IOP Publishing Physics in Medicine and Biology

Phys. Med. Biol. **52** (2007) 7137–7152 doi:10.1088/0031-9155/52/23/024

Real-time prediction of respiratory motion based on local regression methods

D Ruan[1], **J A Fessler**[1] **and J M Balter**[2]

[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

E-mail: druan@eecs.umich.edu

Received 18 July 2007, in final form 3 October 2007
Published 16 November 2007
Online at stacks.iop.org/PMB/52/7137

Abstract

Recent developments in modulation techniques enable conformal delivery of radiation doses to small, localized target volumes. One of the challenges in using these techniques is real-time tracking and predicting target motion, which is necessary to accommodate system latencies. For image-guided-radiotherapy systems, it is also desirable to minimize sampling rates to reduce imaging dose. This study focuses on predicting respiratory motion, which can significantly affect lung tumours. Predicting respiratory motion in real-time is challenging, due to the complexity of breathing patterns and the many sources of variability. We propose a prediction method based on local regression. There are three major ingredients of this approach: (1) forming an augmented state space to capture system dynamics, (2) local regression in the augmented space to train the predictor from previous observation data using semi-periodicity of respiratory motion, (3) local weighting adjustment to incorporate fading temporal correlations. To evaluate prediction accuracy, we computed the root mean square error between predicted tumor motion and its observed location for ten patients. For comparison, we also investigated commonly used predictive methods, namely linear prediction, neural networks and Kalman filtering to the same data. The proposed method reduced the prediction error for all imaging rates and latency lengths, particularly for long prediction lengths.

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

1. Introduction

Current developments in radiotherapy such as cyberknife and intensity-modulated-radiotherapy (IMRT) offer the potential of precise radiation dose delivery for moving objects. Accurate target volume tracking is necessary for conformal treatment plans to fully utilize their

capacity. Image-guided radiotherapy needs to consider system latencies resulting from image acquisition, communication delay, data processing and mechanical processing. For treatment over multiple fractions, or long procedures, the diagnostic radiation dose can be significant, so it is desirable to reduce the image acquisition rate. To address this issue, hybrid tumor tracking approaches that combine episodic radiographic imaging and continuous monitoring of external surrogates have been investigated (Ozhasoglu and Murphy 2002, Murphy 2004, Schweikard *et al* 2000, 2004, Murphy *et al* 2002). There are two active areas of research related to hybrid tracking: (1) study of feasibility and effective use of external surrogates (including the placement mechanism) such as thermistors, thermocouples, strain gauges, pneumotachographs and infrared skin markers (Ahn *et al* 2004, Keall *et al* 2006, Vedam *et al* 2003, Hoisak *et al* 2004, Tsunashima *et al* 2004, Koch *et al* 2004, Mageras *et al* 2004); (2) prediction algorithms (Vedam *et al* 2004, Sharp *et al* 2004, Wu *et al* 2004). In particular, even if perfect information about the current state is assumed, the lag between observing tumor location and treatment delivery still necessitates having predictors that can 'look ahead' enough, yet behave reasonably well even for relatively low input sampling frequencies.

This study belongs to the second category where we are interested in predicting target motion located in the lung area or its vicinity. Such motion is mainly caused by breathing, and exhibits semi-periodicity as observed in normal breathing signals. This is a very active research area (Sharp *et al* 2004, Gierga *et al* 2004, Jiang *et al* 2003, Vedam *et al* 2004). The semi-periodic structure of the breathing signals make explicit modeling challenging, since parametric models often fail to capture local variations. On the other hand, overly flexible models that depend only on temporally local information fail to use correlated historical information. Among the most investigated methods are linear predictors with various covariate lengths, neural networks and Kalman filters (KFs).

We propose a prediction method based on local weighted regression. Adopting a classic approach in modeling dynamical systems, we first generate an augmented state with the most current observation and one or more preceding samples. This augmented state is designed to capture the local dynamics about the time point of interest, and it is used as the covariate for the predictor system. For a pre-specified 'look-ahead' length, the target response pattern of the predictor is obtained from the training data. These state–response pairs form a scatter plot in a high-dimensional space where we apply locally weighted regression. Intuitively, the predictor infers its response map from the behavior of its neighbors in this state space, since it is probable that they are. The regression weights are designed to reflect the 'distance' between the state of interest and the training samples.

For the purpose of real-time tracking and prediction, we adaptively adjust the inference weights to incorporate the decaying temporal correlation among response patterns with longer time lags[3].

We discuss the proposed methods in detail in section 2. The method is applied to clinical real-time position management (RPM) data (RPM Varian Medical System, Palo Alto, CA) that is described in section 3. We report the test results and the comparison to alternative methods in section 4. Finally, we discuss future directions in section 5.

2. Methods

In this section, we propose a prediction method based on locally weighted regression. For simplicity, we describe the model in terms of scalar locations, i.e. 1D observation. The extension to vector observation is straightforward. Section 2.2 first introduces a primitive

[3] In fact, this corresponds to augmenting the state with the time index as an extra dimension.

version that ignores the change of temporal correlation with time lag, and sections 2.3.2 and 2.3.3 extend it to include time indexing.

2.1. Model setup and general notations

Let the continuous scalar function $f : \Re \to \Re$ denote a motion index signal. At time instant τ, we are given a set of discrete samples $\{s_i = f(t(i)), i = 1, 2, \ldots, k\}$ of the breathing trajectory prior to τ, with $t(k) < \tau$. For simplicity, we assume that the observed signal is sampled uniformly with frequency ψ Hz, i.e. $t(i+1) - t(i) = 1/\psi$. We assume that the look-ahead length is an integer multiple of the sampling interval $1/\psi$ seconds, and for later convenience, we represent it in the discrete unit, i.e. a look-ahead length L indicates a L/ψ seconds prediction. We use p to denote the state dimension used to capture system dynamics. To draw an analogy to an ordinary differential equation (ODE) based system, $p = 2$ corresponds to a first-order difference system with location and approximate velocity, and $p = 3$ corresponds to a second-order difference system with the addition of acceleration.

2.2. Basic local weighted regression

At current time instant τ, the available observations are $s_1, \ldots, s_K$, where $K \leqslant \tau \times \psi$. Then for any $i \leqslant K - L$ we construct length p state vector $\boldsymbol{x}_i = [s_{i-(p-1)\Delta}, \ldots, s_i]$, and response variable $\boldsymbol{y}_i = s_{i+L}$. The parameter Δ is an integer that indicates the 'lag length' used to generate the augmented state. It should be chosen to properly reflect system dynamics: small lags are more sensitive to dynamical change as well as noise; big lags are more robust to the presence of noise yet average out the system dynamics at the same scale. The set of hyper-pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ form a scatter plot in the $p+1$ dimensional space. Assuming that the state thus constructed conveys all the information about system dynamics, then the scatter plot summarizes the noisy realizations of the prediction map: $g : \Re^p \to \Re$:

$$\hat{\boldsymbol{y}}_i = g(\boldsymbol{x}_i), \tag{1}$$

where the predictor g is a smooth function. This is a reasonable model as we do expect the prediction to vary smoothly with the historical trajectory. Our goal is to predict the target location at time $(\tau + L/\psi)$ seconds, which is equivalent to estimating $\hat{\boldsymbol{y}}_K$.

Respiratory motion is not stationary, in fact, both the system dynamics and its local statistics vary in a semi-periodic fashion. Unfortunately, most existing methods in estimating the prediction map g fail to take this 'phase-dependent' phenomenon into account, with the exception of (Ruan *et al* 2003) and (Wu *et al* 2004) where a discrete set of stage-wise models are constructed and updated adaptively. The idea is to train (or infer) a predictor at a given state with (only) those historical data samples that behave similarly or, vaguely speaking, belong to a similar respiratory stage. Yet the existing stage-wise models require pre-determining the number of discrete stages and often involve segmentation-based training. To circumvent these difficulties, we hereafter provide a means to locally estimate g in the state-space neighborhood of $\boldsymbol{x}_K$, based on local regression (LOESS) from nonparametric methods in the statistical literatures (Cleveland 1979).

Let r be a pre-determined constant[4] that specifies the size of the neighborhood whose members affect the estimate in the scattered $(p+1)$-dimensional space. Let h_K be the distance from $\boldsymbol{x}_K$ to the rth nearest neighbor in terms of Euclidean distance in the p-dimensional subspace, i.e. h_K is the rth smallest number among the distances between $\boldsymbol{x}_i, i = 1, 2, \ldots, K - L$, and $\boldsymbol{x}_K$. Let $\kappa(\cdot)$ be a symmetric kernel function that satisfies

[4] Equivalently, it could also be specified as a ratio with respect to the total number of data points.

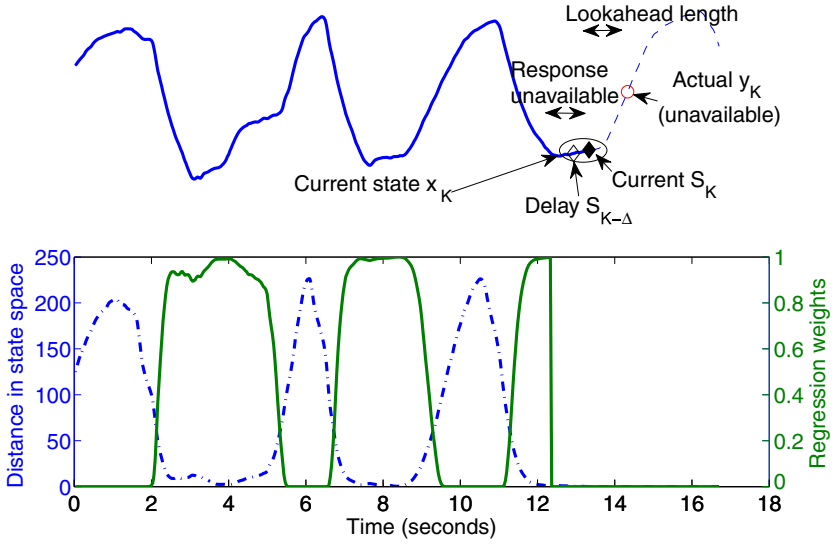


Figure 1. State space distance and local regression weight assignment. Upper subfigure: illustration of prediction quantities with first-order dynamic $\boldsymbol{x}_i = [s_{i-\Delta}, s_i]$, current data point s_K, prediction location $\boldsymbol{y}_K = s_{K+L}$, available covariant–response pair $(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \leqslant K-L}$; lower subfigure: distance map (blue dash-dot line) in the state space $d_i = ||\boldsymbol{x}_i - \boldsymbol{x}_K||$ and inferred regression weights w_i (green solid line).

(i) $\kappa(x) > 0$ for $|x| < 1$ and $\kappa(x) = 0$ for $|x| \geqslant 1$,
(ii) $\kappa(-x) = \kappa(x)$,
(iii) $\kappa(x)$ is a non-increasing function for $x \geqslant 0$.

We select local inference weight according to

$$w_i = \kappa\left(h_K^{-1}||\boldsymbol{x}_i - \boldsymbol{x}_K||\right). \tag{2}$$

Figure 1 illustrates the idea of weighting based on distance in state space. For simplicity, one delay tap is used ($p = 2$), so the state $\boldsymbol{x}_i = [s_{i-\Delta}, s_i]$. The goal is to estimate the response y_K for current state vector $\boldsymbol{x}_K$, from available covariate–response pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i \leqslant K - L$. Note that $K - L < i \leqslant K$ are not used in the regression, since their response values $\boldsymbol{y}_i$ are not yet available at time instant K. Distances between current states $\boldsymbol{x}_K$ and $\boldsymbol{x}_i$ are computed and the kernel function is used to determine the regression weights w_i as in (2). The assigned weights emphasize those training samples that share similar dynamics as the current state of interest as shown in the lower part of figure 1.

We subsequently estimate g locally using a polynomial of degree d, i.e. we use a predictor of the form $g(\boldsymbol{x}) = \sum_{q=1}^{Q} \beta_q z_q(\boldsymbol{x})$, where $Q = (d+1)^p$ and $z_q(\boldsymbol{x}) = \prod_{j=1}^{p} x_j^{d_j}$ and $(d_1, \ldots, d_p) \in \{0, 1, \ldots, d\}^p$ that corresponds to the base-d representation of q.

We estimate the coefficients of the local polynomial by minimizing the weighted local squared error:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{i=1}^{K-L} w_i \left(\boldsymbol{y}_i - \sum_{q=1}^{Q} \beta_q z_q(\boldsymbol{x}_i) \right)^2 \\ &= (Z^T W Z)^{-1} Z^T W Y, \end{aligned} \tag{3}$$

where Z is the $K - L$ by Q matrix with elements $z_{iq} = z_q(\boldsymbol{x}_i)$. The weighting matrix W is a diagonal matrix with $W(i, i) = w_i$. Since the local weight w has a limited bandwidth h_K

as designed in (2), there are only r nonzero diagonal elements in the weighting matrix W. Correspondingly, the outer summation $\sum_{i=1}^{K-L}$ can be equivalently written as supported only on a local neighborhood of radius h, i.e. $\sum_{i:||\boldsymbol{x}_i-\boldsymbol{x}_K||<h}$. Therefore, the data vectors involved have length $r \ll K-L$ rather than $K-L$. It is desirable to choose a small neighborhood size r to decrease the computation cost, yet not overly small to sacrifice the regularity of (3), i.e. the invertibility of $Z^T W Z$.

For subsequent prediction from a given observation $\boldsymbol{x}_K$, we use the estimated polynomial coefficient $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{y}}_K = \sum_{q=1}^{Q} \hat{\beta}_q z_q(\boldsymbol{x}_K). \tag{4}$$

The algorithmic flow chart is as follows.

Algorithm 1 Predict $\hat{\boldsymbol{y}}_K$ from $(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i \leqslant K-L}, \boldsymbol{x}_K$ with local regression.

1: Select r (size of regression neighborhood); obtain h_K from order statistics of $||\boldsymbol{x}_i - \boldsymbol{x}_K||$.
2: Select kernel κ and compute regression weights w_i according to (2).
3: Compute prediction model coefficients $\hat{\beta}$ according to (3). For lag-one state augmentation with a second-order polynomial prediction model, $p = 2$, $d = 2$ and $Q = 9$, so computing $\hat{\beta}$ requires the inversion of a 9×9 matrix $Z^T W Z$ and then multiplying it by a 9×1 vector.
4: Predict the response $\hat{\boldsymbol{y}}_K$ using (4).

2.3. Variations that potentially improve prediction performance

We now describe two design variations that have the potential to improve the prediction performance: using an iterative weighting scheme to increase robustness to outliers in regression (section 2.3.1) and dynamically updating the training atlas to account for temporal variations and/or trends (sections 2.3.2 and 2.3.3).

2.3.1. Robust local weighted regression with iterative weight assignment. It is possible that the training set based on the state-space distance includes abnormal covariate–response pairs due to noisy observation or abrupt (and non-repetitive) changes such as a patient coughing, and thus they may not be 'representative' of the predictor pattern for the given state. To help the local regression method to be robust to such outliers in the $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ pairs, we can diminish the weight of a sample covariate–response pair whenever it is inconsistent with the smooth regression from its neighbors. To quantify such inconsistency, we can compare each response value $\boldsymbol{y}_i$ with its predicted value $\hat{\boldsymbol{y}}_i = g(\boldsymbol{x}_i)$. Intuitively, the distance between the observed response $\boldsymbol{y}_i$ and its estimate $\hat{\boldsymbol{y}}_i$ indicates how different the particular covariate–response pair behaves than its neighbors. (Cleveland 1979) has suggested a robust weighting scheme based on a bi-square function B defined as follows:

$$B(x) = \begin{cases} (1-x^2)^2, & \text{for} \quad |x| < 1 \\ 0, & \text{for} \quad |x| \geqslant 1. \end{cases} \tag{5}$$

Let $e_i = \boldsymbol{y}_i - \hat{\boldsymbol{y}}_i$ be the residual of the observed response from the current fitted value. Let s be the median of $'|e_i|$ for $i = 1, 2, \ldots, K-L$. Define the robustness weights by

$$\delta_i = B(e_i/6s). \tag{6}$$

The original weight $w_i(\boldsymbol{x}_j)$ that determines the 'contribution' of the ith sample covariate–response pair in estimating the jth response $\hat{\boldsymbol{y}}_j$ is then modified to be $w_i(\boldsymbol{x}_j) := \delta_i w_i(\boldsymbol{x}_j)$, reducing the effect of outliers in fitting the other data points. We apply this re-weighting procedure several times, and use the robust adjusted $w_i(\boldsymbol{x}_K)$ in place of w_i in (3) for estimating the local polynomial coefficient $\hat{\boldsymbol{\beta}}$. This is practical since the δ_i values involved in adjusting the local weight depend only on the ith sample fitting quality, and are independent of the predictor. Plugging $\hat{\boldsymbol{\beta}}$ in (4) results in a predicted response value $\hat{\boldsymbol{y}}_K$. Since the estimation of local polynomial coefficients discounts the effect of outlier samples, the result predictor is expected to be robust to outlier behavior in the 'training set' as well. Note that robust local regression could be combined with other methods if needed.

2.3.2. Modified weight assignment with exponentially discounted temporal correlation. Fading memory is present in many natural processes. In breathing trajectories, temporally adjacent sample points tend to be more similar than the sample points further away from one another. To incorporate this property in prediction, we adjust the weights by applying an exponential discount as a function of the temporal distance. Specifically, we modify the weights as follows:

$$w_i(\boldsymbol{x}_j) := \exp(-\alpha|i-j|)w_i(\boldsymbol{x}_j). \tag{7}$$

The positive constant α determines the decaying rate of influence of one sample on another as their temporal distance increases. As a special case, $\alpha = 0$ corresponds to no temporal discounting for the sample contributions, but dynamically adds the new samples into the training atlas as they become available.

2.3.3. Temporally windowed training samples. Alternatively, we can modify the weights using a temporal moving window as follows:

$$w_i(\boldsymbol{x}_j) := \begin{cases} w_i(\boldsymbol{x}_j) & |i-j| < \Gamma \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where Γ is the window size. Here, the only samples that are close enough in time contribute to the local regression with weights determined by (2). The length of the window needs to be chosen long enough to guarantee enough samples for the local regression.

2.4. Baseline methods for comparison

It is desirable to decrease radiation dose due to imaging in image-guided radiotherapy (IGRT). This means we would prefer to predict with low-frequency observation samples (small ψ). On the other hand, it takes time to acquire each observed sample, process it and move the hardware (linac, MLC or cyberknife) accordingly. Thus, a system capable of large look-ahead lengths is preferable. These two requirements are challenges in prediction, and trade-offs between them need to be considered. More specifically, with the look-ahead length determined by the limitation of the system response, we want to determine the smallest measurement rate that still guarantees certain prediction accuracy. We will study the performance of the proposed method when look-ahead lengths and sampling rates are varied, and compare that with some baseline approaches described as follows.

Following (Sharp *et al* 2004), we use some commonly used predictors for baseline comparison. We briefly describe their setups and optimization for free parameters in this section.

- *Most recent sample*
  This method simply uses the last sample value:

$$\hat{\boldsymbol{y}}_K = s_K.$$

  There are no parameters to be estimated.
- *Linear predictor*[5]
  The response is predicted as a linear combination of the previously known positions. This corresponds to a simple model:

$$\hat{\boldsymbol{y}}_K = \boldsymbol{\beta}^T \boldsymbol{x}_K + \beta_0.$$

  Given a training set, and for a fixed history length the optimal coefficients $\boldsymbol{\beta}$, β_0 in terms of mean squared error can be obtained by solving a linear system.
- *Artificial neural networks (ANN)*
  We investigate a multilayer perception (MLP) with two feed-forward layers as the ANN predictor (Lippmann 1987). The first layer takes in a fixed history of samples and a constant value 1, linearly transforms the inputs and then uses a sigmoid function to generate the hidden values. The equation for the first layer is

$$h_j(\boldsymbol{x}) = \frac{1}{1 + \exp\left(-\boldsymbol{\gamma}_j^T \boldsymbol{x} + \gamma_{j,0}\right)}.$$

  The second layer is chosen to be a simple linear system, and the output is given by

$$\hat{\boldsymbol{y}}_i = \boldsymbol{\eta}^T \boldsymbol{h}(\boldsymbol{x}_i).$$

  Parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are estimated from the training set. We use Netlab toolbox (Nabney and Bishop 2003) to implement ANN in Matlab.

We have also implemented a Kalman Filter (KF) for comparison, using the expectation-maximization (EM) method for parameter selection (Murphy 2003), and applied those values for prediction. Our results agree with (Sharp *et al* 2004) that the KF provides inferior performance compared to ANN. For conciseness, we skip reporting them in this paper. A related research worth noting is the adaptive linear filter model introduced in (Vedam *et al* 2004), which can be interpreted as a KF not in the state, but in a linear regression coefficient vector. Unsurprisingly, it shares the limitation of a KF due to the non-stationarity of the respiratory signal.

3. Materials

We used the RPM (RPM Varian Medical Systems, Palo Alto, CA) system to obtain the trajectory of an external fiducial placed on the patient's chest wall. The recorded displacement–time relationship is believed to be highly correlated with superior–inferior diaphragm motion (Vedam *et al* 2004), which is the major source of respiratory motion for tumours in the chest or lung area (the displacements in left–right and anterior–posterior directions are normally on the order of one magnitude lower). To better reflect the behavior of physical superior–inferior motion, the unit-less RPM data were centered and scaled so that their dynamic range matches that with typical SI motion for chest and lung tumours. Table 1 summarizes the RPM data used in our experiment[6]. Figure 2 illustrates two typical breathing trajectories.

[5] The 'linear extrapolation' method described in (Sharp *et al* 2004) is a special case of linear prediction.

[6] The data are adjusted to have globally zero mean; average periods are estimated with the subspace projection method (Ruan *et al* 2006).

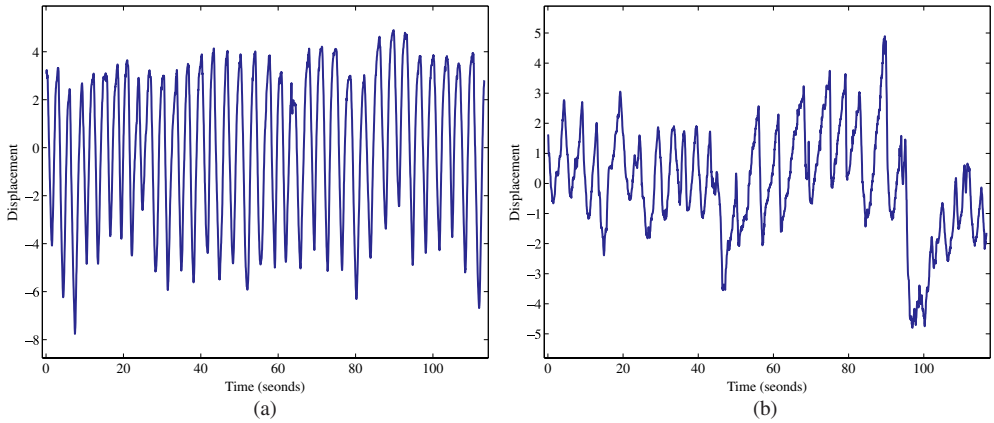


Figure 2. Typical breathing trajectories: (a) rapid yet regular breath, (b) slow yet irregular breath.

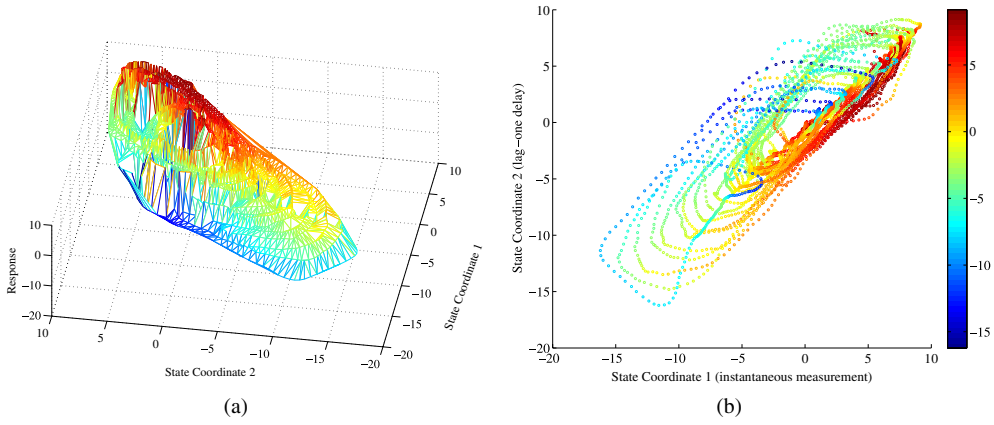


Figure 3. Covariate–response relationship with a lag-one augmented state: (a) three-dimensional Delaunay tessellation plot, (b) two-dimensional scatter plot with color indicating the response value.

Table 1. RPM dataset information.

Subject ID	1	2	3	4	5	6	7	8	9	10
STD	4.96	4.99	3.01	1.99	3.16	1.73	6.27	5.65	2.74	5.29
P-P	25.36	23.65	12.67	11.24	18.72	9.70	28.79	21.89	12.19	21.55

4. Results and discussions

4.1. Scatter plot in augmented space

We first consider a simple $p = 2$ dimensional state vector $\boldsymbol{x}_i = [s_i, s_{i-\Delta}]$. The response variable is of the form $\boldsymbol{y}_i = s_{i+L}$. Figure 3 shows a three-dimensional scatter plot of $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ with the baseline X–Y coordinate reflecting the covariate $\boldsymbol{x}_i$ and the Z coordinate indicating

Table 2. Comparison of prediction performance among static training, dynamic expanding training and updating training with moving window.

Subject ID	1	2	3	4	5	6	7	8	9	10	Average
					Root mean squared error (RMSE)						
Static	9.7	3.6	2.2	1.9	10.8	5.6	4.9	4.2	2.8	4.4	5.0
Expand	3.4	2.8	1.6	1.4	2.5	1.3	4.8	2.6	2.1	3.7	2.6
Update	2.7	2.5	1.4	1.4	2.6	1.3	4.8	2.5	2.1	3.5	2.5
					Mean absolute error (MAE)						
Static	7.5	2.6	1.7	1.4	3.9	2.6	3.7	2.5	2.1	3.1	3.1
Expand	2.6	2.1	1.2	1.1	1.7	1.0	3.5	1.7	1.6	2.6	1.9
Update	2.0	2.0	1.1	1.0	1.7	1.0	3.4	1.7	1.4	2.5	1.8

the corresponding response variable value $\boldsymbol{y}_i$. The covariate–response structure is rather smooth, motivating our use of local regression to predict a response from the samples in the neighborhood of the projection onto the X–Y plane. Roughly speaking, the pattern suggests the existence of a conceived functional g that maps the covariate to the response.

We started with a fine sampling rate of $\psi = 30$ Hz and used only lag-one delay with $\Delta = 12$ that corresponds to 0.4 s to augment the state space. We investigate a look-ahead length of $L = 30$, which is equivalent to a 1 s prediction. We used these parameters as defaults in later experiments (e.g. in section 4.4). This look-ahead length is reported to be difficult by (Vedam *et al* 2004) and (Sharp *et al* 2004) with a wide spectrum of common prediction techniques. In particular, in the comparative study in (Sharp *et al* 2004), the best performance among linear predictors, Kalman filter and artificial neural networks yields a root mean squared error (RMSE) of about 5 mm, with similar data statistics to our rescaled RPM data[7]. Lag-one augmentation corresponds to regression based on the most current sample s_i and one preceding observation $s_{i-\Delta}$, which is the most compact model possible. The temporal lag Δ for augmentation should be chosen to reflect the system dynamics properly and robust enough in the presence of observation noise, and does not have to be unity.

4.2. Local weighted regression without temporal discounting

To illustrate the performance of the simple local weighted regression method described in section 2.2, we conduct two simple experiments with the following configurations: we used the 'tricube' function (Cleveland 1979) as the weighting kernel κ and chose the effective bandwidth so that the local regression is supported on half of the samples. Specifically, we used

$$\kappa(x) = \begin{cases} (1 - |x|^3)^3, & \text{for} \quad |x| < 1 \\ 0, & \text{for} \quad |x| \geqslant 1. \end{cases} \tag{9}$$

The neighborhood size r was chosen to be 200, which is equivalent to about 7 s worth of samples. Accordingly, h_K is the 200th smallest number among $||\boldsymbol{x}_i - \boldsymbol{x}_K||$.

[7] The research conducted in (Sharp *et al* 2004) uses a three-dimensional position, which is presumably more complex than this study. However, since motion in the SI direction dominates the overall respiratory variation in general, we expect the prediction error to be the major contributor to the overall tracking/prediction performance. Rescaling the RPM data to have similar statistics as typical SI motion, we feel it fair and illuminating to compare quantitatively the performance of our predictor to that of the general 3D predictors.

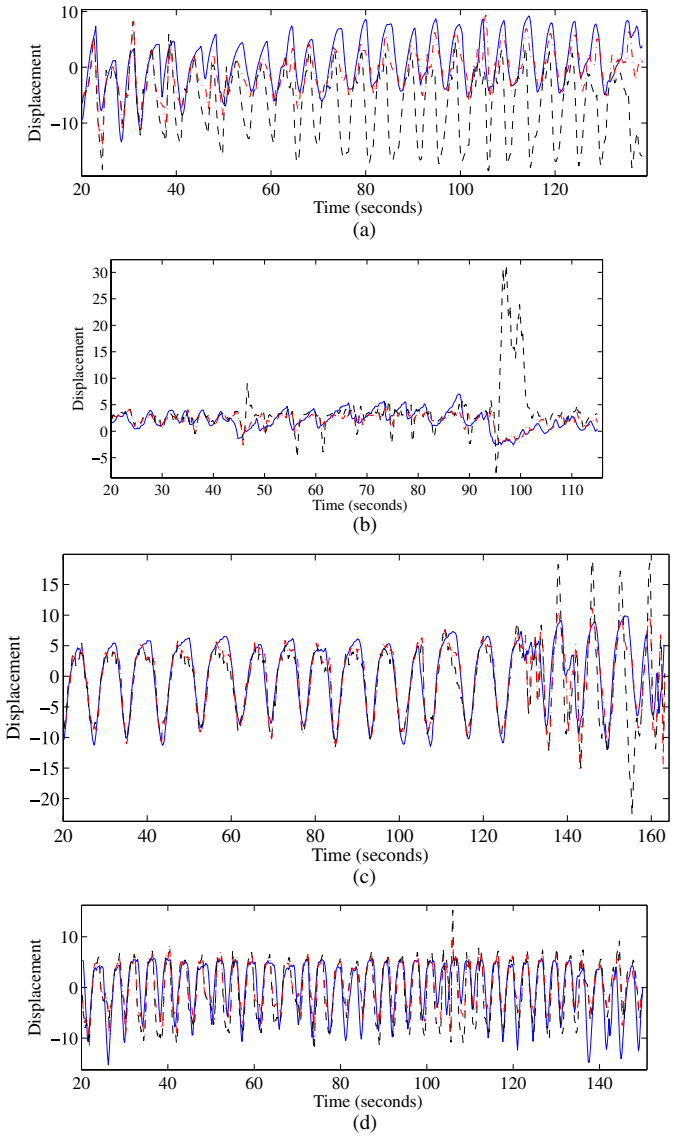


Figure 4. Effect of dynamically updating the training atlas: actual signal time history (blue solid line), prediction from static training (black dash-dot line) and prediction from dynamic expanding atlas (red dashed line). (a) Breathing with mean drifting (patient 1), (b) in the presence of a changing breathing pattern (patient 6), (c) with complicated transient interrupting regular breathing (patient 8), (d) quasi-regular breathing pattern (patient 10).

4.3. Robust local regression with iterative weighting

We investigated the robust iterative weighting of section 2.3.1, but found that iterative weighting did not significantly change the prediction errors in this experiment. This suggests the absence of dramatic outliers in our experimental data.

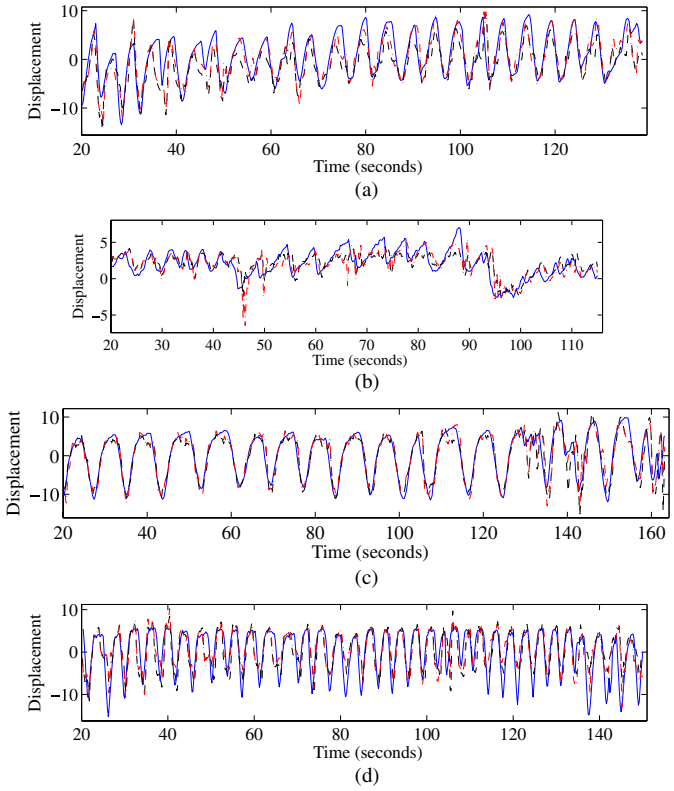


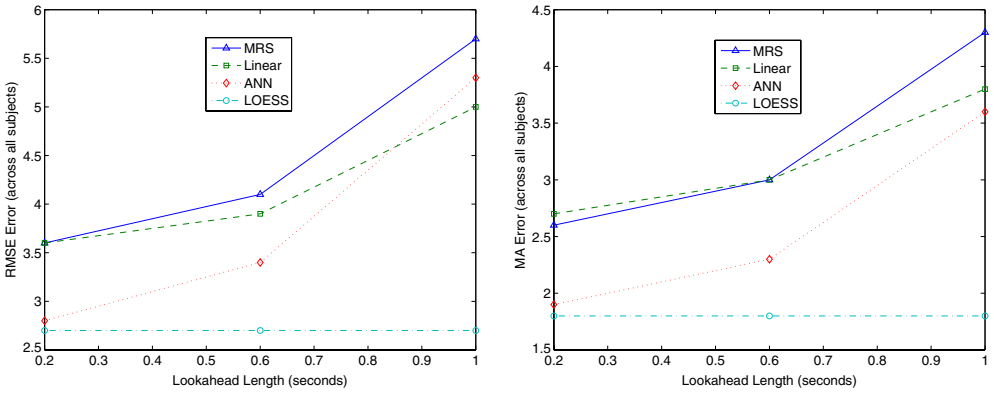
Figure 5. A comparison of prediction performance using dynamic update with moving windowed and expanding training atlas: actual signal time history (blue solid line), dynamic expanding training (black dash-dot line), moving window adaptive training (red dashed line).

4.4. The effect of dynamically updating the training set

If the training set is determined before the treatment process, and is kept the same thereafter, the corresponding local regression structures are also fixed. This is the 'static' inference scenario. It is also possible to 'add' (or 'substitute' the oldest sample with) new samples into the training set during the treatment process, as new responses become available. We refer to the latter approach as 'dynamically updating of the training set'.

The computation for simple local regression is the same regardless of whether we update the training atlas or not, as it uses only the training samples that fall into the neighborhood of the target. On the other hand, when robust local regression with iterative weighting is applied, choosing between static training and dynamic training makes a difference. In the static scenario, the robust weights can be computed offline upon the availability of all the training samples, and are kept the same thereafter. However, if we use dynamic updates, not only does the size of the 'atlas' grow with time, but there would also be changes in the robust weights, since the newly available covariate–response samples can potentially change the regression weights for those existing samples whose supports overlap with those of the new ones. This effect can propagate through the whole atlas.

At the cost of possible additional computation, dynamically updating the training atlas admits new information as the time proceeds. This is particularly valuable when the underlying



(a) RMSE (b) MAE

Figure 6. A collective performance comparison for different look-ahead lengths measured with the (a) root mean squared error (RMSE) and (b) mean absolute error (MAE) with a sampling rate $\psi = 5$ Hz.

system dynamics demonstrate strong temporal variation, such as frequency change or home position drifting, which are both commonly observed. New samples can be added to the training set as simple addition, which corresponds to a collective history case, or substituted for the oldest training sample, as in the windowed training history case. In both cases, experimental results indicate that dynamically updating the training set yields overall superior prediction performance in terms of root mean squared error (RMSE) and mean absolute error (MAE), as we report in sections 4.4.1 and 4.4.2.

4.4.1. Dynamically expanding the training set. Using a discount factor $\alpha = 0$ in (7) to adjust weight for the training samples up to the most currently available one is equivalent to building a collective atlas that includes all previous covariate–response pairs. Of course, new training pairs are entered into the atlas as time proceeds. Table 2 reports the prediction performance for 1 s look-ahead with 5 Hz sampling using this dynamic training structure as opposed to a static 20 s training at the beginning of the fraction. Figure 4 illustrates improved prediction performance by dynamically expanding the training set for breathing traces that either exhibit mean drifting or pattern changes. There is minimal benefit when the breathing pattern is already fairly regular or irregular with no 'trend', and new observations simply add to the already sufficiently dense training atlas. Change detection may be used to locate some local variations, but this imposes extra complexity[8].

4.4.2. Dynamically updating the training set with windowed history. Alternatively, a moving window can be used to update the sample set. This corresponds to substituting the oldest samples with the newly available covariate–response pairs, as discussed in section 4.4.2. We illustrate the effect of this dynamic updating method in table 2 and figure 5. A dynamic windowed history that has a length of 20 s is used in all of our experiments. We used the performance of dynamic expansion as a baseline for the windowed study.

[8] A segmentation-based tracking/prediction model (Wu *et al* 2004) follows similar logic, yet requires further research to improve robustness and automation.

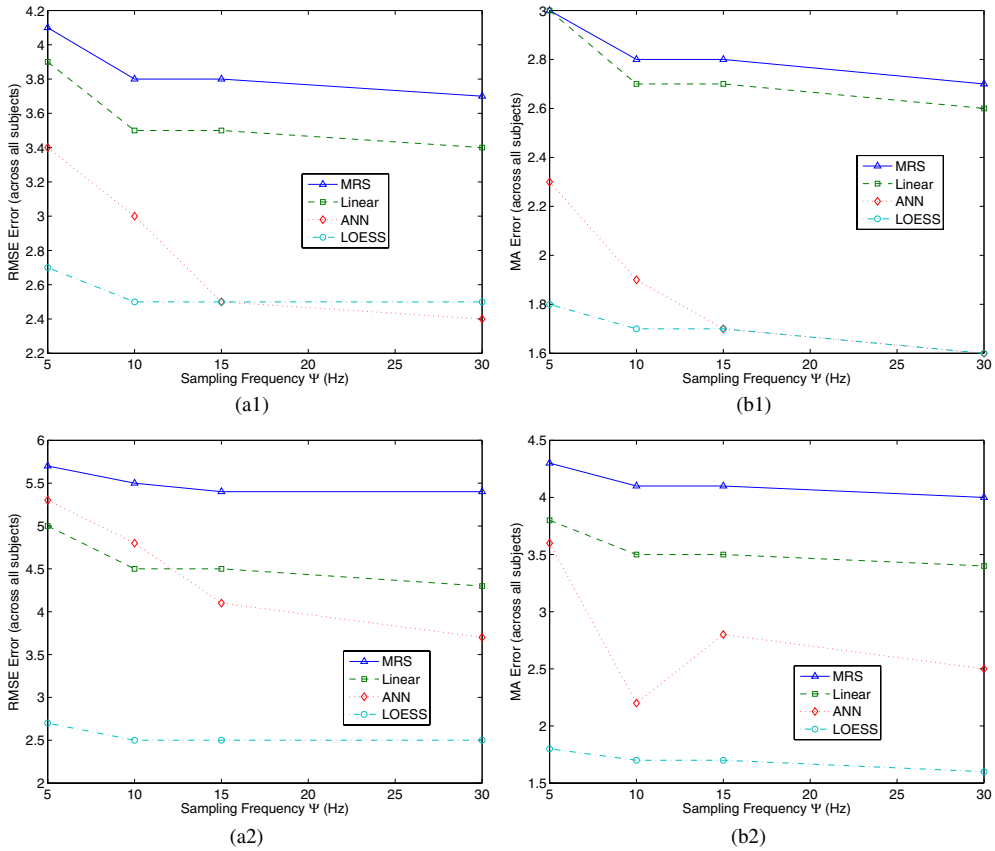


Figure 7. A collective performance comparison for different sampling rates ψ with RMSE (left column [a#]) and MAE (right column [b#]). The top row [X1] illustrates the results from a look-ahead length of 0.6 s and the bottom row [X2] shows the results when the look-ahead length is 1 s.

For the 20 s training window, the overall prediction performance improves upon the previously discussed dynamic expansion. The level of improvement, though, is much smaller than the one we obtained by going from static training to dynamic expansion. Some trade-offs are expected: for long fractions, it is more likely that the later samples are decoupled from the samples acquired at the very beginning of the procedure; thus the moving window method should be favorable. On the other hand, dynamic expansion does not require choosing a window length, and it is almost free of the risk of running into insufficient samples for the local inference; thus it has the advantage of being simple and stable. There is little difference in the prediction performance between the two methods from figure 5 except that in the mean drifting case, the windowed update may be slightly better, which is also reflected quantitatively in table 2.

4.5. The effect of the measurement rate and look-ahead length

We compared the LOESS method using expanding training atlas with the baseline approaches described in section 2.4. In particular, we compared with the most recent sample (MRS),

linear prediction (linear), Kalman filter (KF) and artificial neural networks when the look-ahead length and sampling rates are varied. Figures 6 and 7 report the results in terms of the collective RMSE and MAE across all patients. In general, the prediction errors increase as the sampling frequency ψ decreases and/or the look-ahead length L increases, as expected. Interestingly, the proposed local regression method is insensitive to the sampling rate, and performs almost consistently across different look-ahead lengths. Unlike the most recent sample, linear model or Kalman filter, which not only makes assumptions about the underlying model structure (linearity), but also try to explicitly solve for the model parameters, LOESS makes none of the above assumptions or effort. The nonparametric nature of the regression avoids assuming a fixed model structure, let alone solving for it. The only requirement is consistent behavior (or *existence* of an underlying functional form). A polynomial of sufficient order approximates this underlying function via fitting samples in the neighborhood of the point of interest. This also explains, to some extent, why ANN outperforms the other approaches (Murphy and Dieterich 2006), as it is a combination of local linear perceptrons, with extra nonlinearity provided by the sigmoid activation function. When the look-ahead length is short and the sampling rate is high, linearity holds approximately, and all methods provide reasonably good prediction. However, when we need to look further ahead, linear models are not sufficient to capture the dynamics, even though the response pattern may still be consistent, and that is where LOESS (and ANN) demonstrates its advantage. Figure 6 reports the relationship between the collective prediction error (across all testing subjects) and the look-ahead length for sampling rate $\psi = 5$ Hz and figure 7 illustrates how the collective prediction error changes with different sampling rates. The LOESS approach performs competitively with ANN for look-ahead length 0.6 s, in particular for low sampling frequency, and it demonstrates an obvious advantage for look-ahead length 1 s.

5. Conclusion and future work

In this paper, we have proposed a local regression-based method to predict respiratory motion. We compared the proposed method and conventional approaches such as the most recent sample, linear model, Kalman filtering and artificial neural networks. The proposed method had a lower prediction error than the others for tasks requiring long look-ahead prediction. We have also discussed extensions and variations of the basic method to provide robustness to outliers that may be caused by a low signal-to-noise ratio (SNR) or mis-tracking. We studied the prediction performance with different error metrics (RMSE and MAE) for various combinations of look-ahead length and sampling frequency. The proposed method showed the most advantage for long look-ahead lengths and low sampling rates.

We have discussed the challenge of choosing a good discount factor for weight adjustment in local regression in section 2.3.2 and discussed the two simple cases corresponding to either no forgetting or inference from a windowed historical sample. The proper choice of the temporal discount factor depends on the variation of the underlying breathing pattern, and automatic schemes should be investigated.

As observed in our experiment, various phases of respiratory motion are predicted with different accuracies. Respiratory motion demonstrates obvious non-stationarity: the system variation at the transition phase could be very different than that during extreme tidal stages (end-inhale or end-exhale). From another perspective, if we examine the SNR over a windowed portion of the signal, the SNR would change as the window covers different stages of the breathing: the SNR would be relatively high during transition stage, as the signal variation is big relative to noise, while the SNR is low at the plateau stages, which correspond to the end of inhale or exhale. These observations motivate a potential research topic: if

we aim at homogeneous prediction performance throughout the breathing trajectory, it may be necessary to use adaptive sampling. More precisely, denser sampling may help where prediction uncertainty is big whereas sparser sampling should suffice where prediction is more reliable. This is a topic for future study.

The dynamics of respiratory motions changes over various stages of breathing, and makes general prediction difficult. Models using state-dependent transition probabilities have been investigated for stochastic tracking (Ruan *et al* 2003), and explicit segmentation was also studied (Wu *et al* 2004). Our proposed method uses local kernel regression to capture this variation implicitly by essentially limiting inference to a neighborhood of training samples that are expected to behave similarly. Intuitively, this is almost equivalent to training a local model at each state of interest. Since the state distance (and thus the inference weight) is assigned with respect to the Euclidean distance in state space, it is important that clustering with this distance reflects dynamic similarity. This is expected in most cases, except when home (mean) drifting is high both in frequency and in displacement value. In the exceptional case of dramatic mean drift, samples belonging to different breathing stages may be clustered together. One straightforward remedy would be to incorporate mean drifting compensation in the inference weight. A robust mean tracking algorithm for respiratory motion is provided in (Ruan *et al* 2007) that outputs mean position estimates for both the training samples and the state of interest. We expect improved accuracy by accounting for mean position drifting. We plan to conduct further experiments and analyze this effect in more detail in the future.

Acknowledgments

This work is partially supported by NIH grant P01-CA59827. The authors gratefully acknowledge Dr Gregory Sharp, Dr Paul Keall, Dr Amit Sawant, Dr Raghu Venkat and Vikram Srivasta for valuable discussions, as well as the anonymous reviewers for their helpful comments.

References

Ahn S, Yi B, Suh Y, Kim J, Lee S, Shin S, Shin S and Choi E 2004 A feasibility study on the prediction of tumor location in the lung from skin motion *Br. J. Radiol.* **77** 588–96

Cleveland W S 1979 Robust locally weighted regression and smoothing scatterplots *J. Am. Stat. Assoc.* **74** 829–36

Gierga D P, Chen G T Y, Kung J H, Betke M, Lombardi J and Willett C G 2004 Quantification of respiration-induced abdominal tumor motion and its impact on IMRT dose distributions *Int. J. Radiat. Oncol. Biol. Phys.* **58** 1584–95

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

Jiang S, Pope C, Jarrah K A, Kung J, Bortfeld T and Chen G 2003 An experimental investigation on intra-fractional organ motion effects in lung IMRT treatments *Phys. Med. Biol.* **48** 1773–84

Keall P *et al* 2006 The management of respiratory motion in radiation oncology report of AAPM task group 76 *Med. Phys.* **33** 3874–900

Koch N, Liu H H, Starkschall G, Jacobson M, Forster K, Liao Z, Komaki R and Stevens C W 2004 Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: I. Correlating internal lung motion with skin fiducial motion *Int. J. Radiat. Oncol. Biol. Phys.* **60** 1459–72

Lippmann R P 1987 An introduction to computing with neural nets *IEEE ASSP Mag.* **4** 4–22

Mageras G S *et al* 2004 Measurement of lung tumor motion using respiration-correlated CT *Int. J. Radiat. Oncol. Biol. Phys.* **60** 933–41

Murphy K 2003 Kalman filter toolbox for Matlab http://www.ai.mit.edu/murphyk/Software/Kalman/kalman.html

Murphy M and Dieterich S 2006 Comparative performance of linear and nonlinear neural networks to predict irregular breathing *Phys. Med. Biol.* **51** 5903–14

Murphy M J 2004 Tracking moving organs in real time *Semin. Radiat. Oncol.* **14** 91–100

Murphy M J, Jalden J and Isaksson M 2002 Adaptive filtering to predict lung tumor breathing motion during image-guided radiation therapy *Proc. 16th Int. Congress on Computer-Assisted Radiology and Surgery* pp 539–44

Nabney I and Bishop C 2003 Netlab neural network software http://www.ncrg.aston.ac.uk/netlab/

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

Ruan D, Castanon D A, Sharp G and Jiang S 2003 Real-time tumor tracking with interactive multiple model filter *CenSSIS*

Ruan D, Fessler J A and Balter J M 2007 Mean position tracking of respiratory motion *Med. Phys.* submitted

Ruan D, Fessler J A, Balter J M and Sonke J-J 2006 Exploring breathing pattern irregularity with projection-based method *Med. Phys.* **33** 2491–9

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

Schweikard A, Shiomi H and Adler J 2004 Respiration tracking in radiosurgery *Med. Phys.* **31** 2738–41

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

Tsunashima Y, Sakae T, Shioyama Y, Kagei K, Terunuma T, Nohtomi A and Akine Y 2004 Correlation between the respiratory waveform measured using a respiratory sensor and 3D tumor motion in gated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **60** 951–8

Vedam S S, Keall P J, Docef A, Todor D A, Kini V R and Mohan R 2004 Predicting respiratory motion for four-dimensional radiotherapy *Med. Phys.* **31** 2274–83

Vedam S S, Kini V R, Keall P J, Ramakrishnan V, Mostafavi H and Mohan R 2003 Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker *Med. Phys.* **30** 505–13

Wu H, Sharp G C, Salzberg B, Kaeli D, Shirato H and Jiang S B 2004 A finite state model for respiratory motion analysis in image guided radiation therapy *Phys. Med. Biol.* **49** 5357–72