UNCERTATINTIES IN MODELING, PREDICTION, AND ACTIONS IN RESPONSE TO VARIATIONS IN PATIENT GEOMETRIC MODELS

by

Rojano Kashani

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in The University of Michigan
2009

Doctoral Committee:

Professor James M. Balter, Co-Chair
Professor Alex F. Bielajew, Co-Chair
Professor Edward W. Larsen
Professor William R. Martin
Professor Randall K. Ten Haken

2009

To Mom

With love and gratitude

ACKNOWLEDGMENTS

Many people have contributed to this dissertation and I feel truly lucky to have worked with each and every one of them. I am deeply grateful to Professor James Balter, my thesis advisor, for his support and encouragement over the past six years. His enthusiasm, and dedication to teaching and to his students is truly inspiring, and I feel extremely privileged to have had the chance to work under his direction. Despite his extremely busy schedule, he has always been very generous with his time and very patient with my lack of good time management skills. I especially appreciate the amount of time and energy he put into reading my last minute corrections to various talks and chapters over the past few months. I would also like to offer my sincere thanks to Professor Alex Bielajew for giving me the freedom to choose my research project based on my interests, and for his constructive comments that helped me get a broader perspective of the problems I was trying to solve. I am also very thankful to my committee members, Professor Randall Ten Haken, Professor Edward Larsen, and Professor William Martin, for taking the time to carefully read my dissertation and for providing me with excellent comments which helped me improve the clarity of several part of the dissertation significantly.

I also like to express my deep gratitude to all the faculty of the Department of Radiation Oncology. My heartfelt appreciation goes to Professor Daniel McShan and Dr. Marc Kessler, for helping me develop tools without which my research would not have been possible. They always took the time to help me out even when they were extremely busy. I have also enjoyed working with Dr. Scott Hadley and Dr. Ken Jee on various projects. I learned a great deal from both of them. I am also very grateful to Professor Benedick Fraass, and Dr. Jean Moran for taking an interest in the work of students and for making sure that we get sufficient attention from all faculty members. Their words of encouragement in the past few weeks have helped me gain more

confidence and I truly appreciate that. I also like to thank Dr. Kwok Lam, Dr. Dale Litzenberg, Dr. Martina Hub, Dr. Avi Eisbruch, Dr. Mary Feng, Dr. Annette Kretzler, and Dr. Orit Gutfeld for their contributions to the various projects presented in this dissertation. I also appreciate the help and training that was provided to me by Karen Vineberg, Robin Marsh, Dan Tatro, Cheryl White, and Zhou Shen, as well as the computer group, Xiaoping, Wayne and Eduardo.

My warmest thanks go to my wonderful friends in the Department of Radiation Oncology. They have helped me get through emotional ups and downs in the past six years, and specially in the past few months. I am most grateful to Chrissy Lockhart, Hannan Amro, Shu-Hui Hsu, Auresa Thomas, as well as some of my former collegues, Amit Sawant, Dan Ruan, Anant Gopal, Neelam Tyagi, and Martha Matuszak.

Last but not least, I like to thank my family. My mom gave up her career and retired from her job early so that she could move to the United States and support me through college. For that I am forever in her debt. Her strength of character and her selflessness truly inspires me. I also like to acknowledge my boyfriend Maani, who has been extremely understanding and supportive over the past five years. I also like to thank my brother and his beautiful family for always encouraging me to do my best.

TABLE OF CONTENTS

DEDICATION……ii
ACKNOWLEDGEMENT……iii
LIST OF FIGURES……vii
LIST OF TABLES……xii
CHAPTER I. INTRODUCTION **1**
I.A. Overview of technological advancements in patient positioning 1
I.B. Sources of treatment uncertainty 3
I.C. Geometric variation and its impact on treatment uncertainty 3
I.D. Dissertation purpose and overview 4
CHAPTER II. RESPIRATORY-INDUCED ORGAN MOTION AND ITS MANAGEMENT **9**
II.A. Introduction 9
II.B. Respiratory motion management 10
II.B.1. Breath-hold treatments with Active Breathing Control (ABC) tumor immobilization 10
II.B.2. Forward-planning without margin expansions 18
II.B.3. Discussion and limitations 24
II.C. Summary 25
CHAPTER III. QUANTITATIVE EVALUATION OF IMAGE REGISTRATION ACCURACY **28**
III.A. Introduction 28
III.B. Phantom design 30
III.B.1. Phantom components 30
III.C. System Performance 32
III.C.1. Motion under compression and reproducibility 32
III.C.2. Imaging and analysis 35
III.D. Image registration 37
III.D.1. Method validation 37
III.E. Discussion 53
III.F. Conclusion 55
CHAPTER IV. TRACKING DEFORMATION USING SURROGATES **59**

IV.A. Introduction 59
IV.A.1. Background and purpose 59
IV.A.2. Dimensionality reduction to reduce registration noise 62
IV.B. Model description, implementation, and assessment 62
IV.B.1. Model description 62
IV.B.2. Why use PCA? 68
IV.B.3. Data description 70
IV.B.4. Model implementation and the limits of accuracy 71
IV.B.5. Effectiveness and accuracy of the model in tracking deformation 78
IV.C. Discussion 98
CHAPTER V. UNCERTAINTIES IN ADAPTIVE THERAPY FOR HEAD-AND-NECK CANCERS 102
V.A. Introduction 102
V.A.1. Head and neck cancers and treatment strategies 102
V.A.2. Uncertainties in target definition 103
V.A.3. Geometric variations over the course of treatment 104
V.A.4. Translating observed geometric changes into clinically meaningful metrics 105
V.A.5. Current correction strategies to account for geometric variation 105
V.A.6. The role of image registration and its accuracy in adaptive therapy 106
V.B. Study design 107
V.B.1. Patient data 107
V.B.2. Tracking geometry and dose 108
V.B.3. Accuracy of tracking geometry and dose 112
V.C. Discussion 128
CHAPTER VI. SUMMARY 132

LIST OF FIGURES

Figure II.1: Tumor shrinkage over the course of treatment. a)Tumor at the time of simulation, b) Tumor after 40% of dose was delivered, and c) tumor after 80% of dose was delivered 14
Figure II.2: Axial view of the real and simulated inhale state, indicating sufficient similarity to preserve the characteristics of the inhale lung that would affect dose accumulation. 20
Figure II.3: Simulated GTV volumes and the resulting ITV calculated as a union of the GTVs at exhale, mid-ventilation and inhale 21
Figure II.4. Flowchart of two different planning methods based on ITV and GTV with motion 22
Figure II.5: se volume histograms for the exhale lung and the exhale GTV from ITV plan (blue) and the forward plan (grey) 23
Figure.II.6: Dose difference between the GTV-based forward plan and the ITV plan. It is observed that the forward plan is colder than the ITV plan in regions outside of the GTV volume. The image on the left shows a coronall cut through the center of the tumor. The center image shows an axial cut, 3 mm above the top of the GTV volume and the image on the right is 3mm below the GTV volume. 24
Figure III.1: The diagnostic phantom with the abdominal insert removed (left), the high density foam insert (right) 30
Figure III.2: The set up of the phantom and motion actuator on the CT scanner 31
Figure III.3: Coronal (top) and sagittal (bottomn) views of the CT scan of the phantom with the foam and tumor simulating inserts under compressions (diaphragm movements) of 0, 1, 2, and 3cm (from left to right). 32
Figure III.4 : Histogram of the intensity distribution in a typical section of the foam at the simulated exhale state (dark blue), and the intensity distribution of a sample segment from an image of the left lung at exhale (light blue). 35
Figure III.5: Volumetric views of the surface of the deformable lung insert, the solid tumor-simulating spheres, and the plastic markers, are shown at 1cm compression of the foam (inhale). Z: Superior/Inferior, Y: Anterior/Posterior, X: Right/Left directions. 35
Figure III.6: Axial view of a plane with 4 markers before (left) and after (right) removal of the markers from the image. 37
Figure III.7: The 3D distance between the real marker position and the estimated position based on B-spline transformation, shown as a function of the longitudinal position of the marker relative to the diaphragm. It can be seen that the error in the registration decreases in regions farther away from the diaphrahm. 39
Figure III.8: Colorwash of the mean (left) and standard deviation (right) 3-D error (cm) in the phantom lung, interpolated to all other regions inside the foam. 39

Figure III.9: Inhale to exhale SI motion of markers under 30mm compression as estimated by various registration methods, shown as a function of the marker's distance to diaphragm. The true motion of the markers from inhale to exhale is also shown (pink diamond marks indicate the motion of markers from inhale to exhale, measured manually). 49
Figure III.10. Distribution of 3D error for each registration method 50
Figure III.11. Comparison of 3-D error distribution between different b-splines based registration methods..... 51
Figure III.12: Mean and standard deviation of the 3-D error for each marker calculated over all registration methods. The horizontal axis displays each marker's motion from inhale to exhale in mm. The marker for which the results are shown in red, corresponds to an example with a large mean and a small standard deviation. The image of this marker is shown in figure III.13. 52
Figure III.13. Axial image of the region inside the foam at inhale (left) and exhale (right) where most registration methods performed poorly. 52
Figure IV.1: A uniform grid of voxels on the volumetric image of the left lung, parameterizes the patient geometry. The blue rectangle indicates the region that was used in the calculation of the deformation map. 63
Figure IV.2: Coronal images of the lung at various reconstructed states of a breathing cycle. 64
Figure IV.3: Schematic of creating the displacement vectors from the reference phase (inhale) to all other breathing states, by applying the transformaiton map T_{1N} to the postion vector at phase 1 (end-inhale). The images displayed here correspond to the deforming phantom from chapter III. 65
Figure IV.4. Coronal and sagittal views of the deformation map from inhale to exhale in the SI direction. The blue indicates a motion on the order of 1-2mm, while the red corresponds to displacements of 2cm or higher. 69
Figure IV.5: The position of manually selected landmarks inside the left lung, for evaluation of the model. The three large yellow circles show the possible positions of the three landmarks chosen as the surrogates..... 71
Figure IV.6: Root mean square difference between all data points estimated with SVD 73
Figure IV.7: Maximum error over all landmarks for each breathing state. I1 and I2 refere to image resolution used during registration. SI: Superior-Inferior, AP: Anterior-posterior, RL: Right-Left. 76
Figure IV.8: Registration error in the SI direction for each marker at the first breathing state. Landmaks that show larger differences such as landmakr 50 are useful in determining the effect of registration error on model accuracy. 76

Figure IV.9: Change in the error in estimation of the position of a single landmark (50), using different number of eigenmodes. Negative numbers indicae an improvement in the estimation of the landmark position relative to the registration. 77
Figure IV.10. Lung contours from three different levels in the lung (Top, Mid, Low) were selected for visualization of the first three eigenmodes (Figures IV.11 to IV.19) 79
Figure IV.11. Variation in the mean geometry, described by the first eigenmode, for the contour from the lower lung (Low). The red dots are the contour points at their mean position estimated over the full breathing cycle. The green arrows correspond to the positive variation (+σ), while the blue arrows represent the negative varitation (-σ) from Equation IV.13..... 79
Figure IV.12. Variation in the mean geometry, described by the second eigenmode, for the contour from the lower lung (Low). 80
Figure IV.13. Variation in the mean geometry, described by the third eigenmode, for the contour from the lower lung (Low). 80
Figure IV.14. Variation in the mean geometry, described by the first eigenmode, for the contour from the middle of the lung (Mid). 81
Figure IV.15. Variation in the mean geometry, described by the second eigenmode, for the contour from the middle of the lung (Mid). 81
Figure IV.16. Variation in the mean geometry, described by the third eigenmode, for the contour from the middle of the lung (Mid). 82
Figure IV.17. Variation in the mean geometry, described by the first eigenmode, for the contour from the upper lung (Top). 82
Figure IV.18. Variation in the mean geometry, described by the second eigenmode, for the contour from the upper lung (Top). 83
Figure IV.19. Variation in the mean geometry, described by the third eigenmode, for the contour from the upper lung (Top). 83
Figure IV.20: Percentage of variation described by each eiegnvalue, for different combinations of breathing states used in the model..... 84
Figure IV.21: The left column shows the maximum error in RL, AP, and SI directions (top to bottom) for each breathing state, using different number of eigenmodes in the model. The right column shows the mean error for the corresponding directions..... 86
Figure IV.22: The maximum (top) and mean (bottom) error over all landmarks for all breathing states shown as a function of the number of eigenmodes used in the model. 87
Figure IV.23: Histogram of the error in the estimated landmark positions in the RL direction using different number of eigenmodes over all breathing states. 1 eig indicates that only the first eigenmode was used. 8 eig is when all eigenmodes are included in the model. 88

Figure IV.24: Histogram of the error in the estimated landmark positions in the AP directgion using different number of eigenmodes over all breathing states. 1 eig indicates that only the first eigenmode was used. 8 eig is when all eigenmodes are included in the model. 89
Figure IV.25: Histogram of the error in the estimated landmark positions in the SI direction using different number of eigenmodes over all breathing states. 1 eig indicates that only the first eigenmode was used. 8 eig is when all eigenmodes are included in the model. 89
Figure IV.26: SI error histogram zoomed in on the larger errors, for different number of eigenmodes used in the model. 90
Figure IV.27: Maximum error in the SI direction over all breathing states for each marker when using different numbers of eigenmodes in the model..... 91
Figure IV.28: Maximum error in RL per landmark over all breathing states (blue diamond markers), registration error (red squares), and the error in landmarker relative to the position of landmarks as estimated by the deformation map (green triangles). 92
Figure IV.29: Maximum error in AP direction per landmark over all breathing states (blue diamond markers), registration error (red squares), and the error in landmarker relative to the position of landmarks as estimated by the deformation map (green triangles). 93
Figure IV.30: Maximum error in SI direction per landmark over all breathing states (blue diamond markers), registration error (red squares), and the error in landmarker relative to the position of landmarks as estimated by the deformation map (green triangles). 93
Figure IV.31: The small segments chosen in the deformation map, disaplayed on coronal images of the lung. Upper (left), Mid (middle), and Lower (right). Each section is 6cm in the longitudinal direction and the axial dimensions are the same as before. 94
Figure IV.32: Eigenvalue distribution for each segment of the lung compared to the entire lung 95
Figure IV.33: Error in estimating landmark position at inhale 96
Figure IV.34: Error in estimation of the landmark position for the end-exhale state. 96
Figure IV.35: Model accuracy for intermediate states not sampled by the 4D CT 98
Figure V.1: Schematic of the direct dose accumulation method, where each new sample of patient geometry is mapped to all previous fractions to lookup the fractional dose for accumulation. 110
Figure V.2: Schematic of the propagated dose accumulation method, where each new sample of patient geometry is only mapped to the previous fractions, to lookup the fractional dose for accumulation. 111
Figure V.3 :Contours shown for the target surface estimated using one of the two geometric tracking methods described (yellow: Direct, green: Propagated), compared to the manually drawn contour (pink) on CT scans from fractions 5, 15, and 25 (left to right)..... 114

Figure V.4: Volume overlap index (VOI) for the left and right CTV using the direct and the propagated geomettry tracking methods 115
Figure V.5: Schematic of manual (dark blue) and estimated targets (yellow and red) and the surrounding organs at risk.This diagram shows the relative position of the false positive (yellow on left and red on right) and false negative regions (blue) to the organs at risk. The overlap of these regions with the organs at risk may result in unwanted plan modifications. 115
Figure V.6: Absolute volume of the false positive and false negative volumes for the two methods..... 116
Figure V.7: Specifity and sensitivity of the two estimated CTV volumes for the direct and propagated methods. 117
Figure V.8: Histogram of the 3D error distribution for different fractions 118
Figure V.9: Zoomed view of the 3D error distribution in the region of errors larger than 5mm 119
Figure V.10: Maximum error in the estimated contours using the two different methods shown in each dimension. Light colors show error in propagated method, dark colors show the error in the direct method. 119
Figure V.11: Targets in the head-and-neck geometry (left) and the dose distribution for the optimized plan on the right 123
Figure V.12: DVH for the original CTV63 and the expanded volumes 124
Figure V.13: DVH of CTV70 before and after uniform expansion in 1mm increments 124
Figure V.14: Change in dose based metrics for differnt uniform expansions of the two CTVs ... 125
Figure V.15: Time trend of changes in the dose based metrics for estimated CTVs (CTV63) using both the direct and propagated methods, and the manually delineated CTV63 for three fraction. 127
Figure V.16: Estimated volume of CTV for different fractions over the course of treatment. The red circles indicate the three fractions for which manual contours were available..... 128

LIST OF TABLES

Table II.1: Short term reproducibility of ABC decive. 13
Table II.2. Comparison of the reproducibility of inhale and exhale breath-hold states 15
Table II.3. Reproducibility of tumor position for different degrees of tumor shrinkage 16
Table III.1: Reproducibility of tumor position between repeat compressions. For each diaphragm compression, the position of the center of the tumor in the first scan was taken as the reference and compared to the second scan. The results presented here include all 33
Table III.2: Motion of tumor simulating inserts under different diaphragm compressions. For each diaphragm compression (shown at the top), the actual motion in centimeters and the motion as a percentage of the diaphragm compression are shown. The error in each measurement is shown in paretheses. 33
Table III.3: Error in estimation of marker position based on example alignment results from different image registration algorithms 38
Table III.4: Comparison of registration results for the phantom and a sample patient or an in-house b-spline based method, using the same parameter settings. This table also shows the results for error in registration of the inhale state to other deformation states of the phantom. 41
Table III.5: Summary of registration methods and references 42
Table III.6: Maximum component errors in RL, AP, and SI directions, as well as the mean, standard deviation, and maximum 3-D vector distance for each registration method, shown in random order. 48
Table V.1: Maximum, mimumum and the mean error in estimation of conoutrs/ surfaces of the target area for different fractions 119
Table V.2: Dose based metrics calculated for the CTV63 and CTV70 for different uniform expansions (1-5mm). A min 1% dose indicates the maximum dose delivered to the minimum 1% of the volume. 125

CHAPTER I

INTRODUCTION

I.A. Overview of technological advancements in patient positioning

The past three to four decades have marked an era of significant technological advancement in Radiation Oncology. Starting with the clinical implementation of CT scanners in the early 1970s, better visualization of targets and organs at risk became possible, leading to an understanding of the limitations of two-dimensional planning and the benefits of incorporating CT data into the treatment planning process[1-5]. Subsequently, computerized three-dimensional treatment planning systems were developed that provided the tools necessary for making plans more conformal to the target, while sparing more of the surrounding normal tissue[6]. As a result of the transition from conventional to conformal treatment planning, the need for better geometric accuracy during treatment delivery became apparent. One of the earliest image-based approaches to verifying geometric accuracy at the time of treatment involved comparison of simulation films (typically diagnostic quality radiographs acquired to aid in treatment planning of treatment) with port films (lower quality megavoltage radiographs acquired via projecting the treatment beam through the patient)[7]. This process was generally limited by the quality of the port films and the lack of good registration methods for the alignment of daily images to the reference simulation image[8]. These limitations were partially overcome by the development and implementation of electronic portal imaging devices (EPIDs) in the treatment room, which provided a more efficient way of acquiring online images of the patient setup, resulting in a gradual transition from port films to EPIDs, which continue to be one of the most commonly used localization tools in the clinic[9-10].

Further development of in-room imaging technology, including the integration of kV imaging systems into the design of linear accelerators, which allowed for visualization of soft tissue not previously visible on portal images, resulted in a better understanding of the motion of the target relative to that of the surrogates used for patient positioning[11]. These systems were further enhanced to acquire volumetric images of the patient in treatment position (i.e. cone-beam CT), providing the data necessary for online three-dimensional soft tissue alignment and thus improving setup accuracy[11-12]. In addition to the radiographic imaging tools, the use of other imaging modalities such as ultrasound, and video or surface imaging systems, was also pursued in Radiation Oncology, allowing for verification of patient setup without the additional dose from imaging[13].

Beyond the advancement in patient setup accuracy, improvements have also been made in other aspects of the treatment planning process. Specifically, treatment planning techniques such as intensity modulated radiation therapy (IMRT) and stereotactic body radiation therapy (SBRT) can be utilized to achieve a higher degree of conformality of the dose distribution to the target volume. In IMRT, the beam intensity is modulated such that those segments that go through sensitive structures are lower in intensity and the parts that go through the target primarily are higher in intensity, thus creating a non-uniform dose distribution across the target and the surrounding normal structures. The result is a highly conformal dose distribution with sharp dose gradients between the target and the adjacent normal tissue, assuming all structures are static. However, tumors and the surrounding tissues are rarely static, exhibiting motion on both inter- and intra-fraction time scales, and thus reducing the initial estimates of treatment accuracy. This motion will inevitably result in a deviation of the actual dose distribution at the time of treatment delivery from the planned dose distribution[14]. Consequently, parts of the target may be under-dosed, while some of the surrounding normal tissue could exceed the initially estimated dose limit. The relatively large margins that would be necessary to cover the moving target can result in irradiation of larger volumes of the surrounding normal structures to high doses, thus limiting the

total dose prescribed to the target and reducing tumor control probability. In order to fully understand the implications of geometric uncertainty on treatment outcome, the sources of this uncertainty need to be identified and measurement of the magnitude of the motion and the consequent variations in dose should be considered.

I.B. Sources of treatment uncertainty

Several factors contribute to the overall accuracy of treatment planning and delivery including target definition, beam modeling and dosimetry, and the accuracy of the daily treatment delivery. Target definition accuracy is dependent on the imaging modalities utilized during the patient simulation process, the image quality, and the training and experience of the observer (physician). While CT is the most commonly used imaging modality for patient simulation and is necessary in determining electron density for dose calculation as part of the treatment planning process, a combination of various imaging modalities, such as Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI), may be used to improve target definition depending on the disease site. Aside from the differences between various imaging modalities, the inherent limitation in resolution of any imaging system creates some uncertainty in identifying the edges of the tumor accurately. This is addressed with the addition of a margin around the gross target volume to create a clinical target volume (CTV) to account for microscopic disease, as recommended by the International Commission on Radiation Units and Measurement (ICRU)[15].

I.C. Geometric variation and its impact on treatment uncertainty

A significant source of treatment uncertainty is geometric variation. Geometric variation refers to all changes in position, shape, or size of the tumor or normal tissue, relative to the reference used for treatment planning. Subsequent to the availability of in-room imaging tools was an increasing interest in the measurement of organ motion, as attempted by various research groups[16-25].

Characterization of geometric uncertainty, a necessary step in the development of motion management protocols, goes beyond the simple measurement of organ motion. It requires an understanding of the impact of the different components of the motion, as well as the role of the immobilization technique and the accuracy of measurement tools employed. The measured organ position or motion is very sensitive to the immobilization technique used, as well as on the setup protocol. In addition, the accuracy of a given measurement tool depends on the disease site (i.e. portal images have better accuracy for bone compared to soft tissue). Therefore, the precision of the measurement of geometric uncertainty is just as important as the magnitude of the variation and should be considered in the evaluation of geometric uncertainty. To make this analysis possible, we first need to understand the different components of the motion and the techniques used to compensate for them, so that the residual uncertainties in the measurement and management of this motion can be estimated.

I.D. Dissertation purpose and overview

The purpose of this work is to quantify the uncertainties associated with the measurement and management of patient geometric variations, and to evaluate the resulting limits of accuracy in treatment verification and plan modification. Of the various sources of geometric uncertainty discussed earlier in this chapter, only respiratory-induced organ motion and deformation, and the deformations in target and normal tissue in response to treatment are considered. As will be described in detail for each of the studies presented in the following four chapters, the setup error has been removed from the problem, allowing for the evaluation of the remaining sources of uncertainty beyond the setup correction. Each chapter also provides a discussion of the limitation of the analysis resulting from the specific characteristics of the data used, and the conditions in which the data was acquired for various clinical protocols.

In chapter II, the measurement of esophageal motion between inhale and exhale breathing states, provides the knowledge necessary for determination of the asymmetric margin expansions

necessary to cover the exhale target as it moves over a respiratory cycle. This chapter focuses on the specific problem of respiratory-induced motion and deformation in various thoracic structures and the efficacy of different methods employed throughout the process to ensure target coverage while minimizing dose to the normal tissue. It then examines the residual motion in targets in the lung under multiple breath-holds in one treatment fraction as well as over the treatment course, identifying the uncertainties of various respiratory motion management strategies and the implication on the treatment accuracy.

Chapter III describes the uncertainties in measurement and estimation of deformation, and the accuracy of various deformable registration techniques in aligning scans of the patient at different breath-hold states. It describes the complexity of quantitative evaluation of deformable registration techniques and the limitations in the commonly used techniques such as manually identified landmarks, or contours. It then introduces the design and implementation of a deforming phantom, for an objective test of various image registration techniques. Using this phantom, with a large number of small, easily identifiable implanted markers, the registration error in various regions inside the phantom is evaluated.

Chapter IV describes an alternative method for tracking and monitoring respiratory-induced deformation inside the lung and other thoracic structures using a number of surrogates whose positions can be monitored throughout the treatment. Recognizing the importance of tracking normal tissue deformation, and the uncertainties in the estimation of deformation from registration tools, a patient-specific model of the respiratory-induced motion and deformation is described using a method from multivariate statistics called principal component analysis (PCA). This model, applied to the position of surrogates, is used to provide relatively stable estimates of target and normal tissue positions in the lung or the liver.

In Chapter V, the magnitude of geometric variations in tumors and normal tissues in response to treatment is evaluated, as well as the impact of these configuration changes on the dose

distribution. The sample dataset evaluated here includes head and neck patients with in-room CT scans acquired approximately three times a week during the course of treatment. The focus of this chapter is on finding the most accurate method for tracking geometry and dose by comparing the registration accuracy for two consecutive scans to the registration accuracy for scans that are further apart and show larger changes. This would allow us to determine if the initial geometry defined on the treatment planning CT scans should be directly deformed to each daily CT scan, or if it should be tracked throughout the treatment by ignoring all previous contours once a new one is estimated from image registration.

References

[1]M. Goitein, "Limitations of Two-Dimensional Treatment Planning Programs," Med Phys **9,** 580-586 (1982).

[2]M. Goitein, "The Utility of Computed Tomography in Radiation Therapy: An Estimate of outcome," Int J Radiat Oncol Biol Phys **5,** 1799 (1979).

[3]P. Hobday, N. J. Hodson, J. Husband, R. P. Parker, and J. S. Macdonald, "Computed Tomography Applied to Radiotherapy Treatment Planning: Techniques and Results," Radiology **133,** 477-482 (1979).

[4]J. E. Munzenrider, M. Pilepich, J. B. Rene-Ferrero, I. Tchakarova, and B. L. Carter, "Use of Body Scanner in Radiotherapy Treatment Planning," Cancer **40,** 170-179 (1977).

[5]D. P. Ragan, and C. A. Perez, "Efficacy of CT-Assisted Two Dimensional Treatment Planning: Analysis of 45 Patients," Am J Roentgenol. **131,** 75-79 (1978).

[6]D. L. McShan, A. Silverman, D. M. Lanza, L. E. Reinstein, and A. S. Glicksman, "A Computerized Three-Dimensional Treatment Planning System Utilizing Interactive Colour Graphics," Br J Radiol **52,** 478-481 (1979).

[7]A. G. Haus, S. M. Pinsky, and J. E. Marks, "A Technique for Imaging Patient Treatment Area during a Therapeutic Radiation Exposure," Radiology **97,** 653-656 (1970).

[8]L. E. Reinstein, H. I. Amols, P. J. Biggs, et al. “Radiotherapy Portal Imaging Quality, Report of AAPM Task Group No. 28” 1987

[9]N. A. Baily, R. A. Horn, and T. D. Kampp, "Fluoroscopic Visualization of Megavoltage Therapeutic x Ray Beams," Int.J.Radiat.Oncol.Biol.Phys. **6,** 935-939 (1980).

[10]Boyer, "A Review of Electronic Portal Imaging Devices (EPIDs)." in *Vol. 19,* (Published for the American Association of Physicists in Medicine by the American Institute of Physics], [New York, etc., 1992), pp. 1.

[11]D. A. Jaffray, D. G. Drake, M. Moreau, A. A. Martinez, and J. W. Wong, "A Radiographic and Tomographic Imaging System Integrated into a Medical Linear Accelerator for Localization of Bone and Soft-Tissue Targets," Int.J.Radiat.Oncol.Biol.Phys. **45,** 773-789 (1999).

[12]J. H. Siewerdsen, and D. A. Jaffray, "Cone-Beam Computed Tomography with a Flat-Panel Imager: Effects of Image Lag," Med.Phys. **26,** 2635-2647 (1999).

[13]L. S. Johnson, B. D. Milliken, S. W. Hadley, C. A. Pelizzari, D. J. Haraf, and G. T. Chen, "Initial Clinical Experience with a Video-Based Patient Positioning System," Int.J.Radiat.Oncol.Biol.Phys. **45,** 205-213 (1999).

[14]B. Schaly, J. A. Kempe, G. S. Bauman, J. J. Battista, and J. Van Dyk, "Tracking the Dose Distribution in Radiation Therapy by Accounting for Variable Anatomy," Phys.Med.Biol. **49,** 791-805 (2004).

[15]ICRU Repost 62. Prescribing, Recording, and Reporting photon beam therapy (supplement to ICRU Report 50). Bethesda MD, International Commission on Radiation Units and Measurements, 1999.

[16]J. Pouliot, and A. Lirette, "Verification and Correction of Setup Deviations in Tangential Breast Irradiation using EPID: Gain Versus Workload," Med.Phys. **23,** 1393-1398 (1996).

[17]G. J. Kutcher, G. S. Mageras, and S. A. Liebel, "Control, Correction, and Modeling of Setup Errors and Organ Motion," Semin.Radiat.Oncol. **5,** 134-145 (1995).

[18]D. A. Fein, K. P. McGee, T. E. Schultheiss, B. L. Fowble, and G. E. Hanks, "Intra- and Interfractional Reproducibility of Tangential Breast Fields: A Prospective on-Line Portal Imaging Study," Int.J.Radiat.Oncol.Biol.Phys. **34,** 733-740 (1996).

[19]J. Hanley, M. A. Lumley, G. S. Mageras, J. Sun, M. J. Zelefsky, S. A. Leibel, Z. Fuks, and G. J. Kutcher, "Measurement of Patient Positioning Errors in Three-Dimensional Conformal Radiotherapy of the Prostate," Int.J.Radiat.Oncol.Biol.Phys. **37,** 435-444 (1997).

[20]J. M. Crook, Y. Raymond, D. Salhani, H. Yang, and B. Esche, "Prostate Motion during Standard Radiotherapy as Assessed by Fiducial Markers," Radiother.Oncol. **37,** 35-42 (1995).

[21]J. C. Roeske, J. D. Forman, C. F. Mesina, T. He, C. A. Pelizzari, E. Fontenla, S. Vijayakumar, and G. T. Chen, "Evaluation of Changes in the Size and Location of the Prostate, Seminal Vesicles, Bladder, and Rectum during a Course of External Beam Radiation Therapy," Int.J.Radiat.Oncol.Biol.Phys. **33,** 1321-1329 (1995).

[22]C. J. Beard, P. Kijewski, M. Bussiere, R. Gelman, D. Gladstone, K. Shaffer, M. Plunkett, P. Castello, and C. N. Coleman, "Analysis of Prostate and Seminal Vesicle Motion: Implications for Treatment Planning," Int.J.Radiat.Oncol.Biol.Phys. **34,** 451-458 (1996).

[23]M. van Herk, A. Bruce, A. P. Kroes, T. Shouman, A. Touw, and J. V. Lebesque, "Quantification of Organ Motion during Conformal Radiotherapy of the Prostate by Three Dimensional Image Registration," Int.J.Radiat.Oncol.Biol.Phys. **33,** 1311-1320 (1995).

[24]V. G. Althof, C. J. Hoekstra, and H. J. te Loo, "Variation in Prostate Position Relative to Adjacent Bony Anatomy," Int.J.Radiat.Oncol.Biol.Phys. **34,** 709-715 (1996).

[25]J. M. Balter, K. L. Lam, H. M. Sandler, J. F. Littles, R. L. Bree, and R. K. Ten Haken, "Automated Localization of the Prostate at the Time of Treatment using Implanted Radiopaque Markers: Technical Feasibility," Int.J.Radiat.Oncol.Biol.Phys. **33,** 1281-1286 (1995).

CHAPTER II

RESPIRATORY-INDUCED ORGAN MOTION AND ITS MANAGEMENT

In the introductory chapter, an overview of the issues related to inter- and intra-fraction organ motion and deformation, and the uncertainties associated with its measurement and management, was provided. The focus of this chapter is the specific problem of respiration, looking at the magnitude of motion and deformation in various thoracic structures caused by breathing and the efficacy of different methods employed throughout the process to ensure target coverage while minimizing dose to the normal tissue. The uncertainties of various respiratory motion management strategies are evaluated, and their implication on the treatment accuracy is discussed.

II.A. Introduction

For targets in the thoracic region, tumor motion caused by breathing is a significant source of position uncertainty that impacts all stages of the radiation therapy process, including image acquisition, treatment planning, and radiation delivery. Accurate target localization at the time of treatment delivery allows for the creation of more conformal plans that reduce the volume of the normal tissue irradiated, while higher doses are delivered to the target. To account for respiratory motion and its impact on treatment delivery, various techniques are employed in the clinic. The choice of the motion management technique depends on different factors, such as the availability of technologies at each institution, the treatment planning technique, as well as patient specific characteristics and limitations.

The report of the AAPM task group 76 on the management of respiratory motion in radiation therapy divides methods of accounting for respiratory motion into several categories, including

breath-hold techniques, gating, forced shallow breathing, real-time tracking, and motion-encompassing methods[1]. Some of the techniques that are more extensively described in the literature include gating[2-4], passive or active breath-holding[5-11], and real-time tracking of the tumor for adaptive beam delivery[12-15]. Respiratory gating and real-time tracking of the tumor compensate for the breathing motion by adjusting the delivery system to the breathing motion. In respiratory gating, the patient's breathing is monitored during the treatment delivery and the radiation beam is only turned on during a pre-determined phase or amplitude in the breathing cycle. These methods tend to make simplistic assumptions about the periodicity of the breathing pattern, which have been shown to be insufficient by Ozhasoglu *et al.,* who demonstrated that modeling the breathing pattern for use with compensation methods requires a comprehensive characterization of the individual patient's breathing pattern[16]. Langen and Jones compiled data from various studies of organ motion and its management[17].

II.B. Respiratory motion management

II.B.1. Breath-hold treatments with Active Breathing Control (ABC) tumor immobilization

In breath-hold treatments the patient's breathing is monitored during each treatment session, and the beam is only turned on while the breath is held at a predefined state in the breathing cycle. Based on the report of the AAPM task group 76[1], currently there are various breath-hold techniques that are commonly used in the clinic, including self-held breath-holds with or without respiratory monitoring, active breathing control, and deep-inspiration breath-holds. This section focuses on active breathing control, which is the method currently used in the Department of Radiation Oncology at the University of Michigan.

Several studies focusing on the use of active breathing control (ABC) have shown it to be an effective method in reducing the respiratory motion by temporary suspension of breathing at a desired breathing state. Previous studies have evaluated the reproducibility of tumor immobilization using ABC, by examining different aspects of the breathing motion such as the movement of the chest wall and diaphragm, the change in the volume of the lungs, and the

displacement of the gross target volume (GTV), in lung and liver patients[5-10]. Wong *et al.* evaluated intra-fraction motion and volume changes by scanning the patient twice at inhale, defined as 80% of vital capacity (VC), once at exhale (15% VC) using ABC, and once during free breathing. They evaluated inter-fraction reproducibility by repeating the scan at a second CT session[8]. Dawson et al used hepatic microcoils and fluoroscopy to evaluate the inter-fraction and intra-fraction reproducibility in liver patients. They also evaluated the movement of the diaphragm[5]. Cheung et al showed an inter-fraction reproducibility of 0.3, 1.2, and 1.1mm in Right-Left (RL), Anterior-Posterior (AP), and Superior-Inferior (SI) directions for the center of mass of the gross tumor volume (GTV) from CT scans of patients taken on 5 consecutive days[6]. Another study used deformable alignment (Thin-Plate splines) to register automatically segmented lung volumes from inhale CT scans of the patient acquired in one day, and by moving the patient from the table between the scans taken at different sessions in one day[9].

In this section, the short-term and long-term reproducibility of ABC in immobilizing intra-thoracic tumor position is described. The short-term reproducibility describes the movement of the tumor between different breath-holds in one session, and long-term reproducibility describes the movement of the tumor from one session to the next.

II.B.1.i. ***Study design***

II.B.1.i.a. ***Patients***

Under an institutional review board-approved protocol, ten patients with non-small cell lung cancer were enrolled in this study. Inclusion criteria included ability to maintain a 20 second breath-hold with ABC, which is also the clinical criteria for patient eligibility for ABC treatments.

II.B.1.i.b. ***Data Acquisition***

Patients were scanned three times over the course of treatment. The first scan was the simulation scan acquired prior to the start of the treatment. The second and third CT scans were acquired after completion of 40% and 80% of prescribed total dose respectively (approximately after 4

weeks and 8 weeks of treatment). At each CT scan session, a minimum of four CT scans were acquired; two at inhale (75 – 80 % of vital capacity), one at half the lung volume of inhale scan and one at end exhale. One patient could not reach more than 41% of vital capacity for reproducible breath-hold and was scanned at that level instead. All breath-holds were achieved using the ABC device (Vmax, Sensormedics, Yorba Linda, CA) described by Dawson et al[5].

Occasionally more than one breath-hold was necessary to scan the entire thorax, which resulted in some noticeable discontinuity in some of the CT volumes acquired. Although this may result in some uncertainty in the determination of the motion in the overlap region, the scan regions were selected to make sure that the tumor would be completely scanned in one breath-hold. In 3 of these scans, discontinuities led to significant enough artifacts to exclude the scan and resulting data analysis from the study.

II.B.1.i.c. ***Registration method***

The registration software was developed at the Netherlands Cancer Institute (MATPEL, NKI). For this study we used rigid body transformations, allowing rotation and translation only. To account for setup errors between CT scans acquired at different sessions, CT volumes were registered to the first CT scan based on the spine using a rigid body transformation. Registration results were evaluated visually. Once satisfactory results were achieved for alignment of the spine, the transformation was set as reference and the CT volumes were subsequently registered on the tumor. The local transformation was found by masking the area outside of the region of interest (tumor on initial CT plus a small surrounding region). In aligning the tumor volumes, only translation was allowed to eliminate the possibility of additional errors that could be caused by rotational degeneracy of alignment of mostly convex tumors. Alignments were optimized by a simplex search using normalized mutual information as a figure of merit. All alignments were visually inspected to verify accuracy of regional registration. The movement of the tumor with respect to the spine was measured by placing a reference coordinate at the center of the tumor and measuring the relative transformation of this location.

II.B.1.ii. ***Data Analysis***

II.B.1.ii.a. ***Short-term reproducibility***

Short term reproducibility was evaluated by comparing the repeat inhale CT volumes of the patient acquired in one scanning session. We assumed that the patient does not move significantly between different scans acquired in one session, and therefore there was no need to correct for setup by aligning the patient based on the spine first. The first inhale acquired in each session was chosen as the reference and the second inhale was registered to this scan using only the tumor volume plus a small region around it.

The mean (standard deviation) of the displacement in the center of tumor were found to be 0.2(0.7), 0.3(1.4), and 0.0(1.5) mm in the RL, SI and AP directions. Maximum displacements as high as 1.7, 3.1, and 4.2mm in respective directions were observed, as shown in Table II.1.

Table II.1: Short term reproducibility of ABC decive.

Short-term reproducibility (Inhale)	R/L (mm)	S/I (mm)	A/P (mm)
Mean	0.2	0.3	0.0
Standard Deviation	0.7	1.4	1.5
Minimum	0.0	0.0	0.0
Maximum	1.7	3.1	4.2

The results from the short term reproducibility study were similar to results reported in other studies[6].

II.B.1.ii.b. ***Long-term reproducibility***

Long-term reproducibility was measured by evaluating the relative position change in the center of tumor between same breathing phase CT volumes from two different scanning sessions (two

exhales or two inhales). For consistency, the first inhale scan from the first session was used as the reference for the long term reproducibility of inhale breath holds. Changes in the size and shape of the tumor were observed between the first and the third scans due to tumor volume reduction over the course of treatment (Figure II.1). A maximum reduction of 22mm in the diameter was observed for one case, while for most cases this reduction was approximately 4 to 7mm. Therefore, for long term reproducibility studies, a registration was considered acceptable if the tumor volume at the later session was completely encompassed by the volume of the reference tumor from the first session, although the alignment generally centered the reduced tumor volume within the original (larger) pre-treatment tumor.

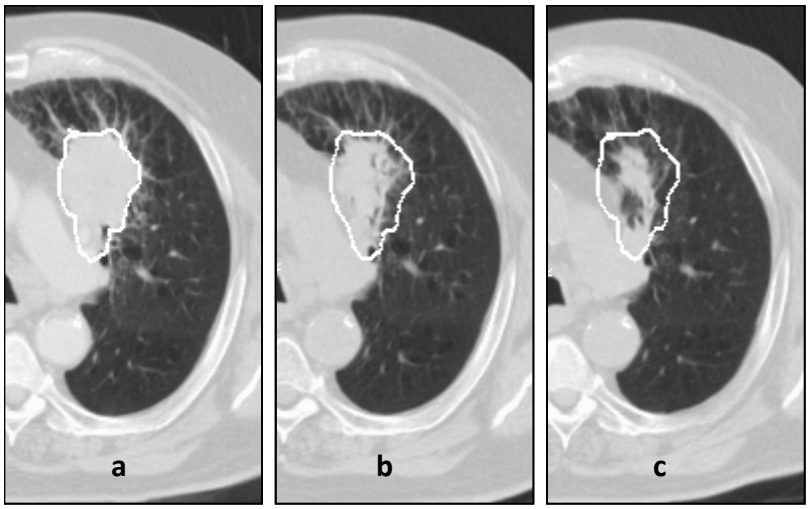

Figure II.1: Tumor shrinkage over the course of treatment. a)Tumor at the time of simulation, b) Tumor after 40% of dose was delivered, and c) tumor after 80% of dose was delivered

The long term mean (standard deviation) of the movement in the center of the tumor was measured similarly (Table II.2) and found to be 0.3(1.6), -0.5(3.8) and -1.3(3.1) mm for inhale and -0.7(1.1), 0.2(2.1), -0.2(2.8), mm for exhale, in RL, SI, and AP directions respectively. No general conclusions can be made based on this data regarding the direction in which the tumor moves during the course of treatment. A maximum displacement of 3.8mm in RL, 9.0mm in SI, and

6.8mm in AP directions was observed for inhale scans. The corresponding values for exhale scans were 3.6mm, 3.3mm, and 7.4mm in RL, SI, and AP directions. These maximum values correspond to a case with a large change in the shape of the tumor, where a good rigid tumor-based alignment was not possible, and a case where the tumor was very small (diameter of approximately 3mm with a volume of approximately 60 mm^3).

Table II.2. Comparison of the reproducibility of inhale and exhale breath-hold states

	Inhale			Exhale		
	R/L (mm)	S/I (mm)	A/P (mm)	R/L (mm)	S/I (mm)	A/P (mm)
Mean	0.3	-0.5	-1.3	-0.7	0.2	-0.2
Standard deviation	1.6	3.8	3.1	1.1	2.1	2.8
Minimum	0.1	0.0	0.4	0.1	0.3	0.2
Maximum	3.8	9.0	6.8	3.6	3.3	7.4

II.B.1.ii.c. ***Comparison of reproducibility of inhale and exhale breath-holds***

Reproducibility of breath-hold at exhale and inhale was also compared, as shown in Table II.2. Due to lack of availability of multiple CT scans at exhale for each session, only the long-term reproducibility of the exhale breath-holds was evaluated and compared to the long term reproducibility of the inhale scans to avoid any bias. Overall, the exhale breath holds seemed to be more stable than inhale. The mean displacement in tumor position in the AP direction showed an increase of about 1mm from exhale to inhale. Also, the standard deviation increased by less than 2mm from exhale to inhale in the SI direction, while the change in the maximum was about 5.5mm in the corresponding direction.

II.B.1.ii.d. ***Impact of tumor size and shape change on breath-hold reproducibility***

The change in reproducibility as a result of changes in shape and size of the tumor was investigated by comparing the pre-treatment CT scans of the patient to the scans acquired after 40% and 80% of the treatment was delivered. There was an insignificant change in tumor volume between the first and second sessions, while the CT scans from the third session showed a noticeable change in the volume and shape of the tumor. The mean and the standard deviation showed a small change between the first and second sessions compared to the first and third sessions, with a maximum change of 1.5mm in the AP direction. This change is considered insignificant compared to the large maximum displacements in both cases.

The effect of changes in shape and size of the tumor on the inter-fraction reproducibility of tumor position was also evaluated by comparing the reproducibility results between the first two sessions, with the reproducibility results between the first and third session (Table II.3). The mean displacement showed a small change in the AP direction (1.5mm), but the standard deviation did not change more than 1mm in any direction. The small value of the standard deviation compared to the large maximums observed, suggests that the differences between the reproducibility from first session to either the second, or third sessions are insignificant.

Table II.3. Reproducibility of tumor position for different degrees of tumor shrinkage

	S1 / S2			S1 / S3		
	R/L (mm)	S/I (mm)	A/P (mm)	R/L (mm)	S/I (mm)	A/P (mm)
Mean	-0.5	-0.5	0.0	0.3	0.2	-1.5
Standard deviation	1.4	2.8	2.5	1.5	3.5	3.3
Min	0.1	0.3	0.2	0.1	0.0	0.4
Max	3.6	5.8	6.3	3.8	9.0	7.4

Previous long term studies have looked at inter-fraction reproducibility by moving the patient from the table between the scans[9], or by repeating the scan several times during consecutive days at the beginning or prior to the start of treatment[6], where no significant change in size and shape of the tumor was present. Here we compared inter-fraction reproducibility over the course of treatment, and therefore the results were different from those reported previously.

II.B.1.iii. ***Discussion***

For both short-term and long-term reproducibility, the standard deviation was observed to be large compared to the mean. This can be attributed to the inter-patient variations in reproducibility. Factors such as differences in stability of breathing, the patient's ability to follow instructions, and the location of the tumor, can affect the reproducibility of each patient's breathing pattern and as a result, the location of the tumor from one breath-hold to the next.

Overall inter-fraction reproducibility had a larger mean and standard deviation compared to intra-fraction (short term) reproducibility by a factor of about two. Although, as mentioned before, the impact of setup uncertainties was accounted for by applying a skeletal alignment prior to evaluation of the reproducibility of tumor position in long term, other factors such as changes in the size and shape of the tumor, as well as changes in the breathing pattern as a result of treatment, are possible sources of discrepancy.

In comparing exhale and inhale reproducibility, slightly better results were observed for exhale breath-holds. Other investigators have reported similar findings. Nonetheless, the reproducibility of inhale, especially when coupled with pre-treatment image guidance, warrants reduced margins coupled with the reduced volume of normal tissue irradiated; as reported in deep-inspiration breath-hold studies[20]. However, the gain of reduced errors due to respiratory motion is small, and the potential reduction in margin depends on the relative contribution of other error sources, which also need to be carefully controlled.

Based on the data presented here, using ABC will not significantly reduce margins for all patients. However, patients who showed a good reproducibility were consistent over the course of treatment. This indicates the importance of more individualized, dosimetric approaches to planning, as opposed to margin-based planning, using population models, as will be discussed further in the next section.

The data collected in this study bring two issues to the forefront. The long-term changes in tumor configuration require thought about changing GTV and CTV relationships, and appropriate strategies for deciding whether plan modification associated with visible tumor volume reduction is warranted. In addition, the availability of this population data on reproducibility empowers the concept of initial planning that is robust to expected residual uncertainties. Such planning has already been demonstrated in the liver[13], and may also benefit lung cancer patients over existing margin-based planning paradigms. In addition, the ability of lung cancer patients to tolerate extended breath held states is reduced compared to patients with breast or intra-hepatic tumors. Modifications are being considered to make the ABC system more tolerable for these patients.

II.B.2. Forward-planning without margin expansions

To ensure coverage of the tumor in the presence of geometric uncertainties caused by internal motion of targets, a margin is added to the clinical target volume. The addition of this margin results in an internal target volume (ITV), which is then used for treatment planning. Though margin expansion is an easy way to account for the motion of the target and guarantee that the desired dose is delivered to a moving GTV, it also results in higher dose to the surrounding normal tissue that moves in and out of the field as the target moves. This is caused by the failure of the margin expansion methods to account for the fact that the moving GTV is not present in all regions of the ITV at all times during a motion cycle (e.g. a breathing cycle). In this section, a forward dose calculation inclusive of motion for treatment planning in the thoracic region is proposed and compared to geometric methods currently recommended by the ICRU[18]. It is

expected that the dose to normal tissue will be lower for this method compared to the conventional margin expansion methods.

II.B.2.i. ***Study design***

II.B.2.i.a. ***Patient model***

Input data requires CT-based models of the patient for the inhale and the exhale breathing states. The CT volume of the patient at exhale is used as the reference. Using this reference scan, and a deformation map between exhale and inhale (with possible intermediate states for non-linear motion), simulated models of the patient at inhale and mid-ventilation states are generated and/or related geometrically to the reference state for dose accumulation. If the intermediate breathing states are available (e.g. from a 4D CT reconstruction), separate deformation maps from the reference phase to these intermediate phases can be estimated, by avoiding the linear motion assumption for simulation of these states as described above. It should be noted, however, that regardless of the availability of CT volumes at these intermediate states, and depending on the application, it is possible to use the reference CT and the deformation map to simulate various states in order to eliminate the error in dose accumulation caused by the uncertainty in the deformation map accuracy. Registration error and uncertainty in geometric alignment will be discussed in more detail in Chapter III.

II.B.2.i.b. ***Deformation map***

A deformable alignment tool based on B-splines is used to find the transformation between the inhale and exhale datasets. In order to eliminate the impact of registration error on dose accumulation, simulated CT volumes were generated by deforming the exhale CT volume according to the deformation map from exhale to inhale. The mid-ventilation dataset was created, using the exhale to inhale deformation map with the assumption that each voxel moves in a linear path from exhale to inhale. This can be done using breath hold CT volumes of the patient, cone beam CT datasets, or respiratory correlated CT datasets at inhale and exhale or other breathing states. Figure II.2 shows an axial cut near the diaphragm for the original inhale scan of the patient

and the simulated inhale constructed from the exhale CT volume. This process will ensure that the registration error is removed from the dose accumulation process and that the difference between the final dose calculations from the different methodologies is not affected by the registration error.

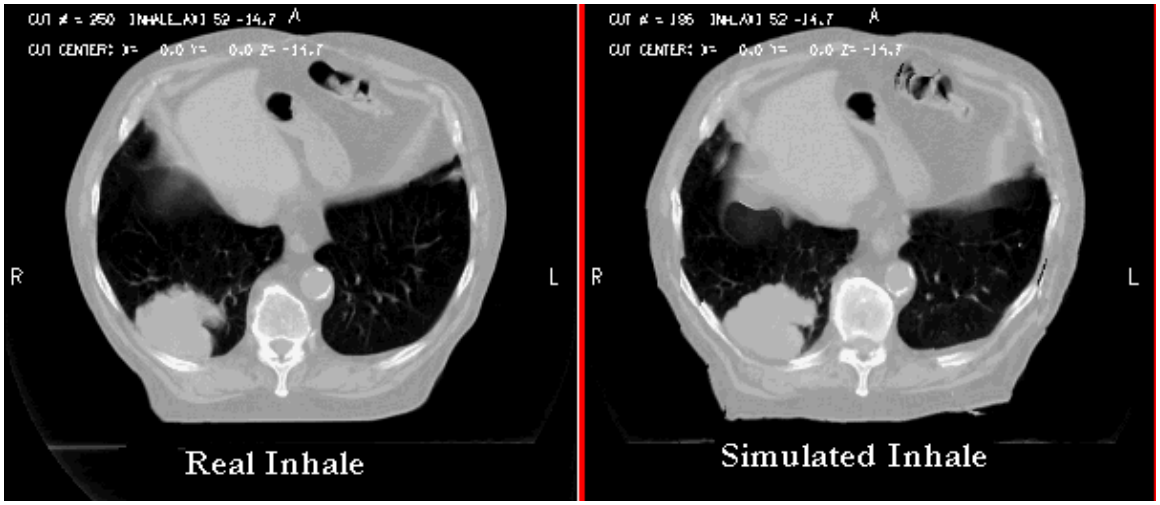

Figure II.2: Axial view of the real and simulated inhale state, indicating sufficient similarity to preserve the characteristics of the inhale lung that would affect dose accumulation.

II.B.2.i.c. ***Target definition***

The reference CT volume of the patient is used to define the gross tumor volume (GTV). Applying the deformation map to this volume at exhale, GTV volumes at inhale and mid-ventilation are generated. These volumes are then added to construct the internal target volume (ITV). Figure II.3.a shows the original exhale GTV in red and the ITV in yellow on a coronal image of the patient at exhale. Figures II.3.b and II.3.c show the simulated mid-ventilation GTV (blue) and simulated inhale GTV (green) on CT images from the corresponding breathing states. Figure II.3.d shows the three GTV volumes used to construct the ITV which is shown in yellow. It can be seen that the automatically constructed GTV volumes closely approximate the actual GTV (inspected visually only).

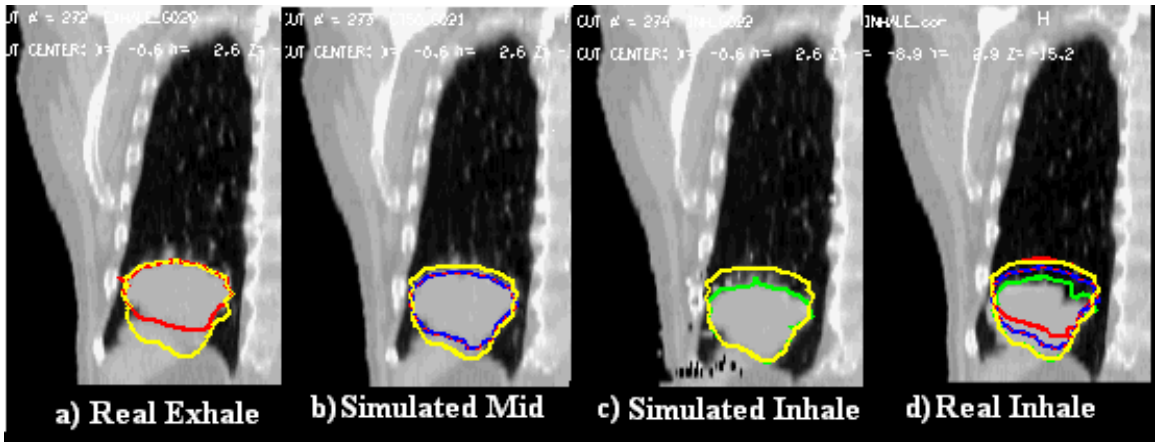

Figure II.3: Simulated GTV volumes and the resulting ITV calculated as a union of the GTVs at exhale, mid-ventilation and inhale

II.B.2.ii. ***Treatment planning method and dose calculation***

The two methodologies are compared in the flowchart shown in Figure II.4. On the left side of this flowchart is the standard ITV based treatment planning, with inclusion of the impact of breathing motion on the total accumulated dose. On the right side the different steps of the alternative forward planning technique is displayed.

II.B.2.ii.a. ***ITV planning***

A conformal treatment plan is created to cover the ITV volume with 95% dose. The initial plan is designed to cover the ITV on a static dose calculation using the reference phase for electron density mapping. In order to compare this method to the GTV-based forward planning method, the effect of breathing motion on the cumulative dose to the target and surrounding normal tissues is accounted for by calculating doses to the density distributions of each simulated breathing state, and then summing doses in the frame of the reference state (exhale) with equal weights.

II.B.2.ii.b. ***Forward planning based on a moving GTV***

In the forward planning method a conformal plan is made to the reference GTV and dose is calculated inclusive of motion as explained above. After evaluating the estimated delivered dose

distribution, the MLC leaves are manually adjusted in regions where the GTV does not receive sufficient dose. This process is repeated (typically three to four times) until the GTV is covered with 95% dose under estimates of breathing.

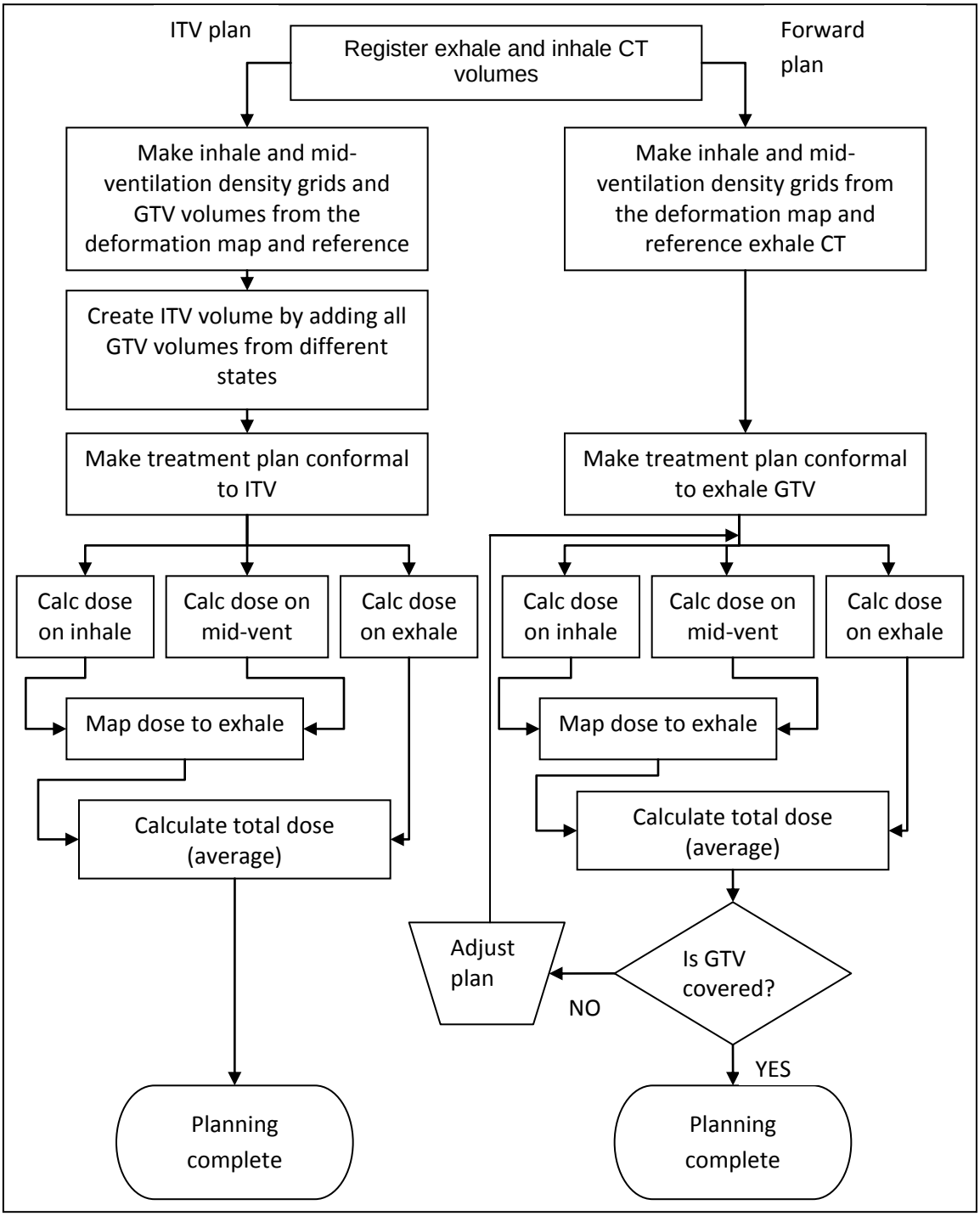

Figure II.4. Flowchart of two different planning methods based on ITV and GTV with motion

II.B.2.iii. ***Data analysis***

The impact of breathing motion on total dose to the GTV and the lungs was accounted for in both planning methods. The beam arrangement was identical in both plans, and adjustments were made by moving the MLC leaves only. The plans produced similar doses to the GTV at exhale, as shown in Figure II.5. The dose volume histograms (DVH) of the lungs showed a decrease of about 3% in the lung volume irradiated to different doses for the forward planning infrastructure compared to the standard ITV planning. A composite plan was also evaluated, in which the total dose inclusive of motion from the ITV plan was subtracted from the dose from the forward plan. The results showed no difference in the GTV region, but the forward plan was colder above and below the GTV in the lung and the liver (Figure II.6). These cold regions are expected, since the MLC leaves are more open around this area for the ITV plan.

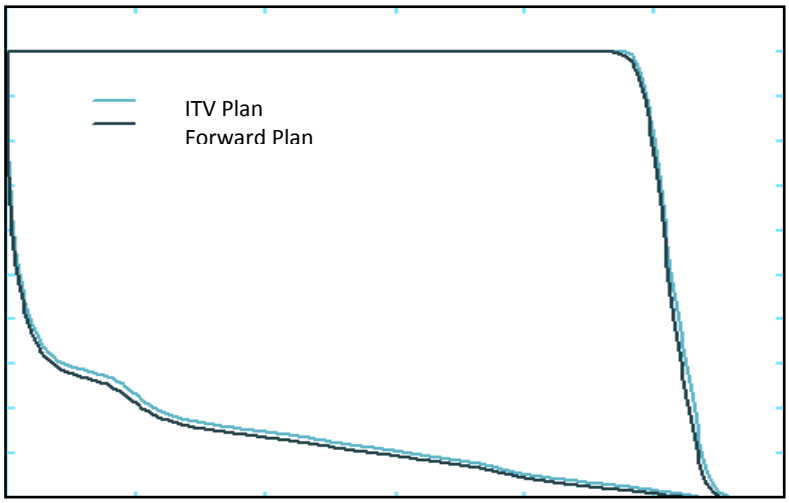

Figure II.5: se volume histograms for the exhale lung and the exhale GTV from ITV plan (blue) and the forward plan (grey)

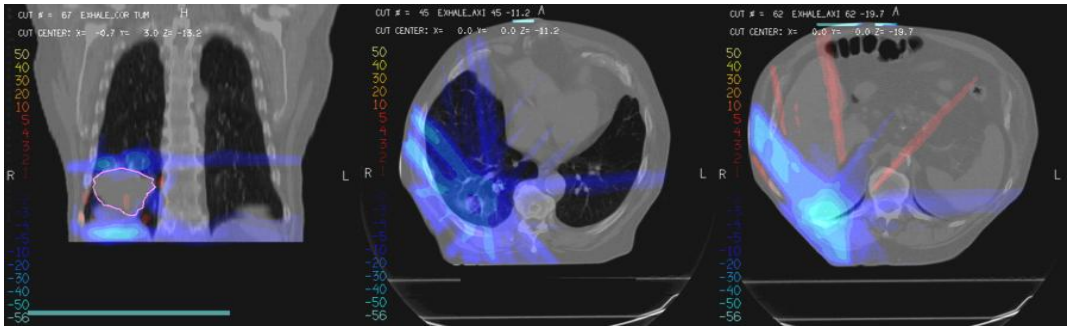

Figure.II.6: Dose difference between the GTV-based forward plan and the ITV plan. It is observed that the forward plan is colder than the ITV plan in regions outside of the GTV volume. The image on the left shows a coronall cut through the center of the tumor. The center image shows an axial cut, 3 mm above the top of the GTV volume and the image on the right is 3mm below the GTV volume.

II.B.3. Discussion and limitations

In this section, we improved our target motion measurement and motion management by moving from simple motion estimation based on rigid alignment to tracking every single voxel in patient geometry, using deformable alignment. Although we removed the effect of the error in the estimation of voxel motion by simulating the deformed inhale and mid-ventilation states for dose calculation for the purpose of this study, in the case of a real patient these errors need to be accounted for. The impact of registration error on estimated dose distribution will be further investigated in chapter V.

This study was limited by the number of breathing states used in calculating the total accumulated dose inclusive of motion. Although there are studies showing that the two breathing states may be sufficient for accurate dose accumulation[20], using more breathing states may have resulted in more accurate results for this comparison. More importantly, in averaging the deformed dose from various breathing states, we made the assumption that the patient spends equal amounts of time at all breathing states, which is not necessarily accurate. More accuracy in the results can be gained by using a more accurate weighting in summing the dose.

II.C. Summary

The studies discussed here, from the measurement of motion, to respiratory motion management with breath-hold devices, to forward planning based on GTV, all show certain levels of uncertainty, and limitations in the highest levels of accuracy achievable. Therefore, more accurate methods of accounting for this uncertainty in the evaluation of the accumulated dose and treatment outcome are necessary. To do this, we first need to accurately measure the geometric error distribution in tracking with deformable alignments, and then estimate its impact on the deformed dose. The accuracy of image registration will be discussed in the next chapter.

References

[1]Keall PJ, Mageras GS, Balter JM, *et al*. "The managment of Respiratory Motion in Radiation Oncology Report of AAPM Task Group 76,"

[2]Liu HH, Koch N, Starkschal G, *et al*. Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part II – Margin reduction of internal target volume. *Int J Radiat Oncol Biol Phys* 2004;60:1473-1483.

[3]Engelsman M, Sharp GC, Bortfeld T, *et al*. How much dose reduction is possible through gating or breath hold?. *Phys Med Biol* 2005;50:477-490.

[4]Ohara K, Okumura T, Akisada M, *et al*. Irradiation synchronized with respiration gate. *Int J Radiat Oncol Biol Phys* 1989;17:853-857.

[5]Dawson LA, Brock KK, Kazanjian S, *et al*. The reproducibility of organ position using active breathing control (ABC) during liver radiotherapy. *Int J Radiat Oncol Biol Phys* 2001;54:1410-1421.

[6]Cheung PC, Sixel KE, Tirona R, *et al*. Reproducibility of lung tumor position and reduction of lung mass within the planning target volume using active breathing control (ABC). *Int J Radiat Oncol Biol Phys* 2003;57:1437-1442.

[7]Remouchamps VM, Letts N, Yan D, *et al*. Three-dimensional evaluation of intra- and interfraction immobilization of lung and chest wall using active breathing control: a reproducibility study with breast cancer patients. *Int J Radiat Oncol Biol Phys* 2003;57:968-978.

[8]Wong JW, Sharpe MB, Jaffray DA, *et al*. The use of active breathing control (ABC) to reduce margin for breathing motion. *Int J Radiat Oncol Biol Phys* 1999;44:911-919.

[9]Sarrut D, Boleda V, Ayadi M, *et al*. Non-rigid registration method to asses the reproducibility of breath holding with ABC in lung cancer. *Int J Radiat Oncol Biol Phys* 2005;61:594-607.

[10]Kashani R, Balter JM, Haymann JA, Short-term and long-term reproducibility of lung tumor position using active breathing control (ABC). *Int J Radiat Oncol Biol Phys* 2006;66:1553-9.

[11]Kim D, Murray B, Halperin R, *et al*. Heald-breath self-gating technique for radiotherapy of non-small-cell lung cancer: A feasibility study. *Int J Radiat Oncol Biol Phys* 2001;49:43-49.

[12]Shirato H, Shimizu S, Kitamura K, *et al*. Four –dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor. *Int J Radiat Oncol Biol Phys* 2000;48:435-442.

[13]Seppenwoolde Y, Berbeco RI, Nishioka S, Shirato H, Heijmen B. Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: a simulation study. Med Phys. 2007 Jul;34 (7):2774-84.

[14]Neicu T, Shirato H, Seppenwoolde Y, Jiang SB. Synchronized moving aperture radiation therapy (SMART): average tumour trajectory for lung patients. Phys Med Biol. 2003 Mar 7;48(5):587-98.

[15]Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, van Herk M, Lebesque JV, Miyasaka K. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. Int J Radiat Oncol Biol Phys. 2002 Jul 15;53(4):822-34.

[16]Ozhasoglu C, Murphy MJ. Issues in respiratory motion compensation during external-beam radiotherapy. *Int J Radiat Oncol Biol Phys* 2002;52:1389-1399.

[17]Langen KM, Jones DT, Organ motion and its management, *Int J Radiat Oncol Biol Phys* 2001;50:265-78

[18]ICRU Repost 62. Prescribing, Recording, and Reporting photon beam therapy (supplement to ICRU Report 50). Bethesda MD, International Commission on Radiation Units and Measurements, 1999.

[19]Hanley J, Debois MM, Mah D, *et al.* Deep inspiration breath-hold technique for lung tumors: The potential value of target immobilization and reduced lung density in dose *escalation. Int J Radiat Oncol Biol Phys* 1999;45:603-611.

[20]Rosu M, Balter JM, Chetty IJ, Kessler ML, et al "How extensive of a 4D dataset is needed to estimate cumulative dose distribution plan evaluation metrics in conformal lung therapy?" Med Phys. **34** 2007 233-45

CHAPTER III

QUANTITATIVE EVALUATION OF IMAGE REGISTRATION ACCURACY

In Chapters I and II, the limitations in accuracy of treatment delivery due to respiratory-induced motion and deformation in the targets and normal structures were discussed. This chapter describes the use of a deformable phantom with a large number of markers, for an objective test of various deformable image registration methods and to investigate potential uncertainties that may be overlooked by commonly-used validation techniques.

III.A. Introduction

Image registration has become an essential part of many stages in the radiation therapy process, due to the increase in use of daily imaging for setup adjustment, and the availability of new scans for re-planning during the course of treatment. In particular, deformable image registration has been used in many applications, ranging from dose accumulation and contour propagation for adaptive therapy[1-13] to generation of analytical models of breathing motion based on deformation maps of the thorax[14]. Some of these applications are extremely sensitive to the results of image registration. For example, when deformable alignment is used for dose accumulation in adaptive therapy, small errors in the deformation map can result in significant changes in the dose at points in high dose gradient regions. If deformable alignment is to be used for these sensitive applications, it is necessary to have a quantitative measure of the accuracy of the resulting deformation maps.

Quantitative evaluation of image registration is a difficult task. Current methods that are used include use of analytically deformed images, some of which consider the biomechanical properties of the patient[15-20]. These studies range in complexity from simulation of simple

deformations applied to the image through random motion of different points, to more complex biomechanical models that simulate physical tissue deformations using finite-element methods[12-17]. However, simulated images do not have all the noise and other imaging artifacts of real images, which may impact the true alignment accuracy. Other validation techniques include a voxel-based evaluation method, proposed by Zhong et al., which automatically detects a region in an image where the registration is not performing well using a finite-element-based elastic framework to calculate the unbalanced energy in each voxel after substitution of the displacement vector field[25]. Use of phantoms with a known physical deformation, or phantoms with easily identifiable markers where motion can be accurately measured, is another method investigated by other groups[15,26]. Studies describing deformable registration methods, as well as those validating a method for a specific anatomical site, have all used one or more of the techniques described here, for validation of their results.

Perhaps the most commonly used registration validation methods are those based on the use of contours and bifurcations manually identified on both image sets for comparison[15, 21-24]. However, precision in identifying these landmarks is greatly dependent on image resolution, slice thickness, and the user. Brock et al reported the average precision in selecting the location of these landmarks to be better than 1 mm, based on repeat measurements done by one individual once a week over a four week period.[8] They also showed a maximum standard deviation of 4.7mm for one bifurcation in the superior/inferior direction. Using anatomical landmarks for evaluation of image registration results also has the disadvantage that those with enough contrast to be accurately identifiable are likely to contribute significantly to the goodness-of-fit metrics used for alignment, and thus have an impact on the registration results. Therefore, evaluation of accuracy at these positions could be biased and not necessarily representative of other locations in the volume. Hence, an objective assessment of image registration is necessary, where the points chosen for validation are not the driving forces in the local deformation parameters, and their scarcity does not mask the difference in global versus local registration accuracy estimates.

This chapter introduces the design and implementation of a deformable phantom for an objective evaluation of various registration techniques. The phantom is embedded with small but easily identifiable reference marks, to calculate the true deformation as the baseline for comparison of all image registration results. The markers are then removed from the images prior to registration to eliminate their possible impact on alignment, as well as any potential bias in the reported results. Using these image volumes, we test our in-house registration techniques, and then describe a multi-institution blind study of other implemented registration techniques.

III.B. Phantom design

III.B.1. Phantom components

A diagnostic thoracic phantom (RS-330, Radiology Support Devices, Long Beach, CA) provided the main shell of the phantom. This phantom has a skeleton, as well as a lung-equivalent insert, with additional tissue-mimicking sections for a mediastinum and various tumor-simulating nodules. An insert for an abdominal cavity, composed of a uniform density material, was removed. In its place, an extension of the existing lung insert was created using high density foam (Figure III.1). This process created a section of the phantom with a rigid skeleton and tissue wall, surrounding a compressible internal cavity.

Figure III.1: The diagnostic phantom with the abdominal insert removed (left), the high density foam insert (right)

Tumor-simulating inserts of varying density and size were embedded in the foam. To facilitate initial studies focusing on imaging and geometric localization, the structures selected were rigid

objects of known shape (balls) and various compositions (hard rubber, plastic, porous foam) to generate varying electron densities. The objects were inserted at various locations by small incisions in the foam, and further held in place with glue.

An actuator-driven diaphragm was created to compress/decompress the foam according to previously defined arbitrary breathing profiles. A single-axis actuator, previously designed for motion experiments and similar to systems described by other investigators[8-11], was employed. The "diaphragm" was created by cutting a 1 cm thick sheet of Lucite to a size slightly smaller than the cavity into which the foam insert was placed. This diaphragm was attached to the driving rod of the actuator via a wooden brace, with a three-point attachment to distribute the force of the actuator uniformly across the Lucite plate (Figure III.2). The phantom was braced cranially to limit the possible rigid body movement of the whole system under compression and relaxation.

Figure III.2: The set up of the phantom and motion actuator on the CT scanner

The actuator system has a control software environment that permits customized positioning, and complex temporal motion profiles to be programmed and executed. Breathing patterns were programmed that represented periodic functions[12], variations on periodicity and amplitude to simulate known but irregular breathing, and finally measured position/amplitude variations from

patients studied under a review board-approved protocol using an external monitoring system (RPM, Varian, Palo Alto CA). Although significantly larger motions could be programmed, the tests performed used diaphragm compressions of up to 3 cm (20% volume reduction of the compressible cavity).

III.C. System Performance

III.C.1. Motion under compression and reproducibility

Important characteristics of this phantom include differential motion of various points under compression, as well as reproducible configuration at the same amount of compression (breathing state). To evaluate reproducibility, CT scans were taken of the phantom under various states of compression (Figure III.3). These scans were repeated in the same session, as well as in a separate session 3 months later to evaluate long-term reproducibility of position at a given compression state.

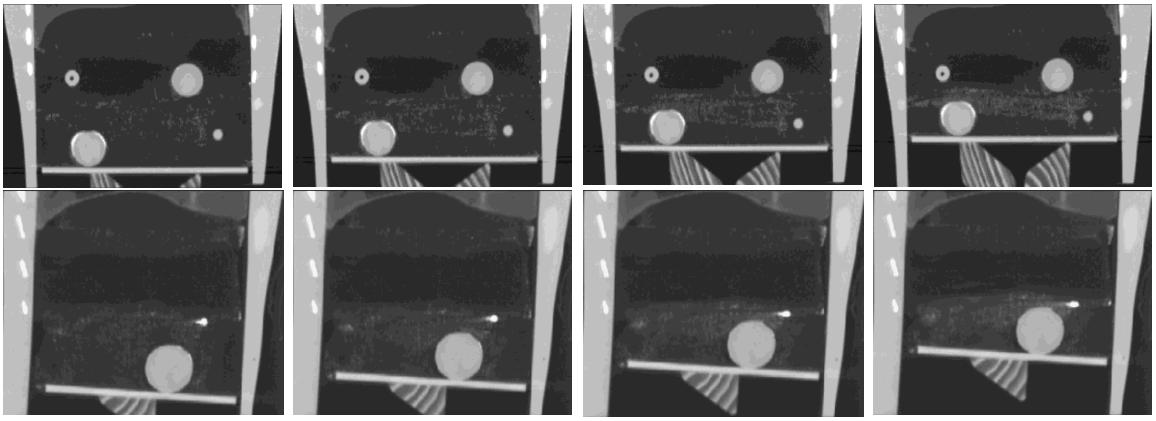

Figure III.3: Coronal (top) and sagittal (bottomn) views of the CT scan of the phantom with the foam and tumor simulating inserts under compressions (diaphragm movements) of 0, 1, 2, and 3cm (from left to right).

Table III.1 shows the reproducibility of tumor positions under repeat compressions. Intermediate tumor positions are reproducible to within 1.3 mm. Table III.2 shows the differential movement of targets in various locations within the phantom. As expected, targets closer to the diaphragm moved farther than those more cranially positioned. Compressions of over 50% of the foam

insert have been demonstrated, with regional target motion ranging from 95% of the actuator motion at the diaphragm to <30% near the “apex” (the interface with the existing lung insert).

Table III.1: Reproducibility of tumor position between repeat compressions. For each diaphragm compression, the position of the center of the tumor in the first scan was taken as the reference and compared to the second scan. The results presented here include all

	Right / Left	Anterior/Posterior	Superior/Inferior
Average (mm)	-0.3	-0.2	-0.2
Standard deviation (mm)	0.6	0.5	1.3
Maximum (mm)	1.3	0.8	2.9

Table III.2: Motion of tumor simulating inserts under different diaphragm compressions. For each diaphragm compression (shown at the top), the actual motion in centimeters and the motion as a percentage of the diaphragm compression are shown. The error in each measurement is shown in paretheses.

Distance to diaphragm (cm)	Compression = 1.2cm Tumor motion (error)		Compression = 2.1cm Tumor motion (error)		Compression = 3.0cm Tumor motion (error)	
	cm	%	cm	%	cm	%
1.31	1.01 (0.05)	84 (4.2)	1.89 (0.04)	90 (1.9)	2.85 (0.10)	95 (3.3)
2.05	0.85 (0.09)	71 (7.5)	1.69 (0.10)	80 (4.8)	2.53 (0.06)	84 (2.0)
4.33	0.55 (0.10)	46 (8.3)	1.24 (0.12)	59 (5.7)	1.87 (0.15)	62 (5.0)
8.53	0.30 (0.07)	25 (5.8)	0.60 (0.10)	29 (4.8)	0.90 (0.15)	30 (5.0)
8.77	0.22 (0.23)	18 (19)	0.53 (0.07)	25 (3.3)	0.83 (0.04)	28 (1.3)
10.55	0.24 (0.16)	20 (13)	0.45 (0.15)	21 (7.1)	0.84 (0.06)	28 (2.0)
10.63	0.25 (0.35)	21 (29)	0.37 (0.16)	18 (7.6)	0.76 (0.06)	25 (2.0)

Repeat reproducibility studies performed 3 months after the first study indicated that the relative local reproducibility stays the same (mean of zero and standard deviation of less than 1.1mm).

III.C.1.i. ***Relative attenuation***

The insert was evaluated for relative attenuation (effective attenuation relative to water) using a commercial CT scanner (HiSpeed, General Electric, Milwaukee WI). Scans demonstrated a relative attenuation of 5-10% of water. To mimic lung attenuation for diagnostic kV imaging, the foam insert was infused with iodine. This was accomplished by soaking the insert in a solution of iodinated contrast agent (Ultravist 300, Berlex Laboratories, Montville NJ) diluted in water in an approximately 5:1 (water:ultravist) mix. After infusion, the foam was allowed to dry. Subsequent scanning demonstrated Hounsfield Units of -800 to -600, more consistent with lung attenuation as reported by previous investigators[7].

The relative attenuation (based on Hounsfield Units) in the foam during the "inhale" (uncompressed) state had a mean value of 0.177 with a standard deviation of 0.143 indicating a large variation in attenuation between voxels. Under 3 cm of compression, these values changed to a mean of 0.196 and standard deviation of 0.139. An example from patient data acquired under a review board-approved protocol, with similar diaphragm movement (approximately 3cm +/- 3mm) showed a change in lung volume of 17% and a change in density from 0.202 at inhale to 0.244 at exhale.

The variations in local intensity distribution of the phantom lung were compared with a typical lung CT image. While not the same as the substructure seen in thoracic images (or lack thereof in scans of the liver), a histogram comparing the two shows reasonably similar variations in intensity (Figure III.4).

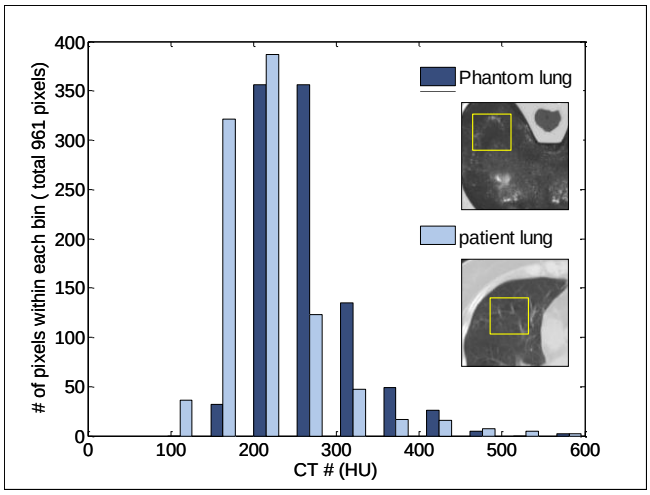

Figure III.4 : Histogram of the intensity distribution in a typical section of the foam at the simulated exhale state (dark blue), and the intensity distribution of a sample segment from an image of the left lung at exhale (light blue).

III.C.2. Imaging and analysis

The foam insert was modified by placing 48 small (2.5 mm diameter) plastic markers at different locations in the foam (figure 5). The phantom was then imaged at 1cm (inhale) and 4cm (exhale) compressions of the foam insert, using a commercial CT scanner (HiSpeed, General Electric, Milwaukee WI). Axial images were acquired with 1mm slice thickness and separation.

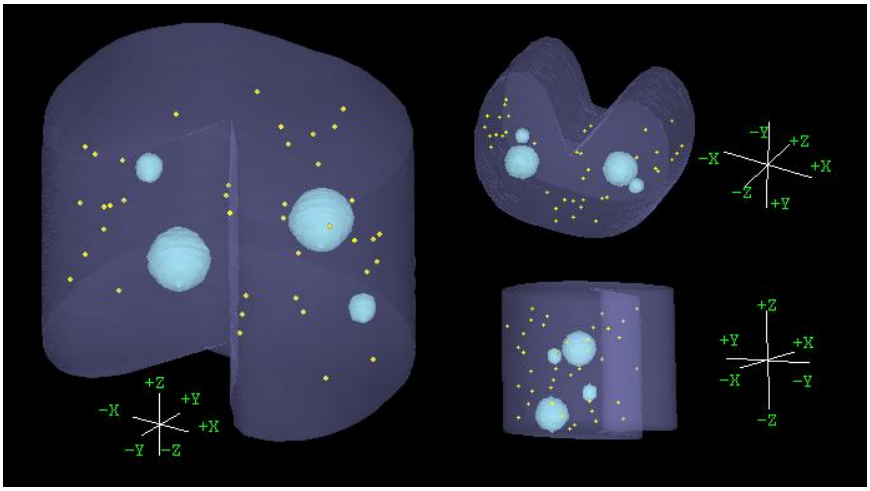

Figure III.5: Volumetric views of the surface of the deformable lung insert, the solid tumor-simulating spheres, and the plastic markers, are shown at 1cm compression of the foam (inhale). Z: Superior/Inferior, Y: Anterior/Posterior, X: Right/Left directions.

III.C.2.i. *Marker localization accuracy*

For each image set, the position of the center of each marker was manually measured by locating the central slice (longitudinal position) and then finding the center of the marker on that slice (Axial position). This introduces an uncertainty in the measurement of the longitudinal position that is limited by the slice thickness of 1mm. In order to evaluate the accuracy of the manual measurement of marker location in the axial plane, a random set of ten markers were chosen and the measurement of their position was repeated 3 times by a single observer. The standard deviation in the measurements for each marker was calculated, and averaged over all markers as a measure of uncertainty in our estimate of the true motion and deformation.

The accuracy of manual reference mark identification, which was determined by repeat measurements of the location of ten randomly chosen reference marks by a single observer, was found to be less than 0.2mm (σ) in all directions. The maximum standard deviation of any single measured point location was 0.3, 0.4, and 0.6 mm in RL, AP, and SI directions. It should be noted that this error applies to identification of reference marks on each of the two image sets. Therefore the overall accuracy in the measurement of the true motion of the reference marks is 0.4, 0.5, and 0.8 mm in RL, AP, and SI directions. Of note, the localization of anatomical landmarks in real patient images may not be as accurate, due to image distortion caused by breathing, as well as a larger slice thickness, which significantly increases the uncertainty in determination of the SI position of the landmark.

III.C.2.ii. *Image processing*

Once the position of the markers was identified on both image sets, the markers were digitally removed from the images prior to image registration. The voxel values in the marker locations were replaced by intensity values of neighboring voxels outside of the markers. Next, Gaussian smoothing with a kernel width of 20 voxels was applied to the intensities of the voxels in the marker locations. This reduces the potential impact of reference marks on the registration

outcome. Figure III.6 shows an original axial image on the left and the image without the markers on the right.

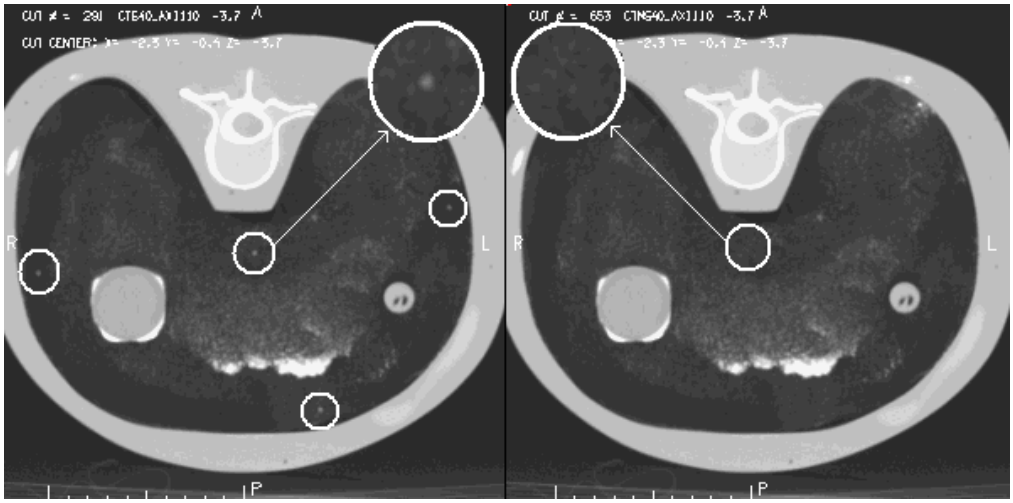

Figure III.6: Axial view of a plane with 4 markers before (left) and after (right) removal of the markers from the image.

III.D. Image registration

III.D.1. Method validation

Image registration was performed on the modified images without the markers, using different in-house algorithms such as rigid and affine transforms, as well as Thin-plate splines and B-splines.[9] One set of images (1cm compression or inhale) was chosen as the reference in all cases, and for each algorithm several combinations of the relevant parameters (resolution, knot spacing, the number of control points) were tested to get visually acceptable results. Once the registration was done, the resulting transform was applied to the position of the reference marks in the reference image set to estimate their location in the other image set. This estimate was then compared to the actual position of the markers measured manually, and the accuracy of the image registration was evaluated.

The results for the average and standard deviation, as well as maximum differences between the manually measured locations of the reference markers and those locations predicted by different

alignment algorithms, are shown in Table III.3. The average 3-D distance between the markers' estimated and actual locations was also calculated.

Table III.3: Error in estimation of marker position based on example alignment results from different image registration algorithms

		RL (cm)	AP(cm)	SI (cm)	3-Ddistance (cm)
Affine	Average	-0.01	0.00	0.05	0.38
	Stdev	0.04	0.04	0.44	0.22
	Max	-0.12	-0.13	0.90	0.90
B-splines	Average	-0.02	-0.01	0.05	0.18
	Stdev	0.08	0.06	0.22	0.16
	Max	-0.42	0.19	0.67	0.81
Thin-plate splines	Average	-0.07	-0.15	-0.14	0.37
	Stdev	0.12	0.19	0.28	0.19
	Max	-0.56	-0.58	-0.74	0.75

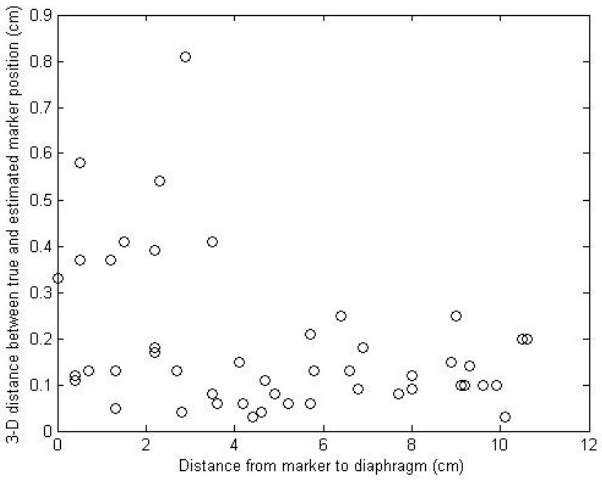

Figure III.7: The 3D distance between the real marker position and the estimated position based on B-spline transformation, shown as a function of the longitudinal position of the marker relative to the diaphragm. It can be seen that the error in the registration decreases in regions farther away from the diaphrahm.

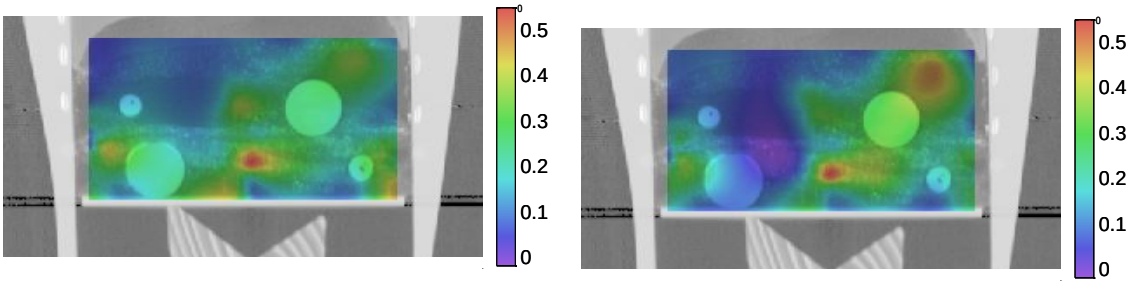

Figure III.8: Colorwash of the mean (left) and standard deviation (right) 3-D error (cm) in the phantom lung, interpolated to all other regions inside the foam.

III.D.1.i. ***Comparison of registration accuracy with and without markers***

To determine if, in the case of the alignment methods tested, a bias would be introduced by the markers in the image, a second set of alignments was performed on the unedited image volumes using our implementation of B-splines (Method 5). In the case of this phantom, comparison of image registration on datasets with and without the markers (with the markers removed) showed no significant difference in the results of the image registration accuracy.

III.D.1.ii. ***Comparison of registration accuracy for different deformation states***

The test was also performed on two additional deformation states of the phantom (1cm and 2cm diaphragm compressions) for the same parameter settings. Similar registration accuracies were obtained for these deformation states of the phantom, indicating that the values reported here are not affected by the deformation state of the phantom.

III.D.1.iii. ***Comparison of registration accuracy in the phantom to a sample lung patient***

In order to estimate whether the magnitude of variations seen in the phantom study are on par with those in clinical image alignment scenarios, the accuracy of alignment in breath-held CT scans of a patient at inhale and exhale states was evaluated using manually identified landmarks at visible bifurcations of vessels and bronchioles as true locations[22]. Deformable alignment was performed using our in-house b-splines based method with the same parameter settings (knot spacing and image resolution) as those used for the phantom alignment.

In comparing the sample patient registration accuracy to the phantom results, we saw no significant changes between the mean or maximum error in the SI direction, with a mean of 0.2mm and 0.5mm, and maximum of 3.1mm and 3.3mm for the phantom and the sample patient respectively (Table III.4). However, the patient data showed larger registration errors in the AP and RL directions, and therefore the 3-D vector error was also larger for the sample patient compared to the phantom. This is expected considering the minimal motion and deformation of the phantom in the AP and RL directions.

Table III.4: Comparison of registration results for the phantom and a sample patient or an in-house b-spline based method, using the same parameter settings. This table also shows the results for error in registration of the inhale state to other deformation states of the phantom.

		RL (mm)	AP (mm)	SI (mm)	3D (mm)
Sample patient	$\bar{d}$	-1.0	-0.8	-0.5	2.2
	σ	1.2	2.1	1.8	2.4
	$d^{\max}$	-3.3	-5.1	-3.3	6.9
Phantom 3cm diaphragm compression	$\bar{d}$	0.0	-0.1	-0.2	0.8
	σ	0.4	0.4	0.8	0.6
	$d^{\max}$	-1.7	-1.3	-3.1	3.1
Phantom 2cm diaphragm compression	$\bar{d}$	-0.2	-0.2	-0.2	0.7
	σ	0.3	0.4	0.6	0.4
	$d^{\max}$	-1.5	-1.4	-2.3	2.3
Phantom 1cm diaphragm compression	$\bar{d}$	-0.3	-0.2	-0.1	0.7
	σ	0.3	0.3	0.7	0.5
	$d^{\max}$	-1.1	-1.0	-2.8	2.8

$\bar{d}$: Mean error, σ : Standard deviation of error , $d^{\max}$:maximum error in estimation of marker position.

III.D.1.iv. ***Multi-institution registration analysis***

The feasibility of applying the developed methodology towards a broader analysis of alignment accuracy was tested by conducting a multi-institution blind study of the accuracy of deformable image registration algorithms.

III.D.1.iv.a. ***Study design***

A total of 8 alignment methods were tested at 6 institutions. Each institution was provided with the modified inhale and exhale images from the phantom (in DICOM format) as well as the coordinates of the markers on the inhale dataset. The position of the markers on the exhale dataset was not provided to participants until after completion of the study, but a single, easily identifiable point was chosen and its coordinates on both datasets were sent to all participants to

ensure the consistency of the coordinate systems at all institutions. At each institution, expert users who were algorithm developers or primary users were asked to align the inhale dataset to the exhale dataset using their technique of choice. There were no restrictions on time or pre-processing (i.e., masking or cropping) of the image sets, but prior assumptions about the nature of motion or deformation inside the foam were not allowed. For example, despite the fact that the foam was compressed in the longitudinal direction only, the registration could not assume that the motion is solely cranial-caudal. Once a satisfactory alignment was achieved, as determined by the user, the resulting deformation map was used to transform the coordinates of the markers on the inhale dataset to estimate their positions on the exhale dataset. Each institution then reported their estimated marker positions for comparison to the manually measured locations.

III.D.1.iv.b. ***Registration methods***

This section provides a brief overview of each registration method tested in this study, as well as references to publications that describe each technique in more detail. Table III.5 summarizes this information.

Table III.5: Summary of registration methods and references

Method	Model	References	Comments
1	Thin-plate splines	22	Cropped to foam
2	Thin-plate splines	31,32,33	No cropping or masking
3	B-splines	34,35	No cropping - masked the vertebrae
4	B-splines	36	No cropping - masked the vertebrae
5	B-splines	30,37	Cropped to foam
6	Demon's algorithm	15,38	No cropping or masking
7	Fluid flow	1,39,41	No cropping or masking
8	Free-form with calculus of variations	40	No cropping or masking

Method 1. Thin-plate splines with manual control point selection and mutual information (MI) as a similarity measure

This method is an implementation based on thin-plate splines, where control points are chosen manually on both image sets as an initial estimate of the transformation between the two geometries. The positions of the homologous points are iteratively manipulated by a Nelder-Mead simplex algorithm to maximize the mutual information between the two image sets, using thin-plate splines as an interpolant. In this study the reference image set (inhale) was automatically cropped to the lungs (foam) prior to registration, and a total of 29 control points were distributed throughout the foam. This implementation of thin-plate splines has been described and evaluated for alignment of inhale and exhale lungs previously[22].

Method 2. Thin-plate splines with automatic control point selection

In this method, the transformation matrix that relates a point on the moving image to its correspondence in the fixed image is found using a thin-plate spline (TPS) deformable model, to model the deformation of the phantom[31-32]. Currently, the TPS method still needs manual placement of control points and this work automates the control point selection by using the SIFT (scale invariant feature transformation) tissue feature searching[33]. Roughly 200 control points are selected based on the prominent tissue features as identified by the SIFT.

Method 3. Multi-resolution B-splines using correlation ratio as a similarity measure

This method is an original implementation of the non-uniform multi-level free-form deformation framework described by Schnabel et. al. with a multi-resolution extension[34]. For this study, a rigid transformation based on the tumor center displacement was applied as an initial step followed by a B-spline registration, with control point (knot) spacing starting at 80mm and going down to 20mm in three steps. The correlation ratio of the image intensities was chosen as the similarity measure[35]. The image pair was thresholded below a Hounsfield value of 0 and above 1000, and the vertebrae were masked out to 1000 HU. Calculation of the similarity measure did not include

intensities below 0 HU, where deformation was also prohibited through B-spline control point adaptation.

Method 4. Single-resolution B-splines using sum of squared differences (SSD) as a similarity measure

This is an implementation of the B-splines, previously described by Hartkens et al.[36]. For this study, the registration was initialized by manually applying a translation and scaling in the longitudinal direction as a starting point for the B-spline based alignment of the two image sets. This method used a multi-resolution approach for the image, with voxel dimensions changing from 3mm to 1.5mm in two steps. Control points were spaced evenly at a constant 20mm interval. The similarity measure used was the sum of squared differences (SSD) between voxel intensities. This method, similar to method 3, masked the majority of the vertebrae to eliminate their signal from driving the cost function.

Method 5. Multi-resolution B-splines using mutual information (MI)

This method is another implementation of the B-splines which has been described previously[30, 37]. In this study, a single control point was used as a starting point for an automatic rigid registration allowing for translations and rotations only. After convergence, the initial rigid registration was used as a starting point for the multi-resolution deformable registration which iteratively changed the weights of the B-splines at each control point (knot), to maximize the mutual information between the reference and the homologous image. A multi-resolution approach was used for both the image and the B-spline knot spacing, with the image resolution changing from 4 voxels to 2 (voxel size of 0.78mm x 0.78mm x 1mm) and the knot resolution starting at 16 voxels and going down to 4 voxels in two steps. The multi-resolution approach helps speed up the convergence and avoids local minima. The reference dataset (inhale) was automatically cropped to the lung prior to registration.

Method 6. 'Demons' algorithm

This method is a grayscale-based fully automatic deformable image registration known as the 'Demons' algorithm[15, 38]. This method uses the intensity information of the image sets and the gradient information to automatically determine the displacement field from one dataset to the other. A multi-resolution approach and a symmetric force are applied to improve algorithm efficiency. In the accelerated Demons algorithm, an 'active force' is used with an adaptively-adjusted strength that is modified during the iterative process. No cropping or masking was used to separate the chest wall from the lung tissue in the registration process.

Method 7. Fluid flow

This is an intensity-based technique which makes use of fluid flow models. In this method the transformation is found by minimizing an energy term which is based on the squared difference of the image intensities and a regularization term. The regularization term is derived from compressible fluid flow equations, as described previously[1, 39]. The registration is driven by a force applied to the fluid at each point, the magnitude and direction of which are determined based on the difference in the intensities of the two images.

Method 8. Freeform deformation with calculus of variations

This is a fully automatic intensity-based free-form deformation with a multi-resolution approach[40]. In this method, the similarity and smoothness criteria are combined into one energy function, which is minimized in the registration process. A set of partial differential equations (PDE) are used to represent the minimization problem, and these equations are iteratively solved using a Gauss-Seidel finite difference scheme.

It should be noted that method 8 was developed and evaluated by a commercial entity. No special treatment was given to this organization for the study, and the results of this specific study should not be interpreted in establishing the superiority or lack thereof for one alignment method over another.

III.D.1.iv.c. ***Data analysis***

III.D.1.iv.c.1. Global evaluation of registration

The error in image registration was defined as the difference between the manually measured exhale marker position and the estimated position based on the deformation map from each registration technique. This difference was measured in three dimensions, right-left (d_{RL}), anterior-posterior (d_{AP}), superior-inferior (d_{SI}), and the 3-D vector distance (d) between the true and estimated marker positions was calculated from these components.

The global accuracy of each registration method was evaluated by calculating the mean ($\bar{d}_k$) and standard deviation (σ_k) of the three-dimensional error (d) over all marker positions for each registration technique (k). Although metrics like the mean and the standard deviation of 3-D error can provide a basic understanding of the behavior of each image registration technique, they do not provide any insight into the distribution of the error in different regions, since similar mean and standard deviation values can result from significantly different distributions in data. Therefore, in this study the frequency distribution of the 3-D error was also evaluated using a differential histogram with 2mm bins, where the percentage of the marker location errors within each bin was calculated. It should be emphasized that no spatial information about the error distribution can be taken from the histogram, and only information on the magnitude distribution is provided.

III.D.1.iv.c.2. Regional evaluation of registration

The behavior of different algorithms in regions with different characteristics was evaluated. The mean and standard deviation of the 3-D error for each marker across all registration techniques was calculated. Markers that show small mean and small standard deviation correspond to regions with intensity and deformation characteristics where the majority of registration methods perform well. Markers that show a large mean error and a relatively small standard deviation indicate regions where most techniques cannot predict the deformation well.

III.D.1.iv.c.3. Comparison of registration methods

This study was not designed to provide a direct comparison of different image registration techniques. Various alignment methods have been optimized to take advantage of different features and/or expected resolutions of shape change for real patient data, and employ vastly different goodness-of-fit metrics as well as search schemes and methods of describing local shape change. The intensity distribution and deformation characteristics of the phantom can be biased towards certain types of algorithms compared to others, and a direct comparison between different registration methods is not possible. The reported results were randomized to completely eliminate any potential inferences made regarding the relative or absolute accuracy of different registration techniques.

One subset of methodologies applied, however, is worthy of some inter-comparison. In this study, three groups used B-splines to characterize deformation. An evaluation of these methods as a group highlights some of the complexities involved in assessing the performance of an alignment technique simply by the mechanism for describing deformation.

III.D.1.v. ***Results***

All deformable registration methods performed generally well, with an average error ($\bar{d}_k$) ranging from 1.5 to 3.9 mm depending on the registration technique. These values are on the same order as accuracies reported in the literature[15-25]. The maximum error, however, showed a wider range, 5.1 to 15.4 mm, indicating non-uniformity in the results of deformable image registration and the potential for large regional inaccuracy in alignment in spite of overall acceptable accuracy. Table III.6 summarizes the results of the different methods, showing the maximum component errors in three dimensions ($d_{RL}^{\max}$, $d_{AP}^{\max}$, $d_{SI}^{\max}$) as well as the mean ($\bar{d}_k$), standard deviation (σ_k) and the maximum 3-D error ($d_k^{\max}$) for each registration technique. The results are randomized and each registration method is identified by a different letter from A to H in the following tables and graphs.

Table III.6: Maximum component errors in RL, AP, and SI directions, as well as the mean, standard deviation, and maximum 3-D vector distance for each registration method, shown in random order.

	Component errors(mm)			3D vector error (mm)		
Method(k)	$d_{RL}^{\max}$	$d_{AP}^{\max}$	$d_{SI}^{\max}$	$\bar{d}_k$	σ_k	$d_k^{\max}$
A	7.7	7.3	9.9	3.6	2.7	10.1
B	1.7	2.7	5.1	1.8	1.1	5.1
C	4.2	1.9	6.7	1.8	1.6	8.1
D	1.5	1.5	6.2	1.5	1.3	6.4
E	1.3	1.0	10.3	2.8	2.3	10.3
F	2.2	3.6	5.1	1.7	1.1	5.5
G	4.5	2.7	4.1	2.3	1.1	6.0
H	4.0	7.3	15.2	3.9	3.0	15.4

$d_{RL}^{\max}$: Absolute maximum component error in RL direction, $d_{AP}^{\max}$: Absolute maximum component error in AP direction, $d_{SI}^{\max}$: Absolute maximum component error in SI direction, $d_k^{\max}$: Maximum 3-D error, $\bar{d}_k$: Mean 3-D error, σ_k: Standard deviation 3-D error

The dominant error was in the SI direction (direction of foam compression) for most markers, as expected. However, errors as large as 8mm in the RL, and 7mm in the AP directions, were observed for some registration techniques. Figure III.9 shows the SI motion of each marker from inhale to exhale (i.e. under 30mm differential compression), plotted against the marker's distance to the diaphragm. The true motion of the markers is also plotted with a line (2nd degree polynomial) through the data to help with visualization of the results. It can be seen that, while some methods perform uniformly throughout the phantom, others do well in some regions while showing large errors in other areas. Therefore, it is clear that the mean and maximum 3-D errors shown in Table III.6 are useful but not sufficient metrics for comparison or evaluation of registration accuracy alone.

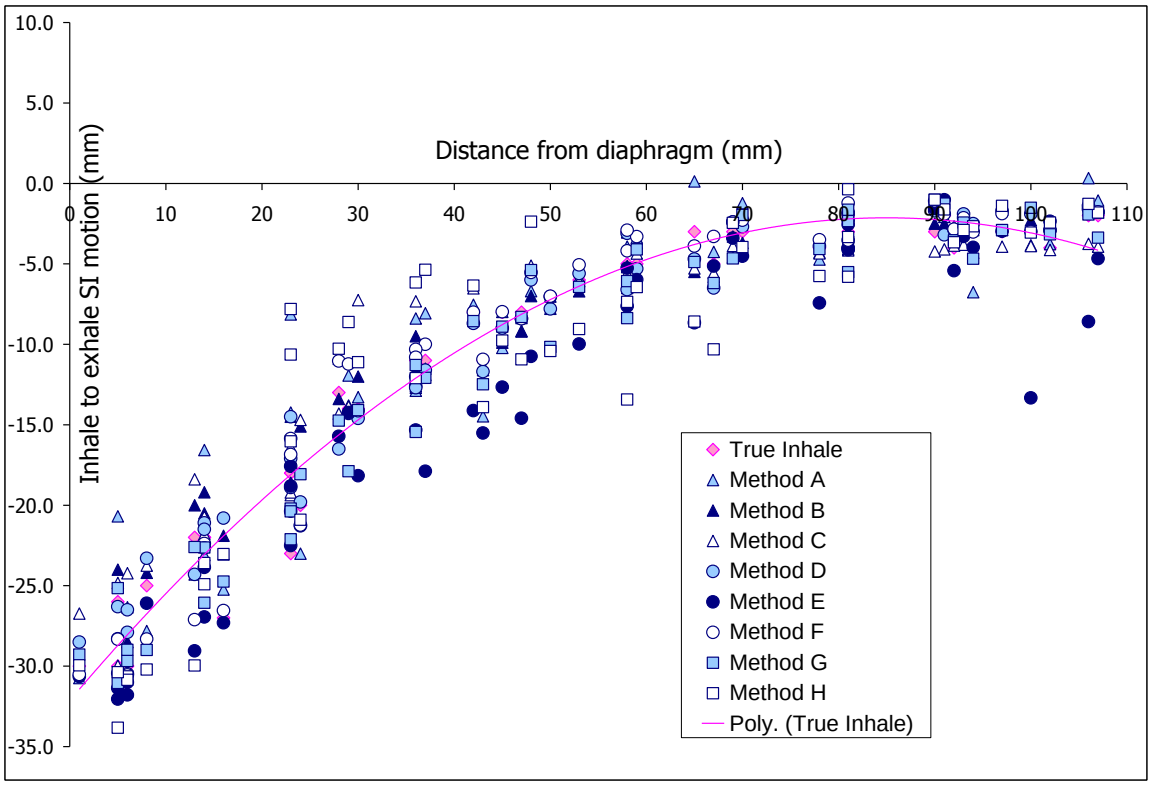

Figure III.9: Inhale to exhale SI motion of markers under 30mm compression as estimated by various registration methods, shown as a function of the marker's distance to diaphragm. The true motion of the markers from inhale to exhale is also shown (pink diamond marks indicate the motion of markers from inhale to exhale, measured manually).

To gain a better understanding of the performance of each method, histograms of the frequency distribution of registration error are plotted in Figure III.10. These histograms show the percentage of the markers (total of 48 markers) that have errors within the limits of each bin.

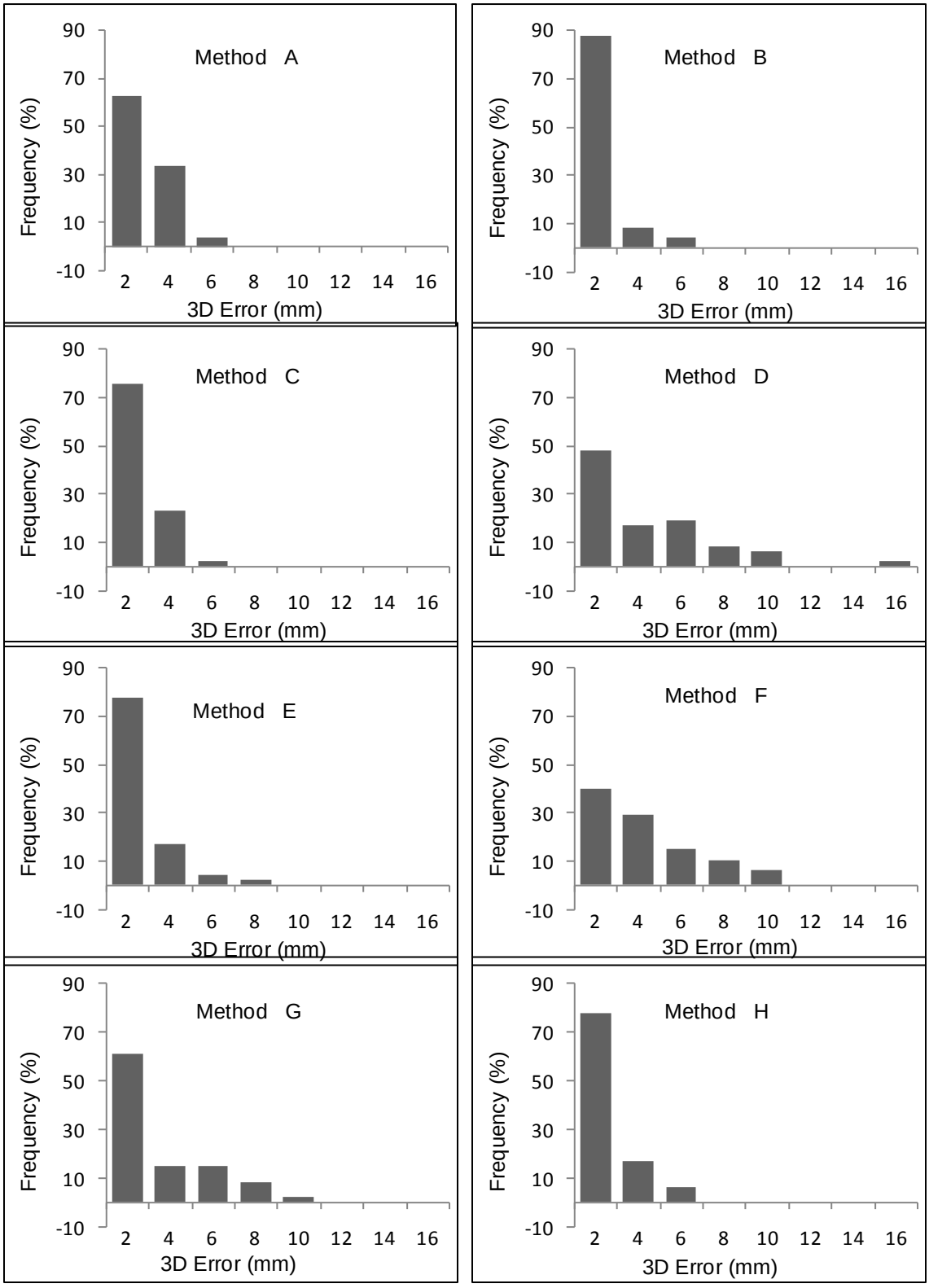

Figure III.10. Distribution of 3D error for each registration method

We also compared the results from three different implementations of B-splines tested in this study. As seen in Figure III.11, there is a significant difference in the frequency distribution of the 3-D error between these three methods. Methods C and F both used a multi-resolution approach for the image and knot spacing, while method H used a multi-resolution approach for the image but a single-resolution knot spacing. The three methods also used different similarity measures in the optimization process as well as differences in the users and the user specified settings. Although no specific conclusion should be made about which method is better, it would appear from these findings that a multi-resolution knot spacing can potentially result in better registration accuracy.

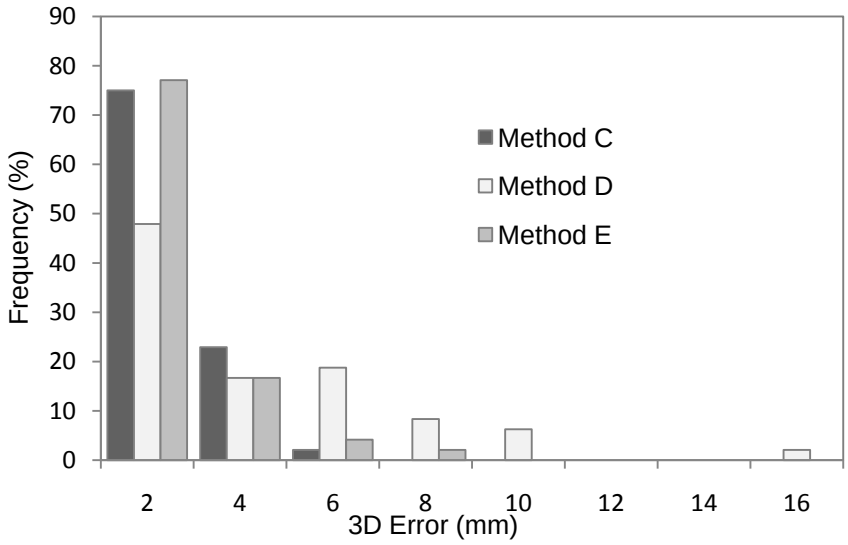

Figure III.11. Comparison of 3-D error distribution between different b-splines based registration methods.

The mean and standard deviation of the error was calculated for each marker over all registration methods, and plotted against the marker's motion from inhale to exhale in Figure III.12. Markers with small mean and small standard deviation error correspond to regions inside the phantom where most registration methods perform well. On the other hand, a large mean with a relatively small standard deviation indicates a region in the foam where most registration techniques fail. One example is the marker shown in red (Figure III.12), with a large mean and a smaller standard deviation. This marker falls in a region with relatively low intensity distribution, located next to a

high intensity region with a large change in intensity between inhale and exhale but without much deformation, as shown in Figure III.13. A slight increase in the average error is observed as the motion in the markers increases from inhale to exhale. However, there are some markers with very small motion between the two deformation states that show large errors.

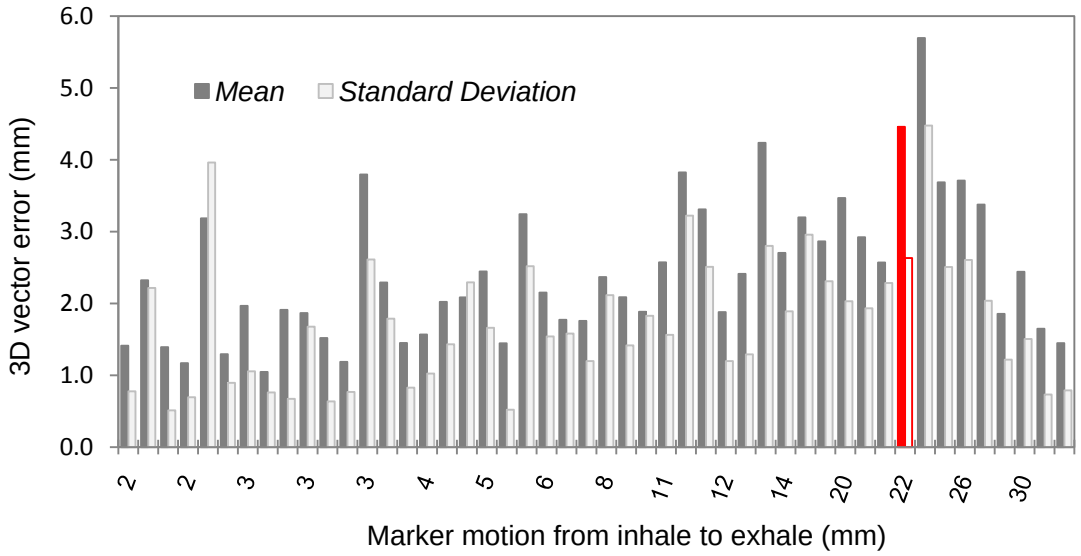

Figure III.12: Mean and standard deviation of the 3-D error for each marker calculated over all registration methods. The horizontal axis displays each marker's motion from inhale to exhale in mm. The marker for which the results are shown in red, corresponds to an example with a large mean and a small standard deviation. The image of this marker is shown in figure III.13.

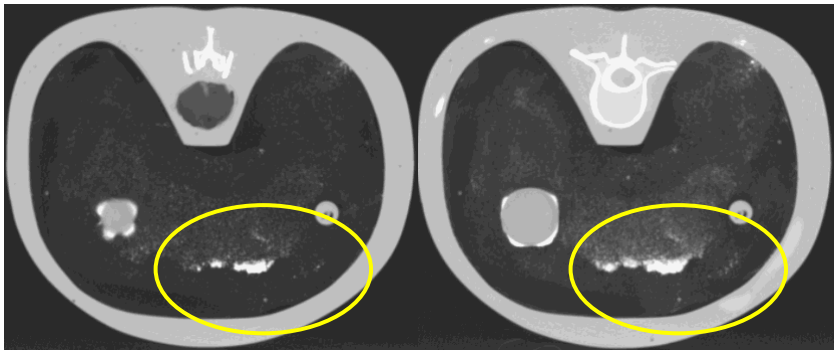

Figure III.13. Axial image of the region inside the foam at inhale (left) and exhale (right) where most registration methods performed poorly.

Of note, the phantom study actually identified a small error in post-processing of deformation results for our in-house alignment method. The multi-institutional analysis reflects the effect of

this variation. However, further analysis of intermediate phantom states as well as comparison with real patient images is based on the corrected process (Table III.4).

III.E. Discussion

In this study we performed a blind test of accuracy of different image registration methods, using a simple deformable lung phantom. The purpose was to objectively evaluate the accuracy of each registration technique and identify potential errors that may be overlooked by other validation methods. The large number of markers distributed throughout the phantom allowed for a better understanding of the variations in the registration error in all regions compared to other methods that use a few bifurcations. Our results showed mean and standard deviation errors, which were on the same order as those reported by other validation studies. However, maximum errors as high as 15mm were observed, suggesting that the sub-voxel accuracies reported based on evaluation of a few bifurcations may not be adequate to represent the overall accuracy of an algorithm. Currently, there is no consensus on how to report the accuracy of the registration. This study suggests that a distribution of residual error could be valuable in evaluating the performance of an algorithm.

Although the phantom was rather simplistic, it had certain properties observed in the lungs, such as density change, a non-deforming moving object within a deforming geometry, and some sliding against the chest wall. However, the microstructure represented in the phantom by the differential deposition of iodine should not be considered equivalent to that manifested by the vascular architecture of the lungs. This difference could potentially bias the performance of some registration methods over others. As a result, a direct comparison of registration accuracy of different techniques was not attempted. Some of the participating investigators stated that their registration techniques would have benefited from more sub-structures that would make the foam more comparable to real lungs. Their main concern was that certain parameter settings that they had optimized for registration of lungs would have to be perturbed to get the best results for the phantom. However, they all felt that the study was fair for the conclusions drawn. One thing that

should be considered here is that the coarse structures in the lung (vessels and bifurcations) are not present, and if the registration technique is dependent on these sub-structures (i.e. these structures are the driving forces of the registration), then the accuracy that is measured at these points cannot be propagated to other regions inside the lungs (e.g. across a tumor or in regions with other structures at different scales such as the mediastinum and bronchi). Such a dependence could thus be a source of bias, and not be reflected in accuracy reports that similarly depend on high spatial frequency content of image signals.

This also suggests that a difference in the registration accuracy of real lung images and the phantom images is expected, and the structural detail in the lung may result in better registration overall, even though our test of the sample clinical data shows errors on the same order as the phantom at least in the SI direction. However, we speculate that the registration error reported for real lungs based on measurement of bifurcations and anatomical landmarks, would underestimate the true error in alignment of the majority of other points in the lung. Therefore, evaluation of the error distribution results presented here would be a useful tool in understanding the limits in accuracy of various deformable alignment tools.

One important observation made in this study was that different implementations, different users, or different parameter settings of the same type of registration, can result in different accuracies. This suggests a need for careful assessment of implementations and standards on user-defined parameters or automation of the registration process. For example, the possibility exists for the algorithms to select and be sensitive to a specific range of intensities, thus yielding different results as these ranges are varied. Future studies will consider guidelines for modifying these ranges and studying sensitivity, as opposed to the optimal application of algorithms by their developers from the current trial design.

Another significant factor in registration accuracy is time, and as a general rule a compromise between time and accuracy has to be made in clinical settings depending on the application. In

this study, the reported registration times ranged from 2 minutes to 37 hours. However, no significant correlation between the registration time and accuracy was observed, mainly due to variations in computer resources.

III.F. Conclusion

In this study the accuracy of different registration methods was evaluated for a phantom with characteristics similar to those of lungs. The results indicated a distribution in the registration error in different regions, which may be overlooked in the standard evaluation techniques which make use of a few anatomical landmarks. Variations in the performance of different implementations, users, and settings of the same type of registration were also observed in this study, all of which suggest the need for careful assessment of potential sources of error in any type of deformable alignment. These results also show that generalization of the reported accuracies should be done carefully. Further improvements to the design of the phantom would be necessary for a more comprehensive evaluation and comparison of the different registration techniques.

References:

[1]B. C. Davis, M. Foskey, J. Rosenman, L. Goyal, S. Chang, and S. Joshi, "Automatic Segmentation of Intra-Treatment CT Images for Adaptive Radiation Therapy of the Prostate," Med. Image Comput. Comput. Assist. Interv. Int. Conf. Med. Image Comput. Comput. Assist. Interv. **8**, 442-450 (2005)

[2]W. Lu, G. H. Olivera, Q. Chen, M. L. Chen, and K. J. Ruchala, "Automatic Re-Contouring in 4D Radiotherapy," Phys. Med. Biol. **51**, 1077-1099 (2006)

[3]T. Zhang, Y. Chi, E. Meldolesi, and D. Yan, "Automatic Delineation of on-Line Head-and-Neck Computed Tomography Images: Toward on-Line Adaptive Radiotherapy," Int. J. Radiat. Oncol. Biol. Phys. **68**, 522-530 (2007)

[4]L. E. Court, L. Dong, A. K. Lee, R. Cheung, M. D. Bonnen, J. O'Daniel, H. Wang, R. Mohan, and D. Kuban, "An Automatic CT-Guided Adaptive Radiation Therapy Technique by Online Modification of Multileaf Collimator Leaf Positions for Prostate Cancer," Int. J. Radiat. Oncol. Biol. Phys. **62**, 154-163 (2005)

[5]L. E. Court, R. B. Tishler, J. Petit, R. Cormack, and L. Chin, "Automatic Online Adaptive Radiation Therapy Techniques for Targets with Significant Shape Change: A Feasibility Study," Phys. Med. Biol. **51**, 2493-2501 (2006)

[6]A. de la Zerda, B. Armbruster, and L. Xing, "Formulating Adaptive Radiation Therapy (ART) Treatment Planning into a Closed-Loop Control Framework," Phys. Med. Biol. **52**, 4137-4153 (2007)

[7]C. Wu, R. Jeraj, W. Lu, and T. R. Mackie, "Fast Treatment Plan Modification with an Over-Relaxed Cimmino Algorithm," Med. Phys. **31**, 191-200 (2004)

[8]A. Mestrovic, M. P. Milette, A. Nichol, B. G. Clark, and K. Otto, "Direct Aperture Optimization for Online Adaptive Radiation Therapy," Med. Phys. **34**, 1631-1646 (2007)

[9]R. Mohan, X. Zhang, H. Wang, Y. Kang, X. Wang, H. Liu, K. K. Ang, D. Kuban, and L. Dong, "Use of Deformed Intensity Distributions for on-Line Modification of Image-Guided IMRT to Account for Interfractional Anatomic Changes," Int. J. Radiat. Oncol. Biol. Phys. **61**, 1258-1266 (2005).

[10]W. Y. Song, E. Wong, G. S. Bauman, J. J. Battista, and J. Van Dyk, "Dosimetric Evaluation of Daily Rigid and Nonrigid Geometric Correction Strategies during on-Line Image-Guided Radiation Therapy (IGRT) of Prostate Cancer," Med. Phys. **34**, 352-365 (2007).

[11]W. Lu, G. H. Olivera, Q. Chen, K. J. Ruchala, J. Haimerl, S. L. Meeks, K. M. Langen, and P. A. Kupelian, "Deformable Registration of the Planning Image (kVCT) and the Daily Images (MVCT) for Adaptive Radiation Therapy," Phys. Med. Biol. **51**, 4357-4374 (2006).

[12]U. Malsch, C. Thieke, and R. Bendl, "Fast Elastic Registration for Adaptive Radiotherapy," Med. Image Comput. Comput. Assist. Interv. Int. Conf. Med. Image Comput. Comput. Assist. Interv. **9**, 612-619 (2006).

[13]U. Malsch, C. Thieke, P. E. Huber, and R. Bendl, "An Enhanced Block Matching Algorithm for Fast Elastic Registration in Adaptive Radiotherapy," Phys. Med. Biol. **51**, 4789-4806 (2006).

[14]J. R. McClelland, J. M. Blackall, S. Tarte, A. C. Chandler, S. Hughes, S. Ahmad, D. B. Landau, and D. J. Hawkes, "A Continuous 4D Motion Model from Multiple Respiratory Cycles for use in Lung Radiotherapy," Med. Phys. **33**, 3348-3358 (2006).

[15]H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an Accelerated 'Demons' Algorithm for Deformable Image Registration in Radiation Therapy," Phys. Med. Biol. **50**, 2887-2905 (2005).

[16]J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, A. Degenhard, M. O. Leach, D. R. Hose, D. L. G. Hill, "Validation of nonrigid image registration using finite-element methods: Application to breast MR images." IEEE Trans. Med. Imag. **22**, 238-247 (2003).

[17]T. Guerrero, G. Zhang, T. Huang, K. Lin, "Intrathoracic tumor motion estimation from CT imaging using the 3D optical flow method," Phys. Med. Biol. **49**, 4147- 4161 (2004).

[18]P. Rogelj, S. Kovacic, J.C. Gee, "Validation of a non-rigid registration algorithm for multi-modal data," Proc. SPIE. 4684, 299–307 (2002)

[19]H. Wang, L. Dong, M. F. Lii, R. de Crevoisier, R. Mohan, J. D. Cox, D. A. Kuban, and R. Cheung, "Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy," Int. J. Radiat. Oncol. Biol. Phys. **61**, 725-735 (2005).

[20]V. Walimbe, R. Shekhar, "Automatic elastic image registration by interpolation of 3D rotations and translations from discrete rigid-body transformations," Med. Image Anal. **10**, 899-914 (2006).

[21]K. K. Brock, M.B. Sharp, L.A. Dawson, S.M. Kim, D.A. Jaffray, "Accuracy of finite element model-based multi-organ deformable image registration," Med. Phys. **32**, 1647-1659 (2005).

[22]M. M. Coselmon, J.M. Balter, D.L. McShan, M.L. Kessler, "Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines," Med. Phys. **31**, 2942-2948 (2004).

[23]E. Rietzel, G. T. Y. Chen, "Deformable registration of 4D computed tomography data," Med. Phys. **33**, 4423-4430 (2006).

[24]J. P. Voroney, K. K. Brock, C. Eccles, M. Haider, L. A. Dawson, "Prospective comparison of computed tomography and magnetic resonance imaging for liver cancer delineation using deformable image registration," Int. J. Radiat. Oncol. Biol. Phys. **66**, 780-791 (2006).

[25]H. Zhong, T. Peters, and J. V. Siebers, "FEM-Based Evaluation of Deformable Image Registration for Radiation Therapy," Phys. Med .Biol. **52**, 4721-4738 (2007).

[26]Y. Y. Chou, O. Skrinjar, "Ground truth data for validation of nonrigid image registration algorithms," IEEE Int. Symp. Biomed. Imag. Macro to Nano 1, 716-719 (2004).

[27]R. Zeng, J.A. Fessler, J.M. Balter, "Estimating 3-D respiratory motion from orbiting views by tomographic image registration," IEEE Trans. Med. Imaging. **26**, 153-163 (2007)

[28]R. Kashani, K. Lam, D.W. Litzenberg, J.M. Balter, "Technical note: A deformable phantom for dynamic modeling in radiation therapy," Med. Phys. **34**, 199-201 (2007).

[29]R. Kashani, M. Hub, M.L. Kessler, J.M. Balter, "Technical note: A physical phantom for assessment of accuracy of deformable alignment algorithms," Med. Phys. **34**, 2785-2788 (2007)

[30]M. L. Kessler, "Image registration and data fusion in radiation therapy," Br. J. Radiol. **79,** S99-S108 (2006)

[31]J. Lian, L. Xing, S. Hunjan, C. Dumoulin, J. Levin, A. Lo, R. Watkins, K. Rohling, R. Giaquinto, D. Kim, D. Spielman, B. Daniel, "Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: A deformable registration and validation study," Med. Phys. **31,** 3087-3094 (2004)

[32]Y. Xie, L. Xing, “Deformable image registration with inclusion of auto-detected homologous tissue features,” Radiotherapy. Oncol. **84**, S109 (2007)

[33]D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comp. Vision. **60**, 91-110 (2004)

[34]J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, D. J. Hawkes, “A Generic framework for non-rigid registration based on non-uniform multi-level free-form deformations,” Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001), 2208 of Lecture Notes in Computer Science, pages 573-581. Springer Verlag.

[35]A. Roche, X. Pennec, M. Rudolph, D. P. Auer, G. Malandain, S. Ourselin, L. M. Auer, N. Ayache, “Generalized correlation ratio for rigid registration of 3D ultrasound with MR images,” Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2000), volume 1935 of Lecture Notes in Computer Science, 567 - 577, Springer Verlag.

[36]T. Hartkens, D. Rueckert, J. A. Schnabel, D. J. Hawkes, D. L. G. Hill, “VTK CISG Registration Toolkit: An open source software package for affine and non-rigid registration of single- and multimodal 3D images,” Proc. Workshop on Bildverarbeitung für die Medizin, Informatik Aktuell, pp. 409-412, Springer Verlag, 2002.

[37]J. Kybic, M. Unser, “Fast parametric elastic image registration,“ IEEE Trans. Image Process. **12**, 1427- 1442 (2003)

[38]J. P. Thirion, "Image Matching as a Diffusion Process: An Analogy with Maxwell's Demons," Med. Image. Anal. **2**, 243-260 (1998)

[39]M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi, "Large Deformation Three-Dimensional Image Registration in Image-Guided Radiation Therapy," Phys. Med. Biol. **50**, 5869-5892 (2005)

[40]W. Lu, M. L. Chen, G. H. Olivera, K. J. Ruchala, and T. R. Mackie, "Fast Free-Form Deformable Registration Via Calculus of Variations," Phys. Med. Biol. **49**, 3067-3087 (2004)

[41]A. Pevsner, B. Davis, S. Joshi, A. Hertanto, J. Mechalakos, E. Yorke, K. Rosenzweig, S. Nehmeh, Y. E. Erdi, J. L. Humm, S. Larson, C. C. Ling, G. S. Mageras, “Evaluation of an automated deformable image matching method for quantifying lung motion in respiration-correlated CT images,” Med. Phys. **33,** 369-376 (2006)
III.F.1.i.a.

CHAPTER IV

TRACKING DEFORMATION USING SURROGATES

IV.A. Introduction

Chapter II described the magnitude and reproducibility of respiratory-induced motion and deformation in the thoracic and abdominal regions. Through a series of motion studies, it was demonstrated that the impact of respiratory motion goes beyond the uncertainty in target positioning and coverage, and can significantly affect the surrounding normal structures. Although the dosimetric effects of this motion were only briefly discussed in Chapter II, there is sufficient evidence in the literature to indicate that a thorough knowledge of the position of normal structures during treatment delivery may be just as important as that of the target, indicating the need to monitor the changes in normal tissue configuration during treatment[1-2]. Recognizing the importance of tracking normal tissue deformation, a method for estimating deformation using the measured position of a few surrogates is proposed in this chapter. A patient-specific model of the respiratory-induced motion and deformation is described using a method from multivariate statistics called principal component analysis (PCA). This model, applied to the position of surrogates, is used to provide relatively stable estimates of target positions in the lung or liver. A discussion of the characteristics of the data is provided, and then the mathematics behind the model, as well as the limitations and uncertainties associated with it, are discussed.

IV.A.1. Background and purpose

While the magnitude and characteristics of breathing motion can be evaluated at the time of simulation or prior to each treatment fraction using imaging techniques such as fluoroscopy and 4D CT, we continue to rely on surrogates for monitoring tumor motion during treatment delivery. These surrogates include breathing traces measured with spirometry or external markers placed

on the surface of the abdomen, and fiducial markers implanted in or near the tumor prior to the start of the treatment course[3-4]. A limitation of spirometry or external surrogates is complete reliance on the assumption that the initial correlation established between the signal from these devices and the internal geometry remains constant over the course of treatment. Contrary to this assumption, several studies have shown that baseline shifts in the position of the tumor and changes in the breathing pattern of the patient can compromise this initially-established correlation, resulting in error in estimating the tumor position[5-6].

Some of these concerns can be addressed by using internal markers. In current monitoring and tracking techniques, a small number of markers are placed in or near the region of interest, and the change in the average position of these markers is used to measure the motion of the target. Although the average position of two or more markers may be a good surrogate for the geometry they encompass, it may poorly estimate the shape or change in the relative position of surrounding normal structures or targets not centered by these fiducials. Therefore, to track the position of more than one point in a deforming geometry using a limited number of surrogates, we need models that can reliably relate the position of the surrogates to the changes in patient configuration.

When using surrogates to track deformation, the first step is to determine the correlation between the surrogate configuration and deformation state of the patient. Surrogate configuration can be defined several ways, depending on the component of the deformation that is being tracked. For example, for three fiducials implanted in the liver to be used for positioning of the patient to the target on a daily basis, surrogate configuration is the average position of these three fiducials relative to the treatment isocenter, or the bony anatomy. When tracking breathing-induced deformation in lungs, surrogate configuration can be defined as the variation in the position of the fiducials relative to their position at some reference breathing state, assuming that the setup component of the motion is already corrected for. To find the correlation between surrogate configuration and deformation state of the patient, we need to first generate an initial estimate of

shape variation from multiple deformation states of the patient by measuring the change in the position of all points in the geometry of interest from a reference state (e.g. inhale or exhale phase of a 4D CT scan) to all available deformed states. Various methods have been used to estimate this motion, including point distribution models (PDM)[19], finite element models (FEM)[8], and deformable image registration[9]. Although all these techniques are reasonably accurate, deformable image registration requires the least degree of user involvement and therefore tends to be the method of choice in many cases. Deformable registration results in a deformation map or a displacement vector field for all voxels from the reference state to each discretely measured deformation state of the patient, creating what can be interpreted as a look-up table for deformation that can be used for geometry tracking during treatment. At the time of treatment the measured surrogate motion will estimate the deformation state of the patient, which in turn allows us to predict which displacement vector field to choose to determine the position of a specific point within that field.

Although this is a simple, straight-forward way of relating the surrogate motion to changes in patient configuration, its application is limited to the discretely sampled deformation states, and the process potentially suffers in local accuracy due to sampling artifacts in 4D CT scanning as well as local uncertainties due to noise in the image and errors in deformable alignment. Local and global registration accuracy of deformable image registration for targets in the thoracic region was discussed in detail in Chapter III. 4D CT sampling artifacts and image noise have been discussed extensively in the literature and are beyond the scope of this chapter[10-14]. However, their indirect impact on marker localization accuracy and registration will be discussed later in this chapter. Furthermore, as mentioned previously, relating different states of breathing is generally done by (deformable) image registration, a technique with variable accuracy throughout the imaged volume. Understanding the dominant components of shape change during breathing in theory may permit the use of these potentially noisy estimates to make relatively stable predictions of shape from limited sample data. Recognizing the potential sources of error, our task is to identify the desired characteristics of the modified data and the assumptions and

limitations of the model in order to identify the best solution for this problem. Using dimensionality reduction, we hope to resolve the issue of redundancy and noise, while preserving the important information for correlating the motion of the surrogate to the patient configuration changes.

IV.A.2. Dimensionality reduction to reduce registration noise

The method used for dimensionality reduction here is a method from multivariate statistics called principal component analysis (PCA). Recently Manke et al. proposed using PCA to find the correspondence between the parameters of an affine model of the respiratory-induced motion of the heart and the position of a few navigators in angiography[7]. Sohn *et al.* also applied PCA to model the inter-fractional organ deformation in the male pelvis, based on multiple CT scans acquired at different times during the course of treatment[8]. This model was further developed by Zhang et al., who characterized an individual patient's lung deformation in relation to the position of a surrogate (i.e. diaphragm position)[9]. They applied the model to 4D CT scans of patients with lung cancer to estimate the motion of tumor and normal tissue during treatment, using the position of the most superior point on the diaphragm as the surrogate[9]. Here we apply a method similar to that described by Zhang et al., using the position of a few implanted markers as the surrogate. Principal component analysis allows us to evaluate the correlated motion and deformation of a large number of points in the geometry of interest (i.e. every voxel) by estimating the dominant modes of variation in the data. The advantage of such a model is that it uses the variations in the entire geometry, analytically connects discretely sampled states, permits extrapolation to unsampled states, and can include temporal weighting of motion patterns. Furthermore, understanding the dominant modes of variation can reduce the sensitivity of the deformation model to local errors in deformation estimation.

IV.B. Model description, implementation, and assessment

IV.B.1. Model description

The process studied here involves three major steps. The first step is to capture an initial estimate of the nature of typical shape change in an individual patient. The next step is to derive

a model that relates the relative position of a subset of surrogate locations to the position and deformation state of the patient. This step results in a coefficient matrix calculated from the registration results and the surrogate positions using statistical models. The last step then involves using this coefficient matrix and the measured motion or position of the surrogates, to calculate the position of various points in the patient's geometry.

Patient geometry is parameterized using a uniform grid of voxels on the reference image volume. The size of the voxels depends on image resolution, slice thickness and the sub-sampling criteria used for image registration. The number of voxels, *M*, depends on the size of the region used in image registration as shown in Figure IV.1.

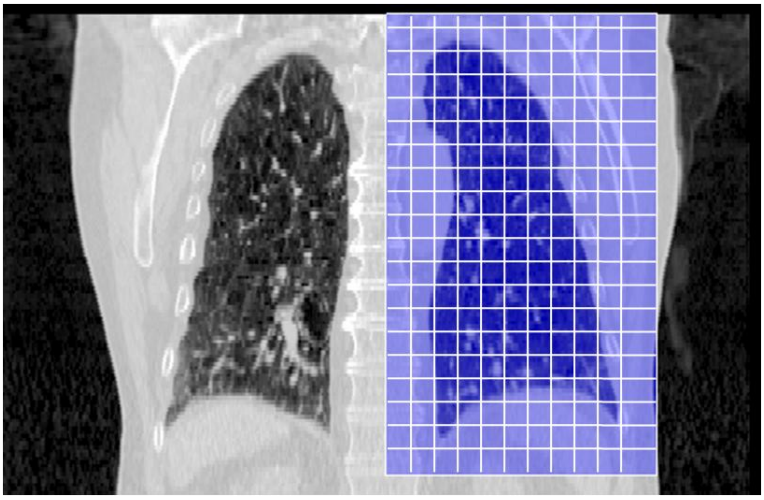

Figure IV.1: A uniform grid of voxels on the volumetric image of the left lung, parameterizes the patient geometry. The blue rectangle indicates the region that was used in the calculation of the deformation map.

The position of the center of each voxel j in 3 dimesnsions can be described by the position vector, $\vec{x}_j(i)$, and thus estimating shape is a 3M dimensional problem. The three elements of this vector can be either the position of a voxel's center at a breathing state i, or the change in the position of that voxel from a reference breathing state, to any other state i in the respiratory cycle, which will be discussed further in the following sections.

The change in the position of each voxel center between various breathing states is estimated from deformable alignment of a reference state (e.g. end-inhale) to other breathing states of a 4D CT reconstruction, resulting in N-1 discrete samples of the motion for each voxel describing the variations in each of the 3M variables over a full breathing cycle. For 4D CT patients in the clinic, N is typically equal to 10 and thus there are 9 variations in any voxel's position from the reference state to the other 9 states of the breathing cycle (Figure IV.2).

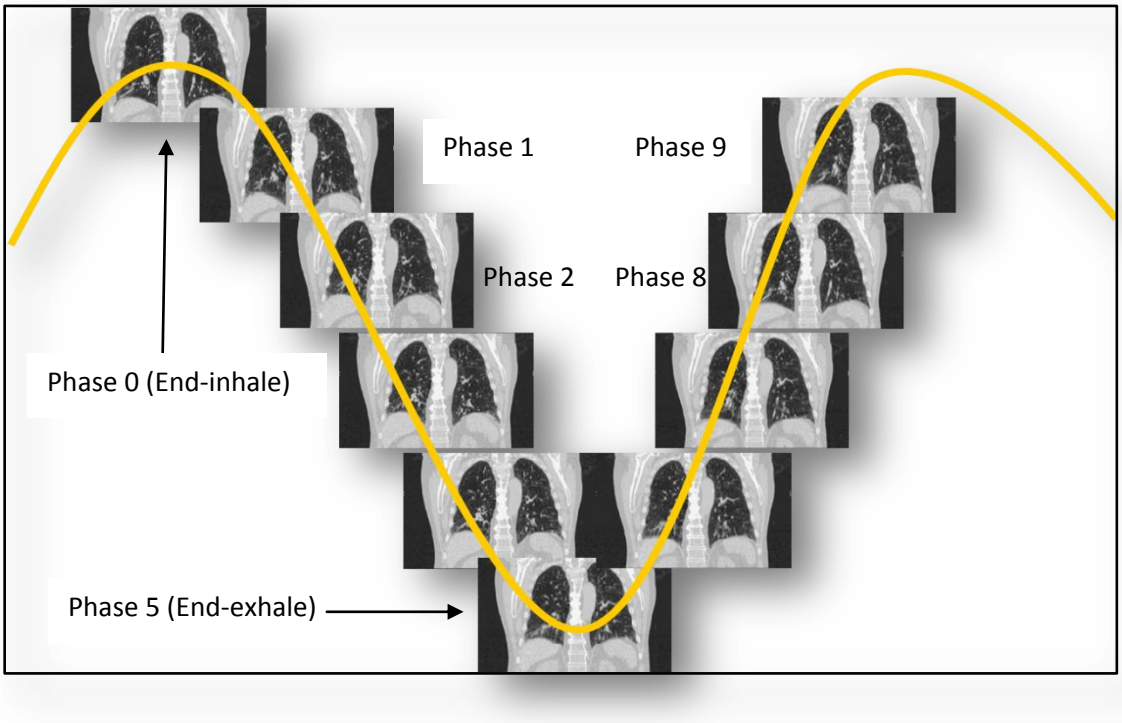

Figure IV.2: Coronal images of the lung at various reconstructed states of a breathing cycle.

The deformation model used here is based on B-splines and is estimated to have an average accuracy of about 2 mm in lung geometry, although with potential error of 5 mm or more in local regions depending on differential signal available for guiding the alignment, as shown in detail in Chapter III. Using deformable alignment, the motion in the center of each voxel is estimated over the full breathing cycle, relative to a reference breathing states. The motion vectors for all voxels are then concatenated to form a displacement vector P_i (Equation IV.1) which represents a single sample of the variation relative to the reference.

$$P_i = [\overrightarrow{x_1}(i), \overrightarrow{x_2}(i), \ldots, \overrightarrow{x_M}(i)] \quad \text{IV.1}$$

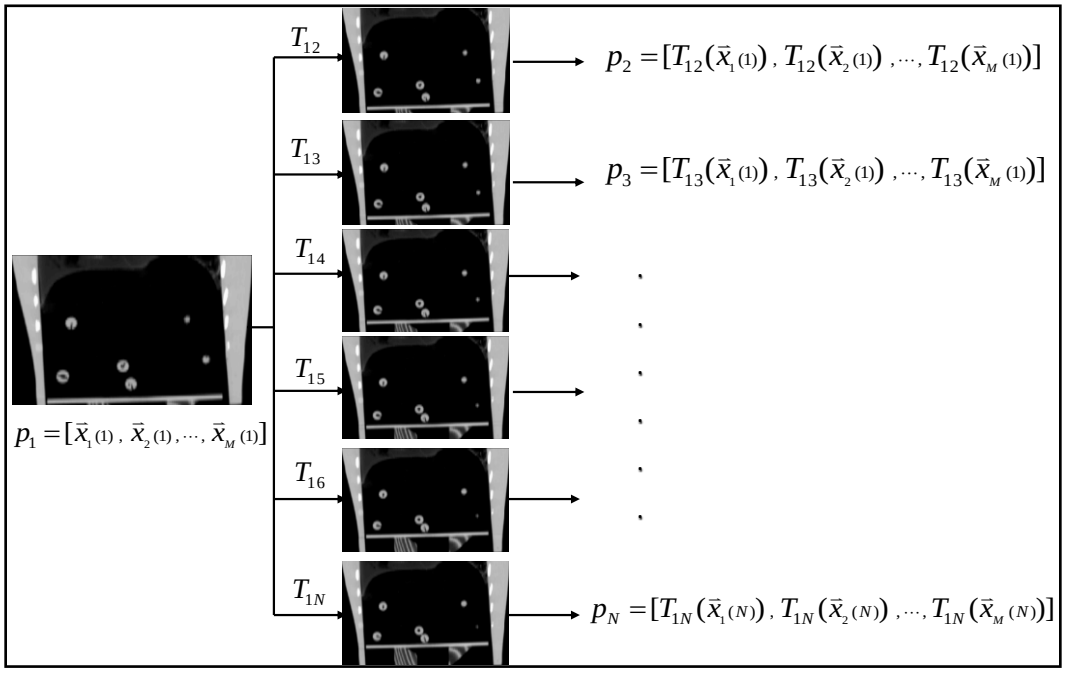

Figure IV.3: Schematic of creating the displacement vectors from the reference phase (inhale) to all other breathing states, by applying the transformaiton map T_{1N} to the postion vector at phase 1 (end-inhale). The images displayed here correspond to the deforming phantom from chapter III.

Patient geometric variation is then decomposed into eigenmodes by subtracting the mean displacement from each displacement vector at breathing state i , and calculating the covariance matrix C of these centered displacement vectors (equation IV.2) .

$$C = \frac{1}{N-1} \sum_{i=1}^{N} (P_i - \bar{P})\,(P_i - \bar{P})^T \quad \text{IV.2}$$

Diagonalization of the covariance matrix results in a set of N-2 non-zero eigenvalues [w_k], and eigenvectors [e_k]. Eigenvectors are 3-D vector fields of correlated displacement of all voxels in the geometry used in modeling, and represent statistically independent modes of variation in the

data. Each eigenvalue represents the amount of variation in the data described by the corresponding eigenmode. Any deformation state of the patient $P(t)$ at an arbitrary breathing state t can be estimated as the weighted sum of K dominant eigenmodes, added to the mean displacement vector field[8-9]:

$$P(t) \approx \bar{P} + \sum_{k=l}^{K} w_k(t)\, e_k \qquad \text{IV.3}$$

For any K eigenmodes used in estimation of deformation in geometry, the optimal weighting on each eigenmode that would result in the best estimation of any of the sampled state, can be found by minimizing the distance between the estimated and measured geometry as described by Sohn *et al.* in equation IV.4 below[8].

$$w_k(i) = (p_i - \bar{p}).\ e_k \qquad \text{IV.4}$$

However, for any un-sampled deformation state of the patient, the weighting factors in Equation IV.3 are unknown. Including the position of the surrogates as variables in the modeling step allows for establishing the relation between the motion of the surrogates and the displacement of the remaining voxels in the geometry at any arbitrary deformation state, providing an estimate of the unknown weighting factors. This is done by adding the displacement vectors for the surrogates at different breathing states, to the end of the voxel displacement vector P_i , so that Equation IV.1 can be re-written as follows:

$$P_i = [\,\overrightarrow{x_1}(i), \dots, \overrightarrow{x_M}(i), \vec{s}_1(i), \dots, \vec{s}_n(i)] \qquad \text{IV.5}$$

In this equation $\vec{s}_n(i)$ is the position of the surrogate at the breathing states used in the model. Rewriting equation IV.5, in matrix form and replacing the displacement vectors with centered displacement vectors, results in equation IV.6[7-9].

$$\tilde{P} = [\,\tilde{x}\;\;\tilde{s}] \qquad \text{IV.6}$$

Also re-writing the sum of weighted eigenvectors (Equation IV.3) in terms of the eigenvalue and eigenvector matrices W and E (Equation IV.7), and then separating it into two sections corresponding to the a) surrogates and b) the rest of the geometry, as shown in Equations IV.8 and IV.9, results in the calculation of the coefficient matrix B (Equation IV.10).

$$\tilde{P} = E\,W \qquad \text{IV.7}$$

$$\tilde{x} = E_x\,W \qquad \text{IV.8}$$

$$\tilde{s} = E_s\,W \qquad \text{IV.9}$$

$$B = E_u\,E_s^{-1}\,\tilde{s}(t) \qquad \text{IV.10}$$

This matrix, which is assumed to be time-invariant, can be used to estimate the motion in any voxel based on the change in the position and configuration of the surrogates at any arbitrary state (Equation IV.12).

$$\tilde{x}(t) \approx E_u\,E_s^{-1}\,\tilde{s}(t) = B\,\tilde{s}(t) \qquad \text{IV.12}$$

IV.B.1.i. ***Patient position/configuration changes***

As shown in equation IV.1 the displacement vector P_i can describe either the position of the center of each voxel at any deformation state i, or the change in the position of the voxel relative to a reference state. The choice of one over the other should not affect the model directly. The more important aspect of patient position and configuration changes is the consideration of the various components of this change. The method presented here only aims at modeling the intra-fractional variations in patient configuration after the patient is setup on the treatment table.

Therefore, the initial setup to the average position of the markers is a necessary step prior to the application of this model.

IV.B.1.ii. ***Surrogate position/configuration***

When the target is monitored or tracked based on the average position of the surrogates, the change in the configuration corresponds to the change in the distance separating the surrogates and also their relative position to a fixed point in the patient and/or treatment reference frame depending on the setup and tracking protocol. In the model presented here, the change in surrogate configuration is the change in the position of each individual surrogate relative to its average position, calculated based on the geometric samples used in the model.

IV.B.2. Why use PCA?

In order to fully understand the reasoning behind choosing PCA for modeling deformation, we need to first understand the characteristics of the deformation and the limitations of the displacement vector fields resulting from deformable registration or the "raw" data. We should also understand the desired characteristics of the modified or processed data. We then need to consider the assumptions and limitations of PCA in order to justify its use in this application.

In the previous section we described how patient geometry is parameterized using the vector $\vec{x}_j(i)$ with the 3 coordinates of the displacement in the position of the center of each voxel from a reference breathing state to other breathing states. If deformable registration was perfect, showing no local error, we could directly link these deformation maps to the position of the surrogates at that deformation state and use them as a lookup table for monitoring or tracking during treatment. In that case, the only problem remaining would have been the estimation of the deformation when the surrogate position is at an un-sampled deformation state, which would be resolved using some interpolation method. However, as shown in Chapter III, registration is not perfect and the resulting displacement vector field is "noisy". In addition, each deformation map contains information on displacement of thousands of voxels whose centers are less than 3mm

apart in any dimension. Considering that it is anatomically impossible for these voxels to be moving independently (except at sliding interfaces), the data is highly redundant. This redundancy can be visualized in the displacement map of all voxels in the lung from inhale to exhale as shown in Figure IV.4. Therefore the two problems with the data that we need to address with PCA are reduction in noise and simplification of redundant information.

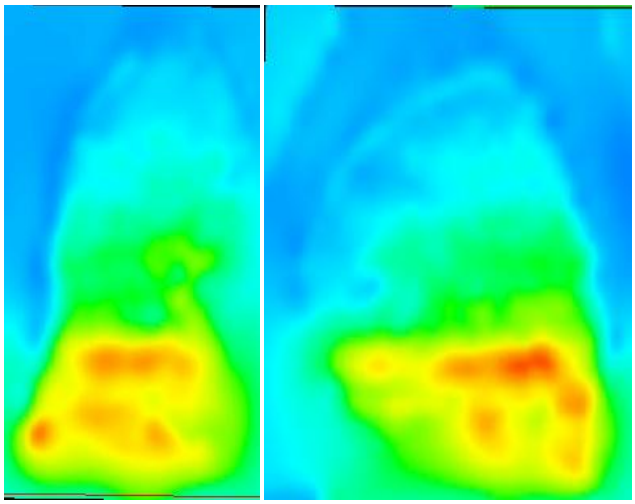

Figure IV.4. Coronal and sagittal views of the deformation map from inhale to exhale in the SI direction. The blue indicates a motion on the order of 1-2mm, while the red corresponds to displacements of 2cm or higher.

Now the question remaining is how we determine what is noise and what is redundancy in the data. Imagine we have a single dynamic variable that we measure multiple times. If the variable has a normal distribution, its statistics are sufficiently captured by the first two moments (mean and variance). Now assuming that the measurement of the variable is reasonably accurate, a large variance would indicate interesting dynamics in the data, while a small variance would correspond to noise. Now if we have two normally distributed variables that we measure multiple times with relatively high accuracy, in addition to the variance of each variable, we can now calculate the covariance between the two, which is a measure of the degree of linear relationship between the two variables. A large covariance in this case indicates a high degree of redundancy, while a small covariance shows lack of correlation between the two variables. We can now generalize this concept to M variables with N measurements each. If we now display these in

matrix form such that each row of the matrix corresponds to the N measurements of a single variable, we would obtain the MxN dimensional matrix described in the previous section. Using Equation IV.2, we can calculate the covariance matrix, which contains the variance in each single variable on the diagonal elements, and the covariance between any two variables i and j, in the off-diagonal elements. In summary, this covariance matrix captures the correlation between all possible pairs of variables, describing the noise and redundancy in the data.

Continuing with the assumption that the quality of measuring this transformation (i.e. image registration) was good enough that a large variance in any single variable would describe the interesting dynamics in the data and a small variance would correspond to noise, this ideal covariance matrix would have large diagonal elements (noise reduction), while the off-diagonal elements that describe redundancy would be zero, indicating no correlation between the different variables. Thus diagonalizing the covariance matrix maximizes the correlation of information between deformation of different voxels. Principal Component Analysis (PCA) is a standard and simple mathematical tool for diagonalizing the covariance matrix. The reason PCA is the easiest way to diagonalize the covariance matrix lies in the underlying linearity assumption of PCA, which frames the problem as a simple change of basis and limits the potential search space for this new basis to linear combinations of the original basis. What this means is that PCA applies a rotation to the data to align the direction of the first basis with the direction of the maximum variance in the data. In addition, PCA has an orthonormality condition, which restricts the search for each new direction to those directions that are orthogonal to the previous ones. The advantage of this condition is that it makes it possible for PCA to be solved using linear algebra decomposition techniques.

IV.B.3. Data description

IV.B.3.i. ***Sample 4D CT patient***

A non-small cell lung cancer patient with a pre-treatment 4D CT scan approved for research was chosen for this study. 57 landmarks were manually identified in the left lung on all 10 states,

representing the possible positions of the electromagnetic transponders and geometric points to be monitored or tracked as shown in Figure IV.5. The accuracy of landmark selection was estimated through repeat measurements of the position of 6 landmarks, which will be described in the next section of this chapter. Different subsets of these landmarks were chosen to represent the position of surrogates for this simulation, and the rest were used for evaluation of model accuracy. These three landmarks are shown with the large yellow dots in Figure IV.5.

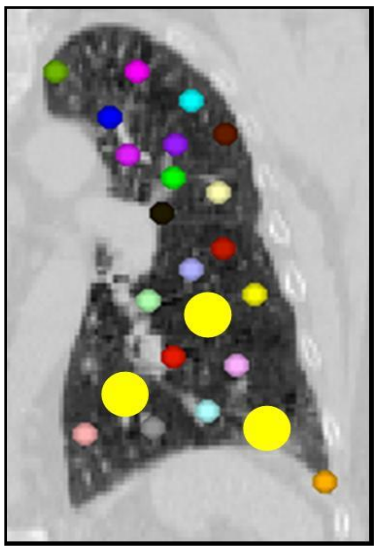

Figure IV.5: The position of manually selected landmarks inside the left lung, for evaluation of the model. The three large yellow circles show the possible positions of the three landmarks chosen as the surrogates

IV.B.4. Model implementation and the limits of accuracy

In addition to the limitations of PCA in modeling deformation in lungs caused by breathing, several other factors contribute to the uncertainties in this model that stem from the implementation method and the tools available for measurement of motion and deformation. These factors include accuracy in calculation of model parameters (i.e. eigenvalues and eigenmodes), the accuracy in manual measurement of landmark positions over various breathing states, and the accuracy of deformable image registration, which is used in finding the correspondence between different samples of patient geometry. The estimated model accuracy is

affected by all these factors, and it is difficult to eliminate or separate out their contribution to the overall error.

IV.B.4.i. *Implementation accuracy*

The principal components of the variation in the data can be calculated using different methods. The first method, which was described with the equations in Section IV.A, is based on the calculation of the eigenvalues and eigenvectors of the covariance matrix IV.3. One limitation of this method is that for the 3MxN dimensional data matrix, where M is on the order of several thousands, the resulting 3Mx3M covariance matrix is too large, and the numerical calculation of its eigenvalues and eigenvectors is not practical. When the number of samples (N) is smaller than the number of variables (3M), it can be shown that the rank of the covariance matrix is smaller or equal to N-1, meaning that the maximum number of non-zero eigenvalues (and thus non-zero eigenvectors) is equal to N-1[8] . Therefore, it can be shown that it is more efficient to calculate the covariance matrix of the transpose data which is an NxN matrix whose eigenvalues are the same as the non-zero eigenvalues of the original covariance matrix C. The eigenvectors of C can then be calculated from the corresponding eigenvectors of smaller covariance matrix as described in the literature[8-9].

Another method that is computationally faster and more efficient is based on the singular value decomposition (SVD) of the initial data matrix with its variables along the columns and the samples along the rows. The mathematical proof of the relation between the two methods has been provided in several publications and will not be discussed here[15]. Both methods were implemented using a commercial software package, MATLAB® (Mathworks, Natick, MA) and tested for accuracy, and speed. Overall the SVD method appeared to be more robust, which is likely caused by specific characteristics of the implementation of various built-in functions in MATLAB. While a theoretical evaluation of the limits of accuracy for each technique may be possible, it is beyond the scope of this work. As a result of this initial investigation, it was decided

that the SVD method would be more appropriate for analysis of the data in this study. The accuracy of SVD was then evaluated numerically as follows.

IV.B.4.ii. ***Computation accuracy***

An easy test of accuracy for the SVD method is to perform SVD on the original data matrix, then multiply the resulting matrices together to see how close the result is to the original data. This test was performed on a high resolution displacement vector field, as well as a subsampled (by a factor of 2) representation of the original data matrix. The results indicated that the larger the matrix size, the lower the accuracy of the SVD calculation (Figure IV.6). This was expected, since SVD is an iterative method, and the MATLAB implementation has preset limits for the number of iterations as well as the stopping criteria. For the lower resolution data used here, using the SVD method is a reasonable approach.

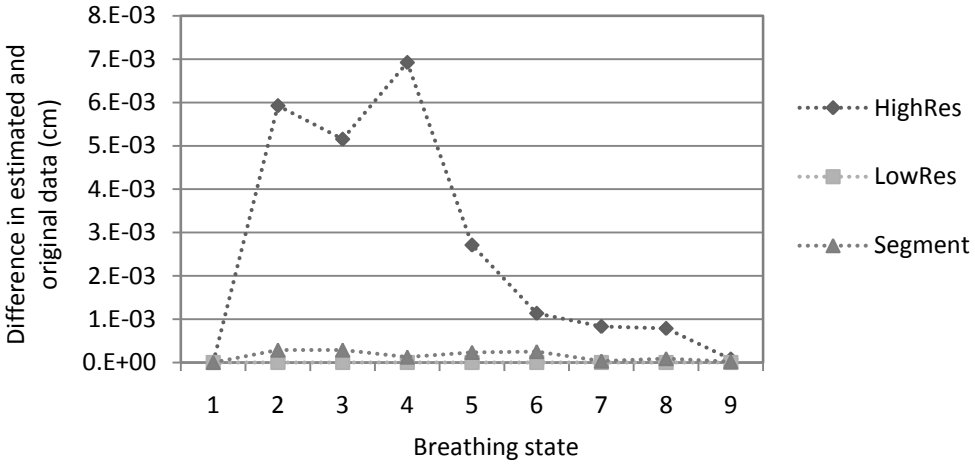

Figure IV.6: Root mean square difference between all data points estimated with SVD

IV.B.4.iii. ***Landmark selection accuracy***

The manually measured position of landmarks selected inside the lung is used as the ground-truth deformation for evaluation of model accuracy. As a result, the accuracy in the measurement of the true position of these markers can have a significant impact on our estimates of model accuracy, and should be evaluated and quantified carefully. The limitations in manual landmark

identification in lung geometry were briefly discussed in Chapter III and estimated for the small markers implanted inside the deforming phantom. For 4D CT images of the lung, the identification of these landmarks is limited due to image quality, especially in regions with larger breathing artifacts in the CT image (e.g. near the diaphragm).

As described earlier, a total of 60 landmarks were identified on all 10 phases of a 4D CT scan of a lung patient. Three of these landmarks were eliminated from the study because of the difficulty in accurately identifying their position in the longitudinal direction for scans at some breathing states. All landmarks were identified by one individual in order to eliminate any possibility of inter-observer variations in the selection of these points. Repeat measurements of the position of 6 landmarks (two in the upper part of lung, two in the middle, and two near the diaphragm) on 4 different breathing states (end-inhale, end-exhale, mid-inhale, and mid-exhale) provided an estimate of landmark identification accuracy in different regions of the lung and at different breathing states. Based on these repeat measurements, the accuracy of manual landmark identification (σ) is on the order of less than 1mm in the RL and AP directions and 1.1mm in the SI directions. It is important to note that, for at least some of the points, errors in position on the order of the voxel size (0.7x.7x.3 mm) are to be expected due to CT sampling effects. Also for landmarks near the diaphragm, the accuracy decreases slightly due to motion artifacts in the image.

IV.B.4.iv. *Impact of deformable registration accuracy on model accuracy*

Earlier in this chapter, the potential of PCA in reducing registration error was described as one of the main benefits of this model. However, the negative impact of registration error on the accuracy of PCA based model should also be considered and investigated. One of the assumptions made about the data in order for PCA to be applicable is that deformable registration is accurate enough that a large variance would correspond to dynamics of variation in the data and not to the error in the alignment. Although this model is expected to be less sensitive to the random errors in registration, which would show up in the less significant eigenvectors that

are not included in the model, this will not be true in case of large systematic errors in the registration. Therefore, the impact of registration accuracy on our estimates of model accuracy needs to be evaluated.

Using the B-splines based registration technique described in Chapter III, two different sets of registrations were performed (different parameter settings for resolution and knot spacing) to map the reference breathing state (end-inhale) to all other states in the 4D CT scan. For the remainder of this chapter, these two registration techniques are referred to as *I1* and *I2*, where 1 and 2 correspond to the final resolution of the image used in the registration. The registration quality was relatively consistent between the two when evaluated visually using overlays of the reference and deformed geometry. However, a quantitative analysis of the results, using the manually identified landmarks for comparison, indicated some differences between the two registration methods in some regions.

Figure IV.7 shows that the maximum error is higher than 0.7cm in the superior-inferior (SI) direction with variations larger than 0.2cm between the two methods for some breathing states (0.5 cm and 0.75cm maximum in the SI direction for breathing state 1 (inhale). Looking more closely at registration accuracy in the SI direction for the first breathing states (Figure IV.8) indicates that the error is relatively constant over all landmarks except for a few that show differences larger than 1mm (e.g. landmark 50)

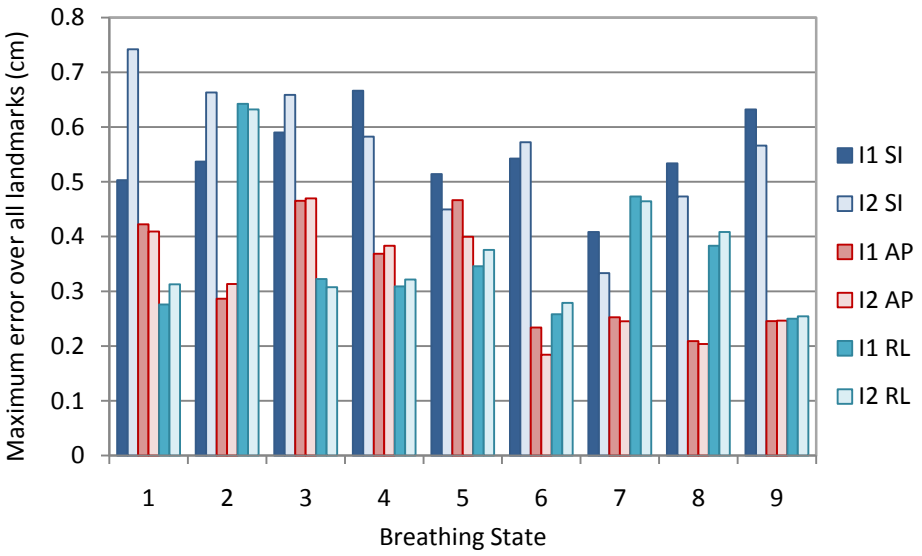

Figure IV.7: Maximum error over all landmarks for each breathing state. I1 and I2 refere to image resolution used during registration. SI: Superior-Inferior, AP: Anterior-posterior, RL: Right-Left.

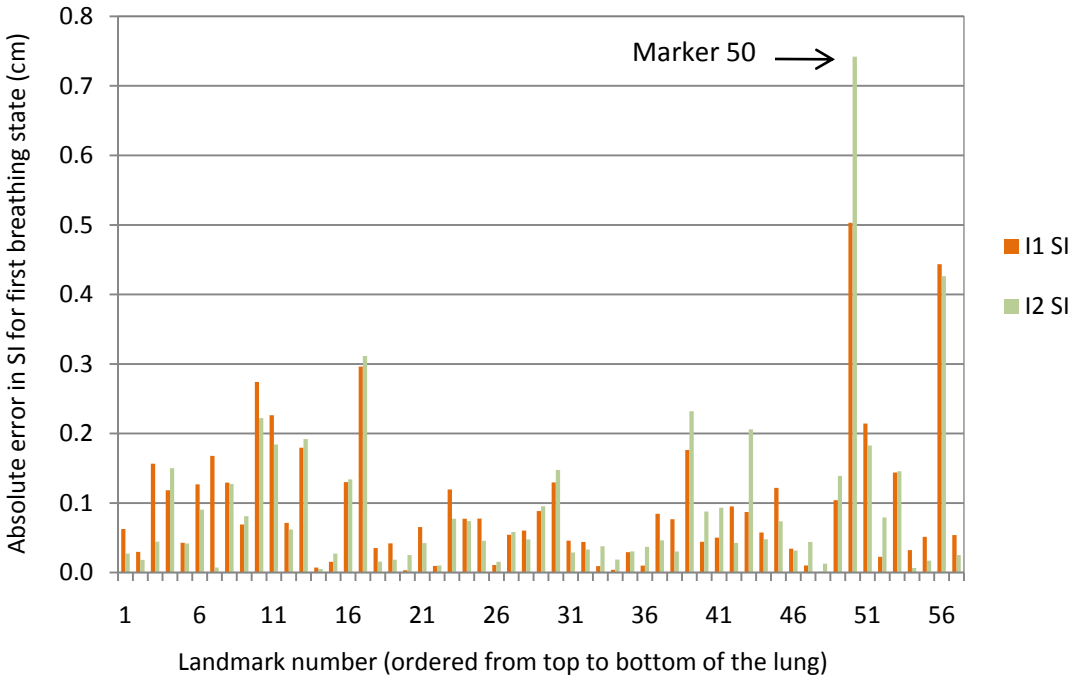

Figure IV.8: Registration error in the SI direction for each marker at the first breathing state. Landmaks that show larger differences such as landmakr 50 are useful in determining the effect of registration error on model accuracy.

These variations in alignment accuracy between the two registration methods represent only a subset of possible errors. The quantitative analysis of registration error confirms our visual observation that no systematic differences exists between the two methods. This prevents us from evaluating the model performance in the presence of large errors in the registration. However, those landmarks that show larger error differences (e.g. landmark 50) may be useful in further analyzing the impact of registration differences on model accuracy.

Keeping all parameters the same, the eigenvectors calculated from the deformation maps created by each registration method were determined. Using 1, 2, or 3 eigenmodes in the model, the error in the model-estimated position of the landmark of interest (50) was determined. It was observed that for the registration method, which showed the smaller error initially (*I1*), there was no significant changes in the accuracy of estimating landmark 50, regardless of the number of eignemodes used in the model. For the registration method with the larger initial registration error (*I2*), the error in estimation of the position of landmark 50 was reduced by about 1.5mm for breathing state 1, but increased by the same amount for some breathing states near the end of the respiratory cycle (Figure IV.9).

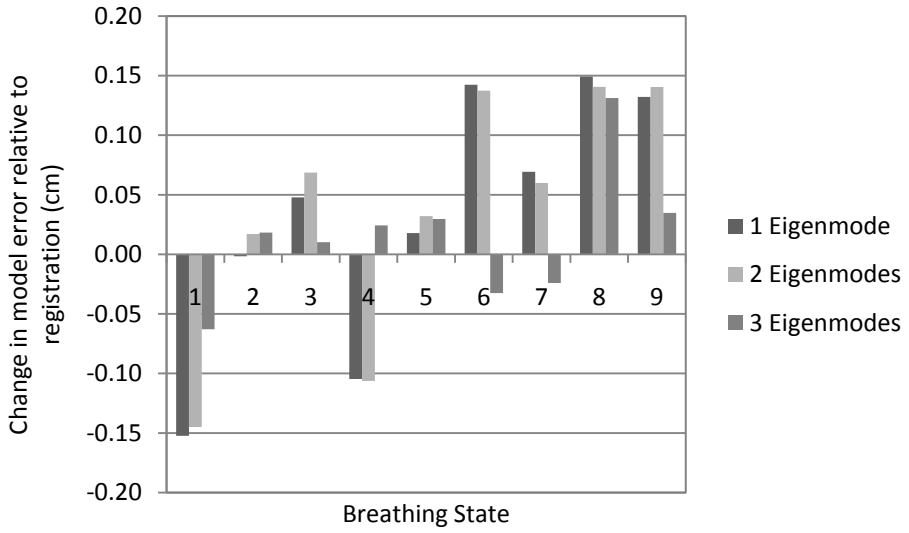

Figure IV.9: Change in the error in estimation of the position of a single landmark (50), using different number of eigenmodes. Negative numbers indicae an improvement in the estimation of the landmark position relative to the registration.

Though this may not appear to be a very interesting finding initially, it is actually in agreement with our expectation of the behavior of the model in response to random versus systematic registration errors. This suggests that the random large error observed for landmark 50 from registration *I2* was removed by the model when the lower ranked eigenmodes (eigenmodes 4 though 8) were eliminated from the model. In addition, we observed that when using the eignmodes estimated from deformation maps from *I1* registration, there was no significant reduction in the error with elimination of higher order eigenmodes. This indicates that this portion of the error contributed to the estimation of the first three eigenmodes, thus suggesting that there is a more systematic error in the registration field.

IV.B.5. Effectiveness and accuracy of the model in tracking deformation

In this section, the accuracy of the model under various conditions that represent situations likely to be encountered in real patients is evaluated.

IV.B.5.i. ***Displacement fields of individual eigenmodes***

The deformation described by each eigenmode about the mean geometry can be estimated by Equation IV.13, where σ_k^2 is the variance described by the corresponding eigenvector (i.e. the eigenvalue $w_k = \sigma_k^2$)[16-18]. Therefore the position (or deformation) vector P in equation IV.13, is the deformation of the mean geometry by eigenvector e_k .

$$P = \bar{P} \pm \sigma_k . e_k \qquad \text{IV.13}$$

The variation about the mean geometry of the lung for the first three eigenmode was estimated and is shown below for contours at three different longitudinal positions in the lung. The approximate position of these contours is displayed in Figure IV.10.

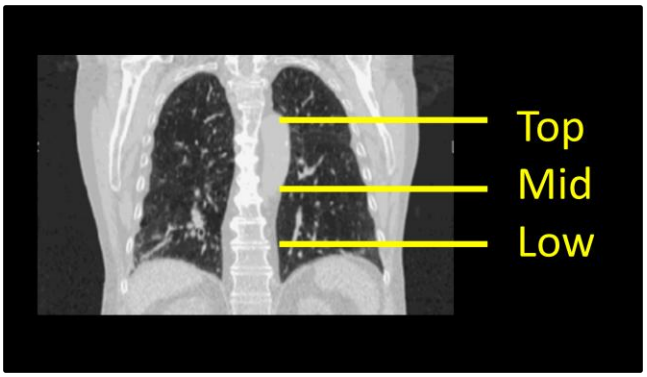

Figure IV.10. Lung contours from three different levels in the lung (Top, Mid, Low) were selected for visualization of the first three eigenmodes (Figures IV.11 to IV.19)

For each longitudinal level identified in Figure IV.10 (Top, Mid, Low) the variation about the mean lung contour described by each of the first three eigenmodes estimated from Equation IV.13 is displayed as arrows for each contour point (Figure IV.11 to IV.19).

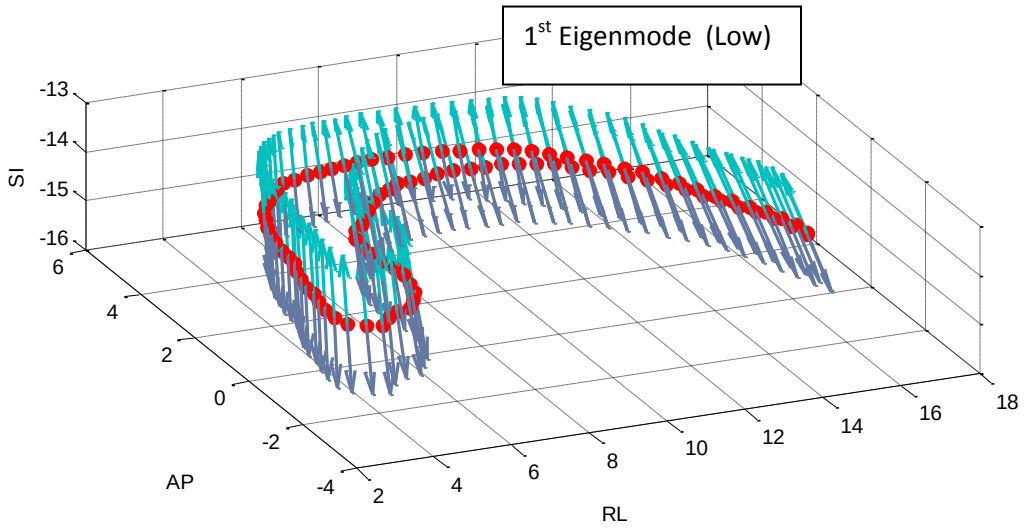

Figure IV.11. Variation in the mean geometry, described by the first eigenmode, for the contour from the lower lung (Low). The red dots are the contour points at their mean position estimated over the full breathing cycle. The green arrows correspond to the positive variation (+σ), while the blue arrows represent the negative varitation (-σ) from Equation IV.13.

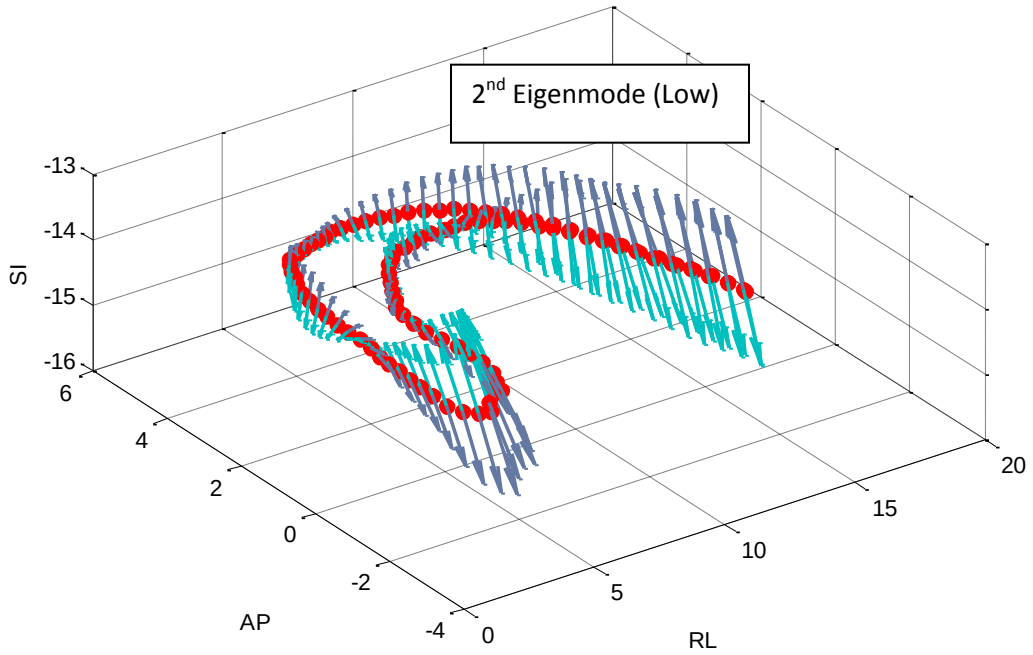

Figure IV.12. Variation in the mean geometry, described by the second eigenmode, for the contour from the lower lung (Low).

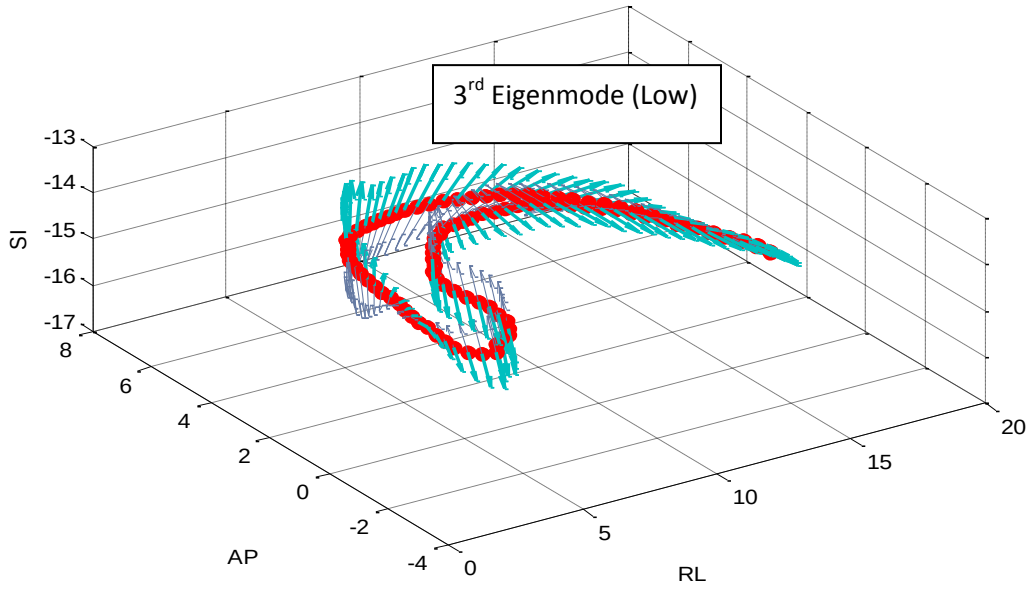

Figure IV.13. Variation in the mean geometry, described by the third eigenmode, for the contour from the lower lung (Low).

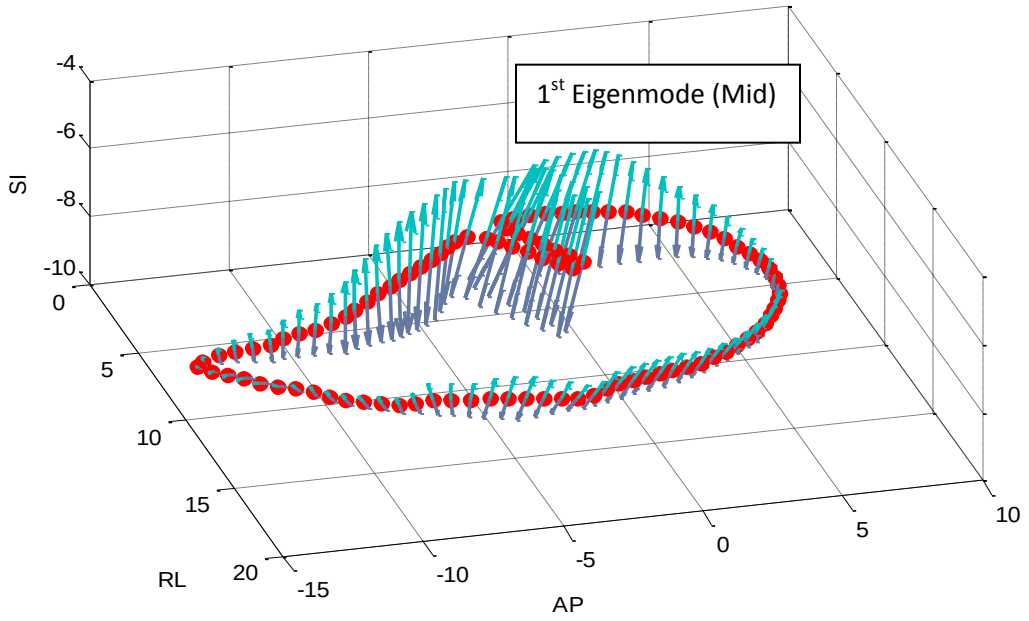

Figure IV.14. Variation in the mean geometry, described by the first eigenmode, for the contour from the middle of the lung (Mid).

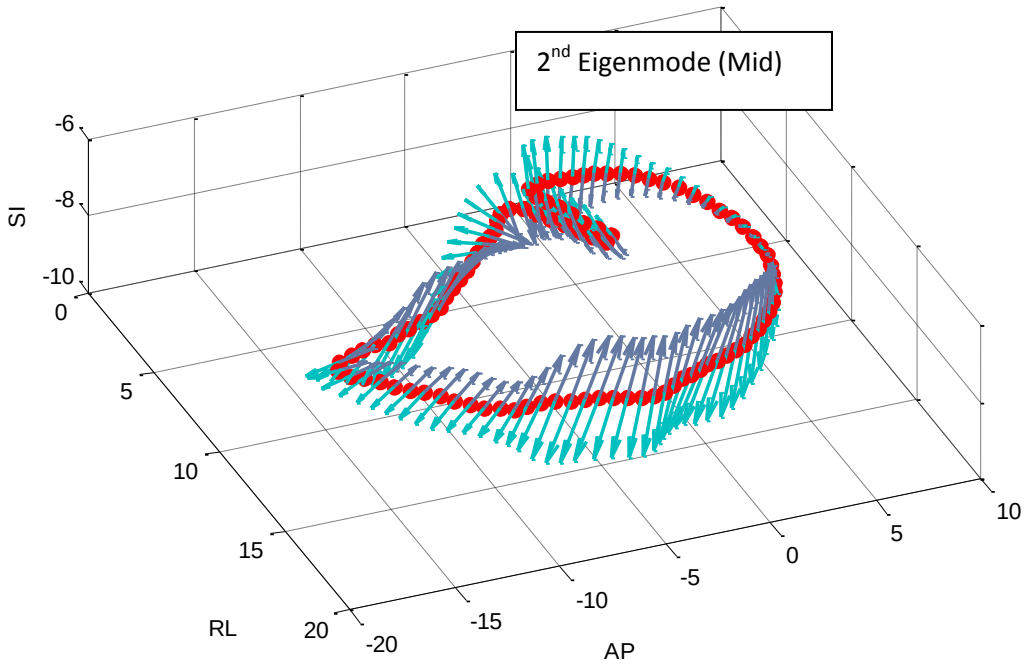

Figure IV.15. Variation in the mean geometry, described by the second eigenmode, for the contour from the middle of the lung (Mid).

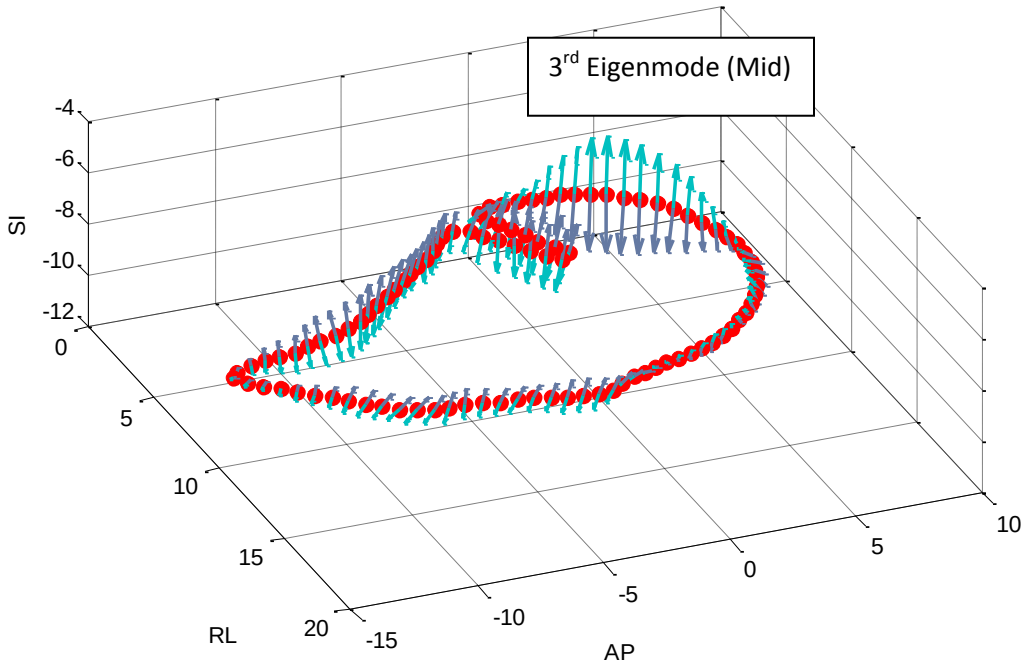

Figure IV.16. Variation in the mean geometry, described by the third eigenmode, for the contour from the middle of the lung (Mid).

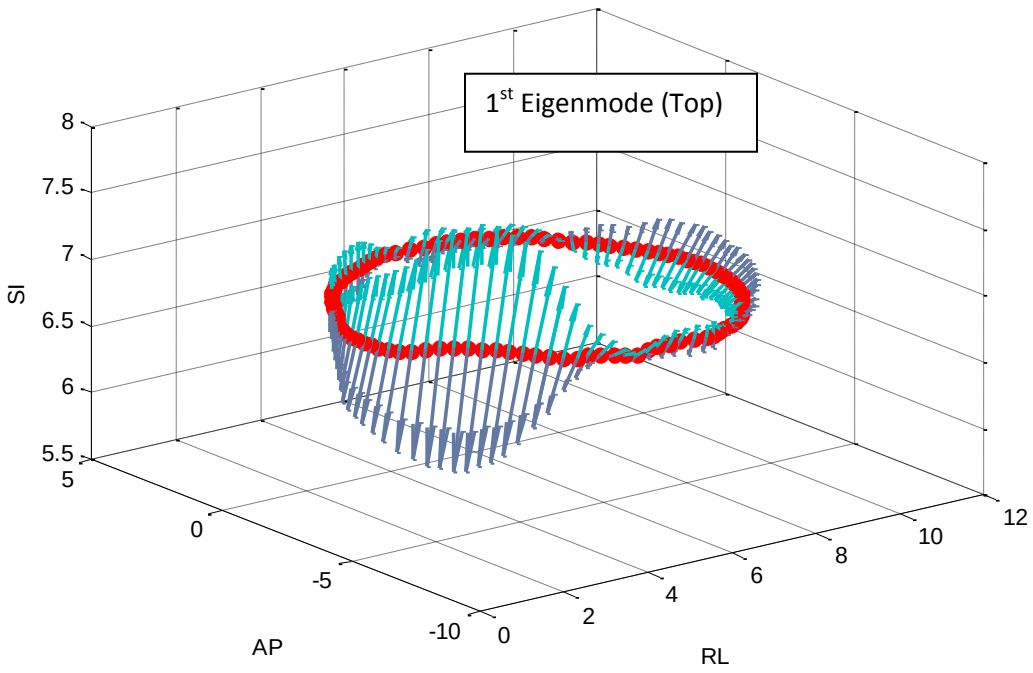

Figure IV.17. Variation in the mean geometry, described by the first eigenmode, for the contour from the upper lung (Top).

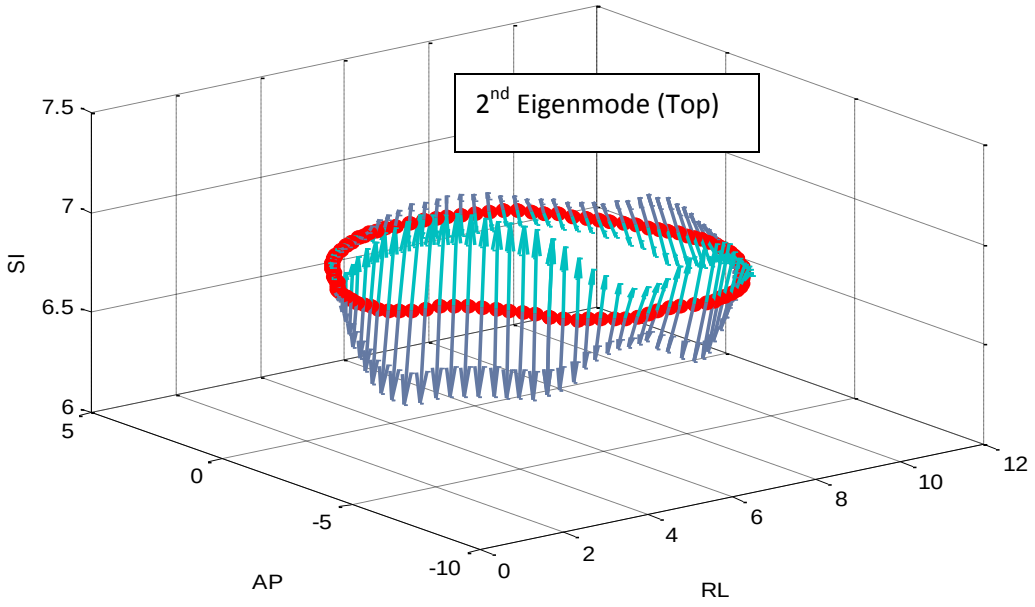

Figure IV.18. Variation in the mean geometry, described by the second eigenmode, for the contour from the upper lung (Top).

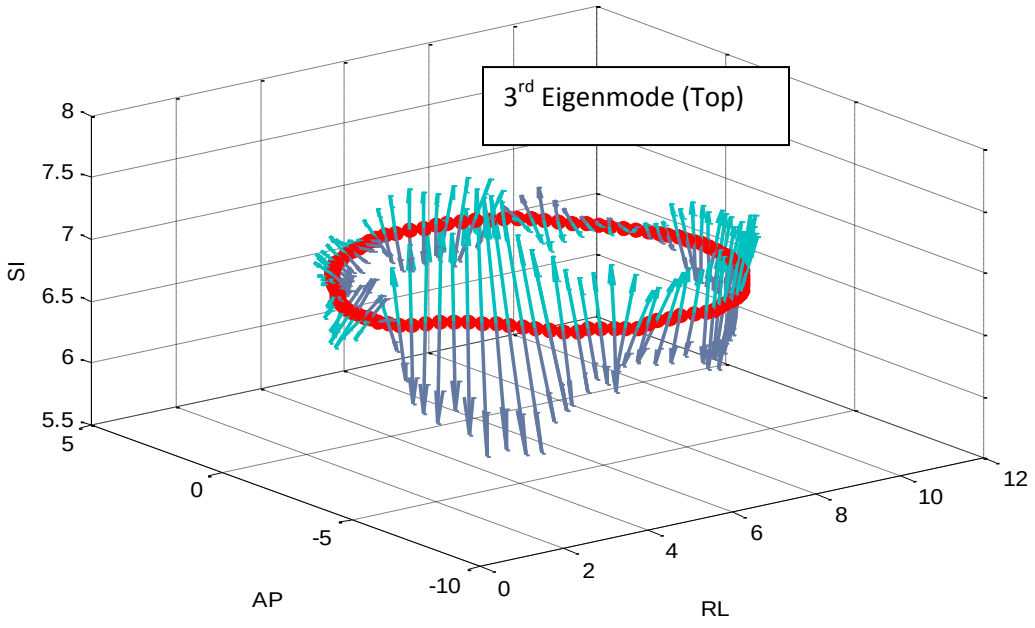

Figure IV.19. Variation in the mean geometry, described by the third eigenmode, for the contour from the upper lung (Top).

IV.B.5.ii. ***Determination of the optimal set of eigenmodes for modeling***

The ranking of the eigenvectors based on their corresponding eigenvalues provides a good estimate of the significance of each eigenmode in describing the variation in the data. The higher ranked eigenvectors and eigenvalues describe the most dominant modes of variation in the data, while the lower ranked eigenvalues and eigenvectors generally describe variations that are not as dominant in the data, as well as noise in the estimation of the geometric change or deformation (deformable registration error). Therefore the optimal set of eigenmodes used for modeling the variations in the geometry depends on the region of interest. Since the objective of our model is to estimate deformation in all points throughout the geometry using a few surrogates, in this initial evaluation of the optimal set of eigenmodes we determine the combination of the eigenmodes that would minimize the cumulative error over all points inside the geometry. A detailed evaluation of the local variations, based on using smaller segments of the geometry in the model, will be discussed later.

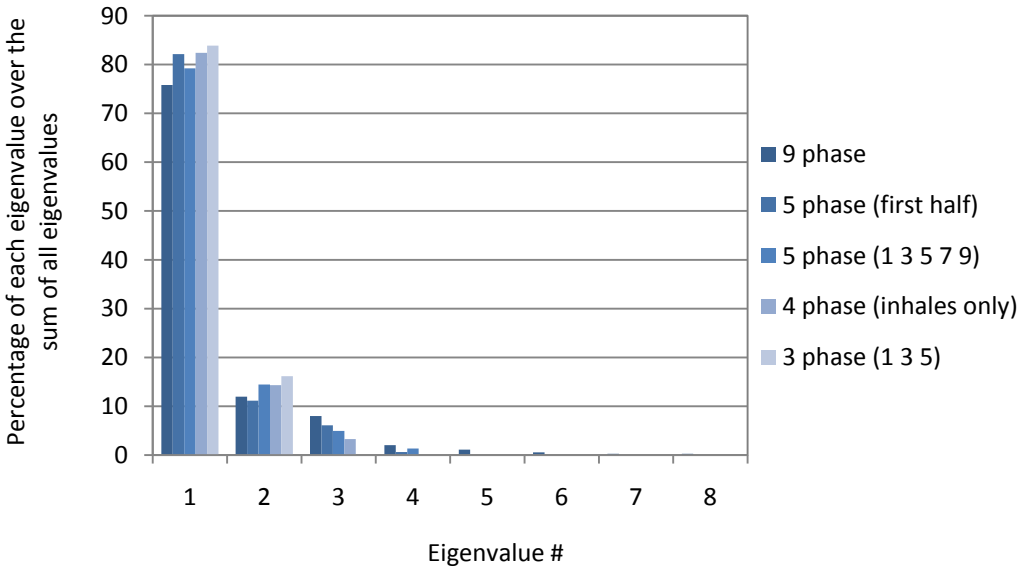

Figure IV.20: Percentage of variation described by each eiegnvalue, for different combinations of breathing states used in the model

Figure IV.20 shows the percentage of each eigenvalue relative to the sum of all eigenvalues for various combinations of breathing states used in the model. This distribution was calculated from

the 4D CT data from the sample patient. It is observed that regardless of the breathing states used, the first two eigenmodes describe more than 85%, and the first three describe more than 95% of the variation in the data. Thus, it is expected that either the first two or the first three eigenvectors should be sufficient in modeling the deformation in the data for most regions of interest. To test this assumption, the positions of all manually identified landmarks were estimated using different number of eiegnmodes and compared to their true positions. All available samples of geometry (i.e. the 9 displacement vector fields from inhale to the other 9 breathing states of the 4D CT scan) were used in calculating the eigenmodes, and then the optimal weights for each eigenvector was calculated and used to estimate the position of the landmarks at all sampled breathing states. The optimal weights were calculated as described in Equation IV.4.

Figure IV.21 shows the mean and maximum error observed in each dimension over all markers at each breathing state, using different numbers of eigenmodes in the model. This figure indicates that the first eigenvector is clearly not sufficient to model the variation in the data. For most breathing states, both the mean and the maximum error decrease when 2 or 3 eigenmodes are used in the model, while in some cases including the 4^{th} or 5^{th} eigenmodes results in a gradual increase in the error. It should also be noted that there is a large variation between the accuracy that can be achieved for different breathing states using different number of eigenmodes, indicating that less dominant variations specific to individual breathing states, are modeled with higher order eigenmodes.

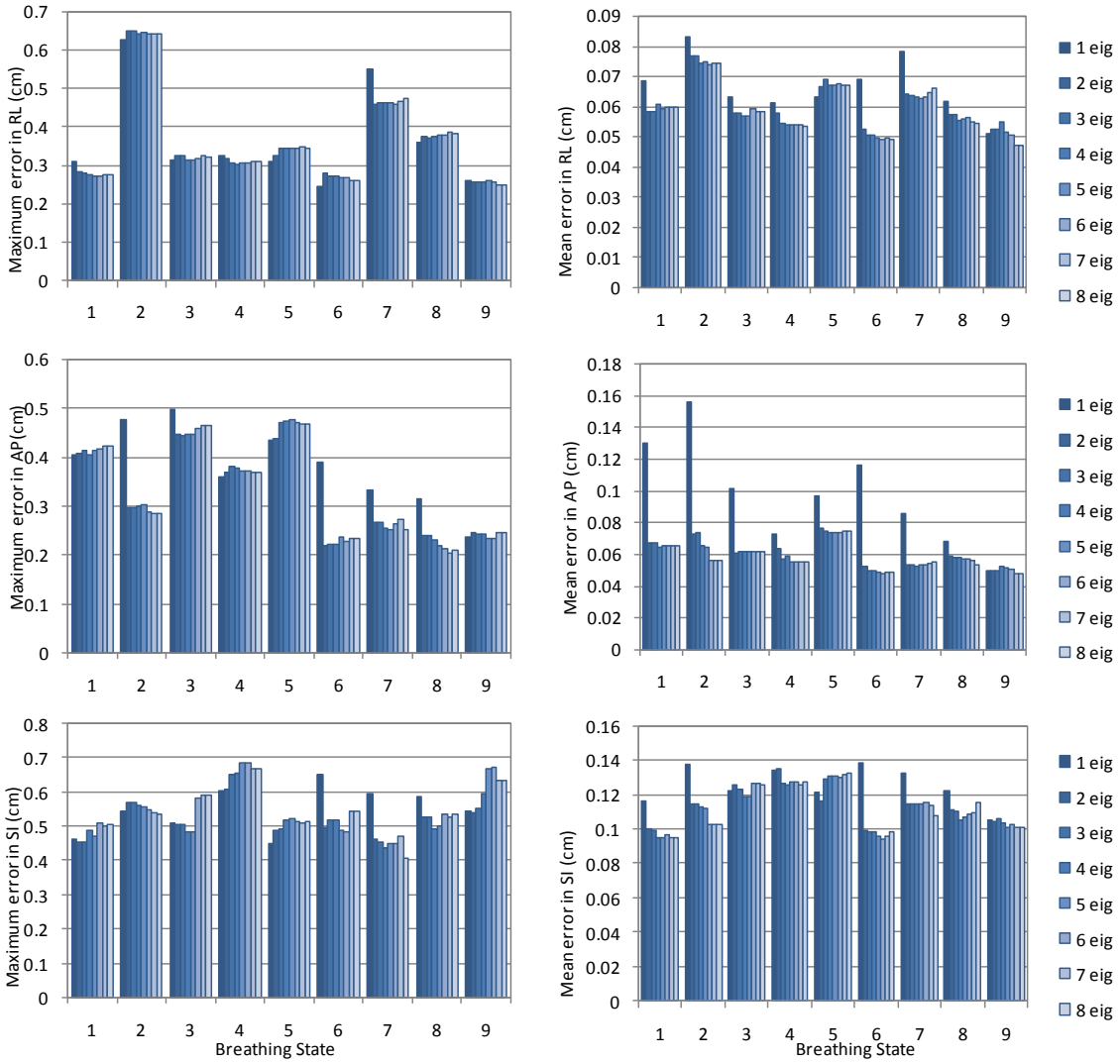

Figure IV.21: The left column shows the maximum error in RL, AP, and SI directions (top to bottom) for each breathing state, using different number of eigenmodes in the model. The right column shows the mean error for the corresponding directions.

In order to determine how many eigenmodes to use in the model, the model accuracy for all breathing states were combined and the mean and maximum error in each direction was calculated over all breathing states and for all markers as shown in Figure IV.22. Although this would ignore some of the characteristics observed in specific breathing states, it is a reasonable approach, considering that in real-time monitoring and tracking of the deformation using surrogates, we would not have any knowledge of the breathing state to use the optimal set of eigenmodes for that state. In fact, it is very likely that the deformation state to be estimated or

measured is one that is not sampled in a 4D CT reconstruction with 10 breathing states. Based on the results shown in Figure IV.22, the mean error shows a clear drop from one to two eigenmodes and then remains stable for all three dimensions. The lowest maximum in the SI direction is observed when using 2 eigenmodes, with a gradual increase in the maximum error as more eigenmodes are added to the model. The error in the AP direction shows a similar drop in the maximum from one to two eigenmodes, although it stays relatively stable after that, while the RL direction shows a slight increase in the maximum error from the first to the second eigenmode.

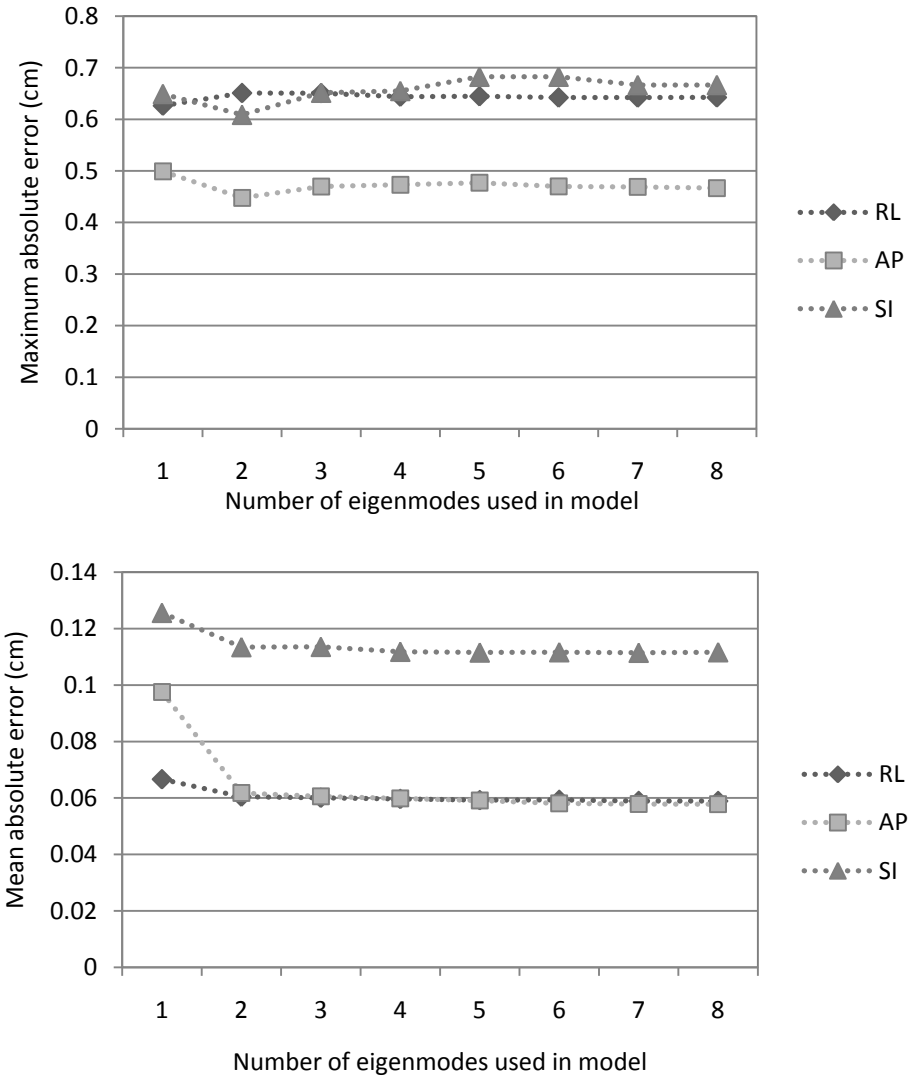

Figure IV.22: The maximum (top) and mean (bottom) error over all landmarks for all breathing states shown as a function of the number of eigenmodes used in the model.

Based on these results, we believe that using the first two eigenmodes would result in the best overall model accuracy.

In addition to the evaluation of the overall maximum and minimum errors for all landmarks and all breathing states, the error distribution over all landmarks was also evaluated by creating the histogram of the error when different numbers of eigenmodes were used in the model. Figures IV.23, IV.24 and IV.25 show the number of landmark positions estimated at any breathing state with an error smaller than or equal to the bin value, for different numbers of eigenmodes.

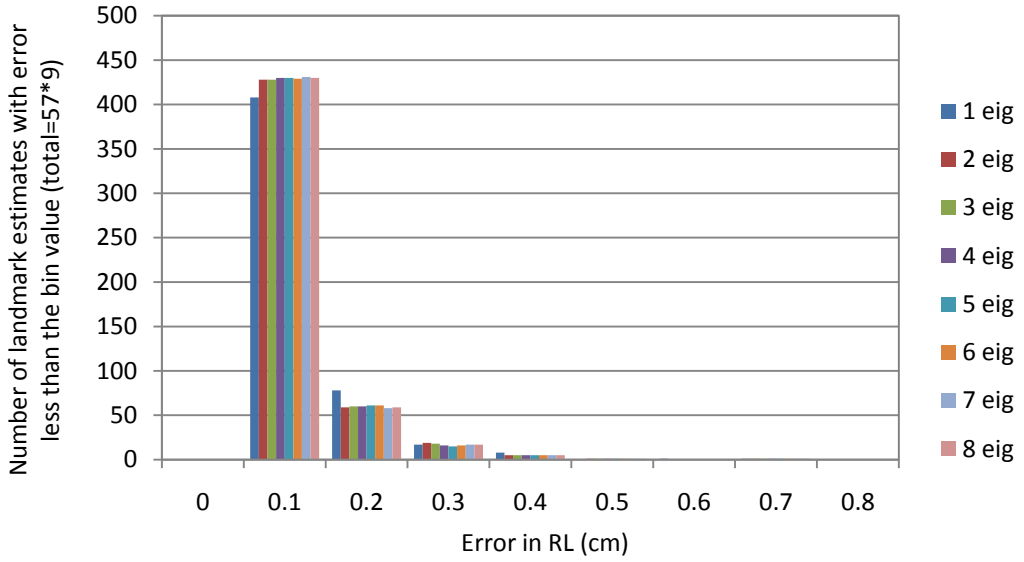

Figure IV.23: Histogram of the error in the estimated landmark positions in the RL direction using different number of eigenmodes over all breathing states. 1 eig indicates that only the first eigenmode was used. 8 eig is when all eigenmodes are included in the model.

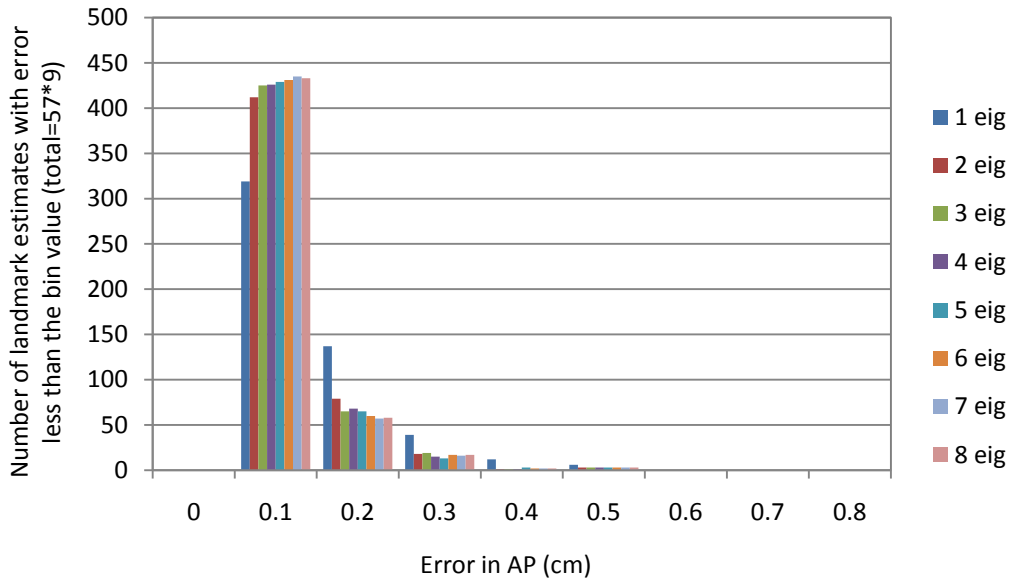

Figure IV.24: Histogram of the error in the estimated landmark positions in the AP directgion using different number of eigenmodes over all breathing states. 1 eig indicates that only the first eigenmode was used. 8 eig is when all eigenmodes are included in the model.

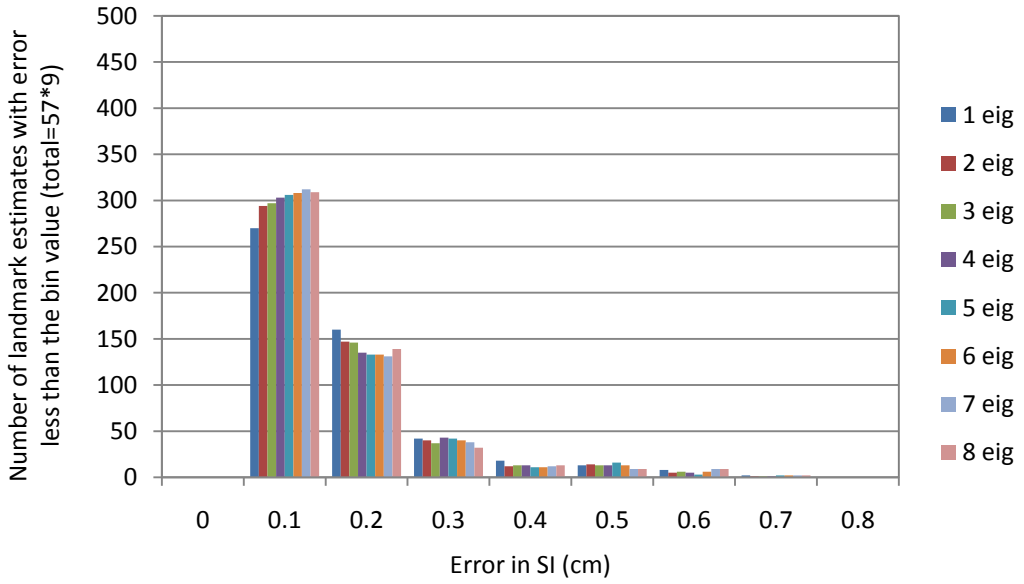

Figure IV.25: Histogram of the error in the estimated landmark positions in the SI direction using different number of eigenmodes over all breathing states. 1 eig indicates that only the first eigenmode was used. 8 eig is when all eigenmodes are included in the model.

Figures IV.23 to IV.25 show that the number of landmarks that fall within the smallest error bin (0.1cm) increases with an increase in the number of eigenmodes included in the model. This

indicates that increasing the number of eigenmodes, improves the overall accuracy so that more landmarks are estimated correctly over more breathing states. However, as we look more closely at the larger error bins in Figure IV.26 it appears that including more eigenmodes also results in more landmarks showing larger errors.

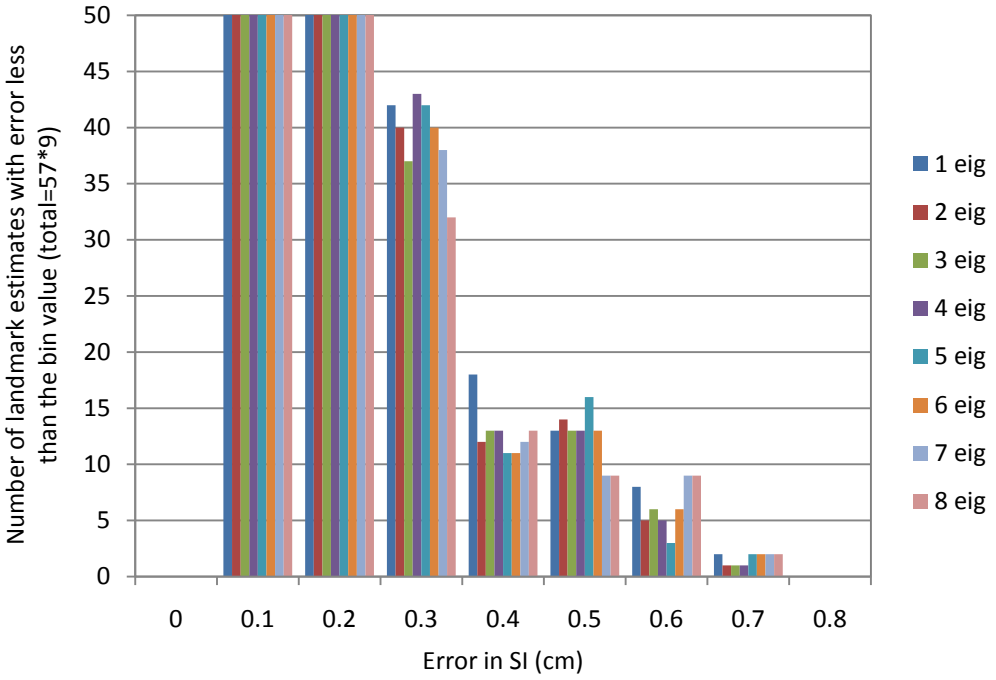

Figure IV.26: SI error histogram zoomed in on the larger errors, for different number of eigenmodes used in the model.

Based on these results, the optimal set of eigenmodes to be used in the model depends on the specific application and a set of predefined objectives for monitoring and tracking. If a larger number of small errors is more tolerable than a few large errors throughout the geometry of interest, then the optimal set of eigenmodes would include only the first few eigenmodes (i.e. 2 to 3). However, if the application can tolerate a few large errors in exchange for a better overall accuracy, then using a larger number of eigenmodes may be beneficial. This would also depend on the spatial distribution of the error relative to critical and normal structures inside the geometry.

Figure IV.27 shows the maximum error for each landmark over all breathing states, using different number of eigenmodes in the model. Ignoring all errors smaller than 3mm, we can see that for all the landmarks with large errors, after an initial drop in the maximum error when going from the first to the second eigenmode, the maximum error increases with inclusion of additional eigenmodes in the model. Although the increase is very small in some cases, it can be as large as 2mm for some landmarks (e.g. landmark 23).

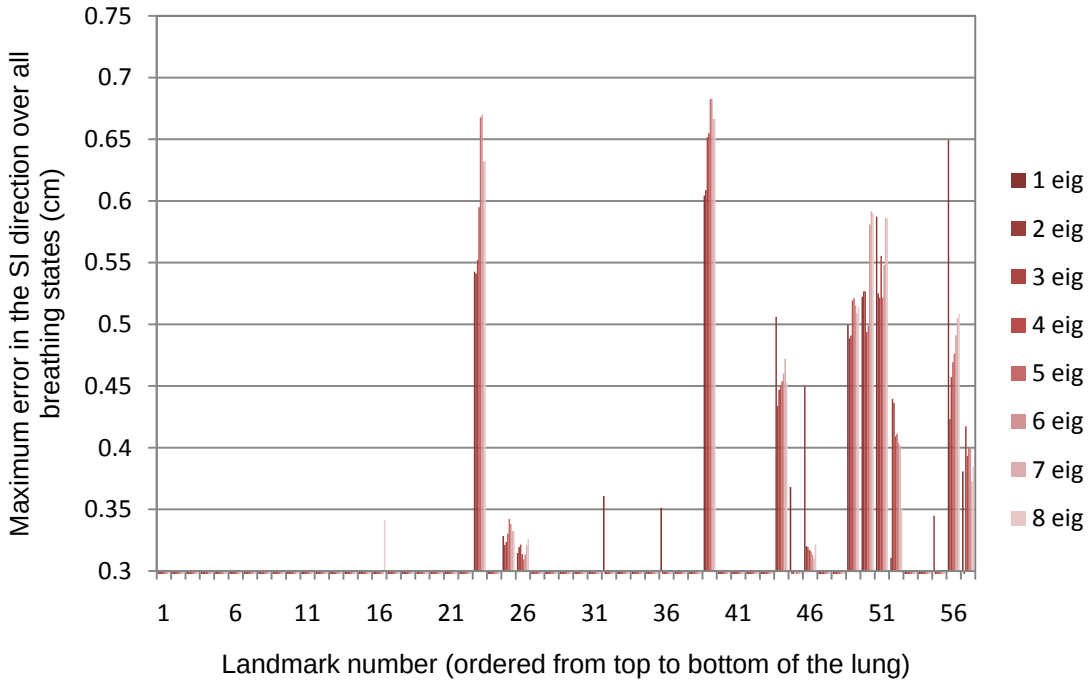

Figure IV.27: Maximum error in the SI direction over all breathing states for each marker when using different numbers of eigenmodes in the model

It can be concluded based on all the graphs provided in this section that the use of two eigenmodes in the model, is sufficient in most cases.

IV.B.5.iii. ***Surrogates and model accuracy***

In the previous section, the inherent limitations of the model implementation and their potential impact on the overall accuracy of the model was evaluated. This section focuses on the efficacy

of this model in tracking the variations in geometry, using surrogates. Thus far, the surrogates have played no role in the model. Instead, depending on the number of eigenmodes that were included in the model, the optimal weighting for each eigenmode was calculated (Equation IV.4) to minimize the difference between the measured (with deformable registration in this case) and the model estimated variations in the sampled states of patient geometry. At this stage, the surrogates are introduced to the problem, and their efficacy and accuracy in determining variation in the geometry is evaluated. The role of the surrogate is to provide frequently updated estimates of the patient deformation state, from which the variations in different regions within the geometry of interest can be estimated using the model.

Selecting three of the landmarks from the lower section of the lung (about 5cm from the diaphragm) to be used as surrogates, the position of all landmarks was estimated and compared to their manually measured position. The model estimated landmark position was also compared with the position of the landmark as estimated by deformable alignment, to simulate the impact of registration error on model accuracy and determine the efficacy of three surrogates in tracking deformation in lung geometry. For each landmark the maximum error in each dimension over all sampled breathing states was calculated (Figures IV.28 – 30).

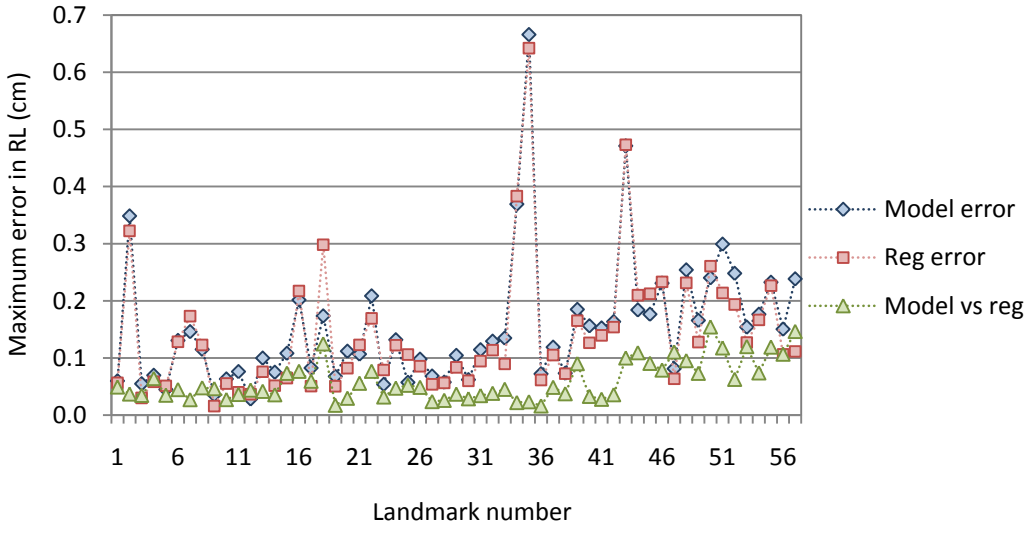

Figure IV.28: Maximum error in RL per landmark over all breathing states (blue diamond markers), registration error (red squares), and the error in landmarker relative to the position of landmarks as estimated by the deformation map (green triangles).

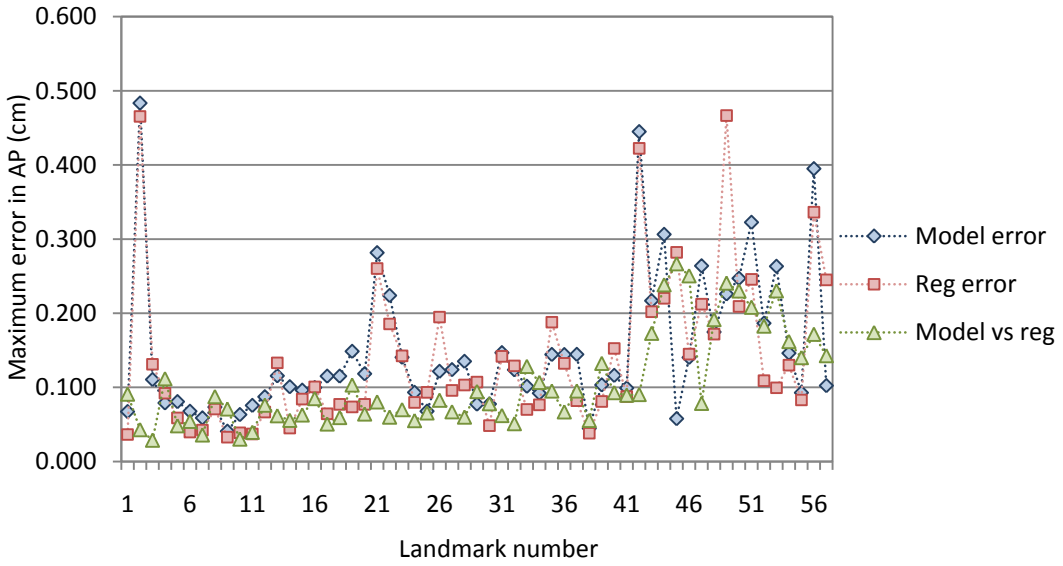

Figure IV.29: Maximum error in AP direction per landmark over all breathing states (blue diamond markers), registration error (red squares), and the error in landmarker relative to the position of landmarks as estimated by the deformation map (green triangles).

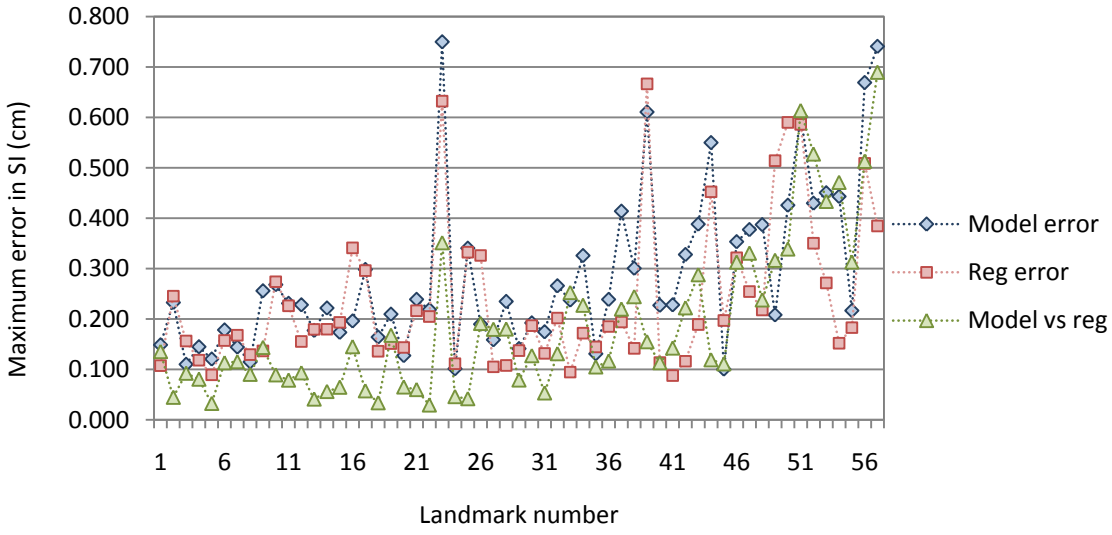

Figure IV.30: Maximum error in SI direction per landmark over all breathing states (blue diamond markers), registration error (red squares), and the error in landmarker relative to the position of landmarks as estimated by the deformation map (green triangles).

From this figure, it can be observed that the model improves the accuracy of landmark position estimation for some landmarks, while it increases the error in some other cases. For example, for landmark 57, the error increases more than 3mm in the SI direction as a result of modeling, while for marker 49 the error is reduced by about the same amount (approximately 3mm). For the same two markers, the error in AP direction is improved by 1.5 to 2mm when using the model and the error in the RL direction remains relatively constant with a change less than 1mm.

IV.B.5.iv. *Impact of the anatomic region used in modeling on the accuracy*

It is hypothesized that using a smaller segment of the deformation map of the lung would improve the accuracy of the model in estimating the position of landmarks in that region. If true, this may benefit cases with highly conformal plans where tracking or monitoring motion or deformation is limited to a small region of interest. Potentially, this may also be used for cases with more than one target, where two separate models (two sets of eigenmodes) can be used simultaneously as the position of the surrogate is updated during treatment.

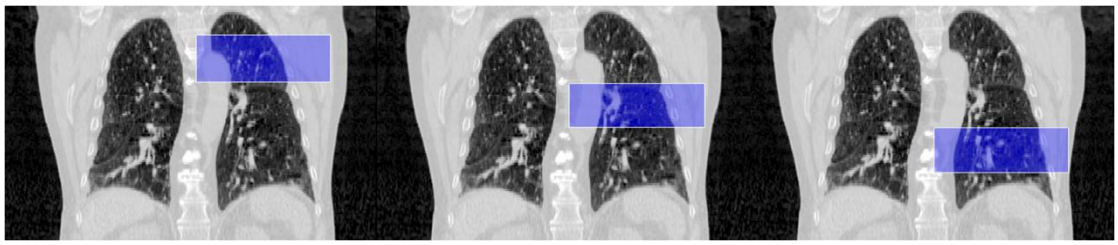

Figure IV.31: The small segments chosen in the deformation map, disaplayed on coronal images of the lung. Upper (left), Mid (middle), and Lower (right). Each section is 6cm in the longitudinal direction and the axial dimensions are the same as before.

To test this, three segments were identified in the lung as shown in Figure IV.31. Each section is 6cm long in the longitudinal direction, while the same limits as the initial case are applied in the axial plane. For each segment, the eigenvalues and eigenvectors were evaluated. Figure IV.32 shows the percentage of variation described by each eigenmode for the three different segments as well as for the whole lung. It is clear that the deformation characteristics of the segments are

quite different compared to each other. The upper part of the lung shows a larger distribution in the eigenvalues, indicating that the motion is not dominant along a single axis. The lower segment of the lung shows that the majority of the variation in this part of the geometry (about 75%) can be described by the first eigenmode, thus indicating a dominant deformation along a single basis.

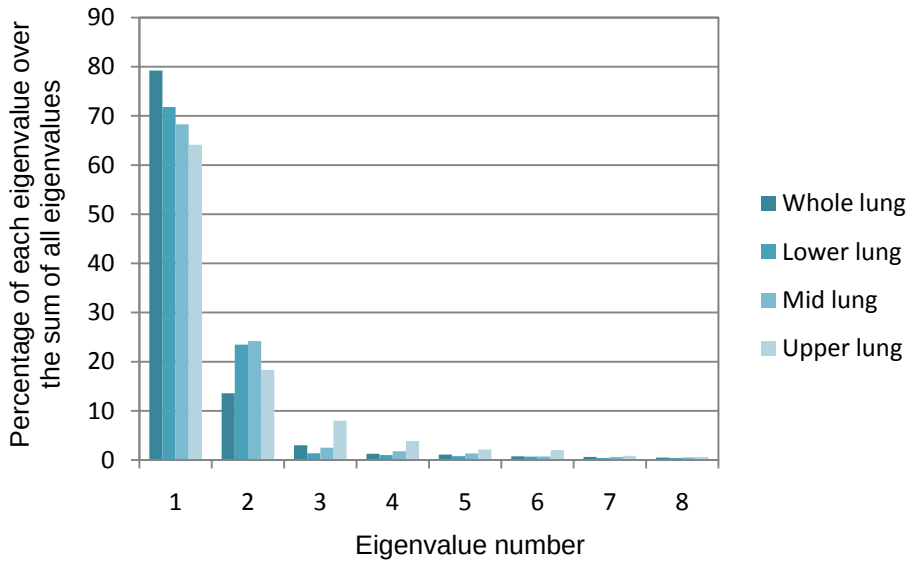

Figure IV.32: Eigenvalue distribution for each segment of the lung compared to the entire lung

Three landmarks were chosen for each segment to be used as the surrogates, and the position of the remaining landmarks in the lung was estimated. Despite our initial expectations, there was little difference observed between the accuracy of the model based on the segments, and that based on the full deformation map, when two eigenmodes were used. The maximum difference observed was less than 1mm.

IV.B.5.v. *Impact of variations in breathing on model accuracy*

The most significant variations observed in tumors over the course of treatment, besides tumor shrinkage, are the change in the baseline position of the tumor and the amplitude of motion[5-6]. In order to evaluate the impact of change in amplitude of tumor motion on the accuracy of the

model, two experiments were performed. In the first experiment, the inhale phases of the 4D CT were eliminated from the modeling step (i.e. phases 1, 2, 8, 9 in Figure IV.2), and the resulting coefficient matrix was estimated and used with surrogate positions at inhale to determine the position of other landmarks at inhale. The results are displayed in Figure IV.33, for breathing state 1. In the second experiment, a similar approach was taken for the end-exhale states, where phases 3, 4, and 5 were eliminated from the model generation step, simulating a change of 0.5cm in the exhale position between the modeling stage and the position at the time of monitoring (Figure IV.34).

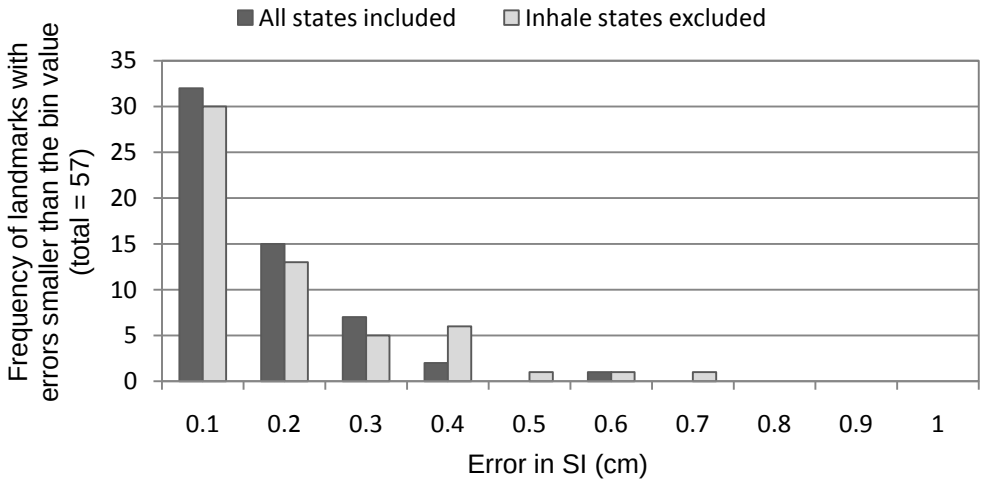

Figure IV.33: Error in estimating landmark position at inhale

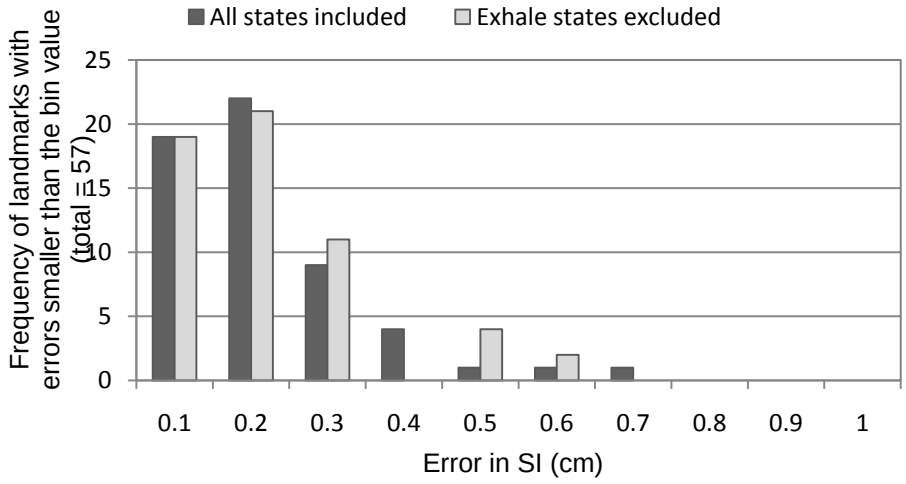

Figure IV.34: Error in estimation of the landmark position for the end-exhale state.

The results were compared to the condition when all breathing states were included in the model. Looking at Figure IV.34, it appears that excluding the exhale from the model has little effect on the estimation of the landmarks' positions at the end-exhale state, as the distribution in the error remains unchanged. However, this is not true for the inhale state, as can be seen in Figure IV.33. This figure shows an increase in the error in estimating the inhale state when the inhale states were eliminated from the modeling stage. Therefore, it can be concluded that a reduction in the amplitude of the breathing signal over time has less effect on the model accuracy than an increase in the amplitude.

IV.B.5.vi. ***Model accuracy in determining deformation in intermediate states not sampled by 4D CT***

For every breathing state sampled in the 4D CT scan, the position of the three surrogates is also known. Therefore, during tracking or monitoring, as the positions of the surrogates get updated, one can refer back the corresponding deformation map to look up the deformation in any other point in the geometry. However, the initial 4D CT is a discrete sample of the patient variations over the full breathing cycle. As a result, if the surrogate positions identified during treatment monitoring do not correspond to any of the sampled positions based on 4D CT, some form of interpolation would become necessary in order to determine the deformation state of the landmarks, based on the two adjacent deformation maps.

In this section, the accuracy of the model in determining landmark motion/deformation for breathing states not sampled with 4D CT, is evaluated and compared with a simple linear interpolation between the two deformation maps. To test this, 4 of the 9 breathing states (phases 2, 4, 6, 8) were eliminated from the modeling stage of the process, and the resulting eigenmodes were used to determine landmark positions when the surrogate was at one of these intermediate states. This was then compared with 1) a linear interpolation between the two surrounding deformation maps, 2) registration error for the intermediate state, 3) model accuracy for that

breathing state when it was included in the model. The results displayed in Figure IV.35 indicate that for some landmarks, the interpolation is more accurate, while for others the model shows better accuracy. Looking at landmark numbers 47 and higher, it appears that for a majority of the landmarks, using the model improves the accuracy compared to interpolation between deformation maps, with a few exceptions at landmarks 56 and 57. These two landmarks are very close to the diaphragm where the breathing artifacts in the 4D CT reconstruction, compromise the quality of deformable registration, and thus reduce the accuracy of the model.

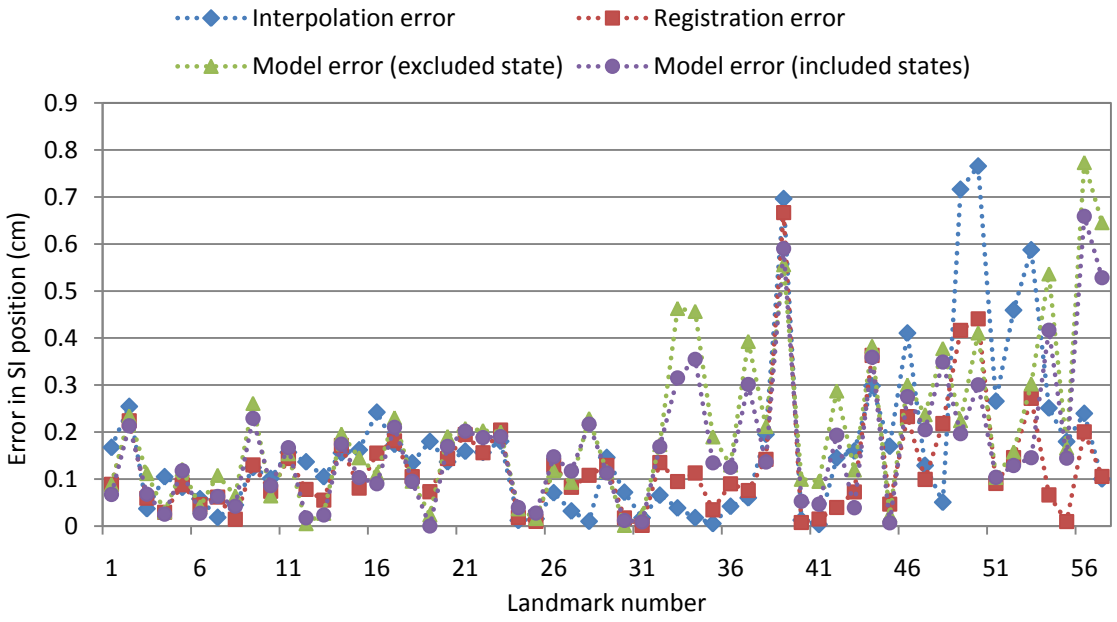

Figure IV.35: Model accuracy for intermediate states not sampled by the 4D CT

IV.C. Discussion

In this chapter, a statistical model of the deformation in the lung over a full breathing cycle was proposed, and the potential of this model in decreasing registration error was evaluated. The performance of the model under various conditions simulating changes in real patient breathing characteristics was evaluated, showing some improvements as well as certain potential short-comings. The model is partially limited by the input data from the deformable image registration, in the sense that if the deformation maps are systematically noisy, the model may not be able to

separate out the dominant motion, from the noise. This model also allows for tracking the deforming geometry of lung, using a few surrogates (implanted markers) whose position can be monitored and updated frequently. One of the benefits of this model is that it is specific to the patient, and thus eliminates the population-based assumptions about the individual patient's breathing characteristics. Such tools may be able to significantly enhance tracking over current methods, in which the rigid movement of the surrogate's average coordinate generally defines a uniform translation to be applied across the body.

References

[1] B Schaly, J A Kempe, G S Bauman, et al. “Tracking the dose distribution in radiation therapy by accounting for variable anatomy,” Phys Med Biol. 2004 Mar 7;49(5):791-805.

[2]M. Rosu, I. J. Chetty, J. M. Balter, M. L. Kessler, D. L. Mcshan, and R. K. Ten Haken, “Dose reconstruction in deforming lung anatomy: dose grid size effects and clinical implications,” Med. Phys. 32, 2487–2495 (2005).

[3]D. A. Low, P. J. Parikh, W. Lu, J. F. Dempsey, S. H. Wahab, J. P. Hubenschmidt, M. M. Nystrom, M. Handoko, and J. D. Bradley, “Novel Breathing motion model for radiotherapy,” Int. J. Radiat. Oncol. Biol. Phys. 63, 921–929 (2005).

[4]Shirato H, Shimizu S, Kitamura K, *et al*. Four –dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor. *Int J Radiat Oncol Biol Phys* 2000;48:435-442.

[5]Hugo G, Vargas C, Liang J, Kestin L, Wong JW, Yan D. Changes in the respiratory pattern during radiotherapy for cancer in the lung. Radiother Oncol. 2006 Mar;78(3):326-31.

[6]Sonke JJ, Lebesque J, van Herk M. Variability of four-dimensional computed tomography patient models. Int J Radiat Oncol Biol Phys. 2008 Feb 1;70(2):590-8.

[7] D. Manke, K. Nehrke, and P. Bornert, “Novel prospective respiratory motion correction approach for free-breathing coronary MR angiography using a patient-adapted affine motion model,” Magn. Reson. Med. 50, 122–131 (2003).

[8] M. Sohn, M. Birkner, D. Yan, and M. Alber, "Modelling Individual Geometric Variation Based on Dominant Eigenmodes of Organ Deformation: Implementation and Evaluation Phys.Med.Biol. **50,** 5893-5908 (2005).

[9]Zhang Q, Pevsner A, Hertanto A, Hu YC, Rosenzweig KE, Ling CC, Mageras GS., A patient-specific respiratory model of anatomical motion for radiation treatment planning., Med Phys. 2007 Dec;34(12):4772-81.

[10]H. H. Liu, N. Koch, G. Starkschall, M. Jacobson, K. Forster, Z. Liao, R. Komaki, and C. W. Stevens, "Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part II–margin reduction of internal target volume," Int. J. Radiat. Oncol., Biol., Phys. **60**(5), 1473–1483 (2004).

[11]J. Leong, "Implementation of random positioning error in computerised radiation treatment planning systems as a result of fractionation," Phys. Med. Biol. **32**(3), 327–334 (1987).

[12]W. A. Beckham, P. J. Keall, and J. V. Siebers, "A fluence-convolution method to calculate radiation therapy dose distributions that incorporate random set-up error," Phys. Med. Biol. **47**(19), 3465–3473 (2002).

[13]J. Duan, S. Shen, J. B. Fiveash, I. A. Brezovich, R. A. Popple, and P. N. Pareek, "Dosimetric effect of respiration-gated beam on IMRT delivery," Med. Phys. **30**(8), 2241–2252 (2003).

[14]E. Nioutsikou, N. S.-T. J. Richard, J. L. Bedford, and S. Webb, "Quantifying the effect of respiratory motion on lung tumor dosimetry with the aid of a breathing phantom with deforming lungs," Phys. Med. Biol. **51**(14), 3359–3374 (2006).

[15]J. J. Gerbrands, "On the Relationships between SVD, KLT and PCA," Pattern Recognit. **14,** 375-381 (1981).

[16]Jackson, J.E., A User's Guide to Principal Components, Wiley, 1988.
[17]Jolliffe, I.T. Principal Component Analysis, 2nd ed., Springer, 2002.

[18]Dryden I L, Mardia, KV, Statistical Shape Analysis,Wiley 1998.

[19] T.F. Cootes, C.J. Taylor, D. Cooper and J. Graham, “Active Shape Models—Their Training and Application,” Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59, Jan. 1995.

CHAPTER V

UNCERTAINTIES IN ADAPTIVE THERAPY FOR HEAD-AND-NECK CANCERS

The previous chapters described the uncertainties in the measurement and management of breathing motion and their impact on the accuracy of treatment delivery under various motion management methodologies. This chapter follows the concept of uncertainty in tracking and monitoring geometric variation, focusing on changes observed in response to treatment (e.g. weight loss, tumor shrinkage, or swelling) in head and neck cancer patients. It explores the uncertainties in tracking geometric and dosimetric variations over a course of fractionated treatment, and the implications of these uncertainties on plan modification and adaptation.

V.A. Introduction

V.A.1. Head and neck cancers and treatment strategies

One of the main concerns in treatment of head and neck cancers with radiation is the large number of noninvolved and critical organs that are in close proximity to the target. Proximity of the spinal cord and brainstem is a challenge in delivering adequate dose to the tumor, while other non-involved radiosensitive tissues such as the parotids and other salivary glands can lose their function over time and thus affect the quality of life in long term survivors. Implementation of intensity modulated radiation therapy (IMRT) allows for improved sparing of the noninvolved tissue, as opposed to previous 3-dimensional conformal[1]. However, the sharper dose gradients possible in IMRT require more accurate selection and delineation of all targeted tissues to avoid the risk of marginal recurrences. The combination of addressing this risk and sparing the noninvolved tissue has been stated as one of the reasons for the substantial variations among

physicians and institutions in their approach to target selection as well as delineation[2], as we discuss further in the next section.

V.A.2. Uncertainties in target definition

In head and neck cancers, the gross tumor volume (GTV) is determined based on contrast enhanced CT images, as well as physical exams. MRI is often used in specific types of head and neck cancer, for example where the proximity of the bones presents artifacts obscuring the target and thus compromising its delineation (e.g. nasopharyngeal cancers), or where the inherent contrast of CT imaging is insufficient. FDG-PET imaging has also been tested in combination with the standard contrast enhanced CT imaging[3], but whether it improves target delineation or not remains to be proved[1]. Physical examination of the patient, which is sometimes accompanied by endoscopy[1], continues to have a significant role in determining the exact extent of the primary tumor. Therefore, a complete reliance on imaging information for target delineation is not recommended.

Almost as important as defining the GTV is selection and delineation of the clinical target volume (CTV), which includes tissue that is expected to contain microscopic disease as well as lymph nodes that do not meet the criteria to be included in the GTV but are at high risk of metastatic disease[1]. In head and neck cancers, the CTVs are defined based on specific criteria such as tumor site, size, stage, and patterns of failure, and their delineation relies on the surrounding non-involved anatomy such as the musculature, bones, or air cavities. Significant variations in the selection as well as delineation of targets have been observed among physicians with different experience levels, as well as variations among different institutions' standards and protocols[4]. These variations have been the topic of several studies that attempt to quantify the uncertainty in target definition. Most of these studies provide the GTV and the staging of the tumor to the participants who are then asked to delineate the CTV[4]. Regardless of the method, most of these studies indicate significant variations among the participants that can be correlated to the

individual's expertise as well as the variations in the institution's standards of practice, or protocols.

Consideration of the uncertainty in the manual delineation of targets and normal tissues is particularly important in the context of this study, in which manual contours are taken as the ground-truth data for evaluation of the accuracy of deformed contours and surfaces estimated by image registration. Whether evaluating changes in the volume of the CTV or GTV over the course of treatment, or the accuracy of various geometric tracking methodologies for dose accumulation, it is important to understand the limits of accuracy of what we accept as the truth (i.e. the accuracy of our ruler must be understood).

V.A.3. Geometric variations over the course of treatment

The need for treatment adaptation in radiotherapy stems from the changes in the patient from the geometric model on which the original plan is based. These variations, which can be due to setup errors, anatomical changes in response to treatment (e.g. weight loss, tumor regression, etc), or daily variations due to spinal articulation, organ filling, or other physiological processes, have received more attention recently due to the increase in the availability of repeat imaging over the course of treatment[5-9] .Although some studies have characterized, with some success, the population estimates of motion in different organs (specific to the institution, localization, and immobilization techniques employed), very little has been done to find similar distributions for the higher order components of motion and deformation. Several studies have evaluated changes in the anatomy over the course of treatment, some of which have attempted to model the variation using mathematical functions or statistical[10-12]. One of the most comprehensive sets of serial imaging data for head and neck was described by Barker et al., who quantified the volumetric and geometric changes in this region using an in-room CT system to acquire images three times a week over the course of treatment[10]. They found an average 1.8% volume reduction per day for the GTV and the lymph nodes. They also observed a reduction in the size of the parotid gland, which resulted in a systematic motion of the center of mass of both parotids medially (i.e. toward

the high dose region). While such studies have provided us with documented and quantified evidence of change, these trends in target volume reduction and position variations should be considered with caution, given the variations in the definition of the target and normal structures on the repeat images of the patient as discussed in the previous section, as well as variations of individual patients within the populations studied.

V.A.4. Translating observed geometric changes into clinically meaningful metrics

Quantifying the changes in patient geometry alone does not provide the information needed to determine if plan modification is necessary. Instead, these decisions should be based on variations in dose distribution as well as the resulting changes in dose-based metrics (e.g. minimum or maximum dose) for the target and normal structures. The threshold for tolerable change for each metric may be influenced by the dose estimation method and its accuracy, as will be discussed in the next two sections. Although the eventual goal of adaptive therapy is to improve the treatment plan beyond what could be achieved at the time of initial treatment planning, the actual objectives set for treatment modification can vary. Most current studies simply focus on achieving the initially set objectives for tumor control and normal tissue sparing. While it is possible to use adaptive therapy for boosting the target dose, or sparing normal tissues not initially spared, the compromise between these two sets of objectives can only be made by the physician and after careful assessment of multiple patients, as well as any individual case. Therefore, in this work, the aim is to retain the initially set objectives through the entire course of treatment, such that a plan modification is triggered only if the initially met objectives are no longer met due to geometric variations. With this aim, the analysis of plan sensitivity to errors in estimation of geometric variation is based on the initial treatment plan.

V.A.5. Current correction strategies to account for geometric variation

In response to the geometric variations observed in head-and-neck patients over the course of treatment, a variety of methods have been proposed for treatment adaptation including field shape and fluence modification, as well as full re-optimization of the plan based on the new[6-9,13-

[15]. The majority of these techniques indicate the potential value of adaptation via simple paper exercises on limited data acquired locally, and not based on evidence of improvement in treatment response, or other metrics of clinical outcome. For example, institutions that acquired daily images have suggested treatment modification on a daily basis[13-15], and those that have weekly repeat scans of the patient studied weekly re-planning approaches[13]. While these may be worthwhile approaches and could potentially prove to be beneficial, currently there is no consensus as to which of these proposed methodologies would be the optimal adaptation technique.

Most correction strategies described in the literature rely on estimating the dose distribution for the new patient model, and comparing this single fraction dose to the original plan. Although this is a valid approach when evaluating the error in the daily dose to the patient, it does not describe the cumulative effects of these variations over the course of treatment. The cumulative effect of the smaller variations in geometry and dose would be less of a concern if daily plan modifications are performed with high levels of accuracy. However, if the daily variation in dose is small enough such that either no action is taken, or the necessary correction is smaller than uncertainties in estimating the variation, it increases the potential for making mistakes. While single fraction based correction strategies may be able to avoid use of deformable image registration for tracking geometry or dose by simply re-defining targets on the new patient model, the preliminary result of a study by Jee *et al.* suggests that corrections made based on accumulated dose may be more accurate, and would require fewer corrections over the course of treatment[16]. It is with the intent of accumulating dose, that image registration accuracy becomes a significant issue in the adaptive therapy process, as will be discussed in the next session.

V.A.6. The role of image registration and its accuracy in adaptive therapy

A critical part of adaptive therapy is identifying the new target to be treated and the normal structures that should be avoided, based on images of the patient acquired during the treatment course. This can be achieved by re-contouring on the new images (manual or automatic

segmentation), or via use of deformable image registration to map the initial target to the new geometry. Once the new targets and normal structures are identified, the variations in the dose distribution relative to that initially planned need to be determined and compared against re-planning criteria. If dose evaluation relies on geometry tracking through multiple treatment sessions (i.e. accumulated as opposed to single fraction dose), then deformable image registration would be necessary to provide estimates of accumulated dose.

Within the context of adaptive therapy, evaluation of tracking accuracy goes beyond simple estimations of image registration error. The spatial distribution of this error relative to the gradient in dose distributions, as well as the dose accumulation process, should also be evaluated. Some components of the error can be measured independently (e.g. registration accuracy), while others depend on other factors and can only be evaluated as part of the overall process. Here we start by evaluating the registration accuracy in the head-and-neck geometry for an example case, followed by the estimation of its impact on the estimated versus true dose.

The aim of this chapter is to address the various sources of uncertainty in tracking geometry and dose, and to evaluate their potential impact on the re-planning decisions made based on the resulting dose based metrics. Two possible dose tracking methodologies are described and compared with manually delineated contours that are used as the ground-truth for change in the patient geometry. Although there is no ground-truth dose distribution available, the difference in the dose calculated for the volumes resulting from each tracking method is evaluated relative to that for the manually drawn contours.

V.B. Study design

V.B.1. Patient data

The patient data used in this study was acquired at The University of Texas M.D. Anderson Cancer Center under an institutional review board (IRB) approved protocol. The eligibility criteria as described by Barker et al., included a new diagnosis of head-and-neck cancer with a gross

primary or cervical nodal disease larger than 4 cm in the maximal dimension[10]. A total of 15 patients were enrolled in the study, one of whom was disqualified because of problems with acquisition of the daily in-room CT scans. The patients were scheduled for in-room CT scans three times a week for a total of 18 to 21 scans, some of which were missed for various reasons. Patients were immobilized using a thermoplastic face mask, and despite the availability of in-room CT scans, setup corrections were made using port films obtained according to their clinical protocol. All patients were treated with conventional or conformal treatment techniques to a dose of 70Gy in 35 fractions.

For the example case used in this chapter, a total of 19 scans were acquired. All in-room CT scans are noncontrast, with 3mm axial slices through the head-and-neck region. For this study, the prescription dose was kept at 70Gy in 35 fractions. For our investigation, targets were re-delineated according to the clinical practice at the University of Michigan, and IMRT plans were created for each case based on current in-house protocols. Patient setup uncertainties were evaluated using the in-room CT scans, since the port films or the resulting shifts were not available.

V.B.2. Tracking geometry and dose

V.B.2.i. *Image registration*

Image registration was performed in two stages. In the first stage, a manual rigid alignment (translation only) of the in-room CT scan to the initial reference scan was performed to accurately position the upper cervical vertebrae (C1 to C2), in order to eliminate the setup error from the patient geometry relative to the beam. This step resembles the online setup correction performed prior to the treatment using port-films or other imaging tools. Once the setup uncertainties were corrected, a deformable image registration was performed to track the variations in the geometry between the two scans. The registration method used for this study is based on B-splines, as described in detail in Chapter III. Either CT scan may be chosen as the reference to be mapped to the other, depending on the specific purpose of aligning the two. Registration takes about 45 –

60 minutes at the full image resolution. The reference image volume is cropped to a region defined by the user, which should include the entire region where dose is to be tracked. The choice of the extent of the image volume used for alignment can affect the registration accuracy to some extent, as it excludes the intensity values from regions outside of the defined box from the calculation of the registration metric. The crop box was kept constant for the registrations performed between various datasets in order to reduce the factors that could affect the registration accuracy.

V.B.2.ii. ***Dose accumulation methodologies***

Two different dose accumulation methods are evaluated here. Both methods estimate the accumulated dose from all previous fractions in the frame of reference of the most recent CT-based geometry. This is based on accepting that all re-planning or plan modification decisions should be based on the deviation of the accumulated dose from the initially planned dose distribution and target and normal tissue dose limits. The two methods differ in the way the total dose is tracked and accumulated on the geometry of the day. The first method, which will be referred to as the "Direct" method from this point on, is based on direct deformation of each previous dataset to geometry of the day. In the second method, the dose is tracked and deformed to the geometry of the day, so that when the next sample of geometry becomes available, the previous accumulated dose is mapped to this new geometry without looking back at all previous samples of the geometry. We will call this method "Propagated". A detailed description of each method follows.

V.B.2.ii.a. ***Direct method***

In the Direct method, when the Nth new geometry sample (CTF_N) becomes available, a total of N deformable registrations are performed. First the initial treatment planning CT scan (*CT*) is mapped to the new CT (CTF_N), and the resulting deformation map is then used to deform all structure surfaces from the initial *CT,* to the *new CT* (CTF_N) for evaluation. This step is equivalent to generating new contours on the new CT geometry, whether through manual delineation by the

physician, or automatic segmentation using anatomical atlases. Each of these methods requires different amounts of time and effort and has different levels of uncertainty associated with it. The resulting structure surfaces can then be used for visual assessment of amount of change in geometry, as well as quantitative evaluation of changes in the volume of targets or displacements in various normal structures relative to their initial position. This single registration would provide a simple method to evaluate the changes in the geometry. However, tracking the dose requires additional N-1 deformable alignments to be performed between the new CT and all previous daily CT scans. Each registration results in a deformation map from the CT scan at fraction N *(CTF_N)* and a CT scan from one of the previous delivered fractions *(CTF_1, CTF_2, .., CTF_{N-1})*, which can then be used to track the dose delivered to any point on the geometry of the new C scan from each previous fraction. These fractional doses (d_1, d_2, or d_{N-1}) can then be added on the new model of patient geometry to calculate the total accumulated dose as described by Equation V.1 and Figure V.1.

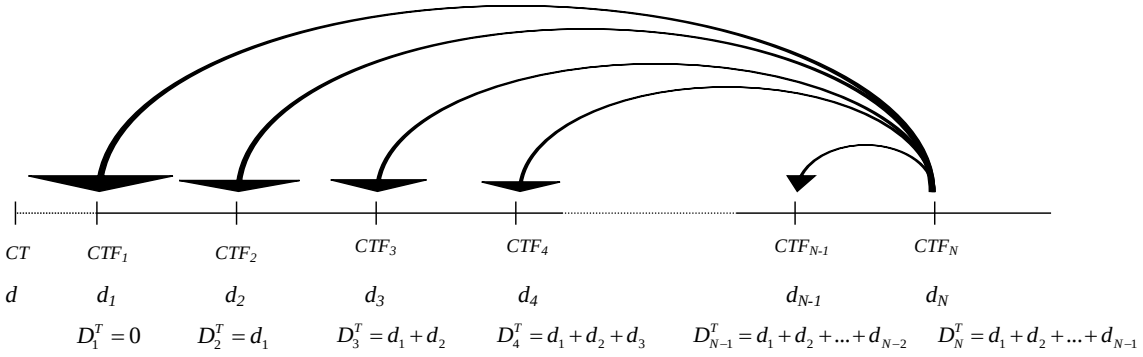

Figure V.1: Schematic of the direct dose accumulation method, where each new sample of patient geometry is mapped to all previous fractions to lookup the fractional dose for accumulation.

$$D_N^T = \ d_1 + \ d_2 + \ \ldots + d_{N-1} \qquad \text{V.1}$$

Here, d_1 is the dose delivered at fraction 1 but deformed to fraction N, using the deformation map estimated from deformable alignment of the CT from fraction N to fraction 1. The total

accumulated dose can then be used to make decisions regarding the need for plan modification or re-planning.

V.B.2.ii.b. ***Propagated method***

In the Propagated method, when the Nth new geometry sample (CTF_N) becomes available, only two deformable registrations are performed. This is done by tracking the accumulated total dose and weighting it by the number of fractions delivered so far. For example at fraction 1, there is no previous dose, and thus the total accumulated dose is equal to the dose delivered at fraction 1. Once a second geometry sample is available, the total dose delivered after fraction 1 is deformed and added to the dose from fraction 2 to create the total dose after fraction 2. At fraction 3, using the propagated method would require that the deformation map is found between the CT from fraction 3 (CTF_3) and the CT from fraction 2 (CTF_2) only, and the total accumulated dose after fraction 2, is mapped to the geometry of the new CT (CTF_3). Therefore at any fraction N, the CT from fraction N-1 is mapped to the new CT. The resulting deformation map is used to deform the contours from fraction N-1 to fraction N. The registration is then performed in the inverse direction, mapping CTF_N to CTF_{N-1} to track the dose accumulated on fraction N-1 geometry, onto the new CT. This process is shown in the schematic in Figure V.2 below.

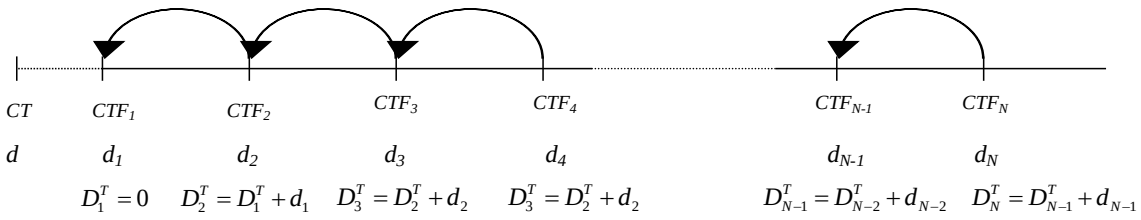

Figure V.2: Schematic of the propagated dose accumulation method, where each new sample of patient geometry is only mapped to the previous fractions, to lookup the fractional dose for accumulation.

$$D_N^T = D_{N-1}^T + d_{N-1} \quad \text{V.2}$$

Here, D_{N-1}^{T} is the total accumulated dose delivered up to fraction N-1, and d_{N-1} is the dose delivered on fraction N-1, both deformed to fraction N, using the deformation map estimated from deformable alignment of the CT from fraction N to fraction N-1.

V.B.2.iii. *Comparison of the direct and propagated methods*

Regardless of the time required to perform a single deformable registration between two image volumes, it is clear from the description of the two dose tracking methodologies above that the direct method would take longer to compute compared to the propagated method, due to the number of registrations required. More importantly, the accuracy of the two methods in tracking dose may be different. Deformable registration may be more accurate in tracking smaller deformations compared to larger variations in the geometry. Therefore, a deformation map estimated from aligning two CT scans taken from successive fractions may be more accurate than one that is estimated for two fractions farther apart that show larger changes in the anatomy. On the contrary, the propagation of the error may become a problem when the same geometry is tracked in stages, using multiple deformation maps with error. For example if the contours from the initial CT are deformed to CT1 with a registration map that has a 3mm error at a specific point, and the resulting contour from CT1 is deformed to CT2 with another deformation map with an additional 2mm error at the same point, the resulting contour on CT2 may have a 5mm error (if both errors in the same direction), thus making the propagated method less accurate than the direct method.

V.B.3. Accuracy of tracking geometry and dose

The accuracy of the tracking method is important in determining the new geometry that is used for calculation of the dose based metrics that are needed for plan re-optimization (e.g. mean and max), as well as the estimation of the accumulated dose distribution. Therefore, any comparison of the accuracy of the two methods should go beyond simple geometric comparisons of surface distance and volume, and include the impact of these errors on the dose distribution and metrics.

This section focuses on each of these aspects, to first quantify the geometric and dosimetric differences and then determine if these variations would affect final changes made to the plan.

V.B.3.i. *Geometric uncertainty*

One problem in evaluating the accuracy of the alignment of subsequent volumetric images of head-and-neck patients undergoing therapy is the scarcity or complete lack of landmarks (other than the bony skeleton) that can be easily and reproducibly identified on the scan. Therefore, despite the limitations and concerns associated with the use of manually delineated contours as the ground-truth information for comparison of automatically delineated or deformed contours, this is the only viable reference for both qualitative and quantitative analysis of the accuracy of the estimated contours or surfaces. In this study the same physician who delineated the contours on the initial treatment planning CT scan also delineated repeat CTV contours on select datasets for comparison.

V.B.3.i.a. *Qualitative analysis of volume accuracy*

Qualitative analysis of the accuracy of different geometric tracking methods is based on visual assessment of the contours or surfaces, comparing the estimated and manually delineated contours on repeat CT scans of the patient. Visual assessment of these contours is very difficult in most regions other than boundaries of air cavities, bone or the skin. Figure V.3 shows the manually drawn CTV (pink), the CTV deformed from the initial scan directly (yellow) and the CTV propagated from each CT to the next available CT (green), on fractions 5, 15, and 25. Based on an initial evaluation of these contours, the direct method appears to be more accurate near the air cavity while the propagated method performs better near the outer edge of the target close to patient surface. This is, however, a single slice and may not represent the error over the entire surface. Clearly, a more quantitative analysis of the error is necessary.

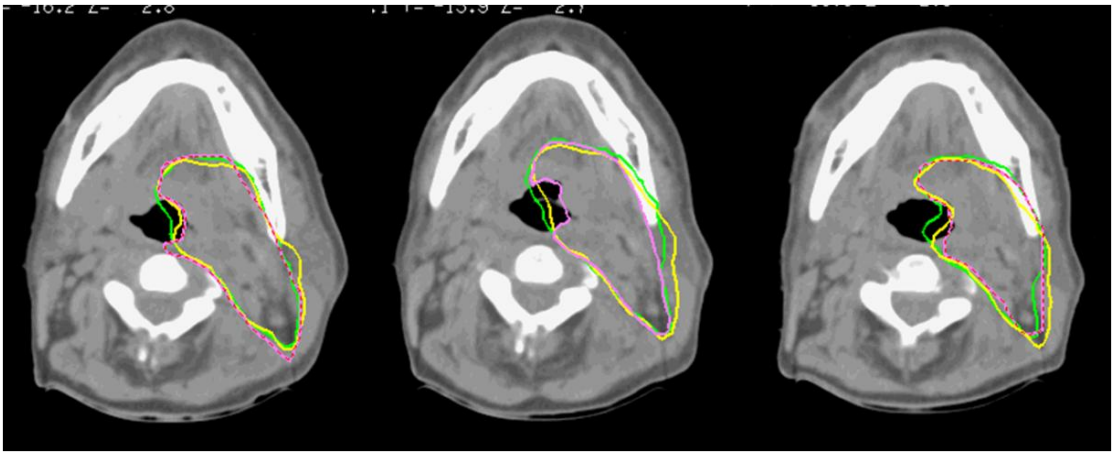

Figure V.3 :Contours shown for the target surface estimated using one of the two geometric tracking methods described (yellow: Direct, green: Propagated), compared to the manually drawn contour (pink) on CT scans from fractions 5, 15, and 25 (left to right).

V.B.3.i.a.1. Volume overlap index (VOI)

Several metrics can be used to compare the estimated surfaces with the manually delineated surfaces (or volumes). One of the most commonly used similarity metrics for comparison of any two surfaces is the volume overlap index (VOI), or Dice's coefficient as described by Equation V.3 below, where *M* is the manually drawn structure surface, and *D* is the deformed surface or volume. VOI describes the number of properly classified voxels in a volume.

$$VOI = 2\,\frac{V(D \cap M)}{V(D) + V(M)} \qquad \text{V.3}$$

Figure V.4 shows the VOI between the manual volume and each estimated volume for the right and left CTVs. The results indicate a high VOI for both estimated volumes and shows very little difference between the two methods. This is expected because one volume is more accurate on one side of the CTV, while the other one shows approximately the same error on the other side. In addition, comparing the VOI values for the right and left CTVs indicates that the right CTV, which appears more accurate visually, has a smaller VOI than the left CTV. This can be attributed to the large volume of the left CTV, which results in the VOI being less sensitive to the relatively

small volume of the non-overlap region. Therefore, VOI has limited applicability for the current study.

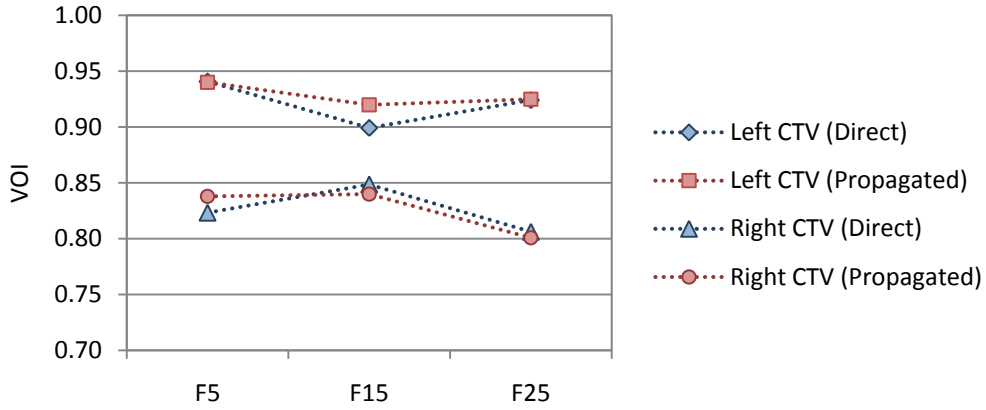

Figure V.4: Volume overlap index (VOI) for the left and right CTV using the direct and the propagated geomettry tracking methods

V.B.3.i.a.2. False positive and false negative volumes

In addition to the volume overlap index, the non-overlap volume was also calculated. The non-overlap volume includes both the regions that were included in the estimated deformed surface of the target but were not within the manually drawn volume (False positive), as well as regions that should have been included in the estimated target volume but were not (False negative) included because of registration errors. The schematic shown in Figure V.5 identifies these regions.

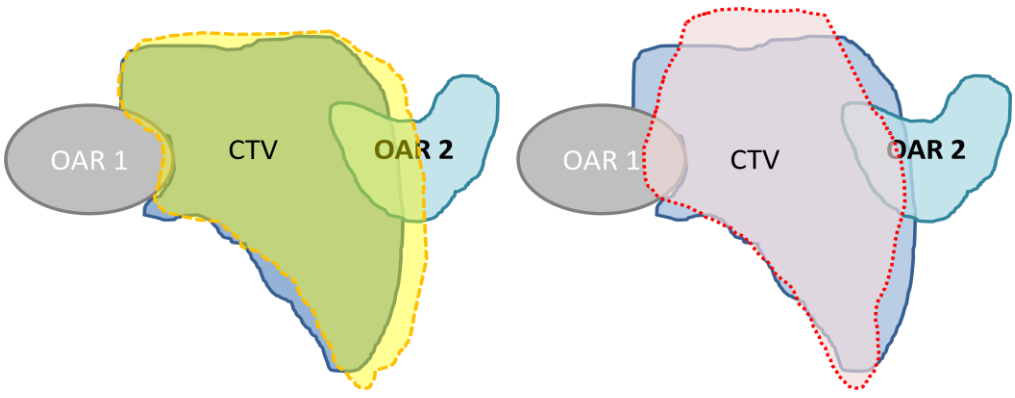

Figure V.5: Schematic of manual (dark blue) and estimated targets (yellow and red) and the surrounding organs at risk.This diagram shows the relative position of the false positive (yellow

on left and red on right) and false negative regions (blue) to the organs at risk. The overlap of these regions with the organs at risk may result in unwanted plan modifications.

The volumes of the false positive and false negative regions were calculated for both estimated CTVs (Figure V.6), which showed little difference between the two methods, with the largest difference of 20cc or 4% of the initial CTV volume (at fraction 15).

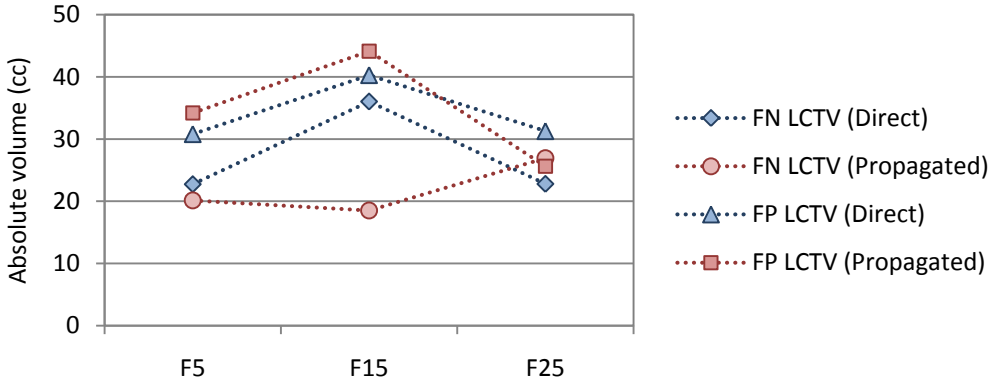

Figure V.6: Absolute volume of the false positive and false negative volumes for the two methods

In order to determine which metrics would be most sensitive to the variations in the manual versus deformed surfaces from the different deformation methods, sensitivity and specifity were also calculated using Equations V.4 and V.5 below. Sensitivity, which describes the ratio of the correctly marked voxels to the total voxels that should have been included in the estimated CTV, (Equation V.4), is similar to VOI except that it is measured relative to the true volume instead of the average of the true and estimated volumes. Specifity, which is calculated as the ratio of correctly marked voxels to the total number of marked voxels, describes the accuracy of the method in excluding sections that are not supposed to be included in the volume. The estimated specifity and sensitivity are shown in Figure V.7.

$$Sensitivity = \frac{V(D \cap M)}{V(M)} \qquad \text{V.4}$$

$$Specifity = \frac{V(D \cap M)}{V(D)} \qquad \text{V.5}$$

Here *D* is the deformed surface estimated using the two different methods (direct, or propagated) and *M* is the manually drawn structure.

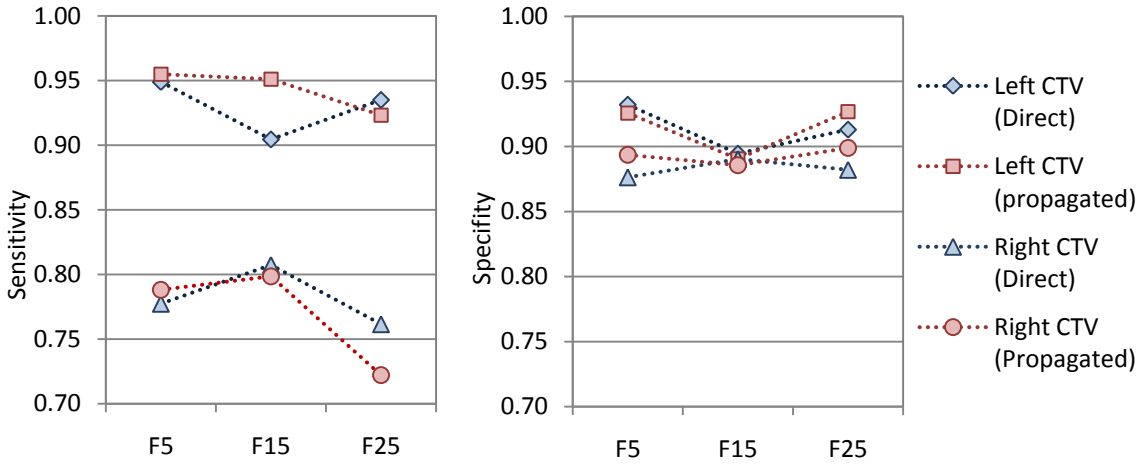

Figure V.7: Specifity and sensitivity of the two estimated CTV volumes for the direct and propagated methods.

From the results displayed in Figure V.7, it is difficult to determine which method is more accurate. In fact, these regions are only relevant and meaningful if they have dosimetric characteristics that are different from the correctly marked section of the target (CTV). This will be described in more detail in the following sections.

V.B.3.i.a.3. Minimum distance between surfaces

The volume indices calculated above do not describe the spatial distribution of the error, which if considered relative to a non-uniform dose distribution in the geometry, may result in significant differences in the estimated delivered dose to the target and normal tissue. To explain this further, imagine the two volumes in Figure V.5 (yellow and red). The magnitude of the two volumes is equal but the shape is different, indicating variations in the local accuracy of the two

estimated surfaces. It can be seen that if a uniform dose distribution is expected, the location of the error would not matter. On the other hand, the presence of non-uniform dose distributions, as well as normal structures that can move in and out of different dose regions, indicates the need for careful assessment of the local variations in error.

Local error was tested by evaluating the minimum distance from each point on the estimated surface of CTV to the manually delineated surface. Evaluating the error distribution in each dimension over all surface points did not show a significant variation in the error distribution between the two methods overall, but showed a small increase in the number of points that had errors larger than 3mm for the propagated method. The 3D distance between these estimated surfaces and the manually drawn surface, was also calculated (**Error! Reference source not ound.**) which showed a similar overall trend, with a higher frequency of larger errors (greater than 5mm) for the propagated method, as shown in the zoomed version of the same plot in **Error! Reference source not found.**.

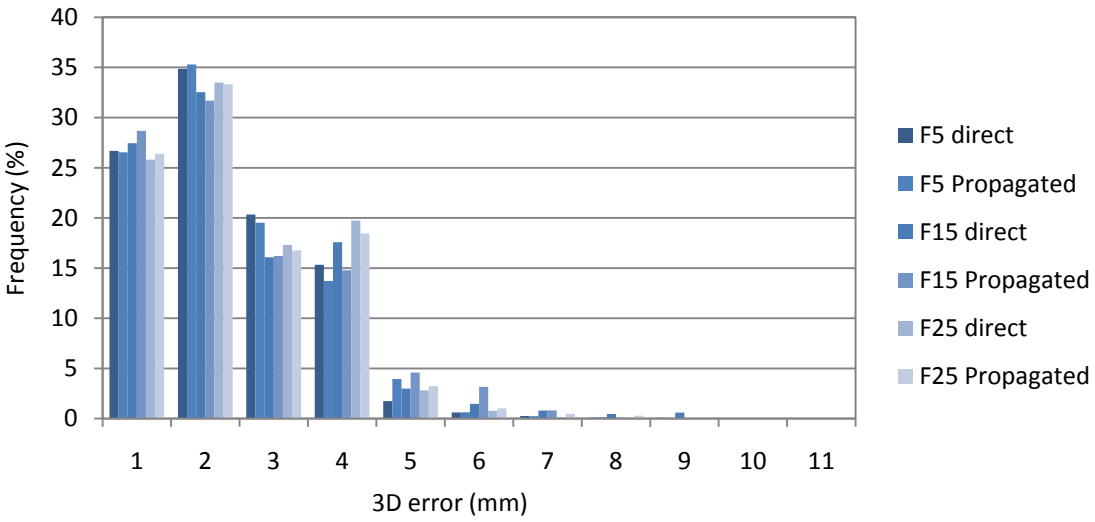

Figure V.8: Histogram of the 3D error distribution for different fractions

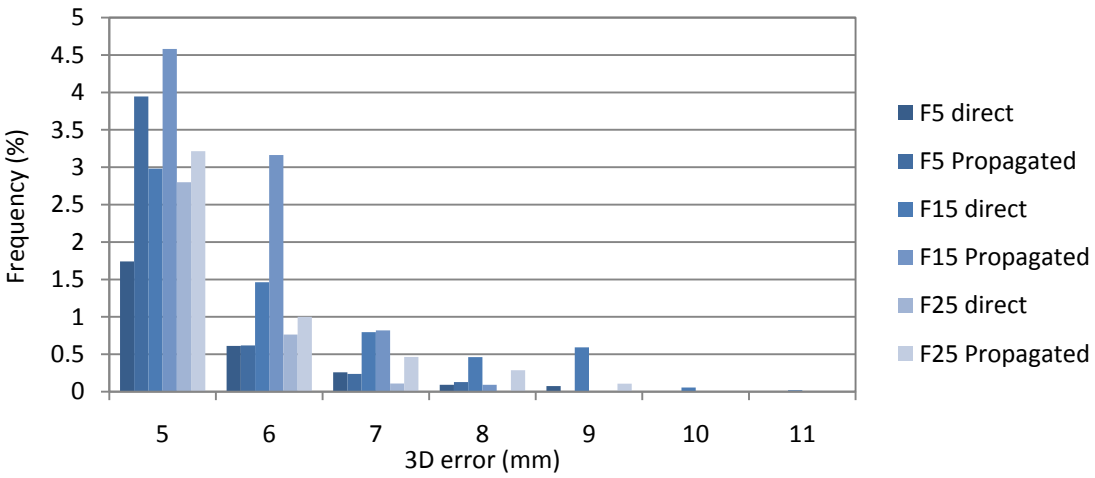

Figure V.9: Zoomed view of the 3D error distribution in the region of errors larger than 5mm

The maximum error in the true versus estimated position of each point on surface of the structure was also evaluated and compared for the two different geometry tracking methods described before. The results are shown in Figure V.10 and Table V.1 below. It can be seen that the maximum errors are higher for the direct method compared to the propagated contours for fraction 5 and 15, but the propagated errors increase for fraction 25.

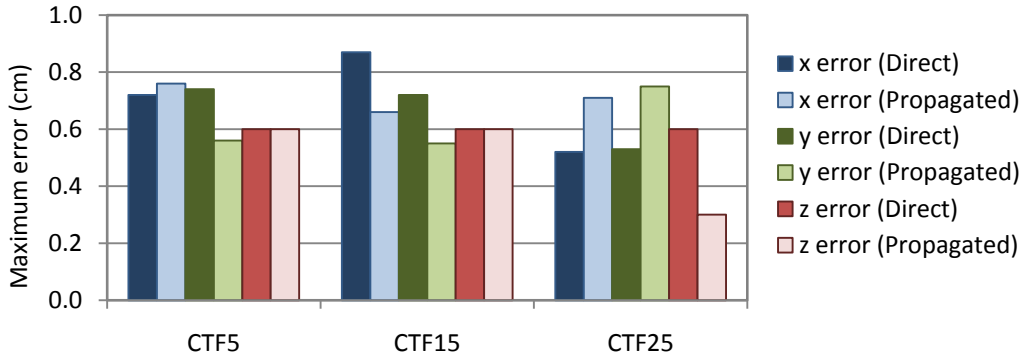

Figure V.10: Maximum error in the estimated contours using the two different methods shown in each dimension. Light colors show error in propagated method, dark colors show the error in the direct method.

Table V.1: Maximum, mimumum and the mean error in estimation of conoutrs/ surfaces of the target area for different fractions

	Max(cm)	Min(cm)	Mean(cm)

		Direct	Propagated	Direct	Propagated	Direct	Propagated
CTF_5	x	0.72	0.76	-0.82	-0.59	-0.01	-0.02
	y	0.74	0.56	-0.85	-0.65	0.02	0.02
	z	0.60	0.60	-0.30	-0.30	0.01	0.00
	3D	0.87	0.78	0.00	0.00	0.18	0.19
CTF_{15}	x	0.87	0.66	-0.78	-0.68	-0.01	-0.01
	y	0.72	0.55	-0.72	-0.55	0.02	0.01
	z	0.60	0.60	-0.60	-0.60	0.00	-0.02
	3D	1.05	0.75	0.00	0.00	0.20	0.19
CTF_{25}	x	0.52	0.71	-0.50	-0.75	-0.01	-0.01
	y	0.53	0.75	-0.53	-0.83	0.01	0.00
	z	0.60	0.30	-0.60	-0.30	0.01	0.00
	3D	0.63	0.89	0.00	0.00	0.19	0.19

Thus far in this chapter the accuracy of two different geometric tracking methodologies have been evaluated using various metrics for comparison of the resulting surfaces as a surrogate for accuracy of tracking and accumulating dose over the course of treatment. The differences observed in the accuracy of the two methods however, are not significant enough to allow us to reach a conclusion regarding the superiority of one method over the other. In fact, such a conclusion should be based on the dosimetric impact of these differences, relative to each other and in comparison to the manually drawn targets. Geometric errors that do not result in significant dosimetric errors, which would affect the decisions about the need for plan modification, or the resulting modifications made to the plan, may be acceptable. Therefore, in the next section of this chapter, the dosimetric impact of these variations in the error from the two methods, will be evaluated.

V.B.3.ii. ***Dosimetric accuracy***

Thus far, several metrics have been described for evaluation of the geometric accuracy of each surface deformation or dose tracking methodology. Looking back at contours in Figure V.3, the spatial variation in the accuracy of each surface is clear. It is the position of these uncertainties with respect to the dose distribution and the surrounding normal structures that dictates whether or not this error would result in an invalid assessment of delivered dose and resulting modification to the plan. The false positive (FP) region, for example, will primarily have an impact on the re-planning process when there is an overlap between this region and any dose limiting or radiosensitive organ in the surrounding region. Figure V.5 shows that the estimated CTV surface (yellow contour on the left) has a false positive region that overlaps an adjacent organ at risk (OAR2). If there is a sharp dose gradient on the boundary of the CTV and the organ at risk, the FP region may fall outside the high dose region, implying that the target is underdosed, thus initiating a plan modification to boost the dose in that region. This in turn would increase the dose to OAR2, creating unwanted increase in risk for toxicity. A similar situation exists for the CTV shown on the right (red) and OAR1. However, the FP region in the top of this CTV is not in close proximity to any normal structures. Therefore, if the target dose in this region initiates a re-plan, the modifications to the plan will not have severe effects on dose to any normal tissue. The false negative region (FN), on the other hand, does not directly impact plan modification, but if it overlaps a normal structure, then changes to the plan that aim at lowering the dose to this organ at risk may result in underdosing this segment of the target.

V.B.3.ii.a. ***Dosimetric sensitivity to geometric error***

In the present study the sensitivity of the treatment plan to geometric variations is defined as the variations in dose-based metrics for targets (CTVs) or normal structures, relative to the initial plan. It should be noted that when planning the treatment, some of the potential uncertainties in patient setup have been incorporated into the planning process through the use of margins added to the CTV to create the planning target volume (PTV). The use of the margin in treatment planning lowers the sensitivity of the target coverage to the variations in the CTV. A more

accurate way of evaluating the dosimetric sensitivity of treatment plans to the geometric variations in targets and normal structures would consider these uncertainties at the time of treatment planning to estimate the variations in the resulting plans that are made based on different samples of geometry, similar to the forward planning technique discussed in Chapter II. In the present study, the initial treatment plan is based on the standard treatment planning techniques utilized in the clinic. Thus no assumptions are made regarding the expected variations in patient geometry over the course of treatment (other than PTV or CTV margins for setup and microscopic disease). The limitations that these assumptions may impose on our estimates of sensitivity will be discussed later.

In order to understand the process described here, the targets and organs at risk in the head-and-neck geometry, as well as the initial treatment planning criteria, are described here. Figure V.11 shows the pre-treatment CT scan of the patient, with the target of interest, Left CTV63, outlined by the physician (yellow). For this geometry, the spinal cord is the only critical structure that limits the dose (D_{max} <45 Gy). Therefore, if the dose to the spinal cord is increased above its limit as a result of geometric variations, it would trigger a plan modification. All other normal structures are radiosensitive and would benefit from limiting the dose delivered to them, which is important in improving the quality of life for long term survivors. However, the sparing of these structures has a lower priority in the optimization process compared to target coverage, thus in this case they are generally not spared. The parotid glands, specifically, have an ideal mean dose limit of 26Gy. They also show a partial volume effect, suggesting that sparing part of the volume of parotid may be beneficial. In the case shown below however, the parotids are obviously not spared in the initial plan.

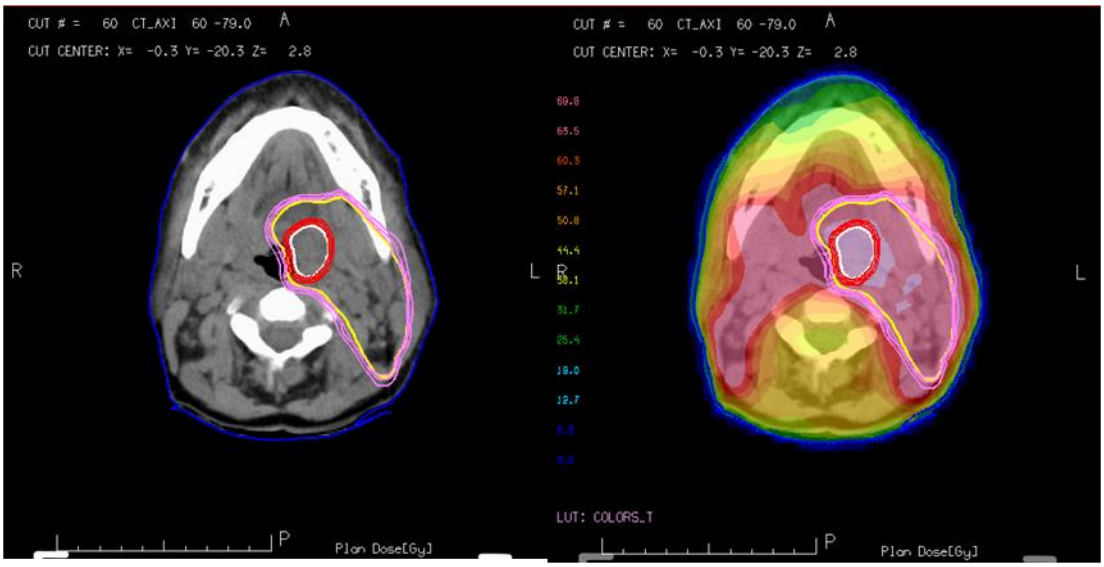

Figure V.11: Targets in the head-and-neck geometry (left) and the dose distribution for the optimized plan on the right

To test sensitivity, the initial target volume (CTV 63) was expanded uniformly in 1 mm increments and the dose to each new volume was evaluated using dose volume histograms (DVH) shown in Figure V.12. These DVHs indicate a degradation in target coverage as the target expands outside of the high dose region. The same test was performed for the higher dose target (CTV 70 shown with white contour in Figure V.11), which is smaller and is encompassed by CTV63, showing similar trends.

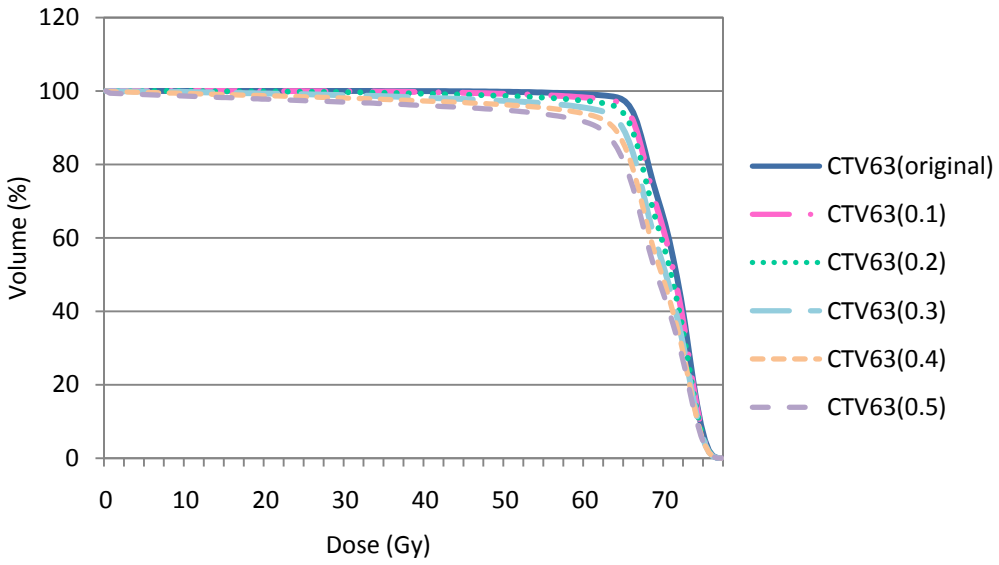

Figure V.12: DVH for the original CTV63 and the expanded volumes

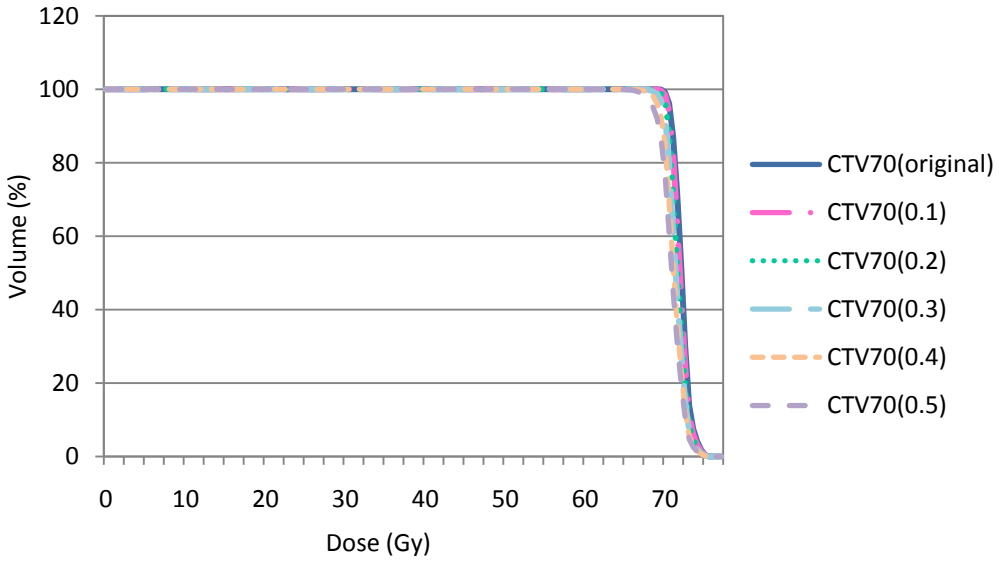

Figure V.13: DVH of CTV70 before and after uniform expansion in 1mm increments

In addition to the DVHs, dose based metrics calculated for these expanded structures are plotted in Figure V.14 and shown in Table V.2. Evaluation of these metrics indicates that the mean and the maximum dose to the target are relatively insensitive to the variations in the geometry, while the minimum dose (max dose to the lowest 1% of the volume) is very sensitive to the small volumetric expansions. This change is especially significant for the larger CTV volume (CTV 63)

which suggests that the regions of target underdose are outside of the of the high dose CTV (CTV 70) area and most likely correspond to sections of the target that are near the spinal cord (Figure V.11) or at the outer edge of the target near the external surface.

Table V.2: Dose based metrics calculated for the CTV63 and CTV70 for different uniform expansions (1-5mm). A min 1% dose indicates the maximum dose delivered to the minimum 1% of the volume.

	CTV 63			CTV 70		
	Max	Mean	Min 1%	Max	Mean	Min 1%
1mm	77.12	70.09	53.00	75.19	71.65	69.25
2mm	77.12	69.55	44.50	75.19	71.39	69.00
3mm	77.12	68.46	27.00	75.19	71.15	68.50
4mm	77.12	67.57	15.00	75.19	70.87	67.50
5mm	77.12	66.39	4.50	75.19	70.60	66.50

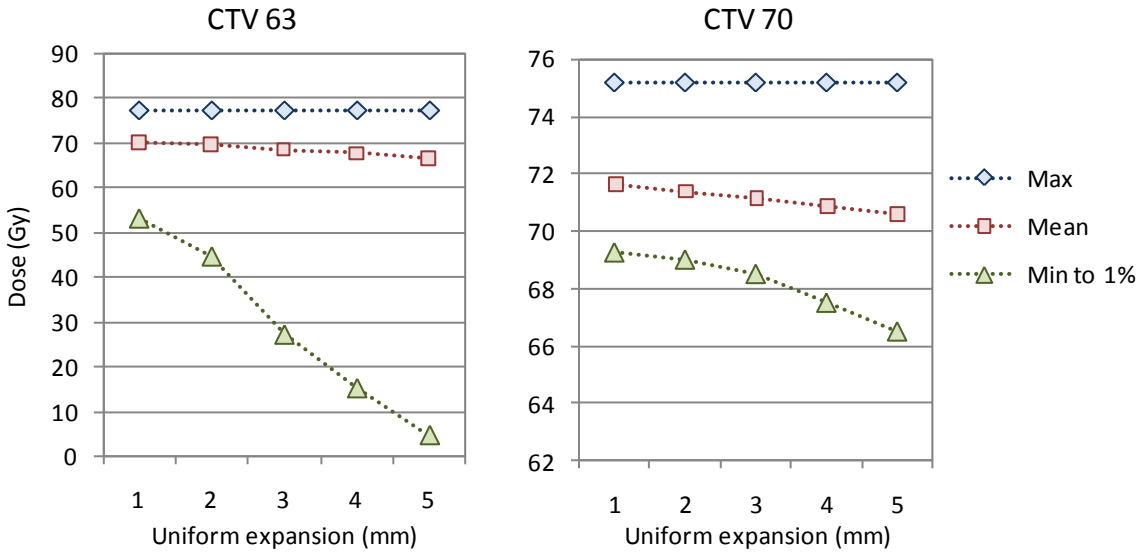

Figure V.14: Change in dose based metrics for differnt uniform expansions of the two CTVs

The results in Figure V.14 and Table V.12 suggest that for small variations in the geometry, the minimum dose to the target is most likely to initiate a plan modification. What this entails is that if these variations are real geometric variations, then the corrections would improve target coverage (i.e. restore it back to the initially planned levels), and in the process it may increase or decrease dose to the normal structures. The compromise between the two would be decided on at the time of the plan modification in a similar manner as done at the time of treatment planning. On the other hand, if these variations are artifacts of registration and not true geometric changes, then the corrections will be performed in the same way, but would result in unnecessary plan modification which may increase dose to normal tissue or reduce target dose by mistake. As described at the beginning of this section, the use of a PTV margin for treatment planning, could limit the sensitivity of the plan to geometric variations in the target volume. This is a known limitation of this test, which requires further investigation in the future.

Unfortunately there is no easy way to determine if an observed change is due to registration error or due to a true geometric variation that requires corretion. We have no knowledge of the truth, other than the manually drawn contours and surfaces identified by the physicians, which was shown to be insufficient without dose. One possible solution is to evaluate the dose for deformed and manual contours, to determine if significant variations exist between the dose-based metrics for the two geometries. Large variations between the two are an indication of the registration error being the cause of change in dose, while small variations suggest that the estimated surface or volume is accurate and that the dose change is caused by variations in the geometry (**Error! eference source not found.**5**Error! Reference source not found.**6). **Error! Reference source not found.** shows the dose based metrics (maximum, mean, and minimum to 1% of volume) calculated on the initial CT scan for each of the estimated CTV volumes from different fractions over the course of treatment. The mean and maximum doses to each of these volumes are relatively stable for both the direct and propagated methods, while the min1% dose fluctuates significantly particularly at the beginning of treatment (fractions 1 to 15). For the three fractions with manually delineated targets (Fractions 5, 15, 25), the same dose based metrics were

calculated and displayed in Figure V.15. There is little difference between any of the measurements for the manual or deformed surfaces, except for the minimum dose at fraction 5, which shows a 15Gy difference between the two volumes. Looking at the volumes for each estimated and manually drawn surface (Figure V.16) it is interesting to see that the largest difference between the volumes of manual and propagated surfaces happens at fraction 15, which showed the smallest difference in the min dose. This emphasizes that the spatial distribution of the error, relative to the dose distribution, is a better surrogate for evaluation of its accuracy for estimating dosimetric changes, than direct global geometric constructs.

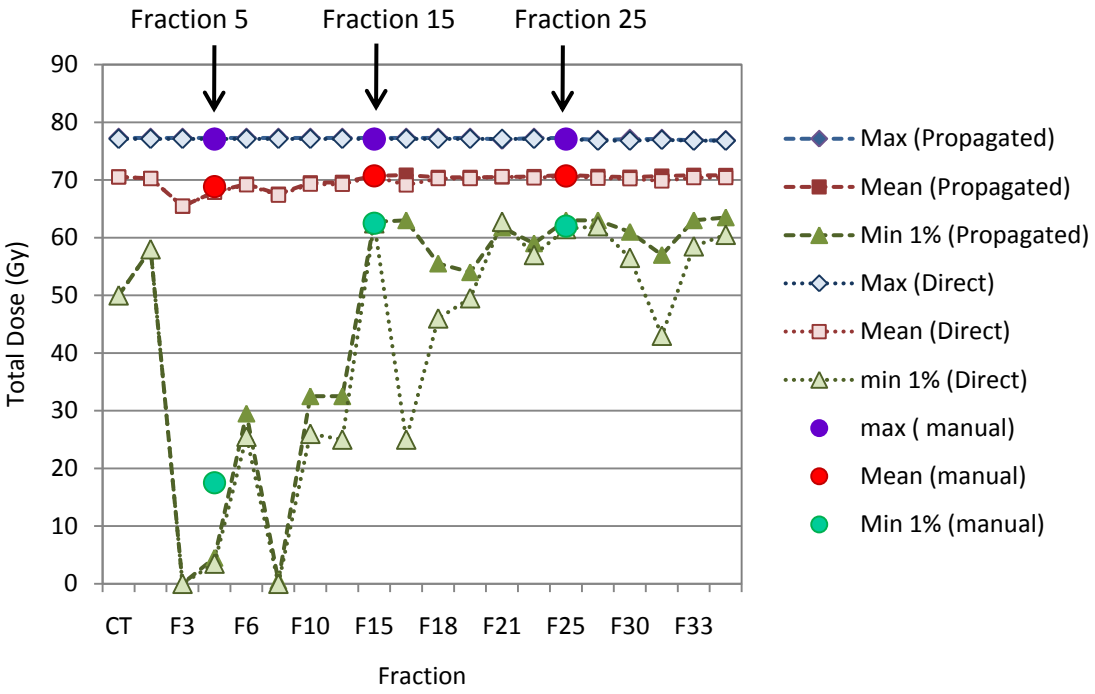

Figure V.15: Time trend of changes in the dose based metrics for estimated CTVs (CTV63) using both the direct and propagated methods, and the manually delineated CTV63 for three fraction.

It is also observed (Figure V.15) that for some fractions, large variations exist in the minimum dose to the CTVs estimated based on the two different geometric tracking methods (Direct vs. Propagated). For some of these fractions (e.g. fraction 16) the difference in the volumes is also large, as shown in Figure V.16. Although there is a generally good agreement between both

volumes, and doses for the surfaces estimated by the two methods, up to fraction 15, the limited number of manual contours available, does not allow us to evaluate fractions where larger differences exist, in order to determine which method is closer to the truth.

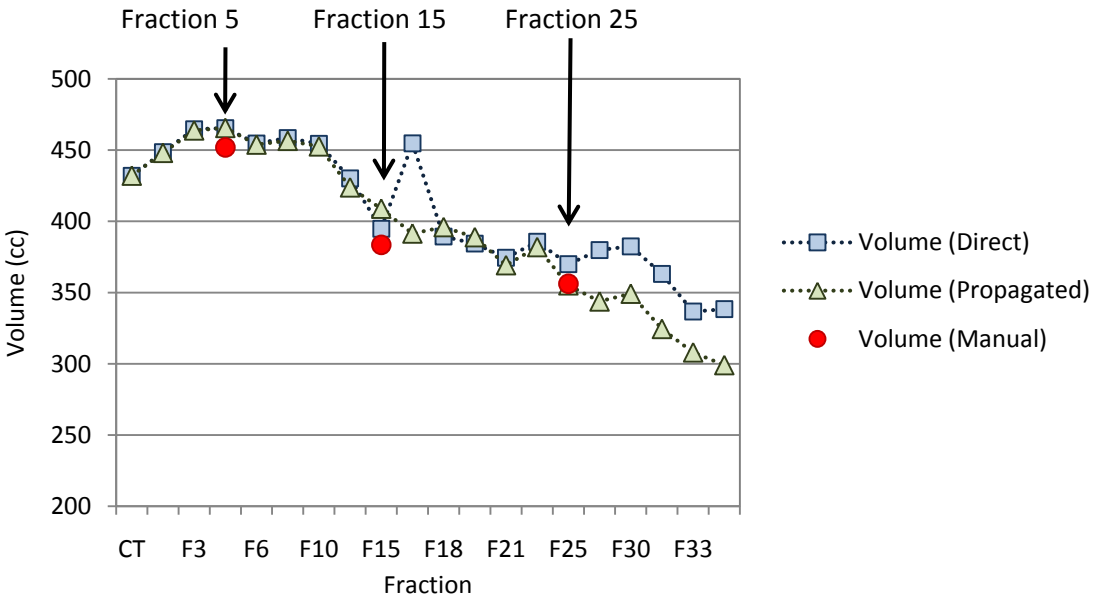

Figure V.16: Estimated volume of CTV for different fractions over the course of treatment. The red circles indicate the three fractions for which manual contours were available.

V.C. Discussion

This chapter evaluated several issues with tracking geometry and dose for use in adaptive therapy. These known issues have not been directly studied or quantified previously. One issue lies in the relation between the visually observed errors in the deformed contours or surfaces, and the actual dosimetric impact of these errors on our estimates of deformed or accumulated dose and re-optimization. This chapter attempted to quantify the visually assessable error in registration, and its impact on the dose metrics and plan modification.

The differences in volume as well as some dose based metrics were evaluated for the two geometric tracking methods, and compared against the ground-truth data from manually delineated targets. Ideally we need to gain a better understanding of the local or spatial

distribution of the error in estimating geometry and dose, so that informed compromises can be made regarding which errors are acceptable and which ones can result in significant mistakes in plan modification.

The main limitation in performing a complete evaluation of the spatial distribution of error in dose is the lack of ground-truth dose data for comparison to the estimates. It was described earlier in this chapter that the closest estimate of the true change in geometry is the manually drawn structure (target or normal tissue). Despite the limitations in the accuracy of manual target delineation, these surfaces are sufficient for evaluating the accuracy in the estimated geometry. However, when it comes to true dose distribution in a deforming geometry, there is no equivalent to the manual contours that can be used as the ground-truth. To clarify this further, in order to determine the total dose in a deforming anatomy, one needs to track every tissue voxel from one fraction to the next, and determine where it falls with respect to the spatial dose distribution, in order to add up the dose to that voxel over time. Therefore, a mapping between the different instances of geometry is absolutely necessary, as described earlier. However, this map is not perfect, and even if targets or other structures are re-defined manually on every sample of the geometry, there is no possible way to manually define or edit the mapping between them. The limitations in establishing the truth based on such standards as manual contours may potentially be calibrated by assessing the local sensitivity of the contour edge locations to relevant expansions in dose metrics. Such sensitivity analyses, when further explored and optimized, my serve as a basis for more robust guidance at initial planning as well as simple rules for triggering meaningful plan adaptation.

References:

[1]Eisbruch A, Foote RL, O'Sullivan B, Beitler JJ, Vikram B. Intensity-modulated radiation therapy for head and neck cancer: emphasis on the selection and delineation of the targets. Semin Radiat Oncol. 12(3):238-49. 2002 Review.

[2] Eisbruch A, Gregoire V. "Balancing risk and reward in target delineation for highly conformal radiotherapy in head and neck cancer," Semin Radiat Oncol. 19(1):43-52. 2009 Review.

[3]Peter H. Ahn, Madhur K. Garg, „Positron Emission Tomography/Computed Tomography for Target Delineation in Head and Neck Cancers" Seminars in Nuclear Medicine,38(2): 141-148, 2008

[4]Chao KS, Bhide S, Chen H, et al. "Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach." Int J Radiat Oncol Biol Phys. 2007 Aug 1;68(5):1512-21

[5]Hansen EK, Bucci MK, Quivey JM, Weinberg V, Xia P. "Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer." Int J Radiat Oncol Biol Phys. 2006 Feb 1;64(2):355-62

[6] Welsh JS, Lock M, Harari PM, Tomé WA, Fowler J, Mackie TR, Ritter M, Kapatoes J, Forrest L, Chappell R, Paliwal B, Mehta MP." Clinical implementation of adaptive helical tomotherapy: a unique approach to image-guided intensity modulated radiotherapy." Technol Cancer Res Treat. 2006 Oct;5(5):465-79.

[7] Lu W, Olivera GH, Chen Q, Ruchala KJ, Haimerl J, Meeks SL, Langen KM, Kupelian PA." Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy." Phys Med Biol. 2006 Sep 7;51(17):4357-74.

[8] Kuo YC, Wu TH, Chung TS, Huang KW, Chao KS, Su WC, Chiou JF. "Effect of regression of enlarged neck lymph nodes on radiation doses received by parotid glands during intensity-modulated radiotherapy for head and neck cancer." Am J Clin Oncol. 2006 Dec;29(6):600-5.

[9] Tomé WA, Jaradat HA, Nelson IA, Ritter MA, Mehta MP. "Helical tomotherapy: image guidance and adaptive dose guidance." Front Radiat Ther Oncol. 2007;40:162-78. Review.

[10] Barker JL Jr, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, Morrison WH, Rosenthal DI, Chao KS, Tucker SL, Mohan R, Dong L. "Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system." Int J Radiat Oncol Biol Phys. 2004 Jul 15;59(4):960-70.

[11] Birkner M, Thorwarth D, Poser A, Ammazzalorso F, Alber M. "Analysis of the rigid and deformable component of setup inaccuracies on portal images in head and neck radiotherapy."Phys Med Biol. 2007 Sep 21;52(18):5721-33. Epub 2007 Sep 4.

[12] Wu Q, Liang J, Yan D. "Application of dose compensation in image-guided radiotherapy of prostate cancer." Phys Med Biol. 2006 Mar 21;51(6):1405-19. Epub 2006 Feb 21.

[13] Woodford C, Yartsev S, Dar AR, Bauman G, Van Dyk J. "Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images." Int J Radiat Oncol Biol Phys. 2007 Nov 15;69(4):1316-22.

[14] Smeenk C, Gaede S, Battista JJ. “Delineation of moving targets with slow MVCT scans: implications for adaptive non-gated lung tomotherapy.”Phys Med Biol. 2007 Feb 21;52(4):1119-34.

[15] Ding GX, Duggan DM, Coffey CW, Deeley M, Hallahan DE, Cmelak A, Malcolm A. “A study on adaptive IMRT treatment planning using kV cone-beam CT.” Radiother Oncol. 2007 Oct;85(1):116-25.

[16] Jee K, KashaniR, Eisbruch JM, et al. “Delivery outcome driven adaptive re-planning strategy for deformation,” ESTRO 2008

CHAPTER VI

SUMMARY

In radiation therapy, the overall accuracy of treatment planning and delivery is highly dependent on accurate target selection, beam modeling and dosimetry, and the accuracy of daily treatments. While target selection and beam dosimetry are unchanged during the course of fractionated treatment for any individual patient, various factors contribute to potentially alter the accuracy of daily treatment delivery. Some of these factors include setup and localization errors, intra-fraction motion due to breathing, and changes in the geometry in response to treatment (e.g. tumor shrinkage and weight loss). These variations are partially accounted for at the time of simulation and treatment planning, through the use of margins to ensure target coverage, or by utilizing immobilization devices that reduce motion and setup error during each fraction. However, these techniques are not perfect and have limits to the accuracy with which they can account for geometric variation. In order to improve these techniques, it is necessary to first understand and carefully measure the geometric changes in the patient, whether caused by breathing motion or in response to treatment, and then develop treatment planning strategies that are robust to the expected changes over time, or develop methods to correct and modify the plan as the treatment progresses. Regardless of the approach, it is important to understand the limitations and uncertainties of the measurement tool that is used for estimation of changes in geometry and dose. This dissertation begins to quantify the uncertainties in measurement and management of patient geometric variations, and to evaluate the resulting limits of accuracy in treatment planning, delivery, and plan modification.

Chapter I provided some necessary background on the advances in various areas of Radiation Oncology, which has provided the possibility to deliver highly conformal plans that are also very sensitive to variations in the patient geometry. The increase in the knowledge of the changes in the patient, and improvements in our capability to quantify these variations based on repeat imaging of the patient during the course of treatment was also discussed.

Chapter II focused on the measurement of respiratory induced motion in several disease sites within the thoracic and abdominal regions, such as lung tumors. The effectiveness of the Active Breathing Control device (ABC) in immobilizing these structures was evaluated, showing breath-hold reproducibility for lung tumors on the order of 3-4 mm. After characterizing the motion of these structures, an alternative treatment planning methodology was proposed, which eliminated the need for margins by iteratively modifying the MLC configuration based on the estimated dose delivered to the target over a full breathing cycle. The advantage of this forward planning method is that it takes into account the fact that the target is not present in all locations that it traverses during a breathing cycle at all times, which leads to large margins and overdosing the surrounding normal tissue. It was demonstrated that this forward-planning methodology can improve the dose to normal tissue, while keeping target coverage the same. However, this improvement was not significant, and the planning process was time consuming.

In Chapter III, the uncertainties in estimating the deformation between multiple samples of geometry were studied. This chapter introduced the design and implementation of a deforming phantom for an objective test of different deformable registration techniques. The phantom was implanted with a large number of small, easily identifiable markers, whose positions were used as the ground-truth deformation to evaluate the registration error in various regions inside the phantom. The images of this phantom were sent to other institutions and the accuracy of their registration techniques was also evaluated, indicating a large distribution in the errors for different methodologies.

Chapter IV proposed the use of surrogates (a few implanted markers), to track and monitor the motion and deformation in various regions inside the lung. A patient-specific model of the respiratory-induced motion and deformation was described, which uses principal component analysis, to determine the most significant modes of variation in lung geometry over a breathing cycle. It was demonstrated that decomposing the deformation maps from a reference breathing state to all other states will result in a set of eigenmodes, the first few of which may be sufficient in describing the full deformation in lung geometry. The lower ranked eigenmodes were associated with errors in the registration and it was shown that eliminating these from the model will improve the accuracy of estimating deformation in the lung.

In Chapter V, the magnitude of geometric variations in tumors and normal tissues in response to treatment was evaluated for head and neck cancer patients. Using deformable image registration, geometries were tracked through the course of treatment, using two different methodologies with differences in workload, to determine if one is superior to the other for tracking and accumulating dose in adaptive therapy. Physician-drawn contours were used as the ground truth for comparison of all deformed contours, and the potential impact of these geometric errors on relevant dose based metrics was investigated. It was demonstrated that the estimated geometric error, by itself, may not be a good surrogate for determining the uncertainty in dose, mainly due to the importance of the relative position of this error and the gradients in dose distribution. The potential impact of this error on tracking changes in dose for treatment adaptation was briefly discussed.

This dissertation has explored the errors in deformable image registration and the potential of deformable models and specifically their higher order descriptions to aid in treatment monitoring and image guidance. It has briefly examined the impact that these errors may have in accumulating dose for adaptive therapy.