Investigation of a breathing surrogate prediction algorithm for prospective pulmonary gating

Benjamin M. White

University of Missouri–Columbia, Columbia, Missouri 65211 and Washington University of St. Louis School of Medicine, St. Louis, Missouri 63110

Daniel A. Low,^{a)} Tianyu Zhao, Sara Wuenschel, Wei Lu, James M. Lamb, Sasa Mutic, Jeffrey D. Bradley, and Issam El Naqa

Washington University of St. Louis School of Medicine, St. Louis, Missouri 63110

(Received 1 June 2010; revised 27 January 2011; accepted for publication 29 January 2011; published 24 February 2011)

Purpose: A major challenge of four dimensional computed tomography (4DCT) in treatment planning and delivery has been the lack of respiration amplitude and phase reproducibility during image acquisition. The implementation of a prospective gating algorithm would ensure that images would be acquired only during user-specified breathing phases. This study describes the development and testing of an autoregressive moving average (ARMA) model for human respiratory phase prediction under quiet respiration conditions.

Methods: A total of 47 4DCT patient datasets and synchronized respiration records was utilized in this study. Three datasets were used in model development and were removed from further evaluation of the ARMA model. The remaining 44 patient datasets were evaluated with the ARMA model for prediction time steps from 50 to 1000 ms in increments of 50 and 100 ms. Thirty-five of these datasets were further used to provide a comparison between the proposed ARMA model and a commercial algorithm with a prediction time step of 240 ms.

Results: The optimal number of parameters for the ARMA model was based on three datasets reserved for model development. Prediction error was found to increase as the prediction time step increased. The minimum prediction time step required for prospective gating was selected to be half of the gantry rotation period. The maximum prediction time step with a conservative 95% confidence criterion was found to be 0.3 s. The ARMA model predicted peak inhalation and peak exhalation phases significantly better than the commercial algorithm. Furthermore, the commercial algorithm had numerous instances of missed breath cycles and falsely predicted breath cycles, while the proposed model did not have these errors.

Conclusions: An ARMA model has been successfully applied to predict human respiratory phase occurrence. For a typical CT scanner gantry rotation period of 0.4 s (0.2 s prediction time step), the absolute error was relatively small, 0.06 ± 0.02 s at peak inhalation and 0.05 ± 0.04 s at peak exhalation. The application of the ARMA model for prospective pulmonary gating has been demonstrated. © 2011 American Association of Physicists in Medicine.

[DOI: [10.1118/1.3556589](https://doi.org/10.1118/1.3556589)]

Key words: 4DCT, quiet respiration, prospective gating

I. INTRODUCTION

One of the principal applications of four dimensional computed tomography (4DCT) has been to provide breathing phase-gated image datasets for treatment planning.^{1,2} The images are typically acquired while the patient undergoes quiet or coached respiration. The image datasets are used to characterize the motion of the tumor and normal lung tissues due to breathing. A challenge in respiratory 4DCT is the lack of breathing amplitude and period reproducibility. This can lead to the acquisition of images that provide either redundant motion information or are acquired during an irregular breath and subsequently do not provide the data necessary for breathing motion characterization.^{3,4} The response in these cases is to repeat the scans, increasing the dose to the patient. A prospective gating algorithm that considered

breathing cycle irregularities and acquired CT data only when the breathing cycle was in user-specified phases, including assessing whether the breath was representative of the patient's normal breathing, would be more accurate and useful for 4DCT acquisition.

In 2005, Low *et al.*⁵ proposed a motion model that used tidal volume and airflow as surrogates for the breathing cycle. Their CT acquisition protocol used ciné mode CT, acquiring 25 CT scans over a 15 s period at each couch position using a reduced mA s to reduce the total CT dose. The relatively large number of scans was selected in order to assure that the patient breathed at least one representative breath during the scan acquisition. Because Low *et al.*⁵ were developing and evaluating a breathing motion model for quiet respiration, coaching was not used.

While the ciné protocol provided adequate image data for

registration and subsequent motion modeling, there remained two issues that limited its utility. First, some scans were acquired during times when the breathing cycle was irregular. This yielded images that were not useful but had still irradiated the subject. Second, the lack of synchronization between the breathing phase and scan acquisition meant that scans could not be acquired at specific phases, such as peak inhalation and exhalation, when breathing motion artifacts were minimal. In principal, prospective pulmonary gating could be used to overcome both limitations.

Prospective pulmonary algorithms have been developed to support phase-based linear accelerator gating using different methods including neural networks,^{6,7} a sinusoidal model,⁸ local regression,⁹ root mean square of a periodic signal,¹⁰ mean positional tracking,¹¹ elliptical shape tracking,¹² and recently an interacting multiple model with a Kalman filtering.¹³ Prediction errors have been the primary metrics used to evaluate prospective gating algorithms.¹⁴

Neural network algorithms use multiple interconnected artificial nodes to model complex biological relationships. These algorithms are highly nonlinear and require a training period for the algorithm to learn which outputs are yielded from particular input combinations. The primary criticism of neural network algorithms is the extensive training period required to cover the typical range of outputs. Kakar *et al.*⁶ used a five-layer neural network combined with fuzzy logic to develop a model that would associate breathing input at an earlier phase with a later phase. They tested the algorithm with 11 patients for coached and uncoached breathing with results showing greater root mean square error (RMSE) for the uncoached cases. In 2008, Murphy and Pokhrel⁷ investigated the need to apply a filter to the neural network. The study incorporated a least mean-squares adjustment to the neural network input from training data to act as a fast back-propagation update (adaptive filtering). This allowed the layers (stages) of the network to be reduced to 2. The algorithm was tested on 30 patient datasets that used both free breathing and coaching. The results showed that without a prediction time step, the algorithm was able to reconstruct the signal exactly, but had greatly increasing error for prediction times from 100 to 500 ms due to the lack of signal reproducibility.

The periodic nature of breathing has led some investigators to approximate the breathing cycle as a sinusoid. Vedom *et al.*⁸ investigated fitting a sinusoidal curve to training data (sinusoidal model) and compared it to a least mean-squares algorithm (adaptive filter model). Five patients were incorporated into the study and each participated in multiple studies so that over 60 sessions were collected with 30 s duration for coached, uncoached, and visual feedback breathing. They concluded that the adaptive filter model had smaller prediction errors than the sinusoidal model and coached breathing had smaller prediction errors than uncoached breathing.

Ruan *et al.*^{9–12} conducted several studies for predicting respiratory motion. In 2007, 12 patients were evaluated for irregularities in breathing traces by comparing the RMSE with a modified cosine signal fit in a least-squares sense.⁹ This method was able to reproduce the shape of the breath-

ing trace but failed to account for breathing irregularities. The group next investigated a local regression method and compared it to a neural network with a Kalman filter.¹⁰ They concluded that the local regression method had smaller prediction errors for long prediction time steps with low sampling rates. The mean position between the peak inhalation and peak exhalation phases was tracked in an attempt to define an average breathing amplitude.¹¹ In this study, the breathing amplitude was graphed against the amplitude at a later time corresponding to a delay proportional to the uniform sampling interval and the discrete time lag. The graph appeared as an ellipse when a sine wave was used to simulate the breathing cycle and had a similar elliptical shape for actual breathing cycles. The ellipse was then used to predict the breathing phase for one breath. The most recent work in this series conducted in 2009 expanded on the previous work to manage breathing irregularities.¹² There were three stages in their process: Real time phase estimation with ellipse tracking, pattern estimation, and filtering. The model was tested on simulated data and 12 patient datasets. While their results were good for the simulated data, patient data had phase prediction errors as large as 50% with significant errors in amplitude.

In 1963, Priban¹⁵ showed that the variation between breath cycles was not entirely a random phenomenon. This work formed the basis for the development of a proper model to predict respiratory motion. For this model, the breathing cycle was a measurement of the tidal volume, allowing the signal to be considered a discrete or finite time signal. Respiration could be estimated using a linear function with respect to previous data samples within the signal.

The goal of the current study is to test an autoregressive moving average (ARMA) approach to predict the respiration amplitude and phase for a clinically relevant range of prediction steps. The algorithm was developed and tested using measured patient breathing cycles. The ARMA model is one of the numerous linear forecasting models that was shown to operate successfully on time series applications.¹⁶

While the ARMA model has not been previously applied to radiation oncology breathing predictions, it has been investigated in magnetic resonance imaging (MRI) image reconstruction.¹⁷ Time series prediction models, such as the ARMA model, should have as few parameters as possible to reduce sources of error.¹⁶ However, the model should also have enough parameters to adequately describe the breathing signal. The main advantage of the ARMA model is in its use of relatively simple autoregressive (AR) and moving average (MA) terms to induce correlation between current and previous data while simultaneously incorporating weighted averages of the previous data. This model will likely perform better than individually using AR and MA terms by utilizing fewer parameters to describe the signal while maintaining sufficient simplicity to avoid overfitting.

II. METHODS

This work was developed to support breathing motion modeling research that utilizes tidal volume as the breathing

surrogate. The surrogate measures the breathing trace as a series of discrete volumetric measurements spaced 10 ms apart. A prediction model was developed to predict the surrogate measured signal amplitude and, consequently, respiratory phase at an advanced time. To quantitatively assess the success of the prediction model in accurately predicting respiratory phase, a comparison was made between the predicted and the observed phases.

II.A. ARMA

The ARMA model is a combination of autoregressive and moving average terms. The autoregressive term represents a predicted signal amplitude A at time $t + \Delta t$, $A_{t+\Delta t}^{\text{AR}}$, as an all pole (rational function with constant numerator) infinite impulse response filter,¹⁶

$$A_{t+\Delta t}^{\text{AR}} = c + \varepsilon_{t+\Delta t} + \sum_{i=1}^p \varphi_i A_{t-(i-1)\Delta t}, \quad (1)$$

where p is the AR model order, φ_i are the model parameters, A_t is the current signal amplitude, $\varepsilon_{t+\Delta t}$ describes the additive white Gaussian noise (AWGN) output error, and c is a constant. This autoregressive function is responsible for the phase of the prediction signal amplitude at time $t + \Delta t$ with respect to the signal amplitude at time t . The moving average term represents the predicted signal amplitude, $A_{t+\Delta t}^{\text{MA}}$, as an all zero (nonrational function) finite impulse response (FIR) filter,¹⁶

$$A_{t+\Delta t}^{\text{MA}} = \mu + \varepsilon_{t+\Delta t} + \sum_{i=1}^q \theta_i \varepsilon_{t-(i-1)\Delta t}, \quad (2)$$

where q is the MA model order, θ_i are the model parameters, $\varepsilon_{t+\Delta t}$ describes the AWGN output error, $\varepsilon_{t-(i-1)\Delta t}$ corresponds to the input white noise errors, and μ is an expectation value, which was set to 0 for this study. The moving average modeled the linear regression of the predicted and observed signal amplitude. Both terms are incorporated into the ARMA model as follows:¹⁶

$$A_{t+\Delta t}^{\text{ARMA}} = c + \varepsilon_{t+\Delta t} + \sum_{i=1}^p \varphi_i A_{t-(i-1)\Delta t} + \sum_{i=1}^q \theta_i \varepsilon_{t-(i-1)\Delta t}. \quad (3)$$

The ARMA model in Eq. (3) is characterized by the doublet (p, q) , where p is the order of the autoregressive term and q is the order of the moving average term.

II.B. ARMA parameter optimization

The Akaike information criterion (AIC) was used to determine the optimal combination of zeros and poles for the ARMA model acting on a breathing trace.¹⁸ The AIC is an analysis tool that measures how well a statistical model represents a physical system. The AIC tool consists of two terms; the number of parameters in the system $k = (p + q + 1)$ and the maximum likelihood L ,¹⁸

$$AIC = 2k - 2 \ln(L). \quad (4)$$

The least-squares definition of maximum likelihood is described as¹⁹

$$L = \prod_{i=1}^n \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp \left(- \sum_{i=1}^n \frac{(A_i^{\text{pred}} - A_i^{\text{obs}})^2}{2\sigma_i^2} \right), \quad (5)$$

where A_i^{pred} is the predicted i th signal amplitude, A_i^{obs} is the observed i th signal amplitude, and σ_i^2 is the variance at the i th sample. Taking the natural log of Eq. (5) yields¹⁸

$$\begin{aligned} \ln(L) &= \ln \left[\prod_{i=1}^n \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \right] - \frac{1}{2} \sum_{i=1}^n \frac{(A_i^{\text{pred}} - A_i^{\text{obs}})^2}{2\sigma_i^2} \\ &= C - \frac{1}{2} \chi^2, \end{aligned} \quad (6)$$

where the first term can be called C , which is considered as a constant if and only if the observations in the dataset are consistent, while the predictions vary (i.e., different model parameters), and χ^2 is the commonly recognized chi-squared expression. As C only varies with the observations, it can be omitted as the intention of using the AIC was to investigate the optimal combination of p and q . Substituting Eq. (6) into Eq. (4) yields a straightforward expression for the AIC in terms of the number of parameters in the system and chi-squared.¹⁸

The final prediction error (FPE) was used as a second metric to identify if the model order was sufficient to describe the breathing signal. In this case, the FPE was incorporated in conjunction with the AIC to aid in the optimization of the number of ARMA parameters. For large sample sizes, the FPE is defined as¹⁹

$$\text{FPE} = f \left(1 + \frac{2d}{m} \right), \quad (7)$$

where f was represented by a Euclidean norm (loss function), d is the number of model parameters ($p + q$), and m is the number of samples in the dataset, which was, on the average, over 45 000 samples. Since p and q are independent variables in the ARMA model, the Euclidean norm will have a Rayleigh distribution (chi distribution with two independent variables). This behavior will display an asymptote in the FPE in a fashion similar to the AIC. The optimal model order is the order that provides the FPE value closest to the asymptote.

II.C. Hamming window filter

Breathing cycles often drift slowly in time, causing difficulties for some predictive algorithms. The breathing trace was considered to have two components; a fast varying signal superimposed on a slowly varying signal (drift). The difference in frequency spectral characteristics of these components was used to design a real-time filter using a FIR design. In order to avoid the spectral leakage problem in the Fourier transform when dealing with discrete signals and ringing effects due to sudden truncations, the signal was mul-

multiplied by a window function to modulate the signal by a single sinusoidal cycle to systematically taper the signal amplitude. This process significantly reduced spectral leakage and ringing effects. The filter selected here to remove slow drift was the Hamming window filter.²⁰ The window output signal at sample i was defined by

$$w(i) = 0.54 - 0.46 \cos\left(\frac{2\pi^*i}{N}\right), \quad 0 \leq i \leq N, \quad (8)$$

where N was the number of samples in the window length. The window length used in this study was 0.1 s or ten samples, selected after testing a range of window lengths on the patient data.

II.D. Signal normalization and gating

The measured signals were proportional to the diaphragmatic expansion during respiration. The signal amplitude was normalized to the 85th percentile tidal volume of the training period to provide a better qualitative view of respiration amplitude and phase. To reduce small-scale undulations of the predicted breathing trace and, consequently, improve the accuracy of the predicted breathing phases, the ARMA prediction curve was smoothed using a fifth order polynomial applied over a 0.5 s sliding window centered on the point being smoothed. The location of peak inhalation was identified for each breath by finding where the slope of the curve changed sign. To avoid inadvertently identifying false peaks, a minimum amplitude difference from the previous peak exhalation was also used when identifying an inhalation peak. The threshold amplitude difference was selected as 30% of the normalized average signal amplitude. The method for identifying peak exhalation was identical to the method to identify inhalation. In order to compare results for different patients, each breathing cycle was evaluated using the signal amplitude percentiles. Normal signal amplitude, which correlates to quiet respiration, was defined as the difference between the 85th and 5th tidal volume percentile. Maximum exhalation was defined at the fifth percentile.

While the ARMA model could be used to determine the peak time and signal amplitude, it was assumed that abnormal breathing peaks, defined either by amplitude or period, would be excluded when gating. A training period for the model was necessary to determine the percentile used for signal amplitude normalization and subsequent gating of the remaining signal. Peaks with a magnitude and period in excess of 1.5 standard deviations of the mean magnitude and period defined in the training period, respectively, were designated as abnormal. The gating window was applied to the filtered breathing trace.

II.E. Data collection and phantom study

The breathing cycle data were acquired using a cylindrical bellows-shaped tube (Lafayette Instrument Co., Lafayette, IN, model 76513NM10) wrapped around the patient's abdomen. The belt was sealed so the air pressure decreased and increased during inhalation and exhalation, respectively. The

air pressure was measured by a pressure transducer that provided a voltage signal proportional to the internal air pressure and the voltage signal was monitored using an analog-to-digital converter that was sampled every 10 ms. The voltage signal from this apparatus has been previously shown to be proportional to tidal volume.^{21,22} The bellows system was used to acquire breathing cycle data from 47 patients while they underwent research 4DCT procedures. Each patient was requested to breathe normally and quietly.

The ARMA algorithm was compared against a commercial prospective breathing algorithm that is used to identify breathing peaks in real time and is installed in a commercial Philips Brilliance 64 slice CT simulator. Because the prospective gating algorithm was being tested and not image acquisition, the images were not analyzed. The scanner required a real-time breathing signal for analysis. To provide these data, the bellows system was attached to a dynamic phantom system developed in our department.²³ The phantom was moved by a distance proportional to the breathing tidal volume. The phantom had submillimeter dynamic positioning accuracy, corresponding to a breathing cycle simulation accuracy of less than a few percent.²³ The CT scanner was operated using the prospective pulmonary axial protocol. The bellows was exercised by the stage during CT acquisition to provide the real-time patient breathing signals. The recorded breathing signal and the tags associated with peak inhalation and peak exhalation phases were exported from the CT scanner at the conclusion of the session. The prediction time step of the commercial algorithm was a fixed 240 ms.

II.F. Patient data

Three of the 47 patient datasets were reserved to determine the order of the autoregressive term (p) and the order of the moving average term (q). Each was selected because it represented the range of breathing frequencies and amplitudes. One dataset had both an amplitude and frequency near the average of the 47 datasets, one had a relatively high frequency and small amplitude, and one had a relatively low frequency and large amplitude. After determination of the ARMA model parameters, these patient breathing datasets were no longer used in the study. The remaining 44 patients were used to test the model. To compare the model to the commercial peak detection prediction, 35 of these patients were selected for use with the CT scanner. The sampling frequency for the ARMA patient datasets was set at 100 Hz. The sampling frequency for the commercial algorithm comparison patient datasets was set at 500 Hz.

II.G. Prediction process

The patient breathing dataset was first filtered with a Hamming window filter and the first 30 s of the signal were used to train the ARMA algorithm. The training period length was determined by increasing the period from 10 to 120 s in 10 s increments. The coefficients of the ARMA polynomials, the volume percentiles, and the phase statistics were determined within the training period. The ARMA sig-

FIG. 1. Workflow of the prediction process.

nal was calculated using the model coefficients and a selected prediction time step while the filtered signal moved in time. When peaks were detected, they were included if they matched a predefined gating window criterion of 1.5 standard deviations from the means of both the period and the amplitude. The prediction error was evaluated for prediction time steps from 0.05 to 1.0 s. The overall workflow of the prediction model development is shown in Fig. 1.

The commercial algorithm was proprietary, so the prediction method for this model was not known. When the ARMA model prediction was compared against the commercial prediction algorithm, a prediction time step of 240 ms was used to coincide with the commercial prediction time step.

II.H. Quantifying prediction error

To aid in interpreting the results, one breath cycle from patient 6 is shown in Fig. 2 where peak inhalation occurs near 0.9 s and the nomenclature used in this work is shown. The sample at t_i is the current time and Δt is the prediction time step, which is the amount of time ahead for which the prediction is made. The algorithm is designed to be used to predict the time at which the patient would be at a specific breathing phase.

FIG. 2. Illustration of the nomenclature used in this work. This is the peak of a single inhalation. The solid and dotted curves indicate the measured and predicted breathing signals, respectively.

An example of peak inhalation is selected to evaluate the algorithm performance. If the algorithm was used to acquire a CT scan at peak inhalation, the prediction time step would be set to half the CT rotation time plus the system latency and time to account for model errors. Once peak inhalation had been predicted by the ARMA model, the scanner would be instructed to acquire images at half of a full gantry rotational period ($0.5 * t_{360^\circ}$) before the predicted peak. The start and stop times are labeled in Fig. 2 as t_{on} and t_{off} , respectively. The minimum duration of the scan is equal to $t_{off} - t_{on}$, which is the time it takes the scanner to perform a full 360° rotation (t_{360°). The difference between the predicted and actual peaks is termed e_t . Positive and negative errors correspond to predicted peaks occurring after and before the actual peak, respectively. If $e_t > 0$, the system would begin scanning but would not have acquired 180° prior to peak inhalation. The acquired image would be suboptimal and the image acquisition might need to be repeated, leading to excess delivered dose. If $e_t < 0$, the scanner might be left on until the patient passed peak inhalation and $t_{360^\circ}/2$ further. Therefore, the error in peak position needed to be as small as possible.

Model performance was evaluated to determine how often the predicted peak was within half the scanner rotational period from the measured peak. In other words, how often would the observed peak occur between t_{on} and t_{off} when the interval $t_{off} - t_{on}$ was centered about the predicted peak. The prediction error relative to the scanner rotational period was represented by η and characterized by

$$\eta = \frac{\Delta t - e_t}{t_{360^\circ}}. \quad (9)$$

There are three values of η that are of particular interest: $\eta = 0, \frac{1}{2}, 1$. In the absence of prediction error ($e_t = 0$), i.e., the prediction peak was in the exact location as the observed peak, the value of η will be $\frac{1}{2}$. If the prediction was too early such that the prediction time error is equal in magnitude to the prediction time step ($e_t = \Delta t$, so $\eta = 0$), the scanner will not capture the desired peak in a single rotation and will need additional rotations to capture the peak. Conversely, the other extreme has the prediction time error in the opposite direction ($e_t = -\Delta t$). In this case, $\eta = 1$, so it would not be possible to capture the measured peak as it occurred much earlier than

FIG. 3. Results of an AIC test for a model development patient.

the predicted peak. The ARMA model was evaluated at increasing prediction time steps, and as long as 95% of the predictions had an interval of $0 < \eta < 1$, the prediction would be more than adequate for image acquisition of the desired respiratory phase.

III. RESULTS

Figure 3 shows an example of an AIC curve for the normal breathing frequency model development patient. The other two patients used for the model development displayed similar trends. The results from the AIC test exhibit an asymptote at -16 with the closest value to the asymptote occurring at high model order combinations of p and q . Selection of the order combination was done by evaluating the AIC, FPE, and the percent fit (less than 5% variation from the original dataset) for a small prediction step of 0.01 s (one sample). Table I summarizes the results of the AIC, FPE, and percent fit for three combinations of the model order for the patient shown in Fig. 3. The selected ARMA parameters of $p=2$ and $q=5$ provided the smallest number of model parameters at the smallest AIC and FPE values which made them the optimal selections.

Using an ARMA ($p=2$, $q=5$) model, the remaining 44 patient datasets were analyzed as a function of prediction time steps in order to find how long in time the algorithm was able to accurately predict specific breathing phases. Fig-

ure 4 shows the prediction time error for predicting peak inhalation (a) and exhalation (b) as a function of the prediction time step for each patient.

The magnitude of the time error typically increased as a function of the prediction time step. While the magnitude of the prediction time error increased with increasing prediction time step, the same relationship could not be seen for the direction of the prediction time error. Furthermore, Fig. 4 shows the greater variability of the prediction time error for maximum exhalation. Patients typically spend more time at exhalation, so the exhalation peak is less well defined than the inhalation peak. For most patients, a prediction time step of 0.4 s resulted in prediction time errors of greater than 0.4 s, implying that if the scanner was going to be triggered such that the midscan time was at peak inhalation, the 0.4 s prediction time did not provide sufficient latitude to ensure that the scanner would image peak inhalation. Figure 5 shows the mean and standard deviation of η as a function of the prediction time step. Peak inhalation has smaller η values than the peak exhalation with considerably less variability for prediction time steps greater than 0.3 s. Figure 6 compares the prediction error for the commercial prediction algorithm and the ARMA algorithm (240 ms prediction time step) at peak inhalation and exhalation for the 35 patient datasets.

In addition to evaluating peak prediction errors, there were two cases that were examined; a false breath and a missed breath. A false breath was defined as a predicted maximum inhalation and corresponding maximum exhalation tag occurring within a measured respiratory cycle. A missed breath was defined as a lack of prediction tags for a measured breath. One of the differences between the ARMA model and commercial model evaluations was the inability of the commercial model to detect small breathing amplitudes in nine patients. In addition, the commercial model inaccurately identified breaths. The commercial algorithm missed four valid breathing cycles and falsely recorded thirteen cycles with four patients having more than one false peak. One patient dataset had both missed and falsely iden-

TABLE I. AIC, FPE, and percent fit results for three particular ARMA parameter combinations for the process outlined in Fig. 1 and Eq. (3).

(p, q)	AIC	FPE	Percent fit ($<5\%$)
			(%)
(1,0)	-10.768	1.472×10^{-5}	68.2
(1,1)	-11.975	4.575×10^{-6}	68.76
(2,5)	-15.654	1.521×10^{-7}	84.51

FIG. 4. Average prediction time error e_t in seconds for peak inhalation (a) and peak exhalation (b) of the ARMA (2,5) prediction as a function of the prediction step. Data are shown for the 44 evaluated patients.

tified peaks. The ARMA (2,5) algorithm did not falsely identify or miss a peak. Figure 7 compares the results of the commercial algorithm and ARMA model predictions. For Fig. 7(a), breath cycle was missed by the commercial algorithm at 37.8 and 40.4 s. Figure 7(b) shows an example of a false breath that was recognized starting at 38.8 s and ending at 39.3 s. A summary of the number of false breaths and missed breaths for the commercial algorithm and the ARMA algorithm is reported in Table II.

IV. DISCUSSION

For a prediction time step equal to half of the time required to perform a full 0.4 s CT gantry rotation (equivalent to 0 s latency and no time to limit the probability that $e_t > 0$), the CT scanner would have sufficient time to have rotated 180° at the time of peak inhalation. With current CT

scanners capable of completing a full gantry rotation in 0.4 s or less, the minimum prediction time step needed to allow gating would be 0.2 s. Using Fig. 4 as a reference, when the prediction step was 0.2 s, the ARMA (2,5) showed an error of 0.06 ± 0.02 s for the 44 tested patients. At peak exhalation, the error was 0.05 ± 0.04 s. The sign of the error was important because if $e_t > 0$, the predicted peak time occurred later than the actual time so images would not correspond to the selected phase. If $e_t \leq 0$, the beam on time would be significantly greater to acquire the images. By keeping 95% of the errors less than 0.2 s for a 0.4 s gantry rotation period, the desired peak will most likely be captured in the scanning window and extra beam on time will be reduced in the process. The maximum prediction time step that would have a 95% probability of capturing the peak was 0.3 s with a 95% confidence criterion for both maximum inhalation and maximum exhalation. The value of η had greater variability for maximum exhalation than inhalation because exhalation exhibited a shallower peak.

In the algorithm comparison, the ARMA (2,5) outperformed the commercial algorithm in every case. Furthermore, the ARMA (2,5) model did not miss or falsely record a breath cycle while the commercial algorithm had numerous instances of both. For peak inhalation, the error in the peak inhalation was 0.03 ± 0.5 s for the commercial algorithm and 0.001 ± 0.01 s for ARMA (2,5). For peak exhalation, the errors were 0.36 ± 1.21 and 0.003 ± 0.01 s for the commercial algorithm and the ARMA (2,5), respectively. The prediction time step for these patients was 0.24 s. Because the

FIG. 5. Average η with the 95% confidence window for 44 patients at increasing prediction time steps. The solid line curve represents maximum inhalation, while the dashed line curve represents maximum exhalation.

TABLE II. Number of patients with missed or falsely tagged breath cycles for both prediction models.

Prediction model	Missed breath	False breath
Commercial algorithm	4	13
ARMA	0	0

FIG. 6. Histograms of the 240 ms prediction error for peak inhalation and exhalation for the ARMA model and commercial algorithm for 35 patients. (a) Peak inhalation for the commercial algorithm. (b) Peak exhalation for the commercial algorithm. (c) Peak inhalation for the ARMA algorithm. (d) Peak exhalation for the ARMA algorithm.

FIG. 7. An example of a patient dataset showing the comparison of the commercial algorithm and the ARMA model predictions.

commercial algorithm had error significantly greater than the prediction time, it was shown to be inappropriate for prospective gating.

V. CONCLUSIONS

An ARMA model, in conjunction with Hamming filtering, has been optimized to predict human respiratory phase and tested at peak inhalation and exhalation phases. The validity and maximum prediction time step of the model were investigated with 44 patient breathing traces measured using a cylindrical bellows-shaped tube. For a gantry rotation period of 0.4 s, the minimum required prediction time step would be 0.2 s. At this prediction time step, the model had relatively small errors. This demonstrates that the model can be used with modern fast CT scanners. The small errors will allow additional time to account for system latency greater than the latency of the CT scanner used in this study. For comparison with a commercial algorithm, the Washington University 4D

Phantom was used to provide real-time breathing signals to a Philips Brilliance 64 slice scanner. Tags representing the peak inhalation and peak exhalation were exported from the operator consol and were compared to the results of ARMA (2,5) for the same breathing trace. The ARMA (2,5) was found to be significantly better than that commercial algorithm. Based on these results, prospective pulmonary gating has been shown to be possible for modern 4D CT simulation.

^{a)}Author to whom correspondence should be addressed. Electronic mail: dlow@mednet.ucla.edu; Also at Department of Radiation Oncology, University of California, Los Angeles 200 Medical Plaza, Suite B265 Los Angeles, CA 90095; Telephone: (310) 983-3205 (office).

¹S. Mutic *et al.*, "Quality assurance for computed-tomography simulators and the computed-tomography-simulation process: Report for the AAPM Radiation Therapy Committee Task Group No. 66," *Med. Phys.* **30**(10), 2762–2792 (2003).

²C. Ritchie *et al.*, "Predictive respiratory gating: A new method to reduce motion artifacts on CT scans," *Radiology* **190**, 847–852 (1994).

- ³A. Berson *et al.*, "Clinical experience using respiratory gated radiation therapy: Comparison of free-breathing and breath-hold techniques," *Int. J. Radiat. Oncol., Biol., Phys.* **60**(2), 419–426 (2004).
- ⁴T. Fox *et al.*, "Free breathing gated delivery (FBGD) of lung radiation therapy: Analysis of factors affecting clinical patient throughput," *Lung Cancer* **56**, 69–75 (2007).
- ⁵D. Low *et al.*, "Novel breathing motion model for radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **63**(3), 921–929 (2005).
- ⁶M. Kakar *et al.*, "Respiratory motion prediction using the adaptive neuro fuzzy inference system (ANFIS)," *Phys. Med. Biol.* **50**, 4721–4728 (2005).
- ⁷M. Murphy and D. Pokhrel, "Optimization of an adaptive neural network to predict breathing," *Med. Phys.* **36**(1), 40–47 (2009).
- ⁸S. Vedam *et al.*, "Predicting respiratory motion for four-dimensional radiotherapy," *Med. Phys.* **31**(8), 2274–2283 (2004).
- ⁹D. Ruan and J. Fessler, "Exploring breathing pattern irregularity with projection-based method," *Med. Phys.* **33**(7), 2491–2499 (2006).
- ¹⁰D. Ruan, J. Fessler, and J. Balter, "Real-time prediction of respiratory motion based on local regression methods," *Phys. Med. Biol.* **52**, 7137–7152 (2007).
- ¹¹D. Ruan, J. Fessler, and J. Balter, "Mean position tracking of respiratory motion," *Med. Phys.* **35**, 782–792 (2008).
- ¹²D. Ruan *et al.*, "Real-time profiling of respiratory motion: Baseline drift, frequency variation and fundamental pattern change," *Phys. Med. Biol.* **54**, 4777–4792 (2009).
- ¹³D. Putra *et al.*, "A multiple model approach to respiratory motion prediction for real-time IGRT," *Phys. Med. Biol.* **53**, 1651–1663 (2008).
- ¹⁴G. Sharp *et al.*, "Prediction of respiratory tumor motion for real-time image-guided radiotherapy," *Phys. Med. Biol.* **49**, 425–440 (2004).
- ¹⁵I. Priban, "An analysis of some short-term patterns of breathing in man at rest," *J. Physiol.* **166**, 425–434 (1963).
- ¹⁶G. Box, G. Jenkins, and G. Rinsel, *Time Series Analysis: Forecasting & Control*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1994).
- ¹⁷M. Smith *et al.*, "Application of autoregressive moving average parametric modeling in magnetic image reconstruction," *IEEE Trans. Med. Imaging* **5**(3), 132–139 (1986).
- ¹⁸H. Akaike, "Information theory and an extension of the maximum likelihood principle," *Proceedings of the Second International Symposium on Information Theory* (Akademiai Kiado, Budapest, 1973).
- ¹⁹L. Ljung, *System Identification: Theory for the User*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1999), pp. 197–224.
- ²⁰A. Oppenheim, R. Schaffer, and J. Buck, *Discrete-Time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1989), pp. 447–448.
- ²¹W. Lu *et al.*, "Comparison of spirometry and abdominal height as four-dimensional computed tomography metrics in lung," *Med. Phys.* **32**(7), 2351–2357 (2005).
- ²²R. Werner *et al.*, "Feasibility of a 4D CT breathing monitoring metric that provides tidal volume without the use of spirometry," *Med. Phys.* **37**(2), 615–619 (2010).
- ²³K. Malinowski *et al.*, "Development of the 4D phantom for patient-specific, end-to-end radiation therapy QA," *Proc. SPIE* **6510**, 65100E (2007).