IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **53** (2008) 4317–4330 doi:10.1088/0031-9155/53/16/007

A margin model to account for respiration-induced tumour motion and its variability

Catherine Coolens[1], Steve Webb[1], H Shirato[2], K Nishioka[2] and Phil M Evans[1]

[1] Joint Department of Physics, Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT, UK
[2] Department of Radiology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

E-mail: Catherine.coolens@rmp.uhn.on.ca

Received 18 March 2008, in final form 29 May 2008
Published 24 July 2008
Online at stacks.iop.org/PMB/53/4317

Abstract

In order to reduce the sensitivity of radiotherapy treatments to organ motion, compensation methods are being investigated such as gating of treatment delivery, tracking of tumour position, 4D scanning and planning of the treatment, etc. An outstanding problem that would occur with all these methods is the assumption that breathing motion is reproducible throughout the planning and delivery process of treatment. This is obviously not a realistic assumption and is one that will introduce errors. A *dynamic internal margin model* (DIM) is presented that is designed to follow the tumour trajectory and account for the variability in respiratory motion. The model statistically describes the variation of the breathing cycle over time, i.e. the uncertainty in motion amplitude and phase reproducibility, in a polar coordinate system from which margins can be derived. This allows accounting for an additional gating window parameter for gated treatment delivery as well as minimizing the area of normal tissue irradiated. The model was illustrated with abdominal motion for a patient with liver cancer and tested with internal 3D lung tumour trajectories. The results confirm that the respiratory phases around exhale are most reproducible and have the smallest variation in motion amplitude and phase (approximately 2 mm). More importantly, the margin area covering normal tissue is significantly reduced by using trajectory-specific margins (as opposed to conventional margins) as the angular component is by far the largest contributor to the margin area. The statistical approach to margin calculation, in addition, offers the possibility for advanced online verification and updating of breathing variation as more data become available.

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

0031-9155/08/164317+14$30.00

1. Introduction

Studies based on daily electronic portal imaging and repeated fluoroscopy or CT imaging show that margin reduction, to the level required for dose escalation, cannot accommodate the errors of treatment set-up and organ motion (Little *et al* 2003, Ten Haken *et al* 1997, Stroom and Heijmen 2002, Hugo *et al* 2007). To successfully create a meaningful margin reduction in the presence of organ motion, one therefore may have to turn to efficient gating of the treatment delivery and/or real-time tracking of the target position.

Motion-gating techniques minimize the range of target motion during irradiation by limiting the delivery to pre-defined periods (duty cycle). The duration of irradiation is usually arranged to be that with the least variation in target position and/or maximal sparing of the surrounding normal tissue. The more technologically advanced approach of tumour tracking is associated with the idea of online synchronization of the treatment delivery to the tumour motion either by closed-loop tracking (i.e. observing the tumour excursion in real-time) or by open-loop tracking (i.e. using prediction algorithms from a baseline position model). Although closed-loop tracking could in theory track an arbitrary trajectory given a margin for feedback lag, these methods still have one common difficulty, which is the presupposition that breathing is reproducible throughout the planning and delivery process of treatment.

Breathing motion is not a robust and 100% reproducible process (Gierga *et al* 2004). In addition, it is easily imagined that the patient, when positioned on the treatment couch, may breathe very differently from imaged for planning due to discomfort and/or anxiety (Shirato *et al* 2004) or other physiological reasons. The role of prediction algorithms to describe target motion has been discussed in this context (Sharp *et al* 2004, Vedam *et al* 2004) but it has not been proven that any prediction algorithm can be robust enough to account for this variability in breathing motion. Furthermore, the idea of performing 4D optimization has been suggested to compensate for organ motion (Trofimov *et al* 2004), which involves the use of modified pencil beam kernels or time-weighted influence matrices to compensate for organ motion. Such an approach will increase the modulation gradients within intensity-modulated beams if not actively restricted, and will therefore be potentially even more prone to the variations in breathing motion. Whatever method is ultimately used for compensating for organ motion, the *variation* of breathing motion is unlikely to disappear for a free-breathing patient and a small margin will always be needed to provide compensation, even if all other treatment uncertainties are ignored. This also applies to 3D conformal treatments.

The variation in the breathing cycle has been commented on before (Vedam *et al* 2004, Zhang *et al* 2004, Nehmeh *et al* 2004, Seppenwoolde *et al* 2002, Ruan *et al* 2008), but never has this irreproducibility been the main subject of investigation with the purpose of deducing and evaluating an appropriate margin size to account for tumour motion uncertainty. This new *margin model* concept (Coolens *et al* 2005) is designed to follow the tumour trajectory and statistically addresses variations in breathing motion, which occur both during and between fractions. The margin for set-up errors was not considered as the model provides a safety margin that follows the tumour path and effectively constitutes the ITV margin (ICRU-62), which needs to be added to the former to compose the PTV from a clinical target volume.

The aim of this paper is therefore

(1) to present the model for calculating a dynamic internal margin (DIM), based on the observed variation in the breathing cycle and
(2) to describe the usefulness of the model in providing clinically relevant margins which are robust against variations in motion-induced target displacements.

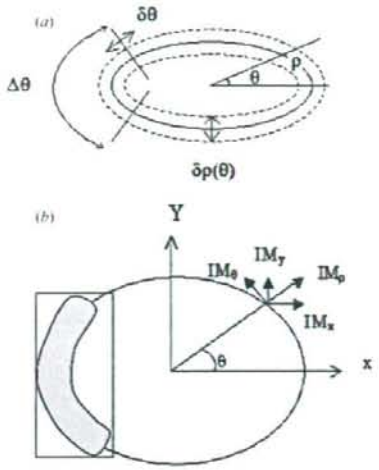

Figure 1. (a) Illustration of the several components that describe the variation in breathing-induced target position. The solid ellipse represents the average breathing motion with hysteresis (described by the polar angle, θ, and polar length, ρ). The two dotted ellipses represent the maximum deviation from this average. When gating the treatment, a gating window, $\Delta\theta$, will be used. For both tracking and gating there is an uncertainty in phase, $\delta\theta$, and a variation in target position per phase, $\delta\rho(\theta)$. (b) Illustration of the polar (i.e. along the curved trajectory) and Cartesian (i.e. rectangular) margin components and the influence on margin size.

2. Method

It has been shown that the respiration-induced motion of lung and liver tumours can display a strong hysteresis effect (Shirato *et al* 2004, Seppenwoolde *et al* 2001), i.e. the breathing forces governing inhalation and exhalation are different. As a consequence, the tumour displacements from inhale to exhale and vice versa do not follow the same path, which results in elliptically shaped tumour trajectories. It will be shown in this study that the hysteresis effect can be exploited to define ITV margins that follow the tumour trajectory and therefore minimize the amount of healthy tissue that otherwise would be irradiated. In addition, this approach allows compensating for the variation in breathing motion over time. An illustration of an elliptical motion trajectory subject to hysteresis is shown in figure 1. This figure will be used to describe the dynamic internal margin model (section 2.1). In the second part of the paper, the model will be applied to data of lung patients to (i) illustrate the effect on margin size and reduction in normal tissue irradiation and (ii) explain how such a dynamic margin model can work in practice for providing safe tumour tracking or gating in the presence of breathing variability.

2.1. Development of the DIM model

Considering the illustration in figure 1, if the patient would breathe perfectly reproducibly the target trajectory would follow an ellipse with zero width (the solid line). In practice though, one revolution following the path of the breathing cycle could take a variable time, depending

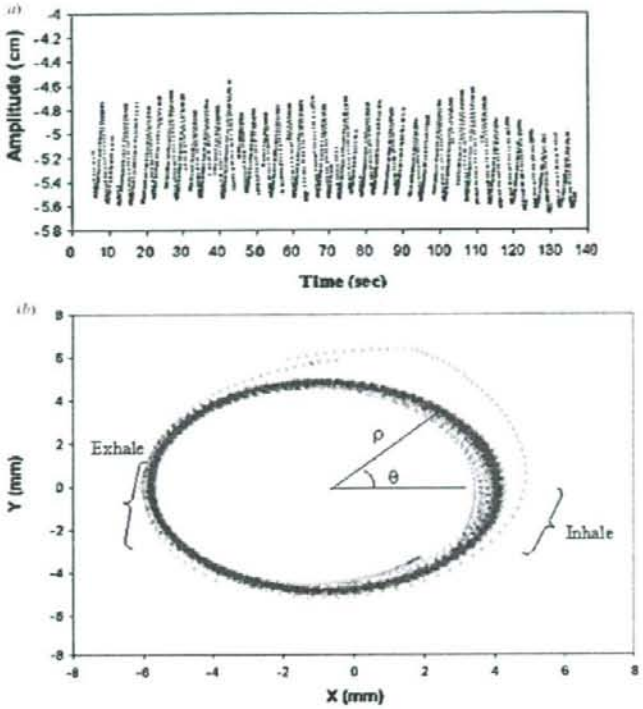

Figure 2. (a) Ant–Post displacement of a liver patient's diaphragm with time. (b) Representation of the data in a polar coordinate system by considering the amplitude as a 2D vector within a Cartesian coordinate system. The inhale phases correspond to the minimum amplitude points in (a). The polar coordinate system is described by the polar angle θ and polar length ρ (= amplitude).

on the period of the specific breathing cycle T. The variation in breathing period over time manifests itself in a certain width in motion amplitude with respiratory phase. This can be seen clearly in the patient data (figures 2 and 3) and has also previously been reported (Shirato *et al* 2004, Nehmeh *et al* 2004). The variation in the length of the semi-major and semi-minor axes of the elliptical trajectory shown in figure 1 indicates the 2D variation in target position due to breathing-induced organ motion. The area between the two dotted ellipses illustrates the motion limits that would contain the measured data points (e.g. 2 standard deviations) around the average breathing motion (solid ellipse). In the case where one is able to track or gate the target, this variation constitutes precisely the required internal margin (IM) for breathing variability. However, considering the limits in tracking accuracy, there will also be some phase uncertainty that has to be accounted for due to time lag and system response. The presented margin model described here treats all these variations statistically to provide confidence limits for variations in breathing from cycle-to-cycle and from fraction-to-fraction. The required measurements to derive the DIM margins for planning could be made at the time of simulation,

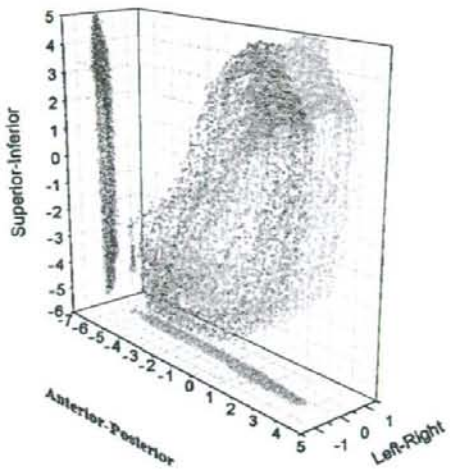

Figure 3. Lung tumour trajectory in 3D and projected onto the three main planes as acquired on day 1, referred to as session 1_1.

i.e. before treatment delivery (Gierga *et al* 2004, Vedam *et al* 2003, Ozhasoglu and Murphy 2002), and updated just before and/or during treatment as extra motion information becomes available.

2.1.1. DIM definitions. The various components involved in the DIM model are also indicated in figure 1. The range in target position at a certain phase θ is given by $\delta\rho(\theta)$ (where (θ) denotes a function of θ), which is twice the standard deviation, σ_ρ, from the average target position so as to give 95% confidence. To account for the fact that the breathing period is irregular (i.e. not constant), the reproducibility in arriving at the same phase as a function of time is defined as $\delta\theta$. This depends on the accuracy of the positional measurements as well as the stability of the breathing period. Similarly, this is determined as twice the standard deviation of the variation in phase angle with time, σ_θ. When gating treatment, the phase window, $\Delta\theta$, to which the treatment is confined, will describe a different variation in target position depending on what range of phases it encompasses. The latter is also referred to as the duty cycle, which is the fractional time of treatment for which the radiation is switched on. In summary, for *tracking* target motion, the envelope of the elliptical trajectories (call this the error or margin area) is therefore determined by $(\delta\rho, \delta\theta)$. When *gating* according to a certain duty cycle, the error area is determined by $(\delta\rho, \Delta\theta + \delta\theta)$. In the first instance the model was developed in 2D. An expansion to a 3D situation will be discussed later, but the only real difference this makes is the expansion from an error area to an error volume. Also an error area may be used in the plane of any treatment beam within good dosimetric accuracy.

2.1.2. DIM model. The model parametrizes both the radial and angular coordinates of tumour motion. The several components of the model (described in section 2.1.1) were designed to

split the angular uncertainty in two parts and therefore allow the margins to be shaped along the tumour trajectory. In the polar coordinate system with unit vectors ($\mathbf{e}_\rho$, $\mathbf{e}_\theta$) the target position at a certain time t_0 can be written as a vector $\mathbf{r} = \rho \cdot \mathbf{e}_\rho$, with ρ the polar length (figure 1). For the situation here, the target positioning error, caused by a variation in motion amplitude $\delta\rho$ and uncertainty in phase, $\delta\theta$, is given by

$$\mathrm{d}\mathbf{r} = \delta\rho \cdot \mathbf{e}_\rho + \rho \cdot \delta\theta \cdot \mathbf{e}_\theta. \tag{1}$$

For gating, a phase-dependence must be included (depending on the gating window):

$$\Theta = \Delta\theta + \delta\theta. \tag{2}$$

The two parameters that define the error area thus are

$$\boxed{\mathrm{IM}_\rho = \delta\rho, \qquad \mathrm{IM}_\theta = \rho \cdot \Theta.} \tag{3}$$

It is possible to express these parameters in the Cartesian coordinates by projecting them onto the x- and y-axis. The 2D Cartesian components to the IM can then be written as

$$\boxed{\begin{aligned} \mathrm{IM}_x &= \delta\rho(\Theta) \cdot \cos\theta - \rho \cdot \Theta \cdot \sin\theta, \\ \mathrm{IM}_y &= \delta\rho(\Theta) \cdot \sin\theta + \rho \cdot \Theta \cdot \cos\theta. \end{aligned}} \tag{4}$$

Both the polar and Cartesian margin components are illustrated in figure 2(b).

2.2. *Application to patient data*

2.2.1. Abdominal motion. To illustrate the model characteristics, real, i.e. irregular, motion data are used. Figure 2(a) shows the anterior–posterior displacement of an infrared marker placed on the abdominal surface of a liver patient as detected by the RPM system (Varian, PA). This one-dimensional data can be transformed into a 2D beam's eye-view motion of a tumour by considering the Ant–Post amplitude displacement as a Cartesian vector and projecting it onto its orthogonal X- and Y-axes with phase ($X = A\cos\theta$, $Y = A\sin\theta$, with A the amplitude and θ the phase at that point). Although this is not a real internal tumour trajectory, it will be assumed to be one for the purpose of portraying the model characteristics in the presence of breathing irregularities and elliptical hysteresis. The positions of inhale and exhale are indicated in figure 2(b). The variation in breathing amplitude is clearly visible through the width of the ellipse-like trajectory. In addition, it can be seen that this width changes with phase due to variations in amplitude. This is important when gating treatment delivery to a certain phase window or duty cycle, as the margin may be phase dependent. Furthermore, a clear deviation from the average breathing trend can be seen in the upper right-hand corner of the plot. This large deviation could be due to a sudden change in breathing pattern, when the patient coughs, for example, or it could constitute the sampling of extremes of a stochastic distribution. This illustrates that having an unexpected change in breathing pattern would be, at best, very hard to account for with motion prediction models and be potentially dangerous to ignore in practice.

2.2.1.1. Variation in target position and respiratory cycle. To calculate the variation in target position at a certain respiratory phase, an ellipse was fitted to the data in figure 2(b) by performing a least-squares minimization of the five parameters that describe it. These parameters were semi-axes a and b, together with the coordinates (x_0, y_0) that describe the origin of the ellipse and a tilt in the ellipse's major and minor axes, ϕ, with respect to the

Cartesian coordinate system. The points on the fitted ellipse were transformed into x- and y-coordinates with the following transformation:

$$\begin{cases} x = x_0 + \cos\phi \cdot \dfrac{x'}{a} - \sin\phi \cdot \dfrac{y'}{b}, \\ y = y_0 + \sin\phi \cdot \dfrac{x'}{a} + \cos\phi \cdot \dfrac{y'}{b}. \end{cases} \tag{5}$$

The variation in motion amplitude, σ_ρ, can then be calculated as the difference between the amplitude at the measured data point (x_i, y_i) and the average amplitude at the point (x, y) with corresponding phase, as derived from the elliptical fit:

$$\sigma_\rho(\vartheta) = \sqrt{x_i^2 + y_i^2} - \sqrt{x^2 + y^2} \qquad \text{with} \qquad \frac{x}{y} = \frac{x_i}{y_i}. \tag{6}$$

The variation in respiratory cycle, i.e. the temporal variation in phase was analysed as follows. First, the running average breathing time, $T_{av,i}$, was determined as the average of the time between the phase midpoints (2π apart), over the last 'i' number of consecutive breathing cycles, i and, with n being the total number of cycles here equal to 25:

$$T_{av,i} = \sum_{k=1}^{i} \tau_i / n. \tag{7}$$

This is done to mimic the online tracking approach where all available data up and until that point are used. Then, the angular variation, σ_θ, i.e. the difference in phase at a certain time t and the time $t + T_{av,i}$, was calculated. For these data, there were 30 phase samples per second and a total study time of 196 s. The variation in respiratory cycle, as illustrated in figure 1, is then defined as

$$\delta\theta = 2\sigma_\theta \tag{8}$$

with the use of 2 standard deviations to give the 95% confidence limit.

2.2.1.2. Assessment of DIM margins versus conventional margins. Making a comparison of the error area obtained from the polar and Cartesian margin components assessed the effect of using the dynamic margin approach. The components IM_x and IM_y were calculated from the previously obtained polar and angular margin components according to equation (4). The area covered by the margins is then given by $2IM_x \times 2IM_y$ (Cartesian) or $2IM_\rho \times 2IM_\theta$ (polar). As the dynamic margins are designed to follow the trajectory of motion, it would be expected that they reduce the amount of normal tissue irradiated compared to conventional margins that consider the maximum extent of motion. Note that this model addresses variations in breathing motion, which occur both during and between fractions. In practice, there will be two types of measurements related to motion. The first one involves fluoroscopy or cone beam CT (CBCT) to estimate the mean tumour position trajectory over time. The second component will address the variability of motion around this mean position to calculate the margin components, as described by the model.

2.2.2. Lung tumour internal trajectories. Data for this part of the study consisted of real-time 3D lung tumour position coordinates from two patients on two treatment days. For ease of use, they will be further referred to as session 1_1 (day 1) and session 1_2 (day 2). Similarly, data for patient 2 will be referenced as session 2_1 (day 1) and session 2_2 (day 2). The target was represented by a single 2 mm gold marker that had been implanted in the tumour and was tracked over a period of up to 3.3 min under fluoroscopy at a frequency of 30 Hz (Seppenwoolde *et al* 2002, Shirato *et al* 2000). Figure 3 illustrates the tumour trajectory in session 1_1.

2.2.2.1. Lung tumour margins. Given the minimal variation in tumour position in the left–right (LR) direction for these particular patients, attention was focused in the first instance on the lateral beam's eye-view in which the more typical hysteresis is visible. Similarly to the margin derivation in section 2.2.1, the margins are derived by first obtaining the mean tumour trajectory, followed by calculating the 95% confidence levels around that mean. However, figure 5 illustrates that to find the mean tumour trajectory a more flexible fitting process might be needed than a purely elliptical path.

The mean trajectory was therefore estimated using a moving average calculation around the centre of mass of the entire data set. As the trajectory is a function of time or breathing phase, the data were sub-sampled in different breathing phases or time intervals. Within each sub-sample (i.e. data bin) a least-squares fit determined the mean positions. The number of sub-sample bins associated with the averaging was determined by balancing the smoothness of fit with the fit residual error.

To calculate the phase component of the margin, it was needed to compare individual phase points to the mean phase calculation. As described in section 2.2.1.1 a running average of the breathing period was calculated (see also equation (8)) to represent the average phase trajectory. The DIM components were then calculated as before through the statistical analysis of the breathing variation around the mean in both amplitude and phase. If there is little, or no, hysteresis, such as, e.g., in the LR direction in figure 3, then the DIM margins may still be along the main axis of motion, with the phase margin set equal to the width of the phase bin. This technique was applied if the range of motion in one direction was smaller than 1.5 mm (see discussion). This situation will be referred to further on in the results section as a 'low' hysteresis trajectory, as opposed to a 'high' hysteresis where the tumour clearly rotates round the centre-of-mass.

2.2.2.2. Margin robustness. To test the model robustness, the variation in tumour motion on day 2 was compared to that for day 1 by overlaying the mean tumour trajectories on top of each other with the associated DIM margins. The aim was to mimic a potential clinical implementation by which the margins and mean trajectory would be determined offline from, e.g., simulator fluoroscopy and/or 4DCT data. Then, the second trace was used as the 'treatment' trace for which an offline or online scenario would be possible. The offline approach could involve, e.g., set-up imaging with CBCT or planar kV in the required beam's eye-view to establish the mean trajectory is (a) still similar to the one used for planning and (b) assess and recalculate the margins on the day for comparison with the plan margins. In an online tracking protocol, it could be envisaged to have the system track and predict not only the tumour position but calculate the margins with increasingly higher statistical significance as data are being acquired.

3. Results

3.1. DIM model characteristics

To illustrate some of the model characteristics, the average target motion in figure 2(b) was estimated by fitting an ellipse to the data (optimal parameters: $a = 4.9$ mm, $b = 4.9$ mm, $x_0 = -0.3$ mm, $y_0 = -0.2$ mm and $\phi = -0.45$ rad). The combination of both polar and angular errors is plotted in figure 4 with the measured data points (only a subset of the points is shown for clarity). This shows the error area as a function of phase in the breathing cycle. As exhale has proven to be the most stable phase in the breathing cycle, the IM components have been calculated for a number of gating window sizes around this phase (table 1). As expected,

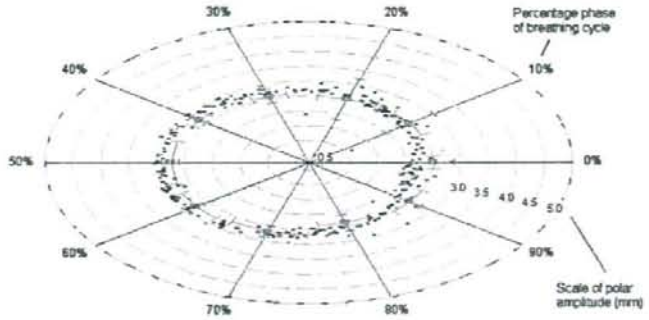

Figure 4. Plot showing the polar (pink) and angular (green) IM for the measured breathing motion (see figure 2) per percentage of the average breathing cycle. Only a subset of the points is shown, for clarity. The internal concentric circles represent the amplitude (of AP displacement) scale of 0.5 mm.

Table 1. Table demonstrating the differences in margin area and internal margins, in polar (IM_ρ) and angular (IM_θ) distance (equation (4)), for different duty cycles around the exhale phase for the data in figure 2(b).

Duty cycle (%)	0	20	30	40	50
IM_ρ (mm)	0.2	0.2	0.3	0.4	0.4
IM_θ (mm)	0.4	4.0	5.6	7.1	8.6
IM_x (mm)	0.4	2.5	3.5	4.4	5.4
IM_y (mm)	0.3	3.1	4.3	5.5	6.7
Polar area (mm^2)	0.3	3.2	6.0	9.9	13.4
Cartesian area (mm^2)	0.4	31.3	61.4	98.2	144.5

the margin size increases significantly with increasing duty cycle (Hugo *et al* 2002). The component in the angular direction is the most crucial contributor to the size of the margin. In addition, a comparison of the Cartesian and polar margin area is given. The error area covered by the polar margins is smaller than by the Cartesian components. This difference becomes even more prominent as the gating window size increases and indicates the benefits of using the DIM model in sparing normal tissues from high-dose irradiation.

3.2. Dynamic margins for lung tumour internal trajectories

3.2.1. High hysteresis. The lateral projections for all four sessions are shown in figure 5 together with the best-fit ellipse, the average trajectory and the bin-width from the centre-of-mass. In this case, however, the number of bins used to average the trajectory was determined by balancing the smoothness of fit with the fit residual error and was chosen to be equal to 40. This corresponds to almost 10 bins per second, given that the average breathing period was approximately 3.6 s (patient 1) and 3.8 s (patient 2).

The associated DIM margins for all four sessions are summarized in table 2 and were calculated on the full extent of the motion to illustrate the total range of motion. The DIM components for different sessions show good agreement, with the exception of the amplitude error in session 2_2. This discrepancy is due to the drift in the tumour trajectory over time. A comparison of the Cartesian and DIM margin sizes (using the margin components of the

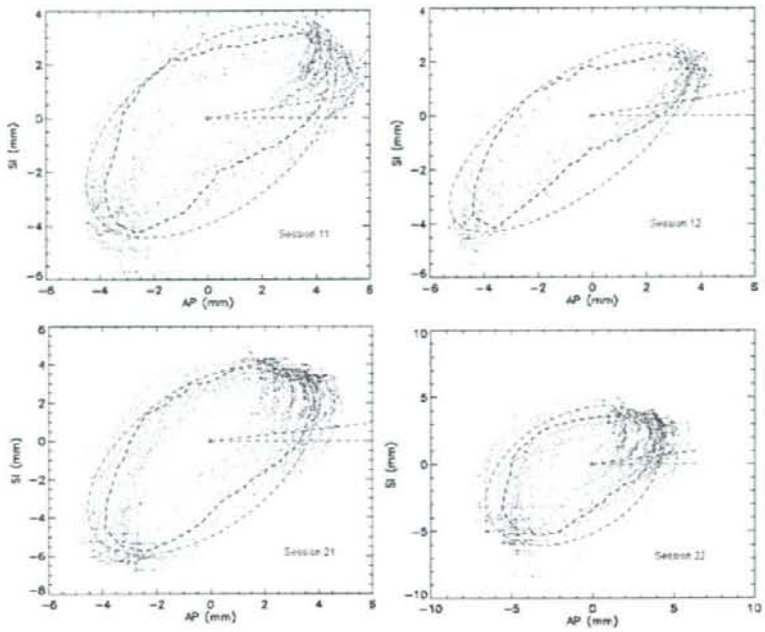

Figure 5. Lateral projection of tumour motion during all sessions. The (red) dotted line represents the average tumour trajectory calculated over a number of bins, the first of which is shown as radiating out from the centre-of-mass (in blue). The green/light grey line shows the best-fit ellipse to the data.

Table 2. Table listing the internal margins in both polar (IM_ρ) and angular (IM_θ) distance calculated over the entire lung tumour trajectory for the different data sessions, together with the polar and Cartesian error areas from these margins at the exhale phase (i.e. tracking of the target).

Session no	IM_ρ (mm)	IM_θ (mm)	Polar area (mm^2)	Cartesian area (mm^2)
1_1	1.8	2.0	14.4	12.6
1_2	1.7	2.0	13.6	12.3
2_1	1.8	2.0	13.6	12.3
2_2	2.4	1.9	18.2	12.3

entire trace but applied at exhale) is also given in table 2 and indicates (1) that in the case of tumour tracking, i.e. small margin values, the Cartesian margins are relatively insensitive to change in tumour amplitude and drifting of the tumour trajectory and (2) the margins are phase dependent as the variation in phase/amplitude at exhale is smaller than the total variation.

Table 3 lists the different margin components for a gated treatment with the impact of using different duty cycles. In this case, the reduction in margin area (and hence normal tissue irradiation) when using the polar margins is significant compared to the Cartesian standard margins.

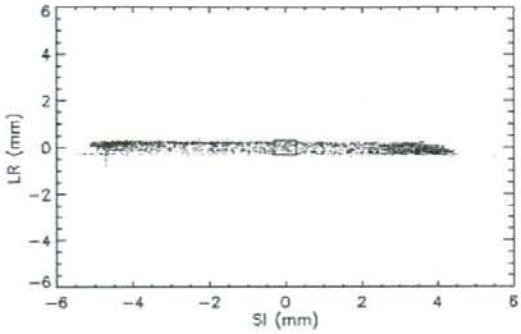

Figure 6. Anterior view of the tumour motion trajectory for session 1_1. The centre-of-mass is marked by a cross (in red). The margin area in that phase bin is indicated by the rectangle and would move along the SI axis with time between inhale and exhale.

Table 3. Table demonstrating the differences in margin area and internal margins, in polar (IM_ρ) and angular (IM_θ) distance (equation (4)), for different duty cycles around the exhale phase for session 1_1 in figure 5.

Duty cycle (%)	0	20	30	40	50
IM_ρ (mm)	1.6	1.7	1.7	1.8	1.8
IM_θ (mm)	2.0	4.0	5.3	6.7	9.2
IM_x (mm)	2.2	2.9	3.3	3.8	4.5
IM_y (mm)	1.4	3.2	4.5	5.8	8.2
Polar area (mm^2)	13.1	27.2	36.9	48.9	65.2
Cartesian area (mm^2)	12.2	36.9	59.3	87.5	147.2

3.2.2. Low hysteresis. The anterior BEV of session 1_1 is shown in figure 6 and displays a 'low' hysteresis trajectory. For this situation, the margins were calculated for ten phase bins (as per standard 4DCT). The amplitude margin was determined to be twice the standard deviation from the centre-of-mass for all data points in the LR direction, resulting in IM_ρ = 0.35 mm. The phase margin was calculated as before from the variation of the running time average and was, on average, equal to 2.2 mm. The resulting DIM margin area therefore was 1.5 mm^2 per phase bin for which the target would be irradiated.

3.2.3. Robustness. To assess the potential of applying the DIM margins from day 1 to account for the variability in breathing in further treatments, the full tumour trajectories of sessions 1_1 and 1_2 are overlaid in figure 7(a). It can be seen that the lung tumour motion follows a similar trajectory with time for both sessions. The DIM margins are shown in figure 7(b) as error bars on top of the average tumour trajectory and cover most of the breathing variability in session 1_2. The mean trajectories show very good similarity but, as discussed earlier, due to the drifting in session 2_2, the mean tumour path appears larger than in sessions 2_1 and potentially a little rotated as well. The drifting of the tumour trajectory with time can then also be a confounding factor in determining the required margins and this means that the 95% confidence levels were not sufficient to provide margins in the event of drifting. Note that the 99% confidence limits are only just covering the mean trajectory.

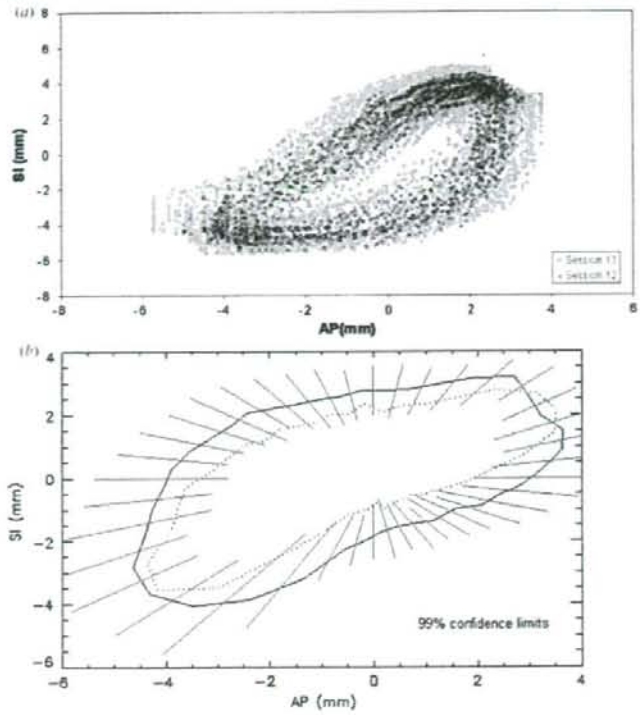

Figure 7. (a) Tumour trajectories in AP–SI view for patient 1. (b) Average tumour trajectories for this patient (session 1_1 as solid line; session 1_2 as dotted line) centred on the original COM coordinates with margin bars from session 1_1.

More advanced statistics and more representative samples of inter-fraction variability (as data become available) should improve this significantly.

4. Discussion

The DIM model was developed for tracking or gating of the treatment delivery, as it is convenient to have a margin that follows the tumour trajectory (i.e. lies along the gating window) so as to maximally spare normal tissues. The internal margin components were calculated in a polar coordinate system, allowing the motion margin to have an axis in the motion direction. The Cartesian components of the IM could be used but this will create a rectangular margin that will in general be larger than the corresponding polar segment. In the case of low or no hysteresis the differences between the Cartesian and DIM margins would be minimal, but applying margins that stretch along the line of motion will provide extra robustness and flexibility to account for potential changes during treatment and smaller margins. Once the statistical margins have been established from pre-planning measurements,

the model could be applied at the treatment stage to verify that the motion uncertainty lies within the planned margins. This could be done either just before beam-on or, in a more advanced situation, even during irradiation with real-time feedback. The flexibility of the DIM model means that data can be readily added as it becomes available during treatment to update the motion statistics and hence margins for an individual patient.

As yet the DIM model was developed in 2D. An extension of the DIM model to a 3D situation would be possible and would need the description of the model components in the general curvilinear coordinates (q_1, q_2, q_3) rather than the polar cylindrical coordinates (ρ, θ, z). The resulting error volume will resemble a (possibly deformed) torus. However, taking fluoroscopy measurements in two orthogonal fields of view could provide information for the incorporation of the third dimension. Figure 3 also indicates that it might not be needed to extend the model to 3D if, e.g., the LR motion is inherently small. In addition, the beam angles could be chosen so that their BEV projection minimizes the tumour motion.

The total number of phase bins (and hence the smallest phase bin size) would depend on the tracking system's feedback accuracy. Similarly, the cut-off value for when to decide on a 'low' or 'high' hysteresis approach will depend on the spatial accuracy of the tracking system as well as the size of the fiducial marker and tumour being tracked and the clinical relevance of such a displacement. Currently, the model can only be accurately applied when there is no drift present that extends beyond the centre-of-mass of the average trajectory. It seems unlikely that breathing coaching could counteract or help alleviate this problem but the drifting could also be incorporated into the DIM model, which will be the topic of further study through different statistics and modelling of the mean trajectory.

Despite the increasing knowledge about lung tumour motion, accurate data on 3D real-time tumour trajectories are still limited and mostly acquired through kV tracking. Practically, the implementation of a DIM model would require internal motion information (e.g. from fluoroscopy), at least until the relationship between internal and external motion can be quantified accurately (Ozhasoglu and Murphy 2002). Alternatively, electro-magnetic systems with implantable transponders hold promise to provide this real-time information (Schweikard *et al* 2000).

The irreproducibility of the breathing cycle is important not only for gating or tracking of the target but also when considering incorporating organ motion into leaf sequencing files. Some researchers have commented on this motion-compensation method with the aim of planning treatment on 4D CT data and creating a leaf segment database from which the MLC controller can pick a segment according to the phase of breathing (Vedam *et al* 2004, Keall 2004). Whether or not this is a technically feasible technique remains to be proven, but it certainly does emphasize the need for reproducible and accurate information on breathing-induced motion and appropriate margins, regardless of what motion compensation method is ultimately used.

5. Conclusion

Attention was turned in this paper to organ motion compensation methods such as gating of treatment delivery and tracking of target position. An issue that has mostly been ignored in the development of these techniques is the fact that breathing motion is not a perfectly reproducible process. The aim of this study was therefore to present a new model that calculates a dynamic internal motion margin, based on the observed variation in the breathing cycle and which minimizes the volume of normal tissue irradiated by shaping them along the tumour trajectory. It determines an error area in polar coordinates based on the hysteresis variation of the breathing cycle over time, i.e. the statistical uncertainty in motion amplitude

and phase reproducibility. This was then developed further for application to real lung tumour trajectories. It was shown that this model proved to be successful in reducing margins for lung tumour motion compared to conventional margins. The phases around exhale were found to have the smallest variation in both phase and amplitude. The phase component dominates the margin size and a larger margin would be needed if a gating window were used. It was shown that the model significantly reduces normal tissue irradiation in this situation and offers the possibility for advanced online verification of breathing motion.

Acknowledgment

The work of the Joint Department of Physics is supported by a programme grant from Cancer Research UK (CR-UK) (under reference grant SP 2312/0201).

References

Coolens C, Webb S and Evans P 2005 A dynamic margin model to account for organ motion variability *Rad. Oncol.* **76** (Suppl. 2) S183

Gierga D P *et al* 2004 Quantification of respiration-induced abdominal tumor motion and its impact on IMRT dose distributions *Int. J. Radiat. Oncol. Biol. Phys.* **58** 1584–95

Hugo G D, Agazaryan N and Solberg T D 2002 An evaluation of gating window size, delivery method, and composite field dosimetry of respiratory-gated IMRT *Med. Phys.* **29** 2517–25

Hugo G D, Yan D and Liang J 2007 Population and patient-specific target margins for 4D adaptive radiotherapy to account for intra- and inter-fraction variation in lung tumour position *Phys. Med. Biol.* **52** 257–74

Keall P J 2004 4-Dimensional computed tomography imaging and treatment planning *Semin. Radiat. Oncol.* **14** 81–90

Little D J, Dong L, Levy L B, Chandra A and Kuban D A 2003 Use of portal images and BAT ultrasonography to measure setup error and organ motion for prostate IMRT: implications for treatment margins *Int. J. Radiat. Oncol. Biol. Phys.* **56** 1218–24

Nehmeh S A *et al* 2004 Quantitation of respiratory motion during 4D-PET/CT acquisition *Med. Phys.* **31** 1333–8

Ozhasoglu C and Murphy M J 2002 Issues in respiratory motion compensation during external-beam radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1389–99

Ruan D, Fessler J A and Balter J M 2008 Mean position tracking of respiratory motion *Med. Phys.* **35** 782–92

Schweikard A, Glosser G, Bodduluri M, Murphy M J and Adler J R 2000 Robotic motion compensation for respiratory movement during radiosurgery *Comput. Aided Surg.* **5** 263–77

Seppenwoolde Y *et al* 2001 3D tumor motion in lung due to breathing and heartbeat, measured during real-time tumor tracking radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **51** 24

Seppenwoolde Y *et al* 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

Sharp G C, Jiang S B, Shimizu S and Shirato H 2004 Prediction of respiratory tumour motion for real-time image-guided radiotherapy *Phys. Med. Biol.* **49** 425–40

Shirato H *et al* 2000 Physical aspects of a real-time tumor-tracking system for gated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **48** 1187–95

Shirato H, Seppenwoolde Y, Kitamura K, Onimura R and Shimizu S 2004 Intrafractional tumor motion: lung and liver *Semin. Radiat. Oncol.* **14** 10–8

Stroom J C and Heijmen B J 2002 Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report *Radiother. Oncol.* **64** 75–83

Ten Haken R K, Balter J M, Marsh L H, Robertson J M and Lawrence T S 1997 Potential benefits of eliminating planning target volume expansions for patient breathing in the treatment of liver tumors *Int. J. Radiat. Oncol. Biol. Phys.* **38** 613–7

Trofimov A *et al* 2004 Development of tempero-spatial (4D) radiotherapy optimisation techniques *Med. Phys.* **31** 1776

Vedam S S *et al* 2003 Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker *Med. Phys.* **30** 505–13

Vedam S S *et al* 2004 Predicting respiratory motion for four-dimensional radiotherapy *Med. Phys.* **31** 2274–83

Zhang T *et al* 2004 Treatment plan optimization incorporating respiratory motion *Med. Phys.* **31** 1576–86

IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. 53 (2008) 2923–2936 doi:10.1088/0031-9155/53/11/011

Inference of hysteretic respiratory tumor motion from external surrogates: a state augmentation approach

D Ruan[1], J A Fessler[1], J M Balter[2], R I Berbeco[3], S Nishioka[4] and H Shirato[5]

[1] Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI, USA
[2] Department of Radiation Oncology, The University of Michigan, Ann Arbor, MI, USA
[3] Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[4] Department of Radiation Oncology, NTT East Japan Sapporo Hospital, Sapporo, Japan
[5] Department of Radiation Medicine, Hokkaido University School of Medicine, Sapporo, Japan

E-mail: druan@eecs.umich.edu

Received 12 September 2007, in final form 15 April 2008
Published 6 May 2008
Online at stacks.iop.org/PMB/53/2923

Abstract

It is important to monitor tumor movement during radiotherapy. Respiration-induced motion affects tumors in the thorax and abdomen (in particular, those located in the lung region). For image-guided radiotherapy (IGRT) systems, it is desirable to minimize imaging dose, so external surrogates are used to infer the internal tumor motion between image acquisitions. This process relies on consistent correspondence between the external surrogate signal and the internal tumor motion. Respiratory hysteresis complicates the external/internal correspondence because two distinct tumor positions during different breathing phases can yield the same external observation. Previous attempts to resolve this ambiguity often subdivided the data into inhale/exhale stages and restricted the estimation to only one of these directions. In this study, we propose a new approach to infer the internal tumor motion from external surrogate signal using state augmentation. This method resolves the hysteresis ambiguity by incorporating higher-order system dynamics. It circumvents the segmentation of the internal/external trajectory into different phases, and estimates the inference map based on all the available external/internal correspondence pairs. Optimization of the state augmentation is investigated. This method generalizes naturally to adaptive on-line algorithms.

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

1. Introduction

Respiratory motion affects tumors in the thorax and abdomen. In particular, breathing is the major reason for intrafractional tumor motion for lung cancer patients. It is important to

monitor such motion during radiotherapy treatment to ensure the accurate delivery of radiation dose in motion-compensated intensity modulated radiotherapy (IMRT). Fluoroscopic imaging or portal imaging can monitor tumor motion during the treatment process. To reduce x-ray exposure, hybrid tumor tracking approaches that combine episodic radiographic imaging and continuous external surrogates have been investigated widely (Ozhasoglu and Murphy 2002, Murphy 2004, Murphy *et al* 2002, Schweikard *et al* 2000, 2004). Using external surrogates to infer internal tumor motion assumes that there is consistent relationship between the internal and external motions.

Hysteresis is typical in lung tumor movements, with the tumor taking a different path during inhale and exhale. Inhalation normally takes longer than exhalation, and the deflating lung volume exceeds the inflating volume at the same trans-pulmonary pressure (Keall *et al* 2006). Respiratory hysteresis makes inferring internal tumor locations from external surrogate signals challenging. Most of the external surrogate systems, such as thermistors, thermocouples, strain gauges, pneumotachographs (Kubo and Hill 1996) and infrared skin markers as applied in the Varian Real-time Position Management™ (RPM) system (Varian Medical Systems, Palo Alto, CA), provide one-dimensional signals, whose instantaneous amplitude (or displacement) alone does not provide sufficient information about the specific breathing stages.

Previous studies about correspondence between the internal tumor motion and external surrogates can be classified into two categories. One class of studies investigates the correlation between the two signals to justify the feasibility of using certain types of surrogates, or compare different surrogate options (including the placement mechanism) (Wade 1954, Vedam *et al* 2003, Ahn *et al* 2004, Hoisak *et al* 2004, Tsunashima *et al* 2004, Koch *et al* 2004, Mageras *et al* 2004). Alternatively, some other studies assume *a priori* the existence of a strong correlation between the internal and external signals, and aim to estimate the correspondence map (Seppenwoolde *et al* 2007). We adopt the latter perspective and study with a general setup the correspondence maps that take the external surrogate trace as the input and output estimates of the internal tumor location, including, but not restricted to linear relations as reflected by the correlation coefficient and its variants. The presence of respiratory hysteresis makes this a challenging problem, as the same external surrogate position can reflect different internal tumor locations during different phases. Existing methods address hysteresis by first separating empirically the breathing trajectories into two distinct 'directions' (inhale versus exhale), and then constructing a piecewise phase-dependent map (Seppenwoolde *et al* 2002, 2007, Low *et al* 2005, Lu *et al* 2005). However, subdividing the breathing into inhale and exhale phases often requires manual intervention, and is infeasible for real-time application, because a breathing 'peak' or 'trough' can only be identified retrospectively.

In this study, we propose to use a simple state augmentation of the external surrogate signal. Augmenting the state space with self-delayed observation bestows the model with 'memory', which is an alternative way to characterize the 'path-dependence' property of hysteretic systems. This procedure captures system dynamics, and embeds the breathing-phase information implicitly into the framework. We then provide the solution to a general class of parametric inference models with the augmented observations. As special cases, we derive optimal solutions for the parameters of linear and quadratic correspondence models. Furthermore, given a training internal/external dataset, we demonstrate a computationally efficient approach to choose a patient-specific (or fraction-dependent) augmentation scheme. Generalization to adaptive correspondence models follows naturally. We test the proposed approach on synchronized recordings of internal gold marker trajectories and external fiducial marker locations (Berbeco *et al* 2005).

Table 1. Description of study participants. Patients 1–3 were brought in for data acquisition purposes only, so there is no prescription dose. Patient 5 was treated twice at the same site, with two months between treatments. The tumor site is indicated using the common anatomical notation for lung segmentation: S1-3 is the upper lobe, S4-5 is the middle lobe and S6-10 is the lower lobe.

Patient	Gender	Age	Tumor Pathology	No of bb's	Tumor site	Prescribed dose (Gy)	No of fractions
1	F	47	Adenocarcinoma	4	R S7	N/A	1
2	F	70	Adenocarcinoma	3	L S6	N/A	1
3	F	71	Adenocarcinoma	2	R S5	N/A	1
4	F	47	Adenocarcinoma	3	R S4	48	8
5	M	81	Squamous cell carcinoma	3	R S6b	48	4
5						40	8
6	M	61	Small cell lung cancer	3	R S10	40	8
7	M	68	Squamous cell carcinoma	3	R S6	48	4
8	M	85	Adenocarcinoma	3	R S8	48	4

Section 2 describes the clinical data used for this test, discusses the challenges caused by hysteresis in converting the external surrogate position directly to the internal tumor location and presents the proposed method. A general correspondence model is formulated with polynomial models as an example. The optimal model parameters are derived and a generalization is given to accommodate adaptivity. Section 3 reports testing results followed by discussions. Section 4 concludes this study with a brief summary.

2. Methods and materials

2.1. Data description

To study the internal/external motion correspondence, we obtained synchronized recordings of internal tumor motion trajectories and external fiducial marker locations. The paired trajectories from eight lung cancer patients were collected with a Mitsubishi real-time radiation therapy (RTRT) system at the Radiation Oncology Clinic at the Nippon Telegraph and Telephone Corporation (NTT) hospital in Sapporo, Japan. Two to four 1.5 mm diameter gold ball bearings (bb's) were implanted in or near the tumor (Shirato *et al* 2003) and these internal markers were tracked in real time with diagnostic x-ray fluoroscopy (Shirato *et al* 2000). External surrogate signals were obtained with the AZ-733 V external respiratory gating system (Anzai Medical, Tokyo, Japan) integrated with the RTRT system. It uses a laser source and a detector, both attached to the treatment couch with the beam placed orthogonal to the patient's abdominal skin surface. The device calculates the change in the surface amplitude by measuring the relative position of the reflected light (Berbeco *et al* 2005) and outputs a one-dimensional relative position measurement of the abdominal surface. The data acquisition rate for the entire system is 30 frames per second. Table 1 describes the study participants. All patients included in this analysis had the peak-to-peak marker motion greater than 1 cm. The KV fluoroscopy + Anzai system took multiple readings for each fraction from several treatment field configurations to account for obscured x-ray views as the gantry rotated. The recording lengths varied between 20 s and 250 s with an average of 82 s. There are in total 128 readings, 46 of which were longer than 100 s.

2.2. A general correspondence model

To minimize diagnostic imaging dose in IGRT systems, it is important to infer the internal tumor location from external surrogates. In principle, we could use a correspondence model

that observes a trajectory $\tilde{r}$ of the scalar external surrogate r up to the time instant n to infer the three-dimensional internal tumor position $\boldsymbol{p} = (x, y, z)$. We denote the collective surrogate information available at time n as $\tilde{r}(n) \triangleq \{r(m) : 0 \leqslant m \leqslant n\}$. However, it is challenging to estimate such a map that estimates the internal tumor position from the complete collection of historical surrogate data, since the length of the input variable grows to infinity as the time progresses. A more practical choice is to use some much more compact quantity $\boldsymbol{r}$ that captures sufficient information from $\tilde{r}$ for inference. With internal and external motions both being smooth, it is reasonable to approximate $\boldsymbol{p}(\boldsymbol{r})$ using polynomials. Therefore, we focus on estimating a class of correspondence models that are linear in their coefficients as follows:

$$\hat{\boldsymbol{p}}(\boldsymbol{r}) = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}), \tag{1}$$

where $\boldsymbol{f}$ is a vector function of the external surrogate $\boldsymbol{r}$; all model parameters to be optimized are contained in the coefficient matrix $\boldsymbol{A}$. In particular, two simple correspondence models, i.e. a linear model and a quadratic model introduced in Seppenwoolde *et al* (2007), are special cases of the form given in equation (1).

Linear models assume each coordinate of internal motion is affine in $\boldsymbol{r} = r(t)$. This corresponds to the case where

$$\boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r \\ 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{A} = \begin{bmatrix} b_x & c_x \\ b_y & c_y \\ b_z & c_z \end{bmatrix}. \tag{2}$$

Quadratic models map the external surrogate to each coordinate of internal motion via a quadratic relation. It can be expressed in equation (1) with

$$\boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r^2 \\ r \\ 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{A} = \begin{bmatrix} b_x & c_x & d_x \\ b_y & c_y & d_y \\ b_z & c_z & d_z \end{bmatrix}. \tag{3}$$

Expression equation (1) is linear in the model coefficients $\boldsymbol{A}$ and yields a closed-form optimal solution in the least-squared error (LSE) sense. Given N sample points $(\boldsymbol{r}_n, \boldsymbol{p}_n), n = 1, 2, \ldots, N$, the solution to the LSE problem

$$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} E(\boldsymbol{A}), \tag{4}$$

where $E(\boldsymbol{A}) = \sum_{n=1}^{N} ||\boldsymbol{p}_n - \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}_n)||^2$, is given by solving the normal equation (Luenberger 1969) and

$$\hat{\boldsymbol{A}} = \boldsymbol{P}^T\boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{F})^{-1}, \tag{5}$$

where

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{r}_1)^T \\ \vdots \\ \boldsymbol{f}(\boldsymbol{r}_N)^T \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}_1^T \\ \vdots \\ \boldsymbol{p}_N^T \end{bmatrix}.$$

The corresponding residual is given by

$$\begin{aligned} \Delta\boldsymbol{P} &\triangleq \boldsymbol{P} - \boldsymbol{F}\hat{\boldsymbol{A}}^T \\ &= (\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T)\boldsymbol{P}, \end{aligned} \tag{6}$$

with the overall residual error (summed over all three dimensions) as

$$\begin{aligned} E(\hat{\boldsymbol{A}}) &= \text{trace}\{\Delta\boldsymbol{P}^T\Delta\boldsymbol{P}\} \\ &= \text{trace}\{\boldsymbol{P}^T(\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T)\boldsymbol{P}\}. \end{aligned} \tag{7}$$

It may be preferable to have simpler models (with fewer free parameters) over more complicated models at the cost of small sacrifice in data fitting performance. This model selection preference can be incorporated into the optimization setting by modifying the objective function as

$$\tilde{E}(\boldsymbol{A}) = E(\boldsymbol{A}) + \lambda R(\#\boldsymbol{A}), \tag{8}$$

where $\#\boldsymbol{A}$ denotes the number of free parameters in the coefficient matrix $\boldsymbol{A}$ and R is a monotonically increasing function that assigns higher costs to more complicated models. The regularization weight λ controls the tradeoff between the data fitting $E(\boldsymbol{A})$ and the preference for lower-order models. A simple example of R would be the linear function $R(\#\boldsymbol{A}) = \#\boldsymbol{A}$, which directly penalizes the number of components in $\boldsymbol{A}$; this is equivalent to the Akaike Information Criterion (Akaike 1974). Using the closed-form optimal solution equation (5) and the expression for optimal residual error equation (7) for a given fixed model structure, the modified objective function can be minimized in two layers. We say two inference models have the same *model structure* if they only differ in parameter values. It follows immediately that all models with the same structure has equal number of degrees of freedom, thus the same complexity regularization $R(\#\boldsymbol{A})$ in equation (8). Therefore, to minimizer over models of different complexity, it is natural to choose the 'best' parameter setting within each model structure (with fixed degrees of freedom thus a constant complexity penalty), and then compare across structures. Within each class, the minimizer of the complexity-penalized objective $\tilde{E}(\tilde{A})$ is the same as that of $E(\tilde{A})$, and can be solved and evaluated efficiently using the closed-form optimal solution equation (5) and expression for optimal residual error equation (7). This motivates the two-layer hierarchical algorithm shown below for finding the optimal solution within K candidate model structures $\mathcal{C} = \cup_{i=1}^{K}\{C_i\}$.

Algorithm 1. Two-layer optimization routine for solving $\hat{\boldsymbol{A}} = \arg\min \tilde{E}(\boldsymbol{A})$ (8).

$\tilde{E} \leftarrow +\infty$; $i_{\text{opt}} \leftarrow 0$; $\tilde{\boldsymbol{A}} \leftarrow []$.
for $i = 0$ to K **do**
  Choose model structure C_i from the collection of models $\mathcal{C}$,
  Compute $R_i = R(\#\boldsymbol{A})$ for structure C_i;
  Compute $\hat{\boldsymbol{A}}_i$ within class C_i according to (5) and its residual error $E(\hat{\boldsymbol{A}}_i)$ from (7).
  if $E(\hat{\boldsymbol{A}}_i) + R_i < \tilde{E}$ **then**
    $\tilde{E} \leftarrow E(\hat{\boldsymbol{A}}_i) + R_i$;
    $i_{\text{opt}} \leftarrow i$;
    $\tilde{\boldsymbol{A}} \leftarrow \hat{\boldsymbol{A}}_i$.
  end if
end for

2.3. *Hysteresis and state augmentation*

Conventional methods that explicitly segment the breathing process into inhale and exhale phases have their limitations, as physical phase transitions (and delays) occur continuously rather than as discrete jumps. To circumvent the intrinsic difficulty of estimating breathing phases, we study the system dynamics directly, expecting them to sufficiently convey phase information. In a discretely observed system, one usually captures the system dynamics with time-lagged samples. For the sake of simplicity and to avoid over-parameterization, we restrict this study to a single lag. The proposed method generalizes to multiple-lag models naturally.

Given a discrete-time external surrogate $r(n)$, $n = 1, 2, \ldots, N$, we augment each external surrogate state with a time τ (in discrete unit) delayed sample, i.e. $\boldsymbol{r}(n) \triangleq (r(n), r(n-\tau))$. This augmentation captures first-order system dynamics, as the difference between $r(n)$ and $r(n-\tau)$ can be regarded as a measure of the average local velocity. As $\boldsymbol{r}$ is uniquely determined by $\vec{r}$, it fits into the general formulation equation (1). We apply the methods provided in section 2.2 to estimate the coefficients for the augmented model. To demonstrate the idea, we establish a linear model that is comparable to equation (2) and a quadratic model analogous to equation (3).

The augmented linear model (in $\boldsymbol{r}$) represents each internal coordinate as a linear combination of $r(n), r(n-\tau)$ and a constant offset, corresponding to

$$\hat{\boldsymbol{p}} = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}), \qquad \text{where} \quad \boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r(n) \\ r(n-\tau) \\ 1 \end{bmatrix} \tag{9}$$

with a 3×3 coefficient matrix $\boldsymbol{A}$.

The augmented quadratic model (in $\boldsymbol{r}$) estimates each internal coordinate as a linear combination of $r^2(n), r(n)r(n-\tau), r^2(n-\tau), r(n), r(n-\tau), 1$, corresponding to

$$\hat{\boldsymbol{p}} = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}), \qquad \text{where} \quad \boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r^2(n) \\ r^2(n-\tau) \\ r(n)r(n-\tau) \\ r(n) \\ r(n-\tau) \\ 1 \end{bmatrix} \tag{10}$$

with a 3×6 coefficient matrix $\boldsymbol{A}$.

In both cases, linearity in $\boldsymbol{A}$ results in the closed-form solution given by equation (5) with the corresponding $\boldsymbol{F}$, respectively.

2.4. Choice of lag length

The delay τ should be chosen properly, since too long a lag provides minimal local dynamic information and too short a lag makes the estimation sensitive to observation noise. For inference purposes, we desire a lag that maximally resolves the ambiguity in the estimated correspondence map. We choose the lag that minimizes the fitting error for training data:

$$\hat{\tau} = \arg\min_{\tau} E(\hat{\boldsymbol{A}}(\tau)), \tag{11}$$

with the objective function E defined in equation (4). The coefficients $\hat{\boldsymbol{A}}$ and the error E depend on τ because $\boldsymbol{f}$ contains both the current external surrogate displacement $r(n)$ and its lagged state $r(n-\tau)$.

Equations equation (6) and equation (7) provide a closed-form expression for $E(\hat{\boldsymbol{A}}(\tau))$ for each given τ. The optimization problem equation (11) simplifies to a simple one-dimensional line search that we solve by searching over an interval with the corresponding delay time between 0 (no lag) and about half of an average breathing period.

2.5. Adaptivity of the correspondence map

Adaptivity may be useful to accommodate gradual changes in the correspondence models, due to drifting or variations in patients' breathing. In the case of linear and quadratic models, the operation in equation (5) involves inverting fairly small matrices (3×3 and 6×6, respectively),