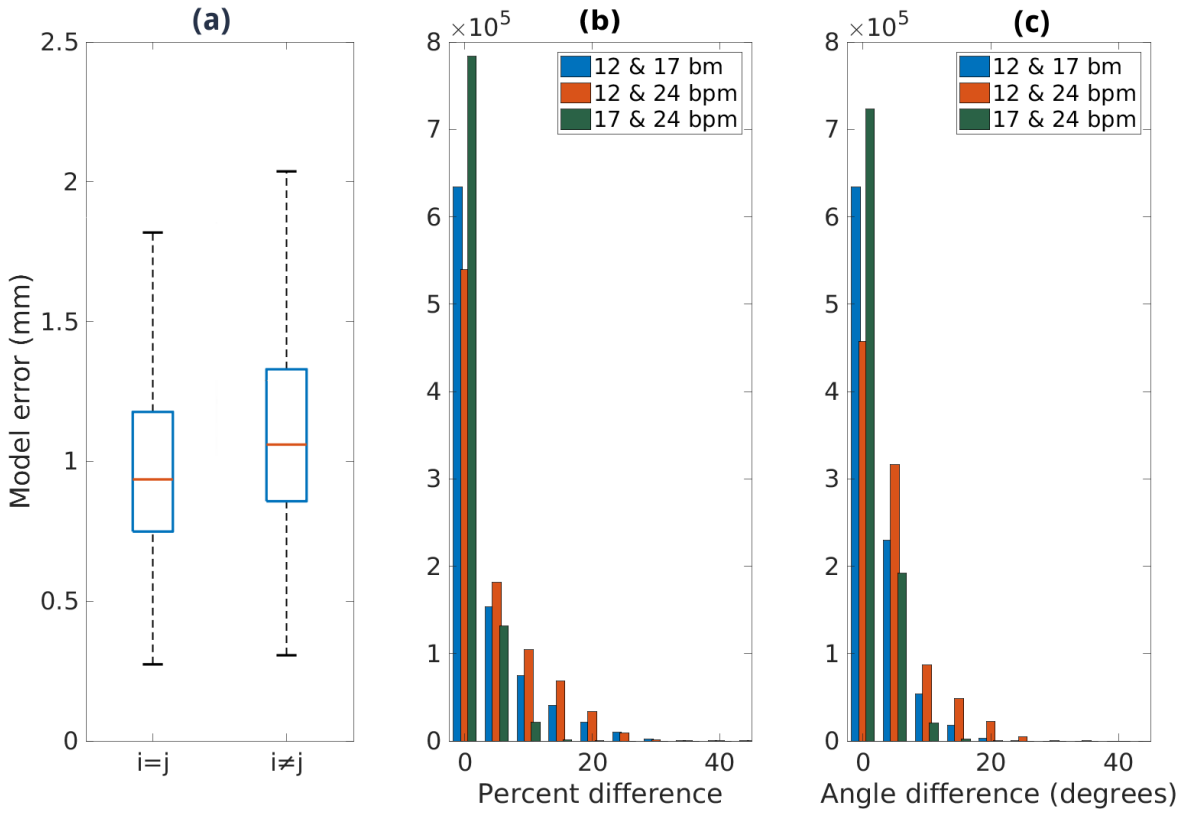


Figure 6.3: (a) Box plots showing the distribution of error for all lung voxels (with motion > 5 mm) for all experiments when using parameter sets $i = 1..3$, and surrogate measurements $j = 1..3$, to predict tissue displacements. Left box plot shows distribution when $i = j$, where parameters and surrogate measurements were from the same experiment. Right box plot shows distribution when $i \neq j$. (b) Histogram of the difference in magnitude, expressed in percent, between values of $\boldsymbol{\alpha}$ calculated for different experiments. (c) Histogram of the angle between pairs of $\boldsymbol{\alpha}$ values for different experiments.

## 6.5 Discussion

The results presented here indicate that the parameters of the 5D motion model are largely unaffected by changes in respiratory rate during image acquisition. For the three exper-

Table 6.2: First row: difference in magnitude, expressed in percent, between the values of $\boldsymbol{\alpha}$ obtained during model fitting for each pair of experiments. Second row: the smallest angle between $\boldsymbol{\alpha}$ values from pairs of experiments, expressed in degrees.

| | | **Respriatory rates (bpm)** | | |
|---|---|---|---|---|
| | | 12 & 17 | 12 & 24 | 17 & 24 |
| **Magnitude difference (%)** | Mean ± std | 5.1 ± 5.9 | 6.3 ± 6.2 | 2.7 ± 2.6 |
| | 95th | 18.1 | 19.7 | 8.4 |
| **Angle (°)** | Mean ± std | 4.8 ± 3.3 | 6.5 ± 5.0 | 3.6 ± 2.6 |
| | 95th | 12.2 | 17.9 | 8.7 |

iments performed, using model parameters determined from data acquired at a different breathing rate resulted in an increase in mean error of 0.14 mm in predicting tissue motion. Figure 6.4c illustrates the effect of using two different sets of model parameters to generate a breathing-gated image at the same phase. Parameters from experiments 2 and 3 were both used to compute the deformation from the end-exhalation reference geometry to full inspiration. The resulting DVFs were applied to the reference image, and both images are shown overlaid. Minor disagreement is present at blood vessels in the right lung and around the heart, potentially due to uncompensated cardiac motion. The diaphragm position is closely matched in both images.

The model parameter $\boldsymbol{\alpha}$, a vector field that relates breathing amplitude to tissue displacement, is largely invariant, to within 5% magnitude and 5 degrees on average, with respect to changes in respiratory rate during image acquisition. That $\boldsymbol{\alpha}$ does not depend on breathing rate is a fundamental hypothesis of the 5D motion model [24], but it has never before been experimentally verified.

Another fundamental hypothesis of the 5D model is that hysteresis motion, behavior causing lung tissues to move along different paths during inhaling and exhaling [102] [103], is caused by the pressure imbalances throughout the lungs. In the case of quiet respiration, these imbalances are hypothesized to be proportional to breathing rate. Therefore, the $\boldsymbol{\beta}$

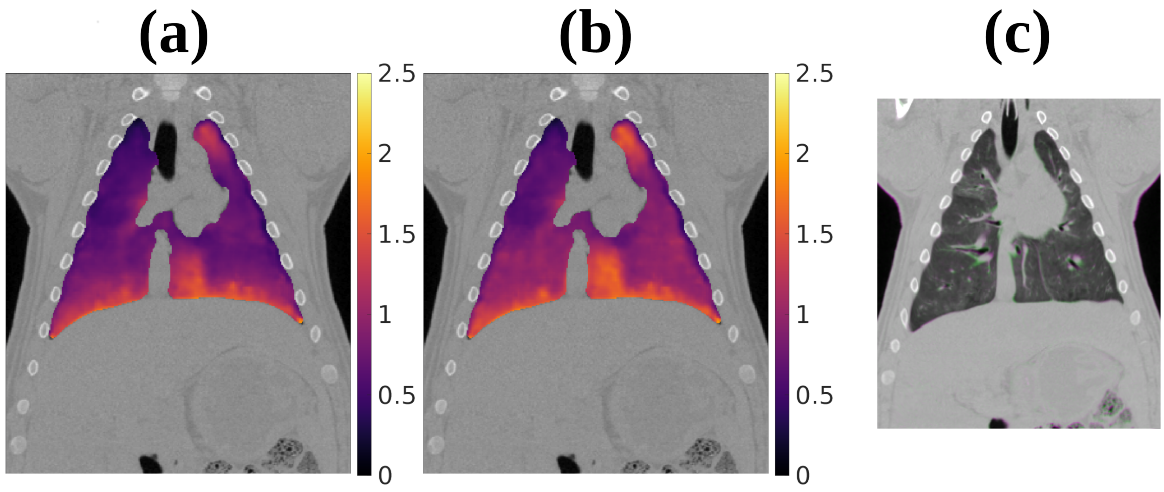


Figure 6.4: (a) Model error (in mm) for experiment 2, which used a respiratory rate of 17 breaths per minute. (b) Error (in mm) using the parameters from experiment 2 to predict the tissue positions observed in experiment 3, given the breathing surrogate measurements for experiment 3. (c) Overlay showing two full inspiration images, created by using the parameters from experiments 2 and 3 to compute the deformation from the end-exhalation reference image to the 95th percentile abdominal height and applying the deformation to the reference image. Green and purple indicate regions of disagreement.

parameter, which relates respiratory rate to tissue displacement, was devised to characterize hysteresis motion. In this study, the small magnitudes of $\beta$ that were computed in all experiments suggest that hysteresis motion was negligible. Remmert et al [104] observed substantial hysteresis using 4D magnetic resonance imaging motion with a dynamic porcine lung phantom. However, the phantom the investigators used, described in [105], employed a water pump and an inflatable balloon to mimic the action of the diaphragm. A possible explanation for the lack of hysteresis motion observed in this study is due to the use of positive pressure mechanical ventilation to drive airflow rather than contraction of the diaphragm or some analogue. Hysteresis is due to the collapse of small airways and surface tension at the alveoli gas-liquid interface that require additional energy to re-inflate the lungs after exhalation to functional residual capacity [106]. With mechanical ventilation, the positive end expiratory pressure prevents mitigates the influence of these factors. The low model

errors in all cases suggest that the measured tissue motion can be almost entirely described by amplitude, and that the small magnitudes of $\boldsymbol{\beta}$ accurately reflect a lack of hysteresis motion rather than a systematic insensitivity of the 5D model. A limitation of the study is that the absence of hysteresis motion in the data precludes meaningful comparison of $\boldsymbol{\beta}$ values between experiments.

A porcine model was necessary in this study because of the radiation dose associated with repeated imaging, as well as the need for closely controlled respiratory rates. While porcine models have been used extensively in the literature in conjunction with CT imaging as approximations of human lung physiology [107] [108] [109], further study is needed to verify that the rate independence of $\boldsymbol{\alpha}$ extends to free-breathing humans.

Periodic breathing patterns were used all experiments in order to ensure that the respiratory rate was consistent within each experiment, and the same maximum amplitude was used in all experiments. The maximum amplitude selected was sufficiently large to encompass tidal volumes typical of quiet respiration. Respiratory waveforms in this study were not intended to mimic those of real patients, where variations in amplitude from one breath to the next are typical. Because of the sampling method 5DCT employs, artificially introducing irregularity into the respiratory waveforms through amplitude modulation used would not have had any effect on the model parameters; the parameter $\boldsymbol{\alpha}$ for each voxel is determined by multiple observations of that voxels position at different amplitudes. Breath-to-breath variation in maximum amplitude do not change the relationship between the surrogate signal and tissue position, which $\boldsymbol{\alpha}$ represents, therefore periodic breathing patterns were sufficient for examining the variation in $\boldsymbol{\alpha}$ with breathing rate.

The 5D model was physiologically motivated; it was formulated based on the hypothesis that lung tissue motion during quiet respiration can be decomposed into two independent functions of breathing amplitude and breathing rate. The results presented here support that hypothesis, but further investigation is needed. In particular, future work will focus on examining the relationship between the parameter and changes in hysteresis motion. Understanding the behavior of parameters in the 5D model with respect to specific changes in the breathing pattern is crucial for developing methods to update the patient-specific model

parameters determined at CT simulation for use with surrogate measurements collected at treatment, and for determining if model parameters are representative of underlying biomechanical tissue properties.

## 6.6 Conclusions

We have demonstrated that in a porcine model the relationship between tissue motion and breathing amplitude, described by the vector field $\boldsymbol{\alpha}$ in the 5D motion model, does not depend on the breathing rate. Model parameters fit using data acquired at one breathing rate can be used to predict tissue positions observed at a different rate from surrogate signal values to within 1.14 mm on average, compared to the 1.00 mm average achieved using parameters determined from the same acquisition.

## 6.7 Acknowledgment

This work was supported in part by funding from Siemens Healthcare. The Tesla K40 used in this work was donated by the NVIDIA Corporation.

# CHAPTER 7

# A prospective gating method to acquire a diverse set of free-breathing CT images for model-based 4DCT

In Chapter 7, we introduce a method to trigger the acquisition of free-breathing helical CT scans based on a respiratory surrogate signal in such a way as to collect a set of images at substantially different breathing phases. The motivation behind the proposed technique is to reduce the number of scans, and consequently the imaging dose, required for accurate respiratory motion characterization for model-based 4DCT. We performed a simulation study to validate our method, and demonstrated that adequate sampling of the respiratory cycle could be achieved in 6 scans, while the current non-prospective protocol uses 25. The achieved reduction in required number of scans, in combination with low-dose free-breathing CT protocols could allow for application of model-based 4DCT to areas other than radiation therapy.

A version of this chapter has been published in Physics in Medicine and Biology: O'Connell, D., Ruan, D., Thomas, D.H., Dou, T.H., Lewis, J.H., Santhanam, A.P., Lee, P.P. and Low, D.A., 2018. A prospective gating method to acquire a diverse set of free-breathing CT images for model-based 4DCT. Physics in Medicine and Biology.

## 7.1 Abstract

Breathing motion modeling requires observation of tissues at sufficiently distinct respiratory states for proper 4D characterization. This work proposes a method to improve sampling of the breathing cycle with limited imaging dose. We designed and tested a prospective free-breathing acquisition protocol with a simulation using datasets from 5 patients imaged with a model-based 4DCT technique. Each dataset contained 25 free-breathing fast helical CT scans with simultaneous breathing surrogate measurements. Tissue displacements were measured using deformable image registration. A correspondence model related tissue displacement to the surrogate. Model residual was computed by comparing predicted displacements to image registration results. To determine a stopping criteria for the prospective protocol, i.e. when the breathing cycle had been sufficiently sampled, subsets of N scans where $5 \leq N \leq 9$ were used to fit reduced models for each patient. A previously published metric was employed to describe the phase coverage, or 'spread,' of the respiratory trajectories of each subset. Minimum phase coverage necessary to achieve mean model residual within 0.5 mm of the full 25-scan model was determined and used as the stopping criteria. Using the patient breathing traces, a prospective acquisition protocol was simulated. In all patients, phase coverage greater than the threshold necessary for model accuracy within 0.5mm of the 25 scan model was achieved in 6 or fewer scans. The prospectively selected respiratory trajectories ranked in the $(97.5 \pm 4.2)$th percentile among subsets of the originally sampled scans on average. Simulation results suggestthat the proposed prospective method provides an effective means to sample the breathing cycle with limited free-breathing scans. One application of the method is to reduce the imaging dose of a previously published model-based 4DCT protocol to 25% of its original value while achieving mean model residual within 0.5 mm.

## 7.2 Introduction

Four dimensional computed tomography (4DCT), also referred to as respiratory- correlated CT, utilizes multiple projections of the same tissues acquired at different points in the breathing cycle to reconstruct a series of breathing-gated volumetric images [33]. 4DCT was developed to improve the accuracy of therapeutic radiation delivery to tumors in the lung and abdomen, which can be displaced by as much as 3 centimeters during treatment due to breathing motion [110] [1].

4DCT has seen widespread adoption in the field of radiation oncology as a tool to manage breathing motion in treatment planning; implementations are offered by the major scanner manufacturers and 4DCT datasets are recognized by many treatment planning systems. It has also shown promise as a functional imaging technique for measuring pulmonary ventilation [111] [112] [113], and more recently as a means to estimate the elastic properties of lung tissue for bio-mechanical modeling [114]. Much of the literature on ventilation imaging and lung elastography relates to the context ofradiation therapy  both techniques are generally proposed as a means to identify healthy lung tissue for preferential sparing in treatment planning. Outside of radiation therapy, 4DCT could potentially be employed to characterize lung function for the phenotyping and response monitoring of chronic obstructive pulmonary disease (COPD) [115] [116], a progressive lung disease which is the third leading cause of death in the United States [117]. COPD is typically diagnosed by measuring the ratio of forced expiratory volume in one second to the full vital capacity (FEV1/FVC). FEV1/FVC is unable to provide information about the physiology causing impaired respiratory dynamics [118]. Single photon emission computed tomography (SPECT) ventilation perfusion scintigraphy (V/Q) is the current gold standard for lung functional imaging, but 4DCT is less expensive, more widely available, and has superior spatial resolution. In most cases, 4DCT effectively characterizes respiratory induced motion, however, commercial implementations suffer from two major limitations: susceptibility to image artifacts if the patient breathes irregularly during acquisition [8], and a high imaging dose  up to four times higher than standard thoracic CT [119]. Multiple strategies have been proposed to reduce 4DCT artifacts

and imaging dose. Castillo et al demonstrated that oversampling led to the fewer artifacts in cine 4DCT, but imaging dose increased [38]. In [120], the authors developed a temporal non-local means approach to reconstruct gated images using fewer projections and acquired with low mAs, howevertheir method does not treat inter-cycle respiratory variability. Pan et al developed a prospective technique to mitigate artifacts in cine 4DCT by stopping irradiation when irregular breathing occurs and selectively re-acquiring images only at the necessary couch positions [121]. Although their method achieved reduced imaging dose compared to repeating an entire scan if severe artifacts occur, the current standard clinical practice, it does not provide a means to lower imaging dose below that of a typical 4DCT.

Respiratory motion model-based 4DCT techniques have been developed that mitigate or avoid the artifacts commonly seen in commercial 4DCT protocols [17]. In contrast to commercial protocols, which generally function by retrospectively sorting projections into bins for independent reconstructions, model- based methods aim to describe the correspondence between internal tissue displacements, typically found using deformable image registration, and an externally measurable breathing surrogate signal using a motion model. Examples of previously published models include polynomial functions of one or more surrogate signals [24], B-splines [122], and principal components of deformation vector fields [99] [123]. More recent developments in respiratory motion modeling include work by McClelland et al., [124], where the authors present a unified framework for image registration and model parameter estimation, and the kernel ridge regression approach proposed by Geimer et al., [125] that used abdominal surface motion as a multidimensional breathing surrogate signal.Our group previously introduced a model-based technique, termed 5DCT, which accurately characterizes tumor motion in the presence of irregular breathing [18] [19]. Our technique has been demonstrated to be accurate to within 2 mm using a landmark analysis [53], and to produce sorting artifact-free images which accurately describe tissue motion to within 2 mm on average in a 16 patient cohort [43].

Model-based 4DCT techniques require images that capture different respiratory states. Unlike other techniques described in [17] that use commercial 4DCT or cine protocols to obtain image data, 5DCT uses repeated fast helical scanning as the patient breathes freely

to sample the respiratory cycle. Although the free breathing images do not correspond to a single respiratory phase as a commercial 4DCT reconstruction would, each individual axial slice can be assigned a breathing amplitude and rate measurement, given by the bellows signal and its derivative at that slices acquisition time. Blurring and doubling artifacts are mitigated by using a fast helical protocol. Repeated scans are performed to collect observations throughout the breathing cycle.

The previously published research protocol acquires 25 scans in succession. Each scan uses a low mA and fast rotation time technique such that the total dose of all 25 scans is similar to that of a commercial 4DCT scan. Although the current protocol ensures a high probability that the breathing cycle will be adequately sampled, redundant image data is acquired because the scans are not selected prospectively. In earlier work, we determined that using unsynchronized scanning required at least 10 scans to guarantee acquisition of images with sufficient variation in the breathing phase to reliably generate accurate motion models [101]. Lowering the number of free-breathing scans would allow for a reduction in imaging dose beyond what is currently achievable with commercial 4DCT. While Thomas et al demonstrated that the number of scans could be reduced without substantially increasing the model residual if the scans were acquired at sufficiently diverse respiratory phases, no method was proposed to select such a set of scans. Ruan et al introduced an objective function to quantify the phase coverage of a set of breathing surrogate signals measured during each scan, which will subsequently be referred to as respiratory trajectories [126]. In this work, we present a novel adaptive method to prospectively select a desirable set of 5 to 10 trajectories by triggering image acquisition based on real-time analysis of the breathing surrogate signal using that objective function. We hypothesize that using the proposed method, the number of scans can be reliably reduced while maintaining acceptable model accuracy. The method was tested by simulating image acquisition using the measured breathing traces of five patients and evaluating the phase coverage of the sampled respiratory trajectories. Although the analysis in this work focuses on 5DCT, the proposed technique could be used in conjunction with any respiratory motion model that can utilize free-breathing CT image data.

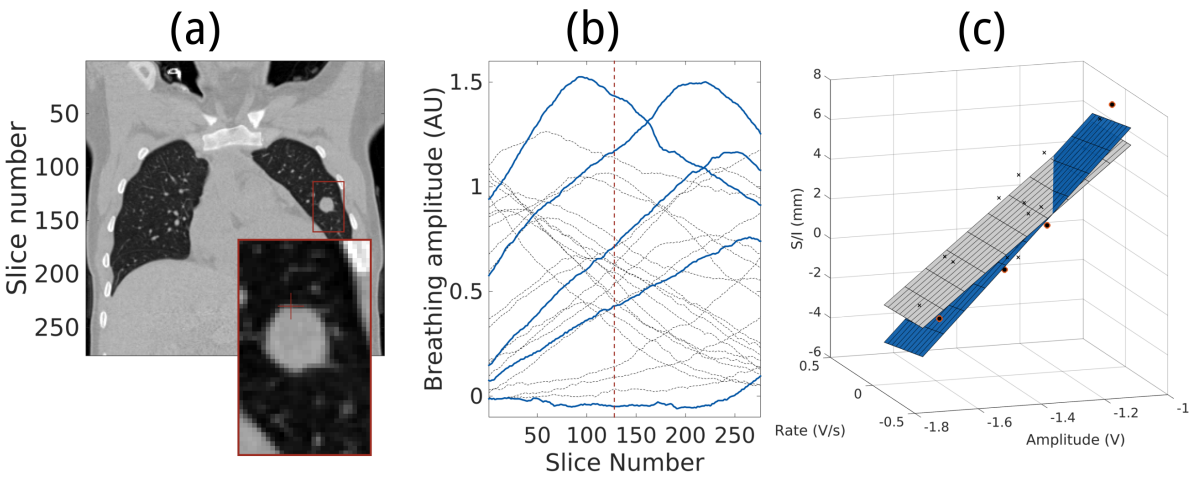


Figure 7.1: (a) Selected voxel in reference geometry marked by '+.' (b) Measured breathing waveforms for 25 scans (dotted), with a subset of 5 'well spread' scans denoted by solid lines. Slice number corresponding to the selected voxel in (a) denoted by vertical line. (c) Superior/inferior voxel displacement given by deformable image registration as a function of breathing amplitude and rate, denoted by 'x.' Plane representing full 25 scan model is lightly shaded. Plane representing reduced 5-scan model is darkly shaded. Displacements which determined reduced model fit are shown as circles.

## 7.3 Methods

We previously found that the phase coverage of a set of respiratory trajectories correlated with model accuracy only for 5 or more scans, and that 10 scans yielded adequate sampling of the breathing cycle. Therefore, the goal of a prospective scanning approach would be to select between 5 and 9 scans with a desirable spread of respiratory trajectories. We had two primary objectives for this study: firstly, to determine the minimum phase coverage for sets of N scans, where $5 \leq N \leq 9$, necessary to achieve mean model accuracy within 0.5 mm of the corresponding 25 scan model; and secondly, to test a prospective protocol and determine the number of scans necessary to reach the previously determined minimum phase coverage. Sections 7.3A-C describe the patient datasets and the process of finding the phasecoverage thresholds. 7.2D and 7.2E describe the prospective method and the simulation used to validate it.

### 7.3.1 Image acquisition and motion modeling

5 lung cancer patients were imaged under the published 5DCT protocol, described in detail in [19], using 64-slice CT scanners (Biograph TruePoint, Definition AS; Siemens Healthcare, Forchhiem, Germany). A low dose technique (40 effective mAs), short rotation time (0.33 or 0.285 s) and a pitch of 1.5 or 1.2 were used to scan each patient 25 times while breathing was simultaneously monitored with an abdominal bellows (Model 76513NM12G; Lafayette Instrument, Lafayette, IN), and a pressure to voltage transducer sampled with LabVIEW (National Instruments, Austin, TX) software. The acquisition protocol was designed such that tissues were imaged for a sufficiently short amount of time (approximately one third of a second) to treat the axial slices of the free-breathing images as measurements of instantaneous tissue positions. Each axial slice was assigned a breathing amplitude and rate according to the bellows signal at the time of acquisition. Importantly, because scans were acquired during free-breathing, the respiratory phase of the images varied by slice rather than an image depicting a single point in the breathing cycle, as in a gated 4DCT reconstruction, the respiratory phase of each scan was described by a curve as shown in Figure 7.1b.

The first scan, selected as the reference geometry, was deformably registered to the remaining 24 scans using the 'deeds' algorithm, proposed in [26] and demonstrated to achieve an average registration error of 1.43 mm over 10 cases using expert selected landmarks. The algorithm recovers large tissue displacements well, and models the sliding motion between the lungs and the rib cage. Additionally, it does not require a lung segmentation. The deformation vector fields (DVF) output by the DIR provided measurements of tissue displacement from the reference geometry to the remaining 24 observed breathing states. The DVFs and corresponding amplitude and rate values were used to solve for the voxel-specific parameters relating tissue motion to breathing amplitude, v, and rate, f, in the 5D motion model [24]:

$$\mathbf{X}(v, f, \mathbf{X}_0) = \mathbf{X}_0 + \boldsymbol{\alpha}(\mathbf{X}_0)v + \boldsymbol{\beta}(\mathbf{X}_0)f \tag{7.1}$$

Here, $\mathbf{X}$ denotes the displaced tissue location, $\mathbf{X}_0$ its original position in the reference

geometry, $\boldsymbol{\alpha}$ relates inhalation depth to surrogate amplitude, and the parameter $\boldsymbol{\beta}$ relates hysteresis motion to the derivative of the surrogate signal.

For each patient, model parameters were solved for using all 25 DVFs to build a 'full' motion model. A cuBLAS batched QR factorization routine was used to find least- squares solutions for model parameters on a graphics processing unit (Tesla K40; NVIDIA, Santa Clara, CA). After fitting $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, equation 7.1 allows for the computation of a DVF to transform the reference image to arbitrary amplitude $v$ and rate $f$, producing a breathing-gated image at any user-selected breathing phase without sorting artifacts; no binning of projections is performed because the model parameters are fit using free-breathing fast helical images.

Model residual was calculated by comparing the tissue displacements computed using equation 1 to the 25 displacements given by DIR, and taking the root mean square of the difference in displacements for each voxel. An overall figure of merit was calculated by averaging the residual values for lung voxels where the measured motion was $\geq 5$ mm. The motion criteria was included to prevent voxels which underwent little to no motion, such as regions of the upper lung, from biasing the accuracy measure. Importantly, the residual is representative of how well the motion model agrees with the DIR and does not account for uncertainty in the registration, ie differences between the measured displacements and the actual physical motion.

### 7.3.2 Quantifying the phase coverage of sampled respiratory trajectories

To faithfully characterize tissue motion with a reduced number of scans it is necessary to image each slice location's tissues at breathing states that are well-distributed throughout the respiratory cycle. We adopt the previously published phase coverage metric described in [126]:

$$M = \sum_{z} \sum_{i=1}^{N-1} (\tilde{\mathbf{v}}_{i+1,z} - \tilde{\mathbf{v}}_{i,z})^{0.5} \tag{7.2}$$

where $\tilde{\mathbf{v}}$ is the vector of the $N$ measured breathing amplitudes for a slice location $z$, shown as values lying on the dashed line in Figure 7.1b, sorted in ascending order.

### 7.3.3 Relationship between phase coverage and model accuracy

In order to determine a stopping criteria for a prospective scanning protocol, it was necessary to relate the phase coverage metric M to model accuracy  given a set of Nscans, what value of M would suggest that sufficient modeling accuracy could be achieved? Thomas et al investigated the correlation between M and accuracy by selecting subsets of N scans, where 2 ¡ N ¡ 15, and solving for the motion parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to fit 'reduced' models. The residual of the reduced models was computed and compared to the 'full' 25 scan model. The authors demonstrated good correlation between M and accuracy for $N > 4$, and additionally that when $N \leq 10$ model accuracy comparable to using 25 scans could be achieved regardless of the M value of the subset. In this study, we fit reduced models for each patient, where $5 \leq N \leq 9$. In order to limit the number of possible combinations of scans, we reduced the candidate pool from 25 to 13 by selecting every other scan. By using every other scan rather than selecting a group from the start or end of the acquisition, temporal biasing was avoided. To further reduce computation, voxel-specific model parameters were only solved for a single coronal slice in the center of the lung.

For each subset, the model residual and M value was computed. Although the reduced models were fit from a limited candidate pool, the residual values were computed by comparing model predictions to measured tissue positions from all 25 scans. Breathing surrogate signals were normalized such that the 5 th percentile inhalation and the 85 th percentile inhalation amplitude mapped to 0 and 1, respectively, in order to facilitate inter-patient comparison. Furthermore, the M value was normalized by the total number of surrogate samples acquired (length of scan multiplied by number of scans).

To determine an acceptable threshold of the normalized M value that was suitable for all patients, we found the maximum M value of a subset of scans which had unacceptable residual (defined as more than 0.5mm increase in mean residual than the full model) for each patient and each number of scans. These maximum values were used as general stopping criteria for the prospective gating approach; in other words the minimum M value which would be acceptable for a given number of scans for any patient.

### 7.3.4 Prospective gating method

The proposed method consisted of three elements: an overall triggering scheme, real- time breathing phase estimation, and a decision on whether or not the patient is currently breathing in a reproducible manner. From previous work, it was observed that sets of trajectories exhibiting a high M value generally resembled a pattern of phase-shifted waveforms and therefore it was hypothesized that adequate phase coverage could be achieved by staggering the phase at which scans are started. A respiratory profiling method [46] was used to estimate the breathing phase at each timepoint. A 'representative breath' [44] was formed based on training data and subsequently updated each new period.

In order to quantify the reproducibility of the breathing pattern, the normalized cross-correlation between the representative breath and the most recent respiratory surrogate measurements was continually computed. An overview of the method is shown in Figure 7.2.

#### 7.3.4.1 Breathing phase estimation

A detailed description of the technique used for real-time breathing phase estimation is given in [46]. In brief, in order to characterize hysteresis motion, an augmented state vector for each time-point $t_i$ was formed by taking the current breathing amplitude, $v(t_i)$ and a previous observation $v(t_i - kd_s)$), where $k$ is the lag expressed in number of samples and $d_s$ is the sampling interval. An ellipse was continually fit to points in the augmented state space within a sliding window encompassing $(v(t_i), v(t_i - kd_s))$ and approximately one period of previous observations. Transition events marking the end of a period were detected using Poincar sectioning. An estimated phase was assigned to each time-point by linearly extrapolating from the most recent marker event.

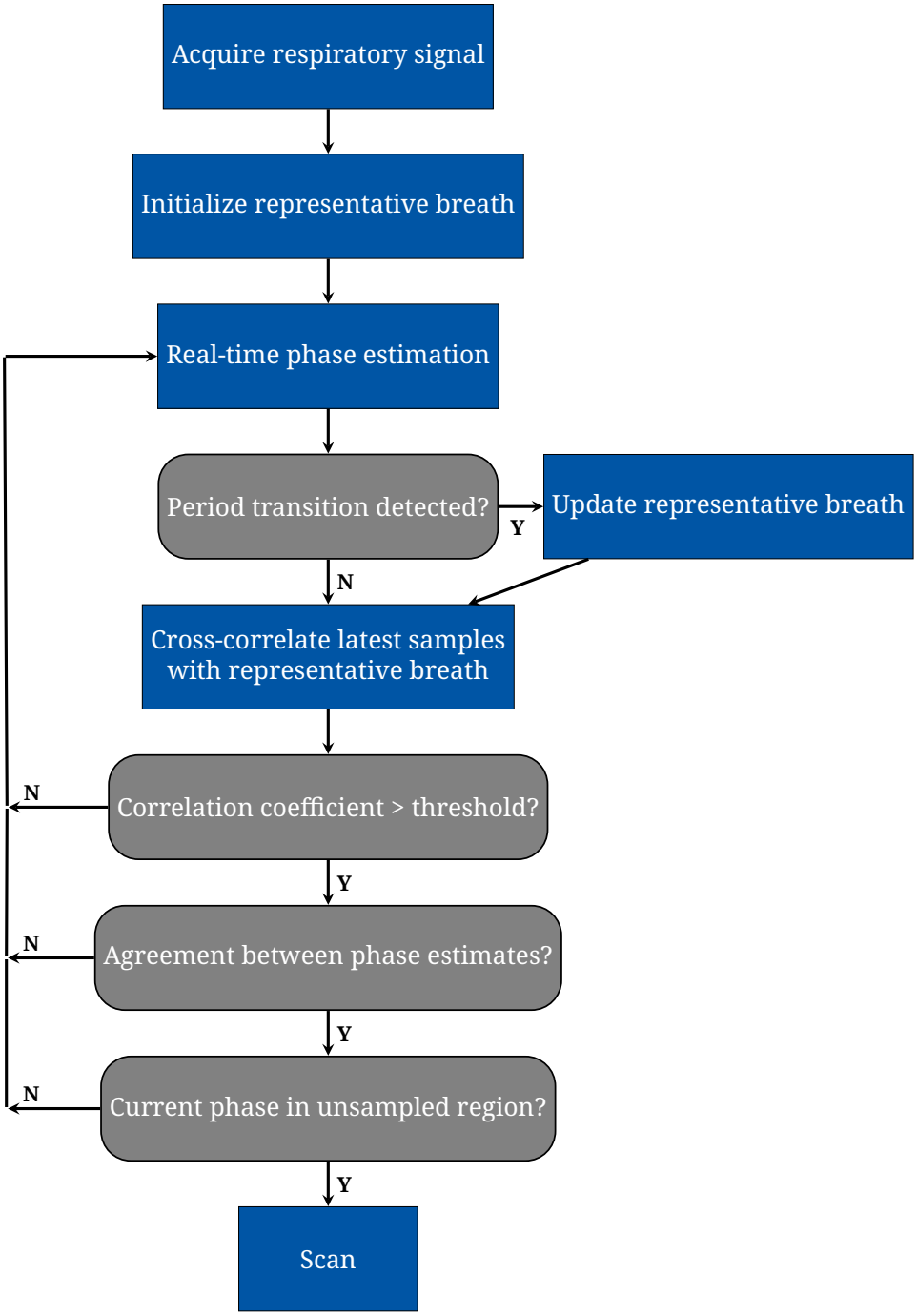


Figure 7.2: Flow chart describing the prospective gating method. The method takes a continuous respiratory surrogate signal as input, and outputs a series of triggers that mark times when helical scans should begin in order to acquire a dissimilar set of images.

#### 7.3.4.2 Reproducibility evaluation

For each patient, the first 30 seconds of the measured breathing trace was used as training data. The trace was segmented into individual breaths and the mean period was taken as the period of the representative breath. The amplitude was found by linearly interpolating the amplitude values at each time-point in the segmenting breaths and computing the weighted average, where the weights emphasized the most recently observed breaths. The resulting waveform was smoothed using Savitzky-Golay filtering [39] to produce the representative breath. After each marker event, the representative breath was updated to include the most recent period. In order to avoid bias due to outliers or incorrectly detected periods, breaths which exceeded 2 standard deviations in amplitude or period from the mean were not used to update the representative breath.

The normalized cross-correlation [127] between the representative breath and the most recent respiratory surrogate signal values, used as the template, was computed. The length of the template was set to half of the breathing period. The correlation coefficient provided an indicator of breathing reproducibility at the current time point. If the coefficient exceeded a preset threshold a scan was eligible to be triggered if the breathing phase criterion was satisfied.

#### 7.3.4.3 Triggering scheme

Sets of respiratory traces with high phase coverage were observed to generally resemble phase-shifted waveforms. A triggering scheme was designed to stagger the starting phase of each free-breathing scan with the aim of collecting a desirable set. The breathing cycle was divided uniformly into ten contiguous phase intervals, with each interval $i$ representing 10% of the breathing period. Initially, all ten phase intervals are marked as 'unsampled.' No restrictions are placed on the phase of the first scan  acquisition is triggered as soon as the breathing reproducibility criterion is met. The interval containing the phase point at which the first scan was triggered is marked as 'sampled' and removed from consideration for triggering subsequent scans. Of the remaining nine intervals, a subset of candidate intervals

is chosen to maximize the minimum distance (defined as the number of intervals), allowing for wrap-around because the representative breath is periodic. There can be more than one candidate interval  if there are multiple unsampled regions with the same minimum distance to a sampled region, then scan $k$ can be triggered in any of them. An example case is shown in Figure 7.3, where the four candidate intervals for the next acquisition are darkly shaded, the sampled intervals are unshaded, and the ineligible intervals are lightly shaded.

Once the set of candidate intervals is determined, acquisition is triggered only when the reproducibility criterion is satisfied and the current phase is within a candidate interval. After five scans are acquired, the phase coverage metric M is computed. If M is satisfactory, acquisition is ended. Otherwise, scanning continues until a satisfactory M is obtained or ten scans acquired.

### 7.3.5 Simulation study

For each of the 5 patients that underwent 5DCT imaging, approximately 3 minutes of breathing surrogate data was measured. The recorded respiratory waveforms were used to test the proposed method in a MATLAB-based simulation. The simulated scan duration was set to match the imaging time of the actual scans. For each patient, breathing trace was sampled at intervals selected using the proposed method. After 5 scans, the M value of the sampled respiratory trajectories was computed. If M was above the threshold acquisition terminated, otherwise acquisition continued until the threshold was reached. In order to model the limitations of an actual hardware implementation, the simulation was performed with a fixed delay between triggering and data collection of 2.0 seconds.

## 7.4 Results

### 7.4.1 Phase coverage and model accuracy

The average full model residual over the 5 patients in this study was 1.06 $\pm$ 0.2 mm. The mean minimum M values for subsets of N scans which achieved a model residual within

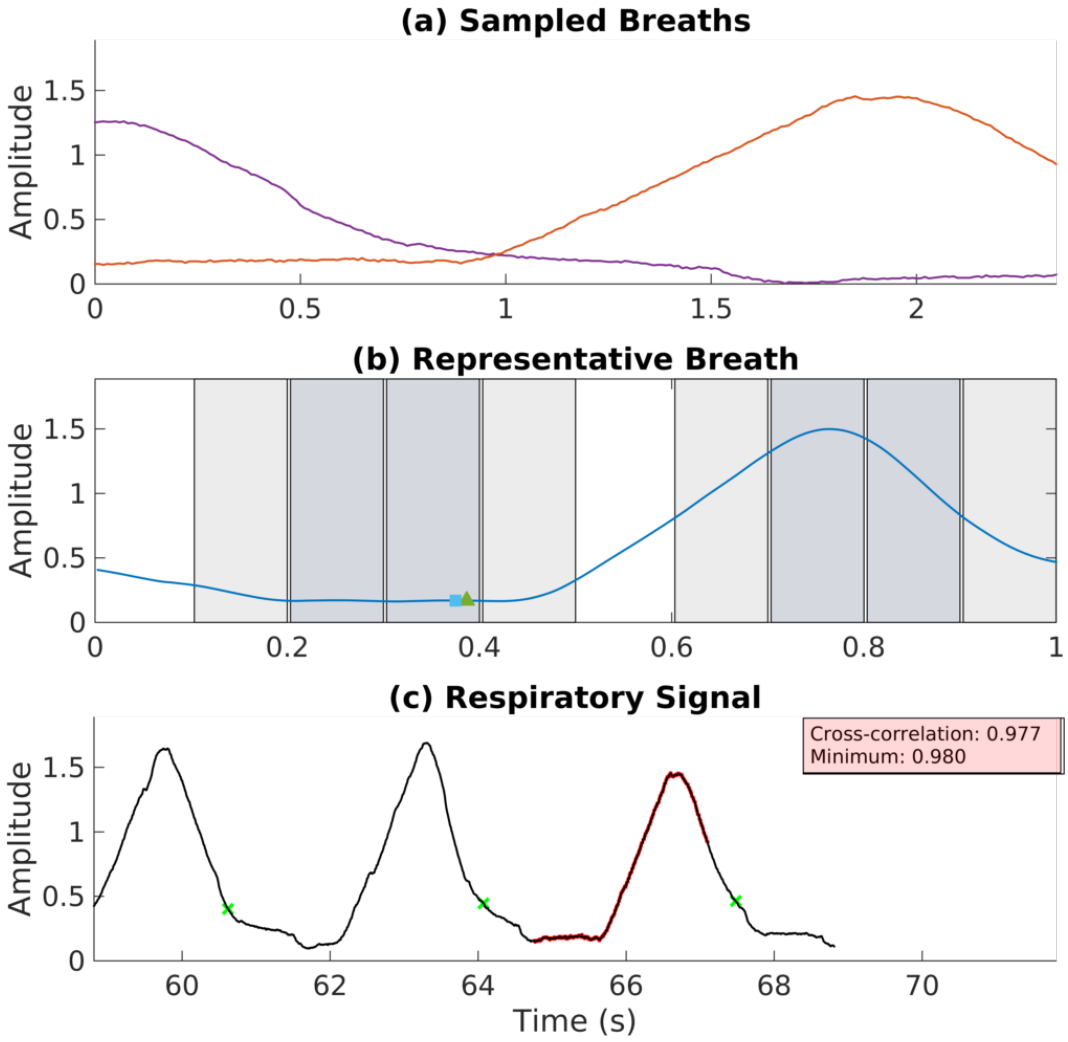


Figure 7.3: Simulated acquisition with the proposed prospective gating method using the breathing trace from Patient 1. (a) Shows two respiratory trajectories which were selected. The representative breath is shown in (b), with the triggering regions corresponding to the two sampled trajectories unshaded. Unsampled regions are shaded lightly, and regions eligible for the next acquisition are shaded darkly. Triangle and square represent the phase estimates using linear extrapolation between detected period transitions and the location of the maximum cross-correlation, retrospectively. Respiratory signal is plotted in (c). Detected period transitions are marked with x, and the most recently acquired respiratory trajectory is shown in bold.

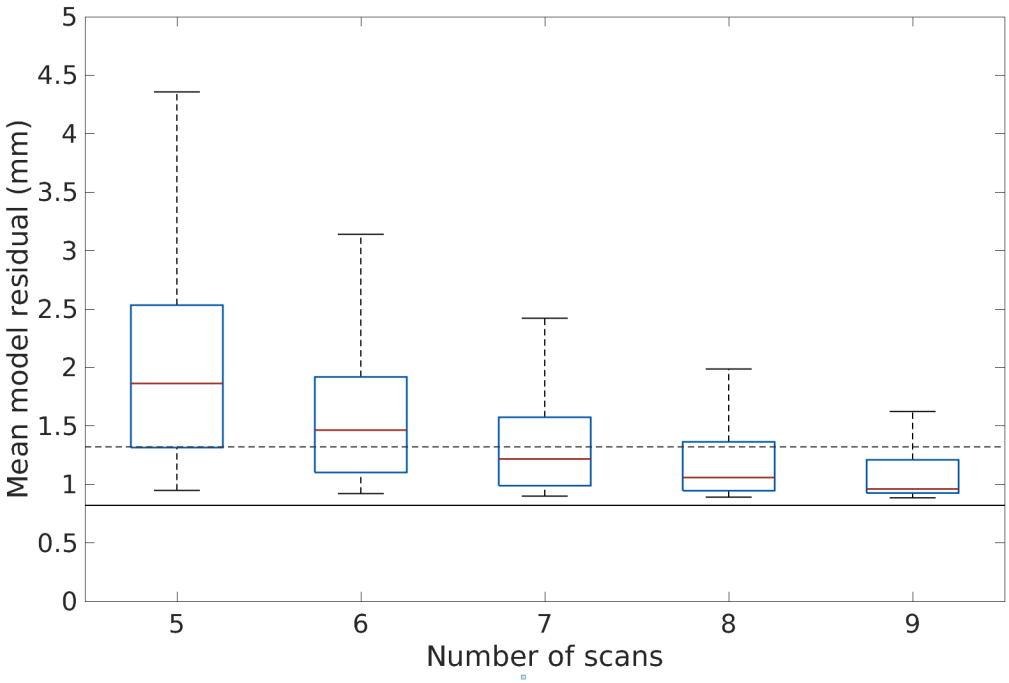


Figure 7.4: Box plots showing the distribution of mean residual for models fit using only subsets of N out of the 25 scans acquired with the previous non-prospective protocol, where $5 \leq N \leq 9$, for Patient 5. Solid black line denotes the minimum residual, achieved when all 25 scans are used in model fitting. Dashed gray line indicates minimum residual plus 0.5 mm, the target residual for the proposed scan selection method.

tolerance were $0.368 \pm 0.020$, $0.339 \pm 0.020$, $0.310 \pm 0.020$, $0.290 \pm 0.020$, $0.268 \pm 0.022$ for $N = 5$, 6, 7, 8, and 9, respectively. These M values corresponded to the 91st, 65th, 39th, 8th, and 1st percentiles on average among the M values of sets of N non-prospectively acquired scans.

Figure 7.4 displays the distribution of residual for reduced models for one patient. Note that for $N = 5$ scans, few of the subsets are able to achieve a residual within tolerance. The proportion of subsets that would yield an acceptable model residual increases as the number of scans increases. When $N = 9$, almost all subsets result in tissue motion being acceptably modeled.

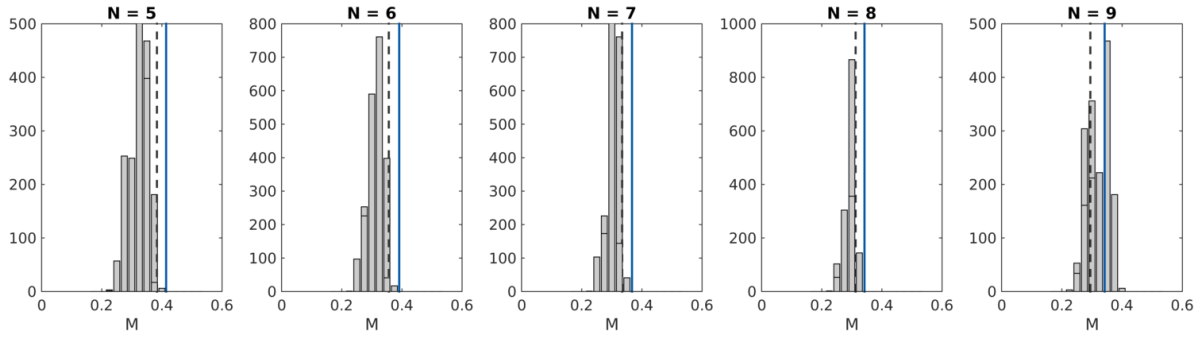


Figure 7.5: Histograms showing the distribution of M values for subsets of N out of the 25 acquired scans for Patient 2. Dashed line indicates the minimum M required for mean model residual within 0.5 mm of the residual when N = 25. Solid line indicates the M value of N scans selected from the patients measured breathing trace using the proposed prospective method.

### 7.4.2 Simulation study

The average minimum M values for N = 5, 6, 7, 8 and 9 were used as stopping criteria in the simulation study. For 4 out of the 5 patients, the prospective method was able to select a set of respiratory trajectories with an acceptable M value in 5 scans. For the remaining patient, acceptable phase coverage was achieved after 6 scans. Percentile scores describing the relative M value with respect to subsets of actually acquired scans for the 5 patients were 90.1, 100.0, 100.0, 99.1, and 98.6.

Figure 7.5 shows the M values for different numbers of prospectively selected scans for one patient, along with M values of subsets of the non-prospectively selected scans, and the minimum M value necessary for sufficient accuracy to provide context. When N = 5,6,7, very few subsets of the non-prospective scans have an M value greater than the threshold. For these cases, the difference in phase coverage achieved by the prospectively selected scans compared to non-prospectively selected is most apparent.

## 7.5 Discussion

We have described a prospective method to acquire a set of free-breathing CT scans such that their associated respiratory surrogate signals are well-spread throughout the breathing cycle in order to facilitate the fitting of a correspondence model between the respiratory surrogate signal and tissue displacement for model-based 4DCT. The results presented here suggest that using the proposed approach, the number of scans required for accurately describing tissue motion with 5DCT can be reduced from the previously published 25 to as few as 6 with less than a 0.5 mm increase in mean residual. The proposed method can be employed to select free-breathing scans for any respiratory motion model.

The primary motivation behind the proposed method is to reduce the imaging dose associated with our model-based 4DCT technique, 5DCT, by eliminating acquisition of redundant data which does not improve modeling accuracy. In this study modeling accuracy was assessed by computing the residual, defined as the root mean square difference in tissue positions given by the model and image registration. While theresidual is indicative of model performance, it does not translate directly to a measurement of error. A thorough assessment of error, including the effects of uncertainty in image registration and the relationship between the respiratory surrogate signal and tissue motion, would require a evaluation such as a landmark analysis.

Uncertainty in the correlation between the respiratory surrogate signal and internal tissue motion could potentially reduce the effectiveness of the prospective scan selection method and degrade the accuracy of the resulting motion model. A major complicating factor in estimating this correlation is respiratory hysteresis. In the real- time phase estimation technique employed in the proposed method, hysteresis is treated by increasing the dimensionality of the surrogate signal and defining each surrogate measurement by its current amplitude as well as the amplitude observed some fixed amount of time ago [45]. Similarly, the 5D model accounts for hysteresis by describing tissue motion as a function of surrogate amplitude as well as its derivative.

In Figure 7.1, the set of 5 scans with the highest M value for patient 5 are shown. The

5D model, Equation 7.1, constrains each voxels motion to lie within a plane with respect to amplitude and rate. The two planes in Figure 1c demonstrate the difference in the full model parameters, fit using all 25, vs. the reduced model, fit using the high M subset of 5. Although overall agreement is good, note that the inclusion of deeper breaths, which drive the M value to be higher, biases the reduced model toward estimating greater displacement. The redundant data incorporated into the full model fitting damps this effect.

We incorporated a delay of two seconds between a trigger signal and the beginning of data acquisition into our simulation. With the current scanner software at our institution (Syngo, Siemens Healthcare; Forchheim, Germany), there is a time lag of 2 seconds between the operator initiating a scan and acquisition. Future work will be centered upon implementing the method in real-time on a workstation separate from the CT console, and triggering will be accomplished by visually signaling to the operator when to manually initialize a scan. Ideally the proposed method could be incorporated into the scanner software to eliminate the delay, similar to commercially available cardiac gating schemes, but we chose to model a more feasible implementation in this study.

The reproducibility evaluation described in section 7.3.4.2 aims to avoid the effects of irregular breathing by preventing imaging under such conditions, but it has obvious limitations. An abrupt change in respiration after triggering but during or shortly before image acquisition could not be protected against. Such variance is unpredictable even for human operators. Although the model-based 4DCT technique used in this work is not susceptible to image artifacts due to irregular breathing, the proposed prospective selection method may require additional scans before the stopping criteria is met compared to a more regularly breathing case, and consequently the potential imaging dose savings would be limited.

Typically, binning-based 4DCT reconstruction methods require an irradiation period equal to or longer than the breathing period in order to reconstruct gated images. In contrast, model-based techniques construct a continuous description of tissue motion from a limited set of observations acquired at different respiratory states. For 5DCT, the model-based approach employed in this work, fast helical free-breathing scans are used as observations. An approximate CTDI vol value for a single free-breathing scan for 5DCT is 2.7 mGy [128].

Simply reducing the number of scans from 25 to 6 allows for the acquisition of 5DCT with less than 25% of the current dose. Further reduction could be achieved by combining the proposed technique with low- dose iterative CT reconstruction [129]. The substantial savings afforded by the proposed technique have the potential to reduce the total imaging dose of model-based 4DCT into the range where it is clinically acceptable to use as a tool for non-radiation therapy applications.

## 7.6 Conclusions

A method to prospectively select a set of free-breathing fast helical CT scans which are well-spread throughout the respiratory cycle has been described and tested in a 5 patient simulation study. Using the proposed technique, respiratory phase coverage which suggests model accuracy within 0.5 mm on average of a 25-scan model could be achieved in 6 or fewer scans.

## 7.7 Acknowledgment

The Tesla K40 used for this research was donated by the NVIDIA Corporation. This work was supported in part by funding from Siemens Healthcare.

Bibliography

[1] Paul J Keall, Gig S Mageras, James M Balter, Richard S Emery, Kenneth M Forster, Steve B Jiang, Jeffrey M Kapatoes, Daniel A Low, Martin J Murphy, Brad R Murray, et al. The management of respiratory motion in radiation oncology report of aapm task group 76a). *Medical physics*, 33(10):3874–3900, 2006.

[2] Minoru Uematsu, Akira Shioda, Atsushi Suda, Toshiharu Fukui, Yuichi Ozeki, Yukihiro Hama, James R Wong, and Shoichi Kusano. Computed tomography-guided frameless stereotactic radiotherapy for stage i non-small cell lung cancer: a 5-year experience. *International Journal of Radiation Oncology* Biology* Physics*, 51(3):666–670, 2001.

[3] Hiroshi Onishi, Hiroki Shirato, Yasushi Nagata, Masahiro Hiraoka, Masaharu Fujino, Kotaro Gomi, Yuzuru Niibe, Katsuyuki Karasawa, Kazushige Hayakawa, Yoshihiro Takai, et al. Hypofractionated stereotactic radiotherapy (hypofxsrt) for stage i non-small cell lung cancer: updated results of 257 patients in a japanese multi-institutional study. *Journal of thoracic oncology*, 2(7):S94–S100, 2007.

[4] Wolfgang A Tomé and Jack F Fowler. On cold spots in tumor subvolumes. *Medical Physics*, 29(7):1590–1598, 2002.

[5] Alexander Chi, Zhongxing Liao, Nam P Nguyen, Jiahong Xu, Baldassarre Stea, and Ritsuko Komaki. Systemic review of the patterns of failure following stereotactic body radiation therapy in early-stage non-small-cell lung cancer: clinical implications. *Radiotherapy and Oncology*, 94(1):1–11, 2010.

[6] René WM Underberg, Frank J Lagerwaard, Ben J Slotman, Johan P Cuijpers, and Suresh Senan. Use of maximum intensity projections (mip) for target volume generation in 4dct scans for lung cancer. *International Journal of Radiation Oncology* Biology* Physics*, 63(1):253–260, 2005.

[7] Jeffrey D Bradley, Ahmed N Nofal, Issam M El Naqa, Wei Lu, Jubei Liu, James Hubenschmidt, Daniel A Low, Robert E Drzymala, and Divya Khullar. Comparison of

helical, maximum intensity projection (mip), and averaged intensity (ai) 4d ct imaging for stereotactic body radiation therapy (sbrt) planning in lung cancer. *Radiotherapy and oncology*, 81(3):264–268, 2006.

[8] Tokihiro Yamamoto, Ulrich Langner, Billy W Loo, John Shen, and Paul J Keall. Retrospective analysis of artifacts in four-dimensional ct images of 50 abdominal and thoracic radiotherapy patients. *International Journal of Radiation Oncology* Biology* Physics*, 72(4):1250–1258, 2008.

[9] W Tyler Watkins, Ruijiang Li, John Lewis, Justin C Park, Ajay Sandhu, Steve B Jiang, and William Y Song. Patient-specific motion artifacts in 4dct. *Medical physics*, 37(6):2855–2861, 2010.

[10] Kwangyoul Park, Long Huang, Havaleh Gagne, and Lech Papiez. Do maximum intensity projection images truly capture tumor motion? *International Journal of Radiation Oncology* Biology* Physics*, 73(2):618–625, 2009.

[11] N Clements, T Kron, R Franich, L Dunn, P Roxby, Y Aarons, B Chesson, S Siva, D Duplan, and D Ball. The effect of irregular breathing patterns on internal target volumes in four-dimensional ct and cone-beam ct images in the context of stereotactic lung radiotherapy. *Medical physics*, 40(2):021904, 2013.

[12] Sara St James, Pankaj Mishra, Fred Hacker, Ross I Berbeco, and John H Lewis. Quantifying itv instabilities arising from 4dct: a simulation study using patient data. *Physics in medicine and biology*, 57(5):L1, 2012.

[13] Sara St James, Joao Seco, Pankaj Mishra, and John H Lewis. Simulations using patient data to evaluate systematic errors that may occur in 4d treatment planning: A proof of concept study. *Medical physics*, 40(9):091706, 2013.

[14] Jing Cai, Paul W Read, Joseph M Baisden, James M Larner, Stanley H Benedict, and Ke Sheng. Estimation of error in maximal intensity projection-based internal target volume of lung tumors: a simulation and comparison study using dynamic magnetic

resonance imaging. *International Journal of Radiation Oncology* Biology* Physics*, 69 (3):895–902, 2007.

[15] Jing Cai, Paul W Read, and Ke Sheng. The effect of respiratory motion variability and tumor size on the accuracy of average intensity projection from four-dimensional computed tomography: an investigation based on dynamic mri. *Medical physics*, 35 (11):4974–4981, 2008.

[16] Jing Cai, Robert McLawhorn, Paul W Read, James M Larner, Fang-fang Yin, Stanley H Benedict, and Ke Sheng. Effects of breathing variation on gating window internal target volume in respiratory gated radiation therapya). *Medical physics*, 37(8):3927–3934, 2010.

[17] Jamie R McClelland, David J Hawkes, Tobias Schaeffter, and Andrew P King. Respiratory motion models: a review. *Medical image analysis*, 17(1):19–42, 2013.

[18] Daniel A Low, Benjamin M White, Percy P Lee, David H Thomas, Sergio Gaudio, Shyam S Jani, Xiao Wu, and James M Lamb. A novel ct acquisition and analysis technique for breathing motion modeling. *Physics in medicine and biology*, 58(11): L31, 2013.

[19] David Thomas, James Lamb, Benjamin White, Shyam Jani, Sergio Gaudio, Percy Lee, Dan Ruan, Michael McNitt-Gray, and Daniel Low. 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(1):191–198, 2014.

[20] David P Gierga, George TY Chen, Jong H Kung, Margrit Betke, Jonathan Lombardi, and Christopher G Willett. Quantification of respiration-induced abdominal tumor motion and its impact on imrt dose distributions. *International Journal of Radiation Oncology* Biology* Physics*, 58(5):1584–1595, 2004.

[21] EC Ford, GS Mageras, E Yorke, and CC Ling. Respiration-correlated spiral ct: a

method of measuring respiratory-induced anatomic motion for radiation treatment planning. *Medical physics*, 30(1):88–97, 2003.

[22] Tinsu Pan. Comparison of helical and cine acquisitions for 4d-ct imaging with multislice ct. *Medical physics*, 32(2):627–634, 2005.

[23] George TY Chen, Jong H Kung, and Kevin P Beaudette. Artifacts in computed tomography scanning of moving objects. In *Seminars in radiation oncology*, volume 14, pages 19–26. Elsevier, 2004.

[24] Daniel A Low, Parag J Parikh, Wei Lu, James F Dempsey, Sasha H Wahab, James P Hubenschmidt, Michelle M Nystrom, Maureen Handoko, and Jeffrey D Bradley. Novel breathing motion model for radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 63(3):921–929, 2005.

[25] René Werner, Benjamin White, Heinz Handels, Wei Lu, and Daniel A Low. Technical note: Development of a tidal volume surrogate that replaces spirometry for physiological breathing monitoring in 4d ct. *Medical physics*, 37(2):615–619, 2010.

[26] HP Heinrich, Mark Jenkinson, Michael Brady, and Julia A Schnabel. Mrf-based deformable registration and ventilation estimation of lung ct. *Medical Imaging, IEEE Transactions on*, 32(7):1239–1248, 2013.

[27] Mattias P Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V Gleeson, Sir Michael Brady, and Julia A Schnabel. Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, 2012.

[28] Keelin Murphy, Bram van Ginneken, Stefan Klein, Marius Staring, BJ de Hoop, Max A Viergever, and Josien PW Pluim. Semi-automatic construction of reference standards for evaluation of image registration. *Medical image analysis*, 15(1):71–84, 2011.

[29] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of

deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6): 567–585, 1989.

[30] CD Cook, JM Sutherland, S Segal, RB Cherry, J Mead, MB McIlroy, and CA Smith. Studies of respiratory physiology in the newborn infant. iii. measurements of mechanics of respiration. *Journal of Clinical Investigation*, 36(3):440, 1957.

[31] Yvette Seppenwoolde, Hiroki Shirato, Kei Kitamura, Shinichi Shimizu, Marcel van Herk, Joos V Lebesque, and Kazuo Miyasaka. Precise and real-time measurement of 3d tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 53(4):822–834, 2002.

[32] René WM Underberg, Frank J Lagerwaard, Ben J Slotman, Johan P Cuijpers, and Suresh Senan. Benefit of respiration-gated stereotactic radiotherapy for stage i lung cancer: an analysis of 4dct datasets. *International Journal of Radiation Oncology Biology Physics*, 62(2):554–560, 2005.

[33] Wei Lu, Parag J Parikh, James P Hubenschmidt, Jeffrey D Bradley, and Daniel A Low. A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4d ct. *Medical physics*, 33(8):2964–2974, 2006.

[34] R Werner, J Ehrhardt, T Frenzel, D Säring, W Lu, D Low, and H Handels. Motion artifact reducing reconstruction of 4d ct image data for the analysis of respiratory dynamics. *Methods of Information in Medicine*, 46(03):254–260, 2007.

[35] UW Langner and PJ Keall. Accuracy in the localization of thoracic and abdominal tumors using respiratory displacement, velocity, and phase. *Medical physics*, 36(2): 386–393, 2009.

[36] Chiara Gianoli, Marco Riboldi, Maria Francesca Spadea, Laura Lavinia Travaini, Mahila Ferrari, Riccardo Mei, Roberto Orecchia, and Guido Baroni. A multiple points method for 4d ct image sorting. *Medical physics*, 38(2):656–667, 2011.

[37] René Werner, Christian Hofmann, and Tobias Gauer. Optimized projection binning for improved helical amplitude-and phase-based 4dct reconstruction in the presence of breathing irregularity. In *Medical Imaging 2016: Physics of Medical Imaging*, volume 9783, page 978313. International Society for Optics and Photonics, 2016.

[38] Sarah J Castillo, Richard Castillo, Edward Castillo, Tinsu Pan, Geoffrey Ibbott, Peter Balter, Brian Hobbs, and Thomas Guerrero. Evaluation of 4d ct acquisition methods designed to reduce artifacts. *Journal of applied clinical medical physics*, 16(2):23–32, 2015.

[39] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[40] Jianwen Luo, Kui Ying, and Jing Bai. Savitzky–golay smoothing and differentiation filter for even number data. *Signal Processing*, 85(7):1429–1434, 2005.

[41] Mattias Paul Heinrich, Mark Jenkinson, Bartlomiej W Papież, Michael Brady, and Julia A Schnabel. Towards realtime multimodal fusion for image-guided interventions using self-similarities. In *International conference on medical image computing and computer-assisted intervention*, pages 187–194. Springer, 2013.

[42] Mattias P Heinrich, Mark Jenkinson, Michael Brady, and Julia A Schnabel. Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 115–122. Springer, 2012.

[43] Tai H Dou, David H Thomas, Dylan P O'Connell, James M Lamb, Percy Lee, and Daniel A Low. A method for assessing ground-truth accuracy of the 5dct technique. *International Journal of Radiation Oncology Biology Physics*, 93(4):925–933, 2015.

[44] Benjamin White, Tianyu Zhao, James Lamb, Sara Wuenschel, Jeffrey Bradley, Issam El Naqa, and Daniel Low. Distribution of lung tissue hysteresis during free breathing. *Medical physics*, 40(4), 2013.

[45] Dan Ruan, Jeffrey A Fessler, and James M Balter. Mean position tracking of respiratory motion. *Medical physics*, 35(2):782–792, 2008.

[46] D Ruan, Jeffrey A Fessler, JM Balter, and PJ Keall. Real-time profiling of respiratory motion: baseline drift, frequency variation and fundamental pattern change. *Physics in Medicine & Biology*, 54(15):4777, 2009.

[47] Carl De Boor. Calculation of the smoothing spline with weighted roughness measure. *Mathematical Models and Methods in Applied Sciences*, 11(01):33–41, 2001.

[48] Hong Ge, Jing Cai, Chris R Kelsey, and Fang-Fang Yin. Quantification and minimization of uncertainties of internal target volume for stereotactic body radiation therapy of lung cancer. *International Journal of Radiation Oncology Biology Physics*, 85(2): 438–443, 2013.

[49] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[50] Kelly H Zou, Simon K Warfield, Aditya Bharatha, Clare MC Tempany, Michael R Kaus, Steven J Haker, William M Wells, Ferenc A Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology*, 11(2):178–189, 2004.

[51] Janid P Blanco Kiely, Arthur Olszanski, Stefan Both, Daniel A Low, and Benjamin M White. Quantitative early decision making metric for identifying irregular breathing in 4dct. *Medical physics*, 42(10):5654–5660, 2015.

[52] Iris E van Dam, John R van Sörnsen de Koste, Gerard G Hanna, Rebecca Muirhead, Ben J Slotman, and Suresh Senan. Improving target delineation on 4-dimensional ct scans in stage i nsclc using a deformable registration tool. *Radiotherapy and oncology*, 96(1):67–72, 2010.

[53] Dylan P O'Connell, David H Thomas, Tai H Dou, James M Lamb, Franklin Feingold, Daniel A Low, Matthew K Fuld, Jered P Sieren, Chelsea M Sloan, Melissa A Shirk,

et al. Comparison of breathing gated ct images generated using a 5dct technique and a commercial clinical protocol in a porcine model. *Medical physics*, 42(7):4033–4042, 2015.

[54] Long Huang, Kwangyoul Park, Thomas Boike, Pam Lee, Lech Papiez, Timothy Solberg, Chuxiong Ding, and Robert D Timmerman. A study on the dosimetric accuracy of treatment planning for stereotactic body radiation therapy of lung cancer using average and maximum intensity projection images. *Radiotherapy and Oncology*, 96(1): 48–54, 2010.

[55] Jochem WH Wolthaus, Christoph Schneider, Jan-Jakob Sonke, Marcel van Herk, José SA Belderbos, Maddalena MG Rossi, Joos V Lebesque, and Eugène MF Damen. Mid-ventilation ct scan construction from four-dimensional respiration-correlated ct scans for radiotherapy planning of lung cancer patients. *International Journal of Radiation Oncology Biology Physics*, 65(5):1560–1571, 2006.

[56] Matthias Guckenberger, Juergen Wilbert, Thomas Krieger, Anne Richter, Kurt Baier, and Michael Flentje. Mid-ventilation concept for mobile pulmonary tumors: internal tumor trajectory versus selective reconstruction of four-dimensional computed tomography frames based on external breathing motion. *International Journal of Radiation Oncology Biology Physics*, 74(2):602–609, 2009.

[57] JWH Wolthaus, J-J Sonke, M Van Herk, and EMF Damen. Reconstruction of a time-averaged midposition ct scan for radiotherapy planning of lung cancer patients using deformable registration. *Medical physics*, 35(9):3998–4011, 2008.

[58] Thanasis Loupas, WN McDicken, and Paul L Allan. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *IEEE transactions on Circuits and Systems*, 36(1):129–135, 1989.

[59] Frederique Crete, Thierry Dolmiere, Patricia Ladret, and Marina Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In

*Human vision and electronic imaging XII*, volume 6492, page 64920I. International Society for Optics and Photonics, 2007.

[60] IEC IEC. 62304: 2006 medical device software–software life cycle processes. *International Electrotechnical Commission, Geneva*, 2006.

[61] Diane Kelly, Alan Wassyng, and Colin G Orton. The most suitable person to establish quality assurance guidelines for the generation and use of noncommercial clinical software is a medical physicist. *Medical Physics*, 41(9), 2014.

[62] Phillip Cosgriff, Paul Ganney, Allan Green, David Willis, and Richard Trouncer. In house development of medical software: Regulations and standards. 2016.

[63] Isaac I Rosen. Writing software for the clinic. *Medical physics*, 25(3):301–309, 1998.

[64] David B Gersten and Steve G Langer. Programming in the small. *Journal of digital imaging*, 24(1):142–150, 2011.

[65] Greg Salomons and Diane Kelly. Software safety in radiation therapy. *Journal of Medical Physics/Association of Medical Physicists of India*, 38(1):1, 2013.

[66] Greg J Salomons and Diane Kelly. A survey of canadian medical physicists: software quality assurance of in-house software. *Journal of applied clinical medical physics*, 16 (1):336–348, 2015.

[67] M Saiful Huq, Benedick A Fraass, Peter B Dunscombe, John P Gibbons, Geoffrey S Ibbott, Arno J Mundt, Sasa Mutic, Jatinder R Palta, Frank Rath, Bruce R Thomadsen, et al. The report of task group 100 of the aapm: Application of risk analysis methods to radiation therapy quality management. *Medical physics*, 43(7):4209–4262, 2016.

[68] D Allan Wilkinson and Matthew D Kolar. Failure modes and effects analysis applied to high-dose-rate brachytherapy treatment planning. *Brachytherapy*, 12(4):382–386, 2013.

[69] Ivan Veronese, Elena De Martin, Anna Stefania Martinotti, Maria Luisa Fumagalli, Cristina Vite, Irene Redaelli, Tiziana Malatesta, Pietro Mancosu, Giancarlo Beltramo, Laura Fariselli, et al. Multi-institutional application of failure mode and effects analysis (fmea) to cyberknife stereotactic body radiation therapy (sbrt). *Radiation Oncology*, 10(1):132, 2015.

[70] Jennifer C O'Daniel and Fang-Fang Yin. Quantitative approach to failure mode and effect analysis for linear accelerator quality assurance. *International Journal of Radiation Oncology Biology Physics*, 98(1):56–62, 2017.

[71] Jane Cleland-Huang and Mona Rahimi. A case study: Injecting safety-critical thinking into graduate software engineering projects. In *Proceedings of the 39th International Conference on Software Engineering: Software Engineering and Education Track*, pages 67–76. IEEE Press, 2017.

[72] Alistair Cockburn. Writing effective use cases, the crystal collection for software professionals. *Addison-Wesley Professional Reading*, 2000.

[73] Alistair Mavin, Philip Wilkinson, Adrian Harwood, and Mark Novak. Easy approach to requirements syntax (ears). In *Requirements Engineering Conference, 2009. RE'09. 17th IEEE International*, pages 317–322. IEEE, 2009.

[74] Matthias Guckenberger, Markus Weininger, Juergen Wilbert, Anne Richter, Kurt Baier, Thomas Krieger, Buelent Polat, and Michael Flentje. Influence of retrospective sorting on image quality in respiratory correlated computed tomography. *Radiotherapy and oncology*, 85(2):223–231, 2007.

[75] Gitte Fredberg Persson, Ditte Eklund Nygaard, Carsten Brink, Jonas Westberg Jahn, Per Munck af Rosenschöld, Lena Specht, and Stine Sofia Korreman. Deviations in delineated gtv caused by artefacts in 4dct. *Radiotherapy and Oncology*, 96(1):61–66, 2010.

[76] Gitte Fredberg Persson, Ditte Eklund Nygaard, Per Munck af Rosenschöld,

Ivan Richter Vogelius, Mirjana Josipovic, Lena Specht, and Stine Sofia Korreman. Artifacts in conventional computed tomography (ct) and free breathing four-dimensional ct induce uncertainty in gross tumor volume determination. *International Journal of Radiation Oncology Biology Physics*, 80(5):1573–1580, 2011.

[77] D O'Connell, N Shaverdian, AU Kishan, DH Thomas, TH Dou, JH Lewis, JM Lamb, M Cao, S Tenn, P Lee, et al. Comparison of lung tumor motion measured using a model-based 4dct technique and a commercial protocol. *Practical Radiation Oncology*, 2017.

[78] JB Battles and RJ Lilford. Organizing patient safety research to identify risks and hazards. *BMJ Quality & Safety*, 12(suppl 2):ii2–ii7, 2003.

[79] Ivar Jacobson. *Object-oriented software engineering: a use case driven approach*. Pearson Education India, 1993.

[80] Karen Allenby and Tim Kelly. Deriving safety requirements using scenarios. In *Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium on*, pages 228–235. IEEE, 2001.

[81] Robyn R Lutz. Targeting safety-related errors during software requirements analysis. *Journal of Systems and Software*, 34(3):223–230, 1996.

[82] Nancy G Leveson. Safeware. *System Safety and Computers. Addison Wesley*, 1995.

[83] Alan Simpson and Joanne Stoker. Will it be safe?an approach to engineering safety requirements. In *Components of System Safety*, pages 140–164. Springer, 2002.

[84] Romaric Guillerm, Hamid Demmou, and Nabil Sadou. Information model for model driven safety requirements management of complex systems. In *Complex Systems Design & Management*, pages 99–111. Springer, 2010.

[85] Jéssyka Vilela, Jaelson Castro, Luiz Eduardo G Martins, and Tony Gorschek. Integration between requirements engineering and safety analysis: A systematic literature review. *Journal of Systems and Software*, 125:68–92, 2017.

[86] Donald Firesmith. Engineering safety requirements, safety constraints, and safety-critical requirements. *Journal of object technology*, 3(3):27–42, 2004.

[87] Robyn R Lutz. Software engineering for safety: a roadmap. In *Proceedings of the Conference on the Future of Software Engineering*, pages 213–226. ACM, 2000.

[88] Glenford J Myers, Corey Sandler, and Tom Badgett. *The art of software testing*. John Wiley & Sons, 2011.

[89] Patrick Mader, Paul L Jones, Yi Zhang, and Jane Cleland-Huang. Strategic traceability for safety-critical projects. *IEEE software*, 30(3):58–66, 2013.

[90] RTCA (Firm). SC 167. *Software considerations in airborne systems and equipment certification*. RTCA, Incorporated, 1992.

[91] ISO Standard. Iso 26262 road vehicles–functional safety. 2011.

[92] Enno Ruijters and Mariëlle Stoelinga. Fault tree analysis: a survey of the state-of-the-art in modeling, analysis and tools. *Computer science review*, 15:29–62, 2015.

[93] Sohag Kabir. An overview of fault tree analysis and its application in model based dependability analysis. *Expert Systems with Applications*, 77:114–135, 2017.

[94] C Michael Holloway. Safety case notations: Alternatives for the non-graphically inclined? In *System Safety, 2008 3rd IET International Conference on*, pages 1–6. IET, 2008.

[95] Anis Baklouti, Nga Nguyen, Jean-Yves Choley, Faida Mhenni, and Abdelfattah Mlika. Free and open source fault tree analysis tools survey. In *Systems Conference (SysCon), 2017 Annual IEEE International*, pages 1–8. IEEE, 2017.

[96] Keith R Britton, George Starkschall, Susan L Tucker, Tinsu Pan, Christopher Nelson, Joe Y Chang, James D Cox, Radhe Mohan, and Ritsuko Komaki. Assessment of gross tumor volume regression and motion changes during radiotherapy for non–small-cell

lung cancer as measured by four-dimensional computed tomography. *International Journal of Radiation Oncology Biology Physics*, 68(4):1036–1046, 2007.

[97] Mischa Hoogeman, Jean-Briac Prévost, Joost Nuyttens, Johan Pöll, Peter Levendag, and Ben Heijmen. Clinical accuracy of the respiratory tumor tracking system of the cyberknife: assessment by analysis of log files. *International Journal of Radiation Oncology Biology Physics*, 74(1):297–303, 2009.

[98] Qinghui Zhang, Alex Pevsner, Agung Hertanto, Yu-Chi Hu, Kenneth E Rosenzweig, C Clifton Ling, and Gig S Mageras. A patient-specific respiratory model of anatomical motion for radiation treatment planning. *Medical Physics*, 34(12):4772–4781, 2007.

[99] Ruijiang Li, John H Lewis, Xun Jia, Tianyu Zhao, Weifeng Liu, Sara Wuenschel, James Lamb, Deshan Yang, Daniel A Low, and Steve B Jiang. On a pca-based lung motion model. *Physics in Medicine & Biology*, 56(18):6009, 2011.

[100] Benjamin M White, Tianyu Zhao, James M Lamb, Jeffrey D Bradley, and Daniel A Low. Physiologically guided approach to characterizing respiratory motion. *Medical physics*, 40(12), 2013.

[101] DH Thomas, D Ruan, P Williams, J Lamb, BM White, T Dou, D OConnell, P Lee, and DA Low. Is there an ideal set of prospective scan acquisition phases for fast-helical based 4d-ct? *Physics in Medicine & Biology*, 61(23):N632, 2016.

[102] Eugene N Bruce. Deflation-related variability of breathing pattern persists with intact upper airway. *Respiration physiology*, 106(3):273–283, 1996.

[103] JD Escolar and A Escolar. Lung hysteresis: a morphological view. *Histology and histopathology*, 19(1):159–166, 2004.

[104] G Remmert, J Biederer, F Lohberger, M Fabel, and GH Hartmann. Four-dimensional magnetic resonance imaging for the determination of tumour movement and its evaluation using a dynamic porcine lung phantom. *Physics in Medicine & Biology*, 52(18): N401, 2007.

[105] Jurgen Biederer and Martin Heller. Artificial thorax for mr imaging studies in porcine heart-lung preparations. *Radiology*, 226(1):250–255, 2003.

[106] Daniel C Grinnan and Jonathon Dean Truwit. Clinical review: respiratory mechanics in spontaneous and assisted ventilation. *Critical Care*, 9(5):472, 2005.

[107] Christopher S Rogers, William M Abraham, Kim A Brogden, John F Engelhardt, John T Fisher, Paul B McCray Jr, Geoffrey McLennan, David K Meyerholz, Eman Namati, Lynda S Ostedgaard, et al. The porcine lung as a potential model for cystic fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 295 (2):L240–L263, 2008.

[108] K Markstaller, B Eberle, H-U Kauczor, A Scholz, A Bink, M Thelen, W Heinrichs, and N Weiler. Temporal dynamics of lung aeration determined by dynamic ct in a porcine model of ards. *British journal of anaesthesia*, 87(3):459–468, 2001.

[109] Nicholas J Screaton, Harvey O Coxson, Steve E Kalloger, Elisabeth M Baile, Yasutaka Nakano, Melanie Hiorns, and John R Mayo. Detection of lung perfusion abnormalities using computed tomography in a porcine model of pulmonary embolism. *Journal of thoracic imaging*, 18(1):14–20, 2003.

[110] Elizabeth A Barnes, Brad R Murray, Donald M Robinson, Lori J Underwood, John Hanson, and Wilson HY Roa. Dosimetric evaluation of lung tumor immobilization using breath hold at deep inspiration. *International Journal of Radiation Oncology Biology Physics*, 50(4):1091–1098, 2001.

[111] Y Vinogradskiy, L Schubert, D Brennan, R Castillo, E Castillo, T Guerrero, M Miften, B Kavanagh, and MK Mary. Clinical validation of 4dct-ventilation with pulmonary function test data. *International Journal of Radiation Oncology Biology Physics*, 90 (1):S61–S62, 2014.

[112] Thomas Guerrero, Kevin Sanders, Edward Castillo, Yin Zhang, Luc Bidaut, Tinsu Pan, and Ritsuko Komaki. Dynamic ventilation imaging from four-dimensional computed tomography. *Physics in Medicine & Biology*, 51(4):777, 2006.

[113] Yevgeniy Vinogradskiy, Richard Castillo, Edward Castillo, Susan L Tucker, Zhongxing Liao, Thomas Guerrero, and Mary K Martel. Use of 4-dimensional computed tomography-based ventilation imaging to correlate lung dose and function with clinical outcomes. *International Journal of Radiation Oncology Biology Physics*, 86(2): 366–371, 2013.

[114] Katelyn Hasse, John Neylon, and Anand P Santhanam. Feasibility and quantitative analysis of a biomechanical model-guided lung elastography for radiotherapy. *Biomedical Physics & Engineering Express*, 3(2):025006, 2017.

[115] Keelin Murphy, Josien PW Pluim, Eva M van Rikxoort, Pim A de Jong, Bartjan de Hoop, Hester A Gietema, Onno Mets, Marleen de Bruijne, Pechin Lo, Mathias Prokop, et al. Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic ct. *Medical physics*, 39(3):1650–1662, 2012.

[116] Tokihiro Yamamoto, Sven Kabus, Tobias Klinder, Cristian Lorenz, Jens Von Berg, Thomas Blaffert, Billy W Loo Jr, and Paul J Keall. Investigation of four-dimensional computed tomography-based pulmonary ventilation imaging in patients with emphysematous lung regions. *Physics in Medicine & Biology*, 56(7):2279, 2011.

[117] Donna L Hoyert, Jiaquan Xu, et al. Deaths: preliminary data for 2011. *Natl Vital Stat Rep*, 61(6):1–51, 2012.

[118] Jonas Jögi, Marie Ekberg, Björn Jonson, Gracijela Bozovic, and Marika Bajc. Ventilation/perfusion spect in chronic obstructive pulmonary disease: an evaluation by reference to symptoms, spirometric lung function and emphysema, as assessed with hrct. *European journal of nuclear medicine and molecular imaging*, 38(7):1344–1352, 2011.

[119] Yuka Matsuzaki, Keisuke Fujii, Motoki Kumagai, Ichiro Tsuruoka, and Shinichiro Mori. Effective and organ doses using helical 4dct for thoracic and abdominal therapies. *Journal of radiation research*, 54(5):962–970, 2013.

[120] Zhen Tian, Xun Jia, Bin Dong, Yifei Lou, and Steve B Jiang. Low-dose 4dct reconstruction via temporal nonlocal means. *Medical physics*, 38(3):1359–1365, 2011.

[121] Tinsu Pan, Rachael M Martin, and Dershan Luo. New prospective 4d-ct for mitigating the effects of irregular respiratory motion. *Physics in Medicine & Biology*, 62(15):N350, 2017.

[122] Jamie R McClelland, Jane M Blackall, Ségolene Tarte, Adam C Chandler, Simon Hughes, Shahreen Ahmad, David B Landau, and David J Hawkes. A continuous 4d motion model from multiple respiratory cycles for use in lung radiotherapy. *Medical Physics*, 33(9):3348–3358, 2006.

[123] Yuan Xu, Hao Yan, Luo Ouyang, Jing Wang, Linghong Zhou, Laura Cervino, Steve B Jiang, and Xun Jia. A method for volumetric imaging in radiotherapy using single x-ray projection. *Medical physics*, 42(5):2498–2509, 2015.

[124] Jamie R McClelland, Marc Modat, Simon Arridge, Helen Grimes, Derek DSouza, David Thomas, Dylan OConnell, Daniel A Low, Evangelia Kaza, David J Collins, et al. A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images. *Physics in Medicine & Biology*, 62(11):4273, 2017.

[125] T Geimer, M Unberath, A Birlutiu, O Taubmann, J Wölfelschneider, C Bert, and A Maier. A kernel-based framework for intra-fractional respiratory motion estimation in radiation therapy. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 1036–1039. IEEE, 2017.

[126] Dan Ruan, David Thomas, and Daniel A Low. Objective function to obtain multiple representative waveforms for a novel helical ct scan protocol. *Medical physics*, 42(3): 1164–1169, 2015.

[127] John P Lewis. Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123, 1995.

[128] Cynthia McCollough, Dianna Cody, Sue Edyvean, Rich Geise, Bob Gould, Nicholas Keat, Walter Huda, Phil Judy, Willi Kalender, Mike McNitt-Gray, et al. The measurement, reporting, and management of radiation dose in ct. *Report of AAPM Task Group*, 23(23):1–28, 2008.

[129] Stephan Baumueller, Anna Winklehner, Christoph Karlo, Robert Goetti, Thomas Flohr, Erich W Russi, Thomas Frauenfelder, and Hatem Alkadhi. Low-dose ct of the lung: potential value of iterative reconstructions. *European radiology*, 22(12): 2597–2606, 2012.