Inter-fraction variations in respiratory motion models

2011 Phys. Med. Biol. 56 251

(http://iopscience.iop.org/0031-9155/56/1/015)

Phys. Med. Biol. **56** (2011) 251–272 doi:10.1088/0031-9155/56/1/015

Inter-fraction variations in respiratory motion models

J R McClelland[1,3], S Hughes[2], M Modat[1], A Qureshi[2], S Ahmad[2], D B Landau[2], S Ourselin[1] and D J Hawkes[1]

[1] Centre for Medical Image Computing, University College London, UK
[2] Department of Oncology, Guy's & St Thomas's Hospitals NHS Trust, London, UK

E-mail: j.mcclelland@cs.ucl.ac.uk

Received 27 July 2010, in final form 14 October 2010
Published 8 December 2010
Online at stacks.iop.org/PMB/56/251

Abstract
Respiratory motion can vary dramatically between the planning stage and the different fractions of radiotherapy treatment. Motion predictions used when constructing the radiotherapy plan may be unsuitable for later fractions of treatment. This paper presents a methodology for constructing patient-specific respiratory motion models and uses these models to evaluate and analyse the inter-fraction variations in the respiratory motion. The internal respiratory motion is determined from the deformable registration of Cine CT data and related to a respiratory surrogate signal derived from 3D skin surface data. Three different models for relating the internal motion to the surrogate signal have been investigated in this work. Data were acquired from six lung cancer patients. Two full datasets were acquired for each patient, one before the course of radiotherapy treatment and one at the end (approximately 6 weeks later). Separate models were built for each dataset. All models could accurately predict the respiratory motion in the same dataset, but had large errors when predicting the motion in the other dataset. Analysis of the inter-fraction variations revealed that most variations were spatially varying base-line shifts, but changes to the anatomy and the motion trajectories were also observed.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Despite advances in recent years, respiratory motion still causes many problems when trying to treat lung tumours accurately with radiotherapy. One of the main reasons why respiratory motion continues to be a problem is that there can be substantial variations in the motion, both during a single fraction of treatment (intra-fraction variation), and between different fractions of treatment (inter-fraction variation).

[3] Author to whom any correspondence should be addressed.

There have been some studies into intra-fraction variation in respiratory motion (Cai *et al* 2008, Guckenberger *et al* 2007a, 2007b, Purdie *et al* 2007, Seppenwoolde *et al* 2002, 2007, Sonke *et al* 2009), but firm conclusions about the magnitude and type of variation that can occur have yet to be established. One of the main problems is that it is very difficult to image the internal respiratory motion for the time taken to deliver a fraction of treatment (several minutes). Different techniques have been employed to image the intra-fraction motion, such as intermittent 4DCT (Guckenberger *et al* 2007b), Cone-Beam CT (CBCT) (Guckenberger *et al* 2007a, Purdie *et al* 2007), 4DCBCT scanning (Sonke *et al* 2009), MR imaging (Cai *et al* 2008) or using implanted markers imaged with fluoroscopy (Seppenwoolde *et al* 2002, 2007). However, as yet none of these are able to adequately image the 3D motion and deformation of the full region of interest over a timeframe corresponding to a whole fraction of treatment.

Inter-fraction variation has also been studied, either using repeat 4DCT scans (Britton *et al* 2007, Redmond *et al* 2009), or CBCT scans acquired at each fraction for patient setup (Purdie *et al* 2007, Sonke *et al* 2008, 2009). These studies show that there can be considerable variation in the respiratory motion between different fractions (and between planning and treatment), and in some cases it can have a larger clinical impact than the respiratory motion itself (Sonke *et al* 2008). Some studies have found that the majority of the inter-fraction variation can be approximated as 'base-line' shifts (Sonke *et al* 2008, 2009). In these cases the underlying anatomy and the motion trajectories are relatively stable from fraction to fraction, but the base-line position of the tumour (e.g. its average position over the respiratory cycle) relative to other structures varies. Other studies have found that there can also be variations in the underlying anatomy (Britton *et al* 2007), in the motion trajectory (Seppenwoolde *et al* 2002, Britton *et al* 2007) and in the relationship between the motion of the tumour and that of other parts of the anatomy (Redmond *et al* 2009).

Models that describe the motion and deformation of the internal anatomy due to respiration have many potential uses in lung radiotherapy and other applications. They can be used in image reconstruction to remove motion artefacts in CBCT (Rit *et al* 2009), MR (Odille *et al* 2008) and PET (Li *et al* 2006), and to facilitate lower imaging doses in CT (Li *et al* 2005). They can be used when planning radiotherapy for defining targets that account for motion (Colgan *et al* 2008), calculating the effects of motion on dose distributions (Colgan *et al* 2008), and creating more appropriate images for planning (Wolthaus *et al* 2008). If the models relate the internal motion to a respiratory surrogate signal that can be easily measured during treatment, such as the displacement of the skin surface, then the models could potentially be used to plan and guide motion compensated treatments such as gated or tracked treatment (McClelland *et al* 2007). They could also be used to monitor what was actually delivered and to help facilitate adaptive radiotherapy (Webb 2008).

Many of the models of respiratory motion used in the literature are built from the results of deformable registration between the volumes of a 4DCT dataset (Li *et al* 2005, 2006, Rit *et al* 2009, Wolthaus *et al* 2008). Such models have been usefully employed for many applications, but cannot account for any variation in the respiratory motion, and may be inaccurate due to variations in the motion during the 4DCT acquisition. A number of other motion models have also been presented in the literature (Low *et al* 2005, McClelland *et al* 2006, Seppenwoolde *et al* 2007, Yang *et al* 2008, Zhang *et al* 2007). Most of these use deformable registration results to represent the internal motion of the region of interest, and relate these to one or more respiratory surrogate signals.

Low *et al* (2005) presented a '5D motion model' which relates the internal motion measured from Cine CT data to the tidal volume and airflow measured from spirometry using a linear function of two variables. By relating the motion to the value of the surrogate signal (tidal volume) and its rate of change (air flow) this can potentially model some of the

intra-fraction variation. Yang *et al* (2008) extended this model to use deformable registration to describe the internal motion and demonstrated that the models can accurately predict the respiratory motion over 18 s. Zhang *et al* (2007) present a similar model that uses a linear function of two variables to relate the deformable registration results to the current diaphragm position and its precursor position a short time earlier. Although the same mathematical functions are used by Yang *et al* (2008) and Zhang *et al* (2007) to relate the registration results to the surrogate signal, different methods are used to fit them to the data. Zhang *et al* (2007) used principal components analysis to fit the model to the data, while Yang *et al* (2008) used a linear regression approach. There are a number of technical differences between these methods but both appear to produce good results. Zhang *et al* (2007) used 4DCT data so their models are only fitted to, and assessed on, data from a single respiratory cycle. They also assess their models using 4DCT data acquired on another day, and show that the tumour motion can still be predicted relatively well. However, they do acknowledge that all the tumours are located close to the diaphragm, and that the predictions may not be so good for tumours in other parts of the lungs.

Seppenwoolde *et al* (2007) evaluated the motion models used for the Synchrony Respiratory Tracking System, part of the CyberKnife robotic treatment device (Accuray, Inc., Sunnyvale, CA). This uses implanted markers which have a number of associated problems, but does enable a new model to be built prior to and updated during each fraction of treatment, greatly improving its ability to account for variations in the motion. It also means that deformable registration is not required as only the motion of the markers is modelled. A single surrogate signal is generated from several external markers placed on the patient's chest. The model relates the motion of the implanted marker(s) to the value of the surrogate signal using either a linear function or two quadratic functions (one for inhalation and one for exhalation so that hysteresis can be modelled).

We have previously presented a method of generating patient-specific motion models from Cine CT data which related the internal motion, described by a deformable transformation, to the respiratory phase (McClelland *et al* 2006). We showed that by averaging out the breath to breath variations, we could accurately predict the respiratory motion over 20 s. However, the models were only assessed on data acquired during the same session as the data used to build them, so were not subject to any inter-fraction variation.

In this paper we present two important further developments to our models. Firstly, we investigate three different motion models: one relating the motion to the respiratory phase (derived from the surrogate signal), one to the value of the surrogate signal and one to the value and rate of change of the surrogate signal. The second development enables the models to produce smooth continuous predictions that are free of the artefacts often seen between adjacent couch positions in 4DCT datasets. We then evaluate the new models with and without inter-fraction variation, using data acquired at two different times: when the patient attends for their planning scan, and close to the end of radiotherapy treatment (approximately 6 weeks later). We also use the models to analyse the inter-fraction variation that has occurred, so as to gain insights into how to compensate for the variation.

2. Material and methods

2.1. Data acquisition

Data were acquired during two different sessions: one when the patient attends for their planning scan and one close to the end of radiotherapy treatment (approximately 6 weeks later). For each session a full dataset was acquired, consisting of Cine CT data and a

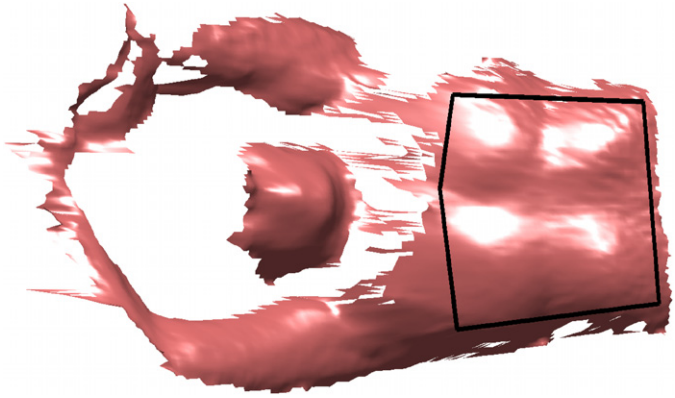


Figure 1. Example of a 3D skin surface acquired by the Vision RT system showing the bounding box (in black) used to help deal with occlusions in the data.

simultaneously acquired respiratory surrogate signal. For both datasets the patient was setup using the skin tattoos, as they would be for RT treatment. Cine CT data were acquired at seven slightly overlapping couch positions covering most of the lung and centred on the tumour. At each couch position a series of 40 Cine CT volumes were acquired over 20 s. Each Cine CT volume contained 12 $\times$ 2.4 mm slices, with a resolution of approximately 1 mm $\times$ 1 mm in-slice.

The respiratory surrogate signal was derived from 3D skin surface data acquired by a stereo camera system (Vision RT, London, UK). A full 3D skin surface was acquired at approximately 10 Hz, and the acquisition was synchronized to the Cine CT acquisition using an external x-ray detector (GM-10 Geiger Counter Radiation Detector, Black Cat Systems, Westminster, MD). A 1D respiratory surrogate signal was generated by finding the volume between each 3D skin surface and the treatment couch. The 3D skins surface did not include the patient's sides and some other areas due to occlusions in the CT scanner room (a much more complete surface can be obtained in the treatment room where multiple camera systems can be combined to cover the surface). Therefore a bounding box was manually defined to encompass as much of the 3D skin surface as possible (figure 1), and the volume enclosed by this bounding box, the skin surface, and the treatment couch was found. We have previously shown that a respiratory signal produced in this way correlates well with spirometry, but is not subject to the drift that often afflicts spirometry signals (Hughes *et al* 2009). To ensure that the respiratory signals were comparable from the two scanning sessions, the two skin surface datasets are rigidly aligned with each other and the mean values of the respiratory signals were normalized to 0.

Three different respiratory parameters were derived from the surrogate signal: the value of the signal, its rate of change (gradient) and the respiratory phase. The respiratory phase was calculated by automatically detecting the local minima and maxima of the smoothed surrogate signal. These would generally correspond to end-exhale and end-inhale points, although when irregular breathing occurred there were occasionally extra maxima and minima present in the surrogate signal that did not correspond to end-exhale or end-inhale. These were manually identified and ignored. All end-exhale points were assigned a phase value of 0%. The average position of end-inhale between end-exhales was calculated (e.g. 40%), and all end-inhales were assigned to this value. This maintained an approximately even distribution of phases

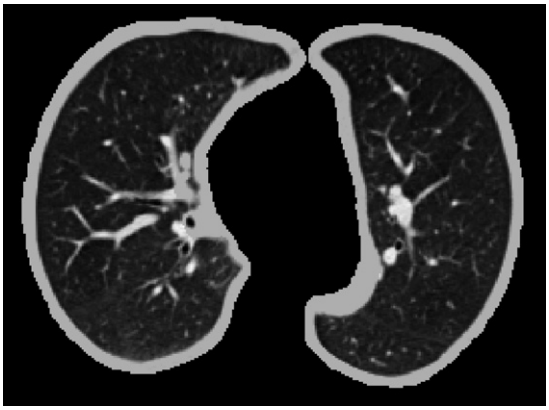


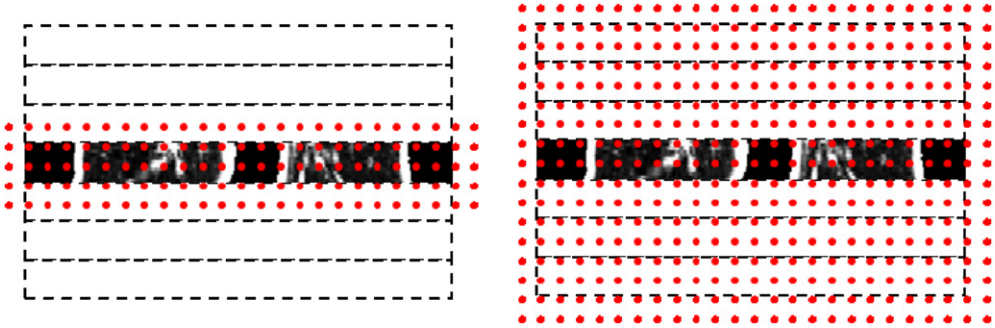
Figure 2. Example of segmented lungs from a Cine CT volume showing the 5 mm 'soft tissue' border that is added to help ensure that the lung border is correctly registered while still allowing for sliding motion.

over the whole breath cycle, even if the patient tended to spend a different amount of time inhaling than exhaling. The phase value corresponding to each Cine CT was then calculated by linearly interpolating the phase values at end-exhale and end-inhale.

2.2. Deformable image registration

Image registration was used to measure the motion and deformation of the lungs. A reference volume was generated by concatenating the Cine CT volumes from each couch position with a phase closest to 0%, i.e. the reference volume was effectively the end-exhale 4DCT volume. Because of the sliding motion that can occur between the lungs and the chest wall, the lungs were segmented prior to registration. First the lungs were automatically segmented using the object extractor tool from Analyze 10 (Biomedical Imaging Resource, Mayo Foundation, Rochester, Minnesota), and then a morphological closing was applied in each dimension so that all the airways and vessels were included in the segmentation. A region was defined 5 mm either side of the lung boundary and any voxels in this region with intensity greater than −100 Hu (corresponding to the low end of soft tissue values) were set to −100 Hu. Voxels inside the boundary region retained their original intensity, while those outside were set to a background value and were ignored from the registrations. This had the effect of adding a border of 'soft tissue' approximately 5 mm wide around the lungs, but was insensitive to small errors in the automatic segmentation (figure 2). This helped to ensure that the edges of the lungs were correctly aligned while still allowing sliding motion to occur.

The reference volume (source) was registered to each of the Cine CT volumes (target). An initial affine registration was performed to account for global motion, followed by a B-spline non-rigid registration to recover the local deformation. The affine registration result was used to initialize the B-spline control point grid used for the deformable registration. In this way the B-spline result includes the affine transformation, and the affine transformation does not need to be modelled separately. The B-spline registrations were performed using an efficient open-source implementation which has recently been developed within CMIC ('niftyreg', http://sourceforge.net/projects/niftyreg/, Modat *et al* 2010b). The control point grid resolution was refined from 40 mm to 20 mm, to 10 mm and finally to 5 mm, with the original image resolution used for all grid resolutions due to the limited number of slices (12)



(a) (b)

Figure 3. Standard control point grid only covering the Cine CT volume being registered (a), and extended control point grid covering all couch positions (b).

in each volume. Normalized mutual information (NMI) was used as the similarity measure and bending energy was used as the penalty term. The registrations were performed using an 'extended control point grid' which covered the full region spanned by all seven couch positions (figure 3). This is so that the model predictions from each couch position can be combined into a single continuous B-spline transformation (see the next section).

2.3. Relating motion to respiratory parameters

The motion models were constructed by fitting functions that relate the lung motion, as described by the B-spline registration results, to the respiratory parameters. A separate function was fitted for each of the control point displacements that define the B-spline transformation, relating it to the respiratory surrogate signal. We have investigated the following three models:

(1) a cyclic B-spline function (McClelland *et al* 2006) with four control points relating the control point displacement to the respiratory phase (figure 4(a)),
(2) a third-order polynomial (cubic) function relating the control point displacement to the value of the respiratory surrogate signal (figure 4(b)),
(3) a 2D first-order polynomial (linear) function relating the control point displacement to the value and rate of change of the surrogate signal (figure 4(c)).

Model 1 relates each control point displacement, CP, to the respiratory phase, p, by the following equation:

$$\mathrm{CP}(p) = \sum_{i=0}^{3} B_i(j)\phi_{i+k(\bmod 4)}$$

where $j = \frac{p}{25\%} - \left\lfloor \frac{p}{25\%} \right\rfloor$, $k = \left\lfloor \frac{p}{25\%} \right\rfloor - 1$, B_i is the ith B-spline basis function (Rueckert *et al* 1999), and $\phi_0, \ldots, \phi_z$ are the coefficients that define the cyclic B-spline and are evenly distributed along the phase axis at phase values of 0, 25, 50 and 75%. This model constrains the motion to follow the same path during every respiratory cycle. This path forms a closed loop, so it is possible to model hysteresis, but it does not allow any intra-fraction variation to be modelled and averages out any which occurs, effectively modelling an average respiratory cycle. Figure 4(d) shows the motion trace of an exemplary landmark point produced using model 1. It can be seen that model 1 constrains the point to go round the same loop during each respiratory cycle.

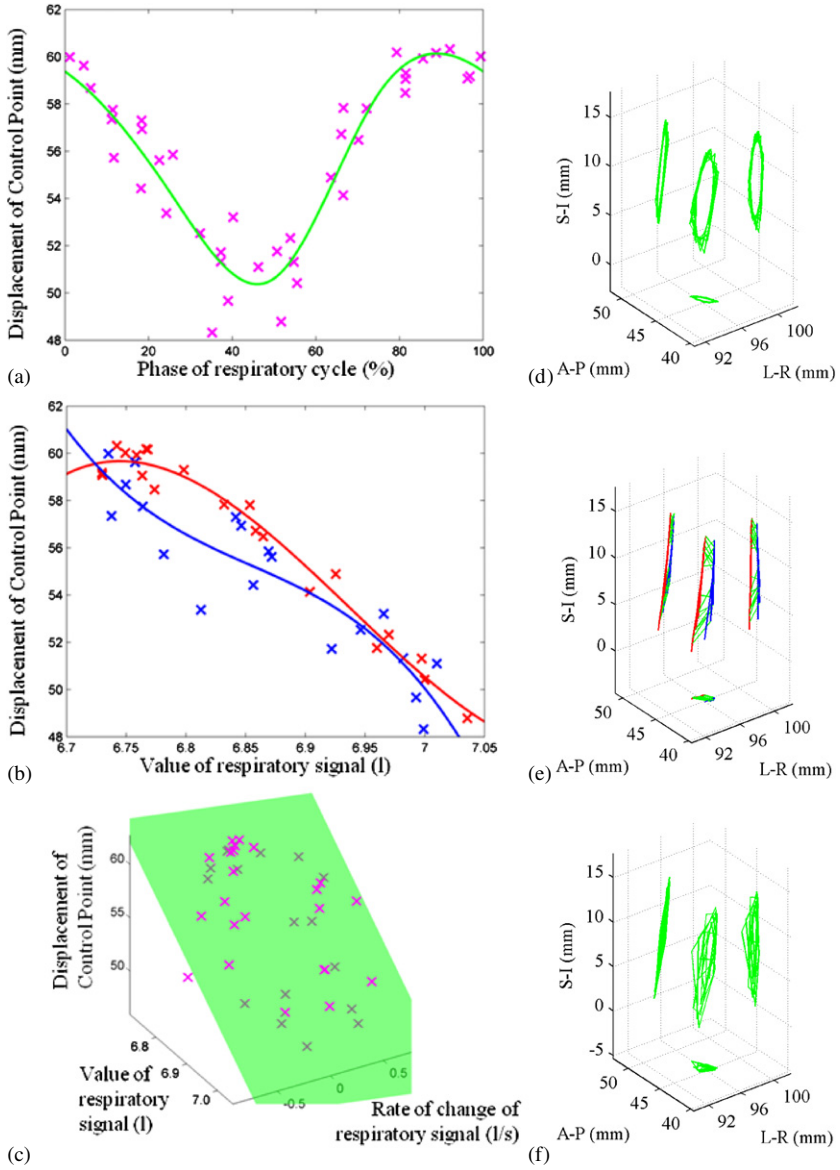


Figure 4. Example plots of the respiratory parameters (horizontal axis) against an arbitrary B-spline control point displacement (vertical axis) showing the results from the individual registrations (crosses) and the models fit to the data for model 1 (a), model 2 (b) (red—inhale model, blue—exhale model) and model 3 (c) and traces showing the 3D motion of an exemplary landmark point for model 1 (d), model 2 (e) (red—inhale model, blue—exhale model, green—'jump' between models) and model 3 (f).

Model 2 relates each control point displacement, CP, to the value of the respiratory surrogate signal, v, by the following equation:

$$\mathrm{CP}(v) = a_3 v^3 + a_2 v^2 + a_1 v + a_0$$

where $a_0, \ldots, a_3$ are the coefficients that define the cubic function. This model also constrains the motion to follow the same path during every respiratory cycle, but this path is formed of two independent trajectories, one for inhalation and one for exhalation rather than a closed loop. The motion can progress a different distance along the trajectories during different breath cycles, allowing some variation in the depth of breathing to be modelled. Hysteresis is modelled as separate models made for the inhalation and exhalation. However, this means that the model of the full respiratory cycle is no longer continuous, and the motion predictions will alternate between the inhalation and exhalation trajectories. Figure 4(e) shows the motion trace of an exemplary landmark point produced using model 2. It can be seen that the motion follows the same trajectories during inhalation (red trace) and exhalation (blue trace) for each breath cycle, but 'jumps' from one trajectory to the other at different points (green trace) as the depth of breathing varied from one breath to another.

Model 3 relates each control point displacement, CP, to the value, v, and the rate of change, r, of the respiratory surrogate signal by the following equation:

$$\mathrm{CP}(v, r) = a_2 v + a_1 r + a_0$$

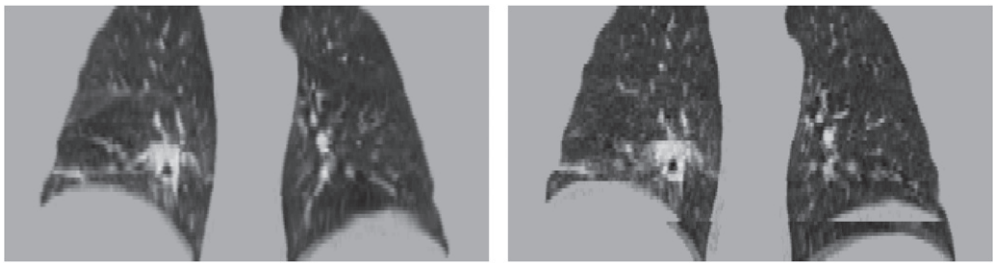
where a_0, a_1, a_2 are the coefficients that define the 2D linear function. This model constrains the motion to lie on a 2D plane, but it can follow a different path on this plane during each respiratory cycle. This enables both hysteresis and some degree of intra-fraction variation to be modelled. Figure 4(f) shows the motion trace of an exemplary landmark point produced using model 3. It can be seen that the motion is constrained to lie on a 2D plane but that it can follow a different path on this plane for each breath cycle.

The parameters for each model were determined by performing a multiple linear regression least-squares fit to the registration results. The registration results from each couch position were modelled separately, so seven models were constructed per dataset. The predictions from each model were then combined into a single continuous B-spline transformation that predicted the motion over all the couch positions. The extended control point grid mentioned in the previous section provided a common control point grid for all the registrations from the different couch positions. The model for each couch position predicted the motion over the whole of the common control point grid, but the prediction was only reliable in a region close to and including the couch position that was modelled. The reliability of the predictions for a particular row of control points depended on the effect those control points had on the registrations. Control points lying within the couch position being registered had a large effect on the registration. Control points outside the couch position but within the spline support area (two control points) had some effect on the registration, and control points outside the spline support area had no effect.

To combine the model predictions into a single transformation a weighted average was taken, with a different weight used for each row of control points and each model. The weight was equivalent to the effect that the control points had on the registrations for that model (calculated from the B-spline basis functions). The combined transformation was then used to predict the deformation field and the deformed volume over all of the couch positions. The deformed volumes predicted in this way were free from the artefacts often seen between adjacent couch positions in 4DCT volumes (figure 5).

2.4. Assessing the models

To quantitatively assess the registration results and model predictions a clinical expert manually located an easily identifiable landmark point (usually a bifurcation in an airway or vessel) in each of the Cine CT volumes. The same landmark was used for all the Cine CT volumes



(a) (b)

Figure 5. Volume predicted by model 1 (a) compared to a phase-sorted 4DCT volume (b) made from the same dataset. It can be seen that the predicted volume is free from the artefacts that are present near the diaphragm in the 4DCT volume.

from the same couch position, and was also located in the reference volumes from both of the datasets for the patient. The Euclidean distance was calculated between the landmark in the Cine CT volumes and the landmark in the reference volumes:

- prior to any transformation to give the target error (TE),
- transformed by the B-spline registration results to give the target registration error (TRE),
- transformed by the model predictions to give the target modelling error (TME).

TME values were calculated for each of the three motion models described above. The motion models were first assessed on the same dataset that they were built from to evaluate the ability of the models to predict the intra-fraction motion and variation (TME_{intra}). The models were assessed using a leave-one-out strategy, where the models were constructed from 39 of the Cine CT volumes, and then used to predict the Cine CT volume that was left out from building the model. This is then repeated for all 40 Cine CT volumes. The models for the other couch positions, which are also required to make the combined model predictions, are constructed using all 40 Cine CT volumes.

The motion models were then assessed using the other dataset from the same patient to evaluate how well the models could predict the motion in the presence of inter-fraction variation (TME_{inter}). For each patient both datasets were used to build and assess the models. Assessing a model built at the end of treatment on data acquired prior to treatment is of course not clinically realistic, but was a valid strategy for this study as we were interested in analysing how the motion has varied and this in effect doubled the data available to us.

The two datasets did not have exactly the same field of view. Current clinical practice at Guy's and St Thomas's hospitals, where the data for this study were gathered, is to form an internal target volume (ITV) which encompasses the tumour during the entire respiratory cycle from the planning 4DCT scan. A CBCT scan is then acquired prior to each fraction of treatment. In this scan the tumour is blurred over the extent of its motion due to the relatively long acquisition time. The treatment couch is then translated to align the ITV from the planning scan with the blurred tumour from the CBCT scan. This can account for base-line shifts of the tumour relative to the skin tattoos and/or bony anatomy used to position the patient for RT treatment. To simulate aligning the patient using a CBCT scan an ITV was formed for both datasets, encompassing the tumour in all Cine CT scans. A rigid registration was then performed (using translations only) to automatically align the two ITVs, and the result was used to align the two datasets.

The TE values were calculated using the reference volume from the same session (TE_{intra}) and using the reference volume from the other sessions, aligned as above (TE_{inter}). These

assessed the error of assuming that the anatomy was always at the reference (end-exhale) position.

As the TRE only measures the accuracy of the registration at one point, the B-spline registration results were also visually assessed to make sure there were no unacceptable errors (approximately greater than 5 mm) in the other regions of the volumes. Any unacceptable registration results were not used to build the motion models. The motion model results were visually compared to the original Cine CT volumes from the same session to assess their ability to model the data.

2.5. Analysing the inter-fraction variation

It is possible that regions of the lung that receive a high dose of radiation may be subject to greater inter-fraction variations than regions that receive a lower dose, as the radiation may damage the lung tissue and affects its mobility. In order to compare the inter-fraction variation for high-dose regions to that of low dose regions, the clinical expert was asked to further identify one landmark point in the high-dose region (adjacent to or very close to the tumour) and three landmark points at a comparable superior-inferior position to the tumour, but in low dose regions. One of these was in the ipsilateral lung, but far from the tumour, and the other two were in the contralateral lung, with one in the anterior half and the other in the posterior half. These landmarks were only located in the two references volumes for each patient. The registration results were then used to find the location of the landmarks in each of the Cine CT volumes that contained the landmarks. As the registrations have been shown to give good results in previous work (Modat *et al* 2010a, mean TRE 0.75 mm) and in the previous experiment (see table 2) this is thought to provide a good approximation of the true landmark location, and greatly reduced the amount of time spent by the clinical expert in identifying the landmark points. The motion models constructed from the other dataset were then used to predict the locations of the high-dose and low-dose landmark points, and the Euclidean distance between the model predictions and the locations given by the registration results was calculated (the inter-fraction registration modelling error, $\mathrm{RME}_{\mathrm{inter}}$). Again, models were built separately from both datasets, but were always compared to the registrations in the dataset not used to build the model, so as to double the amount of data available for analysing the inter-fraction variations.

To further analyse the inter-fraction variation the optimum base-line shift was found for each landmark point (both the landmark points manually located in each Cine CT volume for the previous experiment, and the high-dose and low-dose landmark points used above) by finding the average modelling error ($\mathrm{TME}_{\mathrm{inter}}$ or $\mathrm{RME}_{\mathrm{inter}}$) for each landmark. These were analysed for each patient to study how the base-line shifts varied for different locations in the lung. Additionally, the $\mathrm{TME}_{\mathrm{inter}}$ and $\mathrm{RME}_{\mathrm{inter}}$ values were recalculated after applying the optimum base-line shift to each point individually to enable us to analyse how much of the inter-fraction variation could not be explained by a simple base-line shift.

The volumes predicted by the motion models were also visually compared to each other to further examine the inter-fraction variations that had occurred.

3. Results

Datasets were acquired from six patients with non-small cell lung cancer, whose details are given in table 1. However, one patient (patient 6) had an extremely fast rate of breathing causing many of their Cine CT volumes to contain artefacts such as blurred or double structures. This made it very difficult to assess the registrations and models using landmarks as the landmarks

Table 1. Subject characteristics.

Subject	Histology	Tumour location	TNM
1	Adenocarcinoma	Right lower lobe	T2N0M0
2	Squamous cell carcinoma	Right upper lobe	T3N0M0
3	Squamous cell carcinoma	Left upper lobe	T4N1M0
4	Squamous cell carcinoma	Right upper lobe	T2N0M0
5	NSCLC	Right lower lobe	T1N0M0
6	Squamous cell carcinoma	Right lower lobe	T2N0M0

Table 2. The mean, and 95th and 99th percentiles of the intra-fraction and inter-fraction target error (TE_{intra} and TE_{inter}), the target registration error (TRE) and the intra-fraction and inter-fraction target modelling error (TME_{intra} and TME_{inter}), in mm.

			TME_{intra}				TME_{inter}		
	TE_{intra}	TRE	Model 1	Model 2	Model 3	TE_{inter}	Model 1	Model 2	Model 3
Mean	2.8	1.0	1.2	1.2	1.2	5.5	4.6	4.7	4.6
95th percentile	9.7	2.3	3.0	2.8	2.7	12.9	10.3	11.4	11.3
99th percentile	15.3	3.2	4.5	3.9	3.7	16.6	16.3	15.6	15.1

appeared twice in some volumes. Therefore this patient was excluded from any numerical analysis. The Cine CT volumes from the other patients were also checked for similar artefacts. Only one Cine CT volume from the other five patients was found to have noticeable artefacts, and this volume was not used for the rest of the study.

Occasionally it was not possible to identify the same landmark point in all Cine CT volumes for couch positions near the base of the lungs as the landmark moved out of the Cine CT volumes. As this only occurred in a few cases, these Cine CT volumes were excluded from the numerical analysis (but they were still used in constructing the models). Additionally, there were some couch positions (four in total from all datasets) that contained very little or no lung, and for these couch positions it was not possible to identify a suitable landmark point so they were excluded from the numerical analysis.

3.1. Results of assessing the models

Table 2 gives the mean, and 95th and 99th percentiles for the landmark tracking results. Although the mean target error from the same session (TE_{intra}) was relatively small (2.8 mm), 5% of the TE_{intra}s were almost 10 mm or more, and more than 1% of the TE_{intra}s were more than 15 mm. Closer examination of the results revealed that all patients had some Cine CT volumes with TE_{intra} values close to 10 mm, and hence subject to considerable motion. The mean target error using the reference volume from the other session (TE_{inter}) was approximately double the mean TE_{intra}, indicating that there had been considerable inter-fraction variation. The TRE results indicate that the registrations did very well at recovering the respiratory motion, with a mean TRE of 1 mm and 99% of the TREs 3.2 mm or less. Visual inspection of the registration results confirmed that the vast majority of the registration results did not contain any unacceptable errors in any parts of the volume. A total of 2800 registrations results were visually assessed (patient 6 was not visually assessed due to the blurred Cine CT volumes, therefore 5 patients $\times$ 2 datasets $\times$ 7 couch positions $\times$ 40 Cine CT volumes $=$ 2800 registrations) and of these there were only five that appeared to contain errors larger than

5 mm. These registrations were excluded when constructing the models and when analysing the inter-fraction variation in the next experiment.

The intra-fraction target model error (TME_{intra}) results indicate that the models could accurately predict the motion from the same dataset used to build the models, with a mean TME_{intra} of 1.2 mm for all the models. The percentile values indicate that model 3 performed marginally better than model 2, which performed slightly better than model 1, as would be expected if there were some small intra-fraction variations in the respiratory motion. However, the difference between models 2 and 3 is negligible, and the difference between model 1 and the others is very small considering that model 1 relates the motion to the respiratory phase while the others relate the motion directly to the value of the surrogate signal (and its rate of change for model 3). The TME_{intra} results varied slightly from patient to patient (mean values 0.9–1.7 mm), but the relative performance of the different models was similar for all patients (i.e. model 3 was marginally better than model 2 and slightly better than model 1).

The TME_{inter} results indicate that the overall performance of the three models was also very similar when predicting the motion in the other dataset. However, none of the models were able to adequately predict the motion that occurs in the other dataset with mean errors that are almost 5 mm and a considerable number of larger errors as indicated by the percentile values. There were greater differences in the TME_{inter} results from one patient to another (mean values 2.5–8.9 mm), indicating that the amount of inter-fraction variation was different for different patients. The relative performance of the different models did differ from one patient to another (e.g. for patient 4 model 1 gave the smallest TME_{inter} values, but for patient 5 model 3 gave the smallest TME_{inter} values) but they were generally quite similar to each other, and none of the models gave satisfactory predictions of the motion in the other dataset for any of the patients.

Visual comparison of the Cine CT volumes and the volumes predicted by the models from the same dataset revealed that all three models can reproduce the Cine CT volumes very well, providing there was negligible variation between the few respiratory cycles sampled at each couch position, as was usually the case. When there was appreciable breath-to-breath variation in the data from an individual couch position (which was only noticeable in a few Cine CT volumes from a couple of patients), model 3 generally predicted the Cine CT volumes best and model 1 predicted them worst. However, model 1 predictions were still very good for the majority of the Cine CT volumes even when there was irregular breathing, and the model predictions that did contain noticeable errors usually appeared to be within 5 mm of the Cine CT volumes. Additionally, it was noticed that models 2 and 3 would very occasionally produce predicted volumes that matched the Cine CT data well but were clearly implausible for the other couch positions. This occurred when there was considerable variation in the breathing between the different couch positions, as there was for two of the patients studied. These implausible predictions were caused by the models for the different couch position being fitted to different ranges of surrogate signal values and so needing to extrapolate beyond the fitted data to predict the Cine CT volumes from the other couch positions.

3.2. Results of analysing the inter-fraction variation

Table 3 gives the mean, and 95th and 99th percentiles for the RME_{inter} results for the high-dose and low-dose landmarks. Only the results from model 3 are shown as the results from all three models were quite similar, and the previous experiment has shown that overall model 3 gave the most accurate predictions of the intra-fraction motion and variation within a single dataset. The RME_{inter} results are generally smaller for the high-dose points, indicating that the low-dose landmarks were subject to more inter-fraction variation than the high-dose landmarks. Initially

Table 3. The mean, and 95th and 99th percentiles of the inter-fraction registration modelling error (RME_{inter}), in mm, for the high-dose landmark and the low-dose landmarks for each patient.

		Patient				
		1	2	3	4	5
High-dose landmark	Mean	1.7	1.8	1.8	3.4	9.3
	95th percentile	2.3	2.5	4.7	3.8	13.7
	99th percentile	2.5	2.8	5.3	3.9	14.5
Low-dose landmarks	Mean	4.1	2.4	2.7	5.1	9.0
	95th percentile	5.7	2.9	4.2	7.5	13.7
	99th percentile	6.6	3.1	4.9	8.1	14.5

Table 4. The standard deviation of the optimum base-line shift over all landmarks in the left right (LR), anterior posterior (AP), and superior inferior (SI) directions for each patient.

	Patient				
	1	2	3	4	5
LR	1.7	0.6	2.0	1.6	1.7
AP	2.6	1.3	1.4	1.3	1.6
SI	2.8	3.6	2.6	2.3	5.6

this seems to be the opposite of what might be expected as it was thought that the high-dose regions may have greater inter-fraction variations as the radiation may damage the lung tissue and affects its mobility. However, the high-dose landmarks are obviously located very close to the ITV which has been used to align the datasets from the two sessions, whereas the low-dose landmarks are more distant. This means that there should be relatively small base-line shifts for the high-dose points, whereas it is possible that there may be larger base-line shifts for the low-dose points.

Figure 6 shows the optimum base-line shift for each of the landmark points identified for each patient (again only the results for model 3 are shown for the same reasons as above). It can clearly be seen that for patients 1–3 there is a relatively small base-line shift for the high-dose landmark (green o), whereas there can be larger base-line shifts for the low-dose landmarks (red +). For patient 5, and to a lesser degree for patient 4, there were relatively large base-line shifts for the high-dose landmarks as well as for the low-dose landmarks. This is because the tumour had noticeably shrunk for both patients between the first dataset and the second dataset, and in addition the motion trajectory of the tumour had changed between the two datasets for patient 5. This meant that the ITVs used to align the two datasets were considerably different in size and shape, so points in the high-dose region could still have noticeable base-line shifts even thought the ITVs were aligned.

Table 4 gives the standard deviation of the optimum base-line shifts for each patient and in each direction. From this table and from figure 6 it can easily be seen that there can be a high degree of variability in the optimum base-line shifts for different locations in the lungs. This means that even if the tumour and the high-dose region can be aligned relatively well using a simple translation, other locations in the lung, including the low-dose landmarks that are at the same SI height as the RT beams, may not be well aligned.

Table 5 gives the TME_{inter} results for the Cine CT landmarks from the first experiment, and the RME_{inter} results for the high-dose and low-dose landmarks, after the optimum base-line

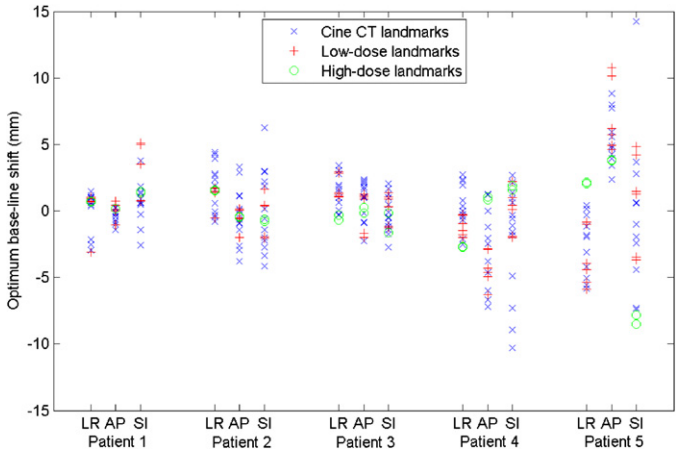


Figure 6. Plot showing the optimum base-line shifts (in mm) in the left right (LR), anterior posterior (AP) and superior inferior (SI) directions for each of the Cine CT landmarks (blue ×), the low-dose landmarks (red +) and the high-dose landmark (green o) for each patient.

Table 5. The mean, and 95th and 99th percentiles of the inter-fraction target modelling error (TME_{inter}) after the optimum base-line shift has been applied, in mm, for the Cine CT landmarks for each patient and the inter-fraction registration modelling error (RME_{inter}) after the optimum base-line shift has been applied, in mm, for the high-dose landmark and the low-dose landmarks for each patient.

		Patient				
		1	2	3	4	5
Cine CT landmarks	Mean	1.2	0.8	1.0	1.7	1.9
	95th percentile	2.4	2.0	2.4	3.5	3.9
	99th percentile	3.4	2.5	3.8	5.1	5.6
High-dose landmark	Mean	0.6	0.6	1.6	0.7	1.7
	95th percentile	1.1	1.2	3.2	1.1	3.7
	99th percentile	1.3	1.3	3.6	1.1	4.3
Low-dose landmarks	Mean	0.8	0.4	1.0	1.6	1.8
	95th percentile	1.6	0.8	2.2	3.2	3.7
	99th percentile	2.2	0.8	4.0	5.3	5.7

shift has been applied to each individual landmark (as above only the results from model 3 are shown). These results indicate that almost all the inter-fraction variations for patients 1–3 were base-line shifts (but as noted above there were different base-line shifts for different locations in the lungs). The results for patients 4 and 5 indicate that for some landmark points the inter-fraction variation could not be explained just by base-line shifts.

Table 5 also shows that for the low-dose landmarks there was a similar or greater amount of inter-fraction variation that could not be explained by base-line shifts than that for the high-dose landmarks. This would indicate that the inter-fraction variations that occurred were not only due to local radiation damage in the high dose region.

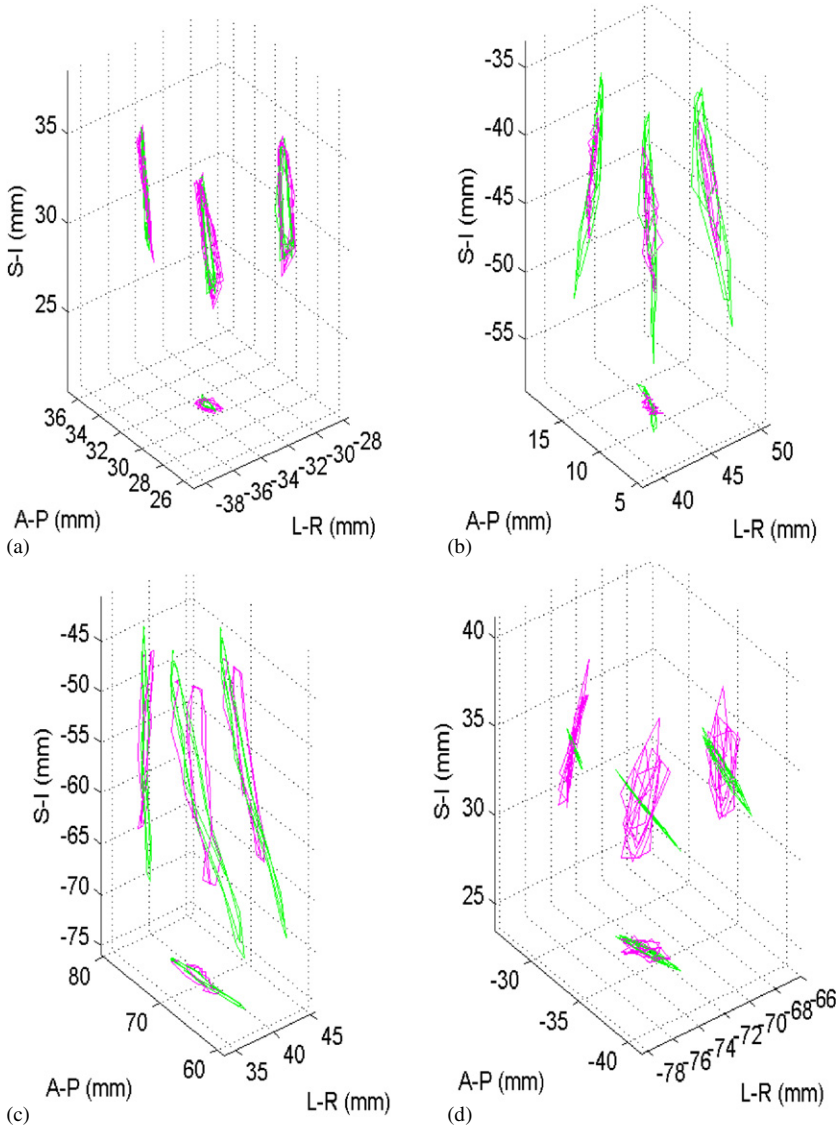
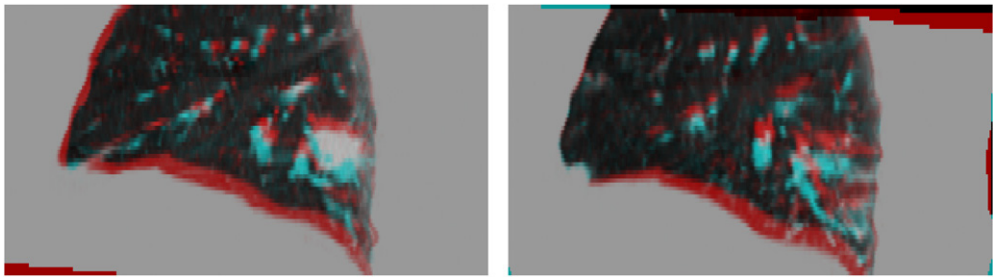


Figure 7. Motion traces for exemplary landmark points for patients 1 (a), 3 (b), 4 (c) and 5 (d). The magenta traces show the landmark trajectories as determined by manually locating the landmarks or using the registration results (depending on whether the landmarks are from the first or second experiment), and the green traces show the landmark traces as predicted by the motion models built from the other dataset.

The traces produced from tracking the landmarks points (with the optimum base-line shifts applied) where examined to determine how the motion varied between the datasets (figure 7). The traces from patients 1 and 2 supported the conclusion that the majority of their inter-fraction variation was due to the base-line shift. Figure 7(a) shows exemplary landmark motion traces for a landmark from patient 1 after the optimum base-line shift has been applied, and it can be seen that the two traces match each other well. The traces from patient 3



(a) (b)

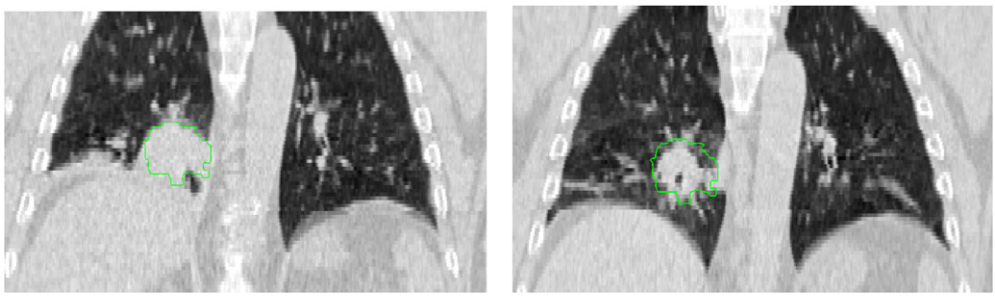
Figure 8. Colour overlays showing a volume predicted by model 3 at end-exhale (red) overlaid on a volume predicted at end-inhale (cyan) for patient 5. It can be seen that this patient was breathing more with their chest during the first session (a) and more with their diaphragm during the second session (b) by the red areas (indicating motion) at the chest and diaphragm in each image respectively. It can also be seen that there was more AP motion during the first session and more SI motion during the second.

revealed that most of their inter-fraction variation was also due to base-line shifts; however, for a few of their landmarks there also appeared to be a noticeable change in the magnitude of the landmark motion (figure 7(b)). The traces for patient 4 showed a similar result, with the motion trajectories being approximately the same shape but of different magnitudes (figure 7(c)). This affected more of the landmarks and had a larger effect for patient 4 than for patient 3. As the motion model should be able to account for (small) changes in the magnitude of the respiratory motion, this indicated that there must have been a change in the relationship between the surrogate signal and the internal motion for these patients. The traces for patient 5 also revealed a change in the magnitude of the respiratory motion for some landmarks, and in addition the shape of the motion trajectory varied noticeably for some landmarks (figure 7(d)).

Visual examination of the volumes predicted by the motion models from the two different datasets supported the above results and conclusions about the types of inter-fraction variation that occurred for the different patients. In addition, they showed that the cause of the different motion trajectory shapes for patient 5 was that the patient was breathing more with their chest during the first dataset and more with their diaphragm during the second dataset (figure 8). Visual examination of the predicted volumes for patient 6, who had been excluded from the numerical analysis due to the blurred Cine CT volumes, showed a large anatomical change where their shrinking tumour caused a previously blocked region of the lung to reopen (figure 9). Generally the predicted volumes from the three different types of model were similar to each other, although models 2 and 3 would occasionally produce implausible predicted volumes when variations in the breathing between the datasets caused one of the models to extrapolate beyond the range of values used to build the model.

4. Discussion

We have presented a framework for modelling respiratory motion that relates the internal motion, described by deformable registrations, to a respiratory surrogate signal derived from the 3D skin surface motion. We have evaluated three different models for relating the internal motion to the surrogate signal. Model 1 related the internal motion to the respiratory phase, model 2 related the motion to the value of the surrogate signal, and model 3 related the motion



(a) (b)

Figure 9. Slice through the reference CT volume from the first dataset (a) and the second dataset (b) from patient 6. The gross tumour volume (GTV) contour from the first dataset has been overlaid on both images (green). It can be seen that there are large anatomical changes between the two images, especially near the base of the right lung.

to the value and rate of change of the surrogate signal. We wanted to investigate all of these models as they are similar to approaches used in the literature and have all shown promise in our initial studies. Respiratory phase is often used to sort 4DCT volumes (Britton *et al* 2007, Redmond *et al* 2009) and was used for our original motion models (McClelland *et al* 2006). The value of the surrogate signal (sometimes referred to as the amplitude of the signal in the literature) can also be used to sort 4DCT volumes which can result in less artefacts than when sorting according to phase (Lu *et al* 2006), although irregular breathing can cause problems with missing data. The value of the surrogate signal is also used by the commercially available CyberKnife system (Seppenwoolde *et al* 2007). Models that use the value and rate of change of the surrogate signal have been proposed in the literature (Low *et al* 2005, Yang *et al* 2008, Zhang *et al* 2007), and have the advantage that they can potentially model breath to breath variations in the respiratory motion.

The landmark tracking results showed that models 2 and 3 were slightly more accurate at predicting intra-fraction motion and variation than model 1. This could be because they were able to model some of the intra-fraction variation that model 1 was not able to, but the differences between the models are too small and the number of datasets too few to draw firm conclusions. Another difficulty is that we only have data over a few breath cycles at each couch position and there is generally very little variation between these few breath cycles, even when there was appreciable intra-fraction variation over the whole acquisition. However, it was not feasible to acquire more CT data at each couch position due to dose considerations. A prospective Cine CT acquisition system which only acquires data for predefined values of the respiratory parameters, such as that proposed by Langner and Keall (2008), could prove very useful for the methods described in this paper. It could be used to efficiently study the intra-fraction variation and the performance of different motion models, and it could also be used to acquire the optimum data for a specific model, sampling the 'average' respiratory motion and any intra-fraction variation that is to be included in the model. The feasibility of implementing a similar system will be investigated in the future, although it is thought that the necessary modifications to the CT scanner will not be possible without the manufactures cooperation and may not be suitable for routine clinical use.

The visual examination of the results revealed a potential problem with models 2 and 3. Shifts and drifts in the respiratory surrogate signal between different couch positions or between the different datasets meant that the models sometimes needed to extrapolate beyond

the range of values used to fit the models. It was not surprising that the extrapolated predictions from the third-order polynomial model were implausible as polynomials are notoriously bad when it comes to extrapolating (the CyberKnife system reverts to a linear model when extrapolating predictions, and uses a second-order polynomial when interpolating predictions). However, we found that the extrapolated predictions from the 2D linear model were also highly implausible. This is not altogether surprising given the nature of the complex deformations being modelled, but shows that great care should be taken when using any of the models to extrapolate predictions of the respiratory motion, and that more work is required to determine how best to minimise extrapolation errors. Although there remains work to be done, our results hold promise that it may be possible to model the majority of the intra-fraction variation in the respiratory motion and relate it to the 3D skin surface or other respiratory surrogate signals.

Our results also show that motion models constructed at the planning stage may be invalid for later fractions of treatment due to inter-fraction variation in the respiratory motion. The datasets were automatically aligned using the ITVs formed from each dataset, to mimic the current clinical practice of aligning the ITV from the planning scan with the (blurred) tumour from a CBCT scan. However, even after aligning the datasets in this way there were still large errors when predicting the motion for some regions of the lungs, including appreciable errors when predicting the motion of the high-dose region close to the ITV in two out of five patients. These results indicate that while aligning the patient with CBCT images of the tumour improves patient setup when compared to aligning them with skin tattoos or bony anatomy, it does not necessarily mean that the RT treatment will have been delivered as planned.

Analysis of the inter-fraction variation showed that the majority of the variation was caused by other regions of the lungs having different base-line shifts to the ITVs (and to each other). This could be an important clinical issue for any regions of the lungs with a dose limit (e.g. the major airways in stereotactic RT), or for multiple targets (e.g. positive nodes), or for other applications (e.g. reconstructing motion free images) as it may not be possible to correct for the different base-line shifts with a single rigid transformation (e.g. moving the treatment couch). Although not studied here, other organs outside the lung that need to be considered in RT plans will likely also have different base-line shifts to the lungs. It may be possible to compensate for base-line shifts using non-rigid deformations rather than simple translations, but to do this correctly the deformations would also have to account for anatomical changes such as tumour shrinkage/growth, and this is a non-trivial problem.

In addition to the base-line shifts, we saw a change in the relationship between the surrogate signal and the internal motion, to some degree, in three patients. For two of the patients (3 and 4), this appeared to be essentially a change in magnitude of the internal motion compared to magnitude of the surrogate signal. For the other patient (5), who was breathing more with their chest in the first dataset and more with their diaphragm in the second, there was a noticeable change in shape of the internal motion trajectories. We also observed a patient (6) with large anatomical changes caused by the opening up of a previously blocked region of the lung.

In the future it may be possible to build models that describe different types of breathing, e.g. abdominal and thoracic, deep and shallow, and can predict the different motion trajectories caused by them, but these will likely require much more data to be acquired to image the different types of motion, or a much better understanding of how the respiratory motion can vary. However, in the foreseeable future it is unlikely that it will be possible to construct models that can account for large anatomical changes or for types of breathing not imaged when building the models. When such changes occur it will be necessary to somehow update or rebuild the motion models with new data acquired at the time of treatment.

Our results showed that the models could predict the motion of the region of the lung receiving a high radiation dose as well as those regions of the lung receiving a lower dose, even after optimum base-line shifts had been applied. This indicates that the high radiation dose had not had a local effect on the lung motion for these patients. This could be because the radiation damage had a more global than local effect on the lung motion, or that the RT treatment has not yet had much effect on the lung motion and the variation observed was natural day-to-day variations in the respiratory motion (some damage to the lungs caused by RT treatment can take a few months to fully develop).

The motion modelling method presented in this paper can utilize any 1D respiratory surrogate signal (e.g. a chest/abdomen marker, a pressure belt, spirometry, internally derived surrogate signals (Carnes *et al* 2009, Korreman *et al* 2006), etc.). The choice of surrogate signal may affect the accuracy of the models, but the inter-fraction variations observed such as base-line shifts and switching from chest to diaphragm breathing would cause similar errors to the model predictions whichever 1D surrogate signal was used. We decided to use a surrogate signal derived from the 3D skin surface as this can be easily acquired during treatment planning and delivery without any extra radiation dose or discomfort to the patient. In addition, we have previously shown that the surrogate signal used is similar to a spirometry signal without the problem of drift often found with spirometry (Hughes *et al* 2009), and spirometry has been shown to correspond to the internal respiratory motion better than the displacement of a point on the abdomen (Hoisak *et al* 2004, Lu *et al* 2005). The 3D skin surface can also be decomposed into several respiratory surrogate signals, for example different points or regions on the skin surface (Hughes *et al* 2009). In the future, models based on multiple surrogate signals may be better able to account for variations in the respiratory motion.

In this study we only registered and modelled the lungs. It is acknowledged that for the models to be utilized for RT planning the rest of the anatomy would also need to be modelled. In principle exactly the same methods could be applied to the rest of the anatomy and a separate model could be created for it, although extra care needs to be taken to ensure there are no gaps or overlaps between the two models (Wu *et al* 2008). However, it is not just the lungs that need segmenting, but also the mediastinum and everything below the diaphragm that moves with the lungs. This is a much harder segmentation task which currently requires a difficult and labour intensive manual segmentation by a clinical expert as automatic methods are still being developed (Vandemeulebroucke *et al* 2010).

The deformable registrations typically took about 45 min for each Cine CT volume (running on a single core of a Sun V20Z Opteron 250 CPU with 4 GB memory). This is far too long to be feasible for clinical use as there were 280 Cine CT volumes that needed registering for each patient. However, as the registrations and their computation time were not the main focus of this work we put little effort into finding optimum registration parameters (e.g. control point spacing) and settled for those that were found to produce suitable results. Based on similar registrations we have recently run, we expect that the registration time can be reduced by about an order of magnitude with little or no detrimental effect on the registration results (Modat *et al* 2010a). Furthermore, we performed all of our registrations on standard CPUs as we have a large cluster of several hundred nodes available in our lab which is well suited to processing large volumes of data. The implementation of the B-spline registration algorithm used for this study (Modat *et al* 2010b) has been written to run on either a CPU or a GPU, and the GPU implementation is typically at least an order of magnitude faster than the CPU version. The affine registrations were comparatively fast taking in the order of a few minutes (and again could be substantially faster if optimal parameters and a GPU were used), as was the model fitting which took about 30 s per model and was implemented in Matlab (MathWorks, Natick, MA).

NMI was used as the similarity measure as it had been implemented and well tested as part of the software used in this study, and was found to produce good results. We also suspected that NMI may be better suited to dealing with the changes in the density of the lung parenchyma that occur during respiration and causes corresponding voxels in different Cine CT images to have different intensity values. In the future we intend to perform an in depth study into the deformable registration of the lungs and will investigate the effects of different similarity measures as well as different constraint terms and different algorithms. It is acknowledged that only using a single landmark point in each Cine CT volume is far from sufficient for assessing the registrations and model predictions. However, as there were so many Cine CT volumes (2800), it was not feasible to identify more points in every volume, and this is one of the reasons why the visual validation was also performed. Future studies will be carried out using many more landmark points but on a much smaller subset of volumes to further assess the accuracy of the registrations.

In general the registration performed very well. Even for the patient who was breathing very fast, and hence had many 'blurry' Cine CT volumes, the registration generally appeared to give a reasonable result (but a higher weight for the penalty term was required), although there was no way to validate them due to the quality of the data. The problem with the data for this patient would also have caused problems with 4DCT volumes or any other motion modelling methods using these data. Excluding this patient (where it is not possible to make a fair assessment) the registration was very robust, only failing in 5 out of 2800 registrations. However, as long as there are some failures occurring, even as few as observed here, an automated method is required to validate the registrations (and model predictions) and to detect the failures if the models were to be used clinically.

It is acknowledged that our method of combining the transformations from the models for the different couch positions could produce folding or other implausible deformation if there are large differences in the motion predicted for adjacent couch positions. This was not observed for any of the patients studied here. Furthermore, it is thought that if the differences in the model predictions are large enough to cause folding then there must have been a considerable amount of variation in the respiratory motion during the acquisition of the adjacent couch positions. It is likely that this would also invalidate 4DCT volumes or any other types of models constructed from the data. It should be noted that our method of combining the model predictions from the different couch positions can also be used to 'remove' small artefacts from standard 4DCT volumes acquired in Cine mode. To do this the 4DCT volume is divided into the individual couch positions, each are registered separately, and then the registration results are combined in the same way as the model predictions in this paper.

In this work we first registered each Cine CT volume separately and then fitted our models to the registration results. An alternative to this approach is to combine the deformable registration and model fitting into a single optimisation. Then all of the Cine CT volumes from all of the couch positions could be registered and modelled simultaneously. However, this would likely be very complex to implement and prohibitively computationally expensive, and it is not clear there would be much to gain over our current approach.

5. Conclusion

We have presented our improved method for generating patient-specific models of the respiratory motion. We have shown that these can model the motion well during a single session, but that inter-fraction variation between sessions invalidates the models. We have used the models to analyse the inter-fraction variation. We found that most of the variations

were 'base-line shifts', and that these shifts were different for different points in the lung. There were also changes to the relationship between the respiratory surrogate signal and the internal motion for three patients, including changes to the shape of the internal motion trajectories for one of the patients. Large anatomical changes were observed in another patient.

Acknowledgments

We would like to acknowledge Matthew Adams and the other CT simulator radiographers at Guy's and St Thomas's hospitals for their assistance with acquiring the patient data, Vision RT for their help with acquiring and processing the 3D surface data, and Tristan Clark from the UCL CS Technical Support Group for assistance and support using the computer cluster for our deformable registrations. This work was supported by the UCL/KCL Comprehensive Cancer Imaging Centre funded jointly from Cancer Research UK and the EPSRC.

References

Britton K R, Starkschall G, Tucker S L, Pan T, Nelson C, Chang J Y, Cox J D, Mohan R and Komaki R 2007 Assessment of gross tumour volume regression and motion changes during radiotherapy for non-small-cell lung cancer as measured by four-dimensional computed tomography *Int. J. Radiat. Oncol. Biol. Phys.* **68** 1036–46

Cai J, Read P W and Sheng K 2008 The effect of respiratory motion variability and tumor size on the accuracy of average intensity projection from four-dimensional computed tomography: an investigation based on dynamic MRI *Med. Phys.* **35** 4974–81

Carnes G, Gaede S, Yu E, Van Dyke J, Battista J and Lee T-Y 2009 A fully automated non-external marker 4D-CT sorting algorithm using a serial cine scanning protocol *Phys. Med. Biol.* **54** 2049–66

Colgan R, McClelland J R, McQuaid D, Evans P M, Hawkes D J, Brock J, Landau D and Webb S 2008 Planning lung radiotherapy using 4D CT data and a motion model *Phys. Med. Biol.* **53** 5815–30

Guckenberger M, Meyer J, Wilbert J, Richter A, Baier K, Muller G and Flentje M 2007a Intra-fractional uncertainties in cone-beam CT based image-guided radiotherapy (IGRT) of pulmonary tumors *Radiother. Oncol.* **83** 57–64

Guckenberger M, Wilbert J, Meyer J, Baier K, Richter A and Flentje M 2007b Is a single respiratory correlated 4D-CT study sufficient for evaluation of breathing motion? *Int. J. Radiat. Oncol. Biol. Phys.* **67** 1352–9

Hoisak J D P, Sixel K E, Tirona R, Cheung P C F and Pignol J-P 2004 Correlation of lung tumour motion with external surrogate indicators of respiration *Int. J. Radiat. Oncol. Biol. Phys.* **60** 1298–306

Hughes S, McClelland J, Tarte S, Lawrence D, Ahmad S, Hawkes D and Landau D 2009 Assessment of two novel ventilatory surrogates for use in the delivery of gated/tracked radiotherapy for non-small cell lung cancer *Radiother. Oncol.* **91** 336–41

Korreman S, Mostafavi H, Le Q-T and Boyer A 2006 Comparison of respiratory surrogates for gated lung radiotherapy without internal fiducials *Acta Oncol.* **45** 935–42

Langner U W and Keall P J 2008 Prospective displacement and velocity-based cine 4D CT *Med. Phys.* **35** 4501–12

Li T, Schreibmann E, Thorndyke B, Tillman G, Boyer A, Koong A, Goodman K and Xing L 2005 Radiation dose reduction in four-dimensional computed tomography *Med. Phys.* **32** 3650–60

Li T, Thorndyke B, Schreibmann E, Yang Y and Xing L 2006 Model-based image reconstruction for four-dimensional PET *Med. Phys.* **33** 1288–98

Low D A, Parikh P J, Lu W, Dempsey J F, Wahab S H, Hubenschmidt J P, Nystrom M M, Handoko M and Bradley J D 2005 Novel breathing motion model for radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **63** 921–9

Lu W, Low D A, Parikh P J, Nystrom M M, El Naqa I M, Wahab S H, Handoko M, Fooshee D and Bradley J D 2005 Comparison of spirometry and abdominal height as four-dimensional computed tomography metrics in lung *Med. Phys.* **32** 2351–7

Lu W, Parikh P J, Hubenschmidt J P, Bradley J D and Low D A 2006 A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4D CT *Med. Phys.* **33** 2964–74

McClelland J R, Blackall J M, Tarte S, Chandler A C, Hughes S, Ahmad S, Landau D B and Hawkes D J 2006 A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy *Med. Phys.* **33** 3348–58

McClelland J R, Webb S, Binnie D M and Hawkes D J 2007 Tracking 'differential organ motion' with a 'breathing' multileaf collimator: magnitude of problem assessed using 4D CT data and a motion-compensation strategy *Phys. Med. Biol.* **52** 4805–26

Modat M, McClelland J and Ourselin S 2010a Lung registration using the NiftyReg package *Medical Image Analysis for the Clinic: A Grand Challenge, Workshop Proc. from MICCAI 2010* pp 33–42

Modat M, Ridgway G R, Taylor Z A, Lehmann M, Barnes J, Fox N C, Hawkes D J and Ourselin S 2010b Fast free-form deformation using graphics processing units *Comput. Methods Programs Biomed.* **98** 278–84

Odille F, Vuissoz P A, Marie P Y and Felblinger J 2008 Generalized reconstruction by inversion of coupled systems (GRICS) applied to free-breathing MRI *Magn. Reson. Med.* **60** 146–57

Purdie T G, Bissonnette J-P, Franks K, Bezjak A, Payne D, Sie F, Sharpe M B and Jaffray D A 2007 Cone-beam computed tomography for on-line image guidance of lung stereotactic radiotherapy: localization, verification, and intrafraction tumor position *Int. J. Radiat. Oncol. Biol. Phys.* **68** 243–52

Redmond K J, Song D Y, Fox J L, Zhou J, Rosenzweig C N and Ford E 2009 Respiratory motion changes of lung tumours over the course of radiation therapy based on respiratory-correlated four-dimensional computed tomography scans *Int. J. Radiat. Oncol. Biol. Phys.* **75** 1605–12

Rit S, Wolthaus J W H, van Herk M and Sonke J-J 2009 On-the-fly motion-compensated cone-beam CT using an a priori model of the respiratory motion *Med. Phys.* **36** 2283–96

Rueckert D, Sonoda L I, Hayes C, Hill D L, Leach M O and Hawkes D J 1999 Nonrigid registration using free-form deformations: application to breast MR image *IEEE Trans. Med. Imaging* **18** 712–72

Seppenwoolde Y, Berbeco R I, Nishioka S, Shirato H and Heijmen B 2007 Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: a simulation study *Med. Phys.* **34** 2774–84

Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, van Herk M, Lebesque J V and Miyasaka K 2002 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.* **53** 822–34

Sonke J-J, Lebesque J and van Herk M 2008 Variability of four-dimensional computed tomography patient models *Int. J. Radiat. Oncol. Biol. Phys.* **70** 590–8

Sonke J-J, Rossi M, Wolthaus J, van Herk M, Damen E and Belderbos J 2009 Frameless stereotactic body radiotherapy for lung cancer using four-dimensional cone beam CT guidance *Int. J. Radiat. Oncol. Biol. Phys.* **74** 567–74

Vandemeulebroucke J, Bernard O, Kybic J, Clarysse P and Sarrut S 2010 Automatic motion mask extraction for deformable registration of the lungs *Proc. XVIth Int. Conf. on the Use of Computers in Radiation Therapy (ICCR) (Amsterdam, 2010)*

Webb S 2008 Adapting IMRT delivery fraction-by-fraction to cater for variable intrafraction motion *Phys. Med. Biol.* **53** 1–21

Wolthaus J W H, Sonke J J, van Herk M and Damen E M F 2008 Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registration *Med. Phys.* **35** 3998–4011

Wu Z, Rietzel E, Boldea V, Sarrut D and Sharp G C 2008 Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations *Med. Phys.* **35** 775–81

Yang D, Lu W, Low D A, Deasy J O, Hope A J and El Naqa I M 2008 4D-CT motion estimation using deformable image registration and 5D respiratory motion modeling *Med. Phys.* **35** 4577–90

Zhang Q, Pevsner A, Hertanto A and Hu Y-C 2007 A patient-specific respiratory model of anatomical motion for radiation treatment planning *Med. Phys.* **34** 4772–81