A comparison of gantry-mounted x-ray based real-time target tracking methods

Tim Montanaro[1], Doan Trang Nguyen[1], Paul J Keall[1], Jeremy Booth[2], Vincent Caillet[1,2], Thomas Eade[2], Carol Haddad[2] and Chun-Chien Shieh[1,*]

[1]ACRF Image X Institute, Sydney Medical School, University of Sydney, Camperdown NSW, Australia, 2006

[2]Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards NSW, Australia 2065

[*]Corresponding author: andy.shieh@sydney.edu.au Rm 215, Biomedical Building, Central Avenue, Eveleigh, NSW 2015

ABSTRACT

Purpose: Most modern radiotherapy machines are built with a 2D kV imaging system. Combining this imaging system with a 2D-3D inference method would allow for a ready-made option for real-time 3D tumor tracking. This work investigates and compares the accuracy of four existing 2D-3D inference methods using both motion traces inferred from external surrogates and measured internally from implanted beacons.

Method: Tumor motion data from 160 fractions (46 thoracic/abdominal patients) of Synchrony traces (inferred traces), and 28 fractions (7 lung patients) of Calypso traces (internal traces) from the LIGHT SABR trial (NCT02514512) were used in this study. The motion traces were used as the ground truth. The ground truth trajectories were used in-silico to generate 2D positions projected on the kV detector. These 2D traces were then passed to the 2D-3D inference methods: interdimensional correlation, Gaussian probability density function (PDF), arbitrary-shape PDF, and the Kalman filter. The inferred 3D positions were compared with the ground truth to determine tracking errors. The relationships between

tracking error and motion magnitude, interdimensional correlation, and breathing periodicity index (BPI) were also investigated.

*Results***:** Larger tracking errors were observed from the Calypso traces, with RMS and 95th percentile 3D errors of 0.84-1.25 mm and 1.72-2.64 mm, compared to 0.45-0.68 mm and 0.74-1.13 mm from the Synchrony traces. The Gaussian PDF method was found to be the most accurate, followed by the Kalman filter, the interdimensional correlation method, and the arbitrary-shape PDF method. Tracking error was found to strongly and positively correlate with motion magnitude for both the Synchrony and Calypso traces and for all four methods. Interdimensional correlation and BPI were found to negatively correlate with tracking error only for the Synchrony traces. The Synchrony traces exhibited higher interdimensional correlation than the Calypso traces especially in the anterior-posterior direction.

*Conclusion***:** Inferred traces often exhibit higher interdimensional correlation, which are not true representation of thoracic/abdominal motion and may underestimate kV-based tracking errors. The use of internal traces acquired from systems such as Calypso is advised for future kV-based tracking studies. The Gaussian PDF method is the most accurate 2D-3D inference method for tracking thoracic/abdominal targets. Motion magnitude has significant impact on 2D-3D inference error, and should be considered when estimating kV-based tracking error.

Keywords: lung radiation therapy, tumor motion, tumor tracking, image guidance, kV imaging

1 INTRODUCTION

During radiotherapy, lung tumors exhibit complex and substantial motion[1]. The standard of care is to map the magnitude of the tumor trajectory during a pre-treatment 4D CT scan and irradiate the motion-encompassing region to ensure the tumor receives the correct dose. 4D CT based margins often underestimate lung tumor motion[2] and there is mounting evidence that 4D CT based margins can also lead to significant tumor underdose[3]. Real-time tumor tracking enables direct adaptation to intrafraction tumor motion using techniques such as gating, multileaf collimator (MLC) adaptation, and robotic couch[4, 5].

Stereoscopic systems, with more than one imager, such as CyberKnife, Novalis Tx and Mitsubishi/Hokkaido RTRT systems can provide direct 3D localisation through triangulation but have not achieved widespread adoption. These systems require high capital cost, with large modifications to the linacs[6]. Another method, incorporating the MV treatment imaging beam can also enable the triangulation of the target position[7, 8]. In practice however, the multi-leaf collimator blocks important components, such as the tumor or markers and the high energy beam reduces the contrast of these images[9].

Currently most modern radiotherapy machines are built with a 2D kV imaging system, which is used to acquired CBCT scans prior to treatment. Utilising this imaging system during treatment would allow for a ready-made option in real-time tumor tracking. This has already been implemented at a number of clinics and is a growing field of research[10, 11].

The major challenge with kV imaging based tracking is that the 3D tumor position needs to be inferred from the 2D imaging information. A number of 2D-3D inference approaches have been proposed for both fiducial marker-based tracking and markerless tracking. These include the interdimensional correlation (IDC) method[6], the Gaussian Probability Density Function (PDF) method[12], the arbitrary-shape PDF method[13], and the respiratory-based Kalman filter method[14]. The IDC method put forward most recently by Chung et al.[6], builds

an interdimensional correlation matrix based off the unambiguously resolved superior-inferior component. The CBCT data and the most recent 2D positions are used to determine the correlation coefficients. The Gaussian PDF[12] method uses recent 2D positions to build a 3D Gaussian PDF that describes the distribution of 3D target positions. For each kV image, the most likely 3D target position is determined by the 3D point with the highest PDF value along the backprojection of the 2D position. The arbitrary-shape PDF[13] method models the distribution of 3D target positions as a superposition of multiple exponential functions. This allows for asymmetric PDFs which can potentially perform better for cases with notable hysteresis motion. The final method, the Kalman filter[15] with respiratory prediction combines the measured 2D position with a respiratory prediction model that incorporates the semi-periodic nature of respiratory motion to determine a 3D position estimate.

Previous studies[6, 12] have used motion data inferred from external motion traces as ground truths. These types of traces will be referred to as "inferred traces". Inferred traces may not accurately represent the true internal motion of lung tumor, as the nature of the methods used to infer the motion from external signals may lead to inter-dimensionally correlated traces. This may result in lower than expected tracking errors since most 2D-3D methods exploit the interdimensional correlation of lung tumor motion. Alternatively, other studies examine kV-based 2D-3D inference for abdominal or prostate tumor motion where the motion amplitude is lower or more regular[16, 17]. Because lung tumors exhibit the most substantial and complex motion[1], the accuracy of 2D-3D inference may be worse than that for abdominal or prostate tumors. A comparison study of the 2D-3D inference methods for lung tumors using true internal 3D motion traces (referred to as internal traces) is thus essential. One way to acquire internal traces is by using the Calypso electromagnetic transponder beacons (Varian Medical System, Palo Alto, CA), which are implanted near the tumor and emit their 3D positions in real-time to an external detector.

The aim of this study is to quantify the accuracy of the abovementioned 2D-3D inference methods when applied to kV based intrafraction tracking. More specifically, the first aim is to investigate if the 2D-3D inference methods perform differently when applied to inferred traces and internal traces. The hypothesis is that internal traces more accurately represent the complexity of internal lung target motion, and would result in larger tracking errors than inferred traces. Results regarding this aim would advise whether it is important to use internal traces when testing kV-based tracking methods. The second aim is to compare the accuracy of the different methods and determine which of the methods are most suitable for intrafraction lung tumor tracking. 160 tumor trajectories estimated by a Cyberknife Synchrony system (Accuray, Sunnyvale, CA) at Georgetown University Hospital using a correlation model[1] were used as the inferred traces. 28 tumor trajectories recorded by a Calypso system from a first-in-world Calypso-guided multileaf collimator tracking lung stereotactic ablative radiotherapy (SABR) trial (NCT02514512)[18] were used as the internal traces. This is the first study to quantify the errors of kV-based tracking resulting from the different 2D-3D inference methods when applied to true internal 3D lung tumor traces. Section 2 describes the four existing inference methods investigated in this study, the clinical data used and explains the experimental procedure. The results are presented in Section 3 and Section 4 discusses the results and how these findings compare to previous studies as well as their impact to the medical physics community.

2 Materials and Methods

2.1 Inference Methods

The following sections provide an overview of the four 2D-3D inference methods investigated in this study. The readers are referred to the original publications for further technical details[6, 12-14]. In this paper, 2D position refers to the projected target position on the

kV imager, and 3D position refers to the actual target position in the 3D space. All of the methods were first initiated from a one minute cone-beam CT (CBCT) scan, either using the 2D target positions from the kV images or the 3D target positions observed from the 4D-CBCT reconstruction. During intrafraction tracking, each model continued to update itself using the most recent tracking history within a certain temporal window. The selection of the optimal temporal window of tracking history is described in Section 2.3.

2.1.1 The Interdimensional Correlation (IDC) method

The IDC method[6] exploits the correlation between motion components in the left-right (LR), superior-inferior (SI), and anterior-posterior (AP) directions. Since the SI motion is always unambiguously resolved on the kV image, it is possible to infer the LR and AP motions by assuming their linear dependencies on the SI motion. Previous studies have shown that SI motion magnitude is the largest of the dimensions for the majority of cases and the motion is predominantly linear[1]. For 93% of lung tumor cases tested in ref.[1] the principal power is linear. The IDC method can be mathematically described as:

$$\begin{pmatrix} \hat{x}(t_j) \\ \hat{z}(t_j) \end{pmatrix} = \boldsymbol{a}_j \hat{y}(t_j) + \boldsymbol{b}_j \hat{y}(t_j - \tau) + \boldsymbol{c}_j, \quad (1)$$

where $\hat{x}(t_j)$, $\hat{y}(t_j)$, and $\hat{z}(t_j)$ are the inferred LR, SI, AP positions for the current frame t_j, $\boldsymbol{a}_j$, $\boldsymbol{b}_j$, and $\boldsymbol{C}_j$ are the parameters (2 by 1 column vectors) describing the linear model, and τ is the phase shift parameter used to account for hysteresis. In this study, $\tau = 0.6$ seconds was used as suggested in the original work by Chung et al[6]. The parameters are determined by minimising the least squared errors between the projected 2D positions estimated from the model and the observed 2D positions on the kV images during CBCT (model initialization) or treatment delivery (continuous model update while tracking).

2.1.2 The Gaussian Probability Density Function (G-PDF) method

The Gaussian PDF (G-PDF) method[12] models the distribution of 3D target positions as a 3D Gaussian PDF. The parameters of the Gaussian PDF at frame j, i.e. the mean vector $\widehat{\boldsymbol{\mu}}_j$ and the covariance matrix $\widehat{\boldsymbol{\Sigma}}_j^{-1}$, are estimated by maximizing the likelihood of observing the 2D positions either during CBCT (model initialization) or within the temporal window of recent tracking history during treatment delivery (continuous model update while tracking). The Gaussian PDF estimated for frame j, $f_j(\boldsymbol{x})$, can be mathematically expressed as:

$$f_j(\boldsymbol{x}) = \exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\widehat{\boldsymbol{\mu}}_j\right)^T\widehat{\boldsymbol{\Sigma}}_j^{-1}\left(\boldsymbol{x}-\widehat{\boldsymbol{\mu}}_j\right)\right)/\sqrt{(2\pi)^3\left|\widehat{\boldsymbol{\Sigma}}_j\right|}. \quad (2)$$

As the covariance matrix encodes information about both the variance and correlation of the underlying random variables, the Gaussian PDF exploits both the ranges and interdimensional correlations of the target motion to infer the 3D position. The inferred 3D position, $\widehat{\boldsymbol{x}}_j$, is determined to be the 3D point with the highest probability along the backprojection of the 2D point from the kV imager:

$$\widehat{\boldsymbol{x}}_j = \mathrm{argmax}_{\boldsymbol{x}}\, f_j(\boldsymbol{x}),\ \ s.t.\ T_j(\boldsymbol{x}) - \boldsymbol{c}_j = \boldsymbol{0}, \quad (3)$$

where $T_j(\boldsymbol{x})$ is the 3D-2D transformation function, and $\boldsymbol{c}_j$ is the measured 2D position for frame j.

2.1.3 The Arbitrary-shape PDF (A-PDF) method

Similarly to the Gaussian PDF method, the arbitrary-shape PDF[13] method also determines the 3D target position as the 3D point with highest PDF value along the backprojection of the 2D position. Instead of a single Gaussian function, the PDF for the current frame j, $f_j(\boldsymbol{x})$, was modeled as a superposition of exponential functions built based on the previously observed 2D positions:

$$f_j(\boldsymbol{x}) = \gamma\prod_{i=j-N}^{j-1}\exp\left(-\left|T_i(\boldsymbol{x})-\boldsymbol{c}_i\right|_2^p/\left|t_j-t_i\right|\right), \quad (4)$$

where i is the index of the past frames, N is the number of frames within the optimal temporal window of recent tracking history (cf. Section 2.3), γ is the normalization constant,

and p is the power operator of the exponential functions and was chosen to be 0.1 as suggested in the original paper by Li et al.[13] In short, each exponential term describes the backprojected trace of a particular measured 2D position convolved with the exponential function with a p-norm distance from the trace as the exponent. The denominator $|t_j - t_i|$, measures the temporal difference between the current and the ith frame to assign a higher weight to the more recent frames as they likely represent the current target positions more closely. Equation 4 allows for complex and asymmetric PDFs that may capture the complexity of lung tumor motion such as hysteresis[19] better. The inferred 3D position is then solved by applying equation (3) as in the Gaussian PDF case.

2.1.4 The Kalman Filter method

The Kalman filter[15] is a common technique for combining partial measurements with prior knowledge to estimate the time evolution of the system. The respiratory-based Kalman filter approach was proposed for lung tumor tracking in the study by Shieh et al[14]. The concept is to combine the observed 2D position (partial measurement) and the assumption that target motion in the thoracic/abdominal region is semi-periodic (prior knowledge) to estimate the 3D target position. For each frame, the assumption of semi-periodicity is implemented by first making a spatial prediction that the distribution of target position is a 3D Gaussian function with its mean, $\boldsymbol{x}_{j,\mathrm{Predicted}}$, and covariance matrix, $\boldsymbol{\Sigma}_{\mathrm{j,Predicted}}$, being the mean and covariance of the previously tracked 3D positions of the same respiratory phase bin (1-10), within the temporal window of recent history. In this study, the respiratory phase was calculated based on the SI component of the measured 2D position using the real-time phase method proposed by Ruan et al.[20] The measured 2D position is modeled as a backprojected trace in the 3D space convolved with a Gaussian kernel that represents the measurement uncertainty. The 3D prediction distribution and the 2D measurement distribution are then combined to yield the final 3D estimate of target position $\hat{\boldsymbol{x}}_j$ by:

$$\hat{\boldsymbol{x}}_j = \boldsymbol{x}_{j,\text{Predicted}} + \boldsymbol{K}_j\left(\boldsymbol{c}_j - T_j\left(\boldsymbol{x}_{j,\text{Predicted}}\right)\right), \quad (5)$$

$$\boldsymbol{K}_j = \boldsymbol{\Sigma}_{\text{j,Predicted}}\boldsymbol{H}_j^T\left(\boldsymbol{H}_j\boldsymbol{\Sigma}_{\text{j,Predicted}}\boldsymbol{H}_j^T + \boldsymbol{R}\right)^{-1}, \quad (6)$$

where $\boldsymbol{K}_j$ is the Kalman gain that determines how much the final estimate is weighted towards the measurement, $\boldsymbol{H}_j$ is the Jacobian matrix describing the 3D-2D transformation function $T_j(.)$, and $\boldsymbol{R}$ is a 2 by 2 matrix describing the measurement uncertainty. In this study, $\boldsymbol{R}$ was set to be the identity matrix multiplied by the square of the kV detector pixel size. Before the start of intrafraction tracking, the Kalman filter was initialized by setting $\boldsymbol{x}_{j,\text{Predicted}}$ for each respiratory phase bin to be the observed target position from the 10-phase 4DCBCT reconstructed images, and setting $\boldsymbol{\Sigma}_{\text{j,Predicted}}$ to be the covariance of $\boldsymbol{x}_{j,\text{Predicted}}$.

2.2 Study design and patient data

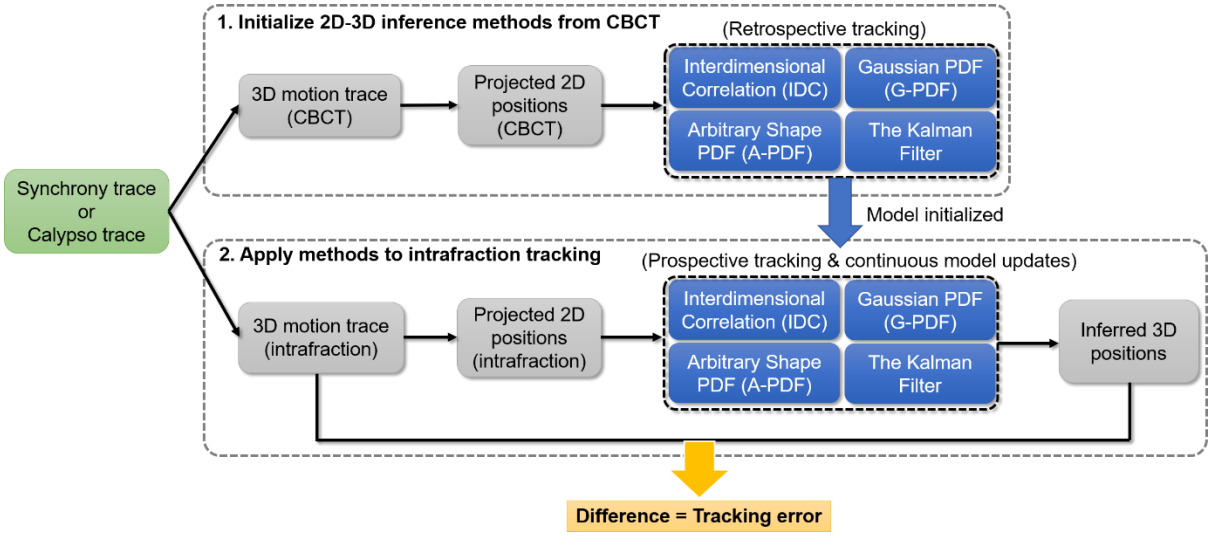


Figure 1. The flowchart of this study. The 3D motion trace refers to either the Synchrony traces (inferred traces) or the Calypso traces (internal traces).

The four 2D-3D inference methods were tested in silico retrospectively using both inferred traces and internal traces from thoracic/abdominal patients. Figure 1 summarizes the workflow of this study. For each of the cases investigated, the 3D motion trace was first separated into a CBCT session and an intrafraction session. 3D motion traces during both sessions were forward projected onto the kV detector coordinate to generate projected 2D positions. Note that this means the geometric match between the projected 2D positions and

the underlying 3D motion trace was exact, excluding any additional source of tracking errors such as 2D segmentation errors from the study. The results of the study thus represented the sole effect of each of the 2D-3D inference methods. The projected 2D positions during the CBCT session were first used to initialize each of the 2D-3D inference methods in a retrospective manner, i.e. the model was built by attempting to match all the projected 2D positions at the same time. Then, each of the initialized 2D-3D inference methods inferred the 3D positions from the intrafraction projected 2D positions in a prospective manner, i.e. each 2D position was passed to the method frame by frame, with the model constantly updating itself with every incoming 2D position. The tracking error was calculated as the difference between the inferred and actual 3D motion trace during intrafraction tracking.

For the inferred traces, 160 thoracic/abdominal tumor trajectories (46 patients) estimated by a Cyberknife Synchrony system (Accuray, Sunnyvale, CA) at Georgetown University Hospital using a correlation model[1] were used. From here on these traces are referred to as the Synchrony traces. Tumor locations include upper, middle, lower lung lobes, the hilum, and retroperitoneum. The duration of the trajectories ranged from 8 to 110 minutes. The mean (maximum) motion range was 2.5 mm (26.5 mm) in left-right (LR), 6.8 mm (56.6 mm) in superior-inferior (SI), and 3.3 mm (37.4 mm) in anterior-posterior (AP). The beginning one minute of each trajectory was used as the motion trace during CBCT, and was down-sampled from 25 Hz to 11 Hz to match the imaging frequency of a standard one minute CBCT scan. Trajectories between the one minute and two-minute time stamps were discarded to mimic the time gap between CBCT setup and the start of intrafraction kV acquisition, i.e. start of treatment. The rest of the trajectories (6-108 minutes) were then used as the intrafraction motion trace.

For the internal traces, 28 lung tumor trajectories (7 patients) recorded by the Calypso system from a first-in-world Calypso-guided multileaf collimator (MLC) tracking lung stereotactic

ablative radiotherapy (SABR) trial (NCT02514512; on-going)[18], the LIGHT SABR trial, were used. In this trial, inoperable stage I non-small cell lung cancer patients were inserted with the Calypso electromagnetic transponder beacons. The 3D positions of the beacons were measured in real-time from CBCT acquisition until the end of treatment delivery, and the motion of the centroid of the beacons were used for MLC adaptation. From here on these traces are referred to as the Calypso traces. Patients with tumors in the upper, middle, and lower lobes were included. The mean (maximum) motion range was 5.0 mm (13.39 mm) in left-right (LR), 12.9 mm (47.3 mm) in superior-inferior (SI), and 8.0 mm (27.2 mm) in anterior-posterior (AP). Prior to each radiotherapy fraction, a one minute pre-treatment CBCT was performed at 11 Hz over a 360-degree angular range. During treatment, intrafraction kV images were acquired at 5 Hz over two 200-degree treatment arcs, which lasted for about four minutes in total. The beacon centroid trajectories during CBCT and treatment delivery were used as the CBCT and intrafraction 3D motion traces, and were down-sampled to 11 Hz and 5 Hz to match the kV imaging frequencies.

The simulated 2D-3D geometry for both the study of inferred traces and the study of internal traces was set to match the kV imaging geometry of the LIGHT SABR trial. The source-to-isocenter distance (SID) was kept at 1000 mm. The source-to-detector distance (SDD) was 1500 mm during CBCT and 1800 mm during treatment delivery. The kV detector was laterally offset by 148 mm during CBCT (half-fan acquisition), and was centered during treatment delivery. Around 40% of the projected 2D positions during CBCT were outside the field-of-view (FOV) of the kV detector (397 mm by 298 mm) due to the half-fan offset, and were excluded from the initialization of each 2D-3D inference method. The gantry rotation speed was 6 degrees/s during CBCT and 1.6 degrees/s during treatment delivery.

2.3 Optimal Temporal Window Length of Recent Tracking History

All four 2D-3D inference methods were constantly updated with every incoming 2D position. At every frame, each of the methods rebuilds the underlying model using the most recent 2D positions within a certain temporal window. A shorter temporal window has a greater risk of overfitting the model with occasional abrupt motion, while a longer temporal window may fail to represent the most recent motion pattern. To allow for a fair comparison, the optimal temporal window for each method was determined according to the lowest root-mean-squared (RMS) values of the 3D tracking errors for the Calypso study. Temporal window length of 20, 40, 60, …, 140 seconds were investigated, which corresponded to 100, 200, 300, …, 700 intrafraction kV frames.

2.4 Impacts of motion parameters

This study also investigated how different motion parameters may affect tracking errors resulting from 2D-3D inference. For each intrafraction motion trace, the impacts of motion magnitude, interdimensional correlation, and breathing periodicity index (BPI) on tracking error in the LR or AP direction were investigated. The motion magnitude in each direction (LR, SI, AP) was defined to be the 5^{th} to 95^{th} percentile range. The LR/AP interdimensional correlation was calculated as the absolute value of the Pearson's correlation coefficient between the LR/AP motion and the SI motion. The BPI[21] was calculated for the LR and AP motion separately as the sum of the five largest Fourier energies divided by the integral of the entire Fourier energy spectrum. A higher BPI value (0-1) indicates higher motion regularity. The impact of motion magnitude, interdimensional correlation, and BPI on tracking errors were quantified by the Spearman correlation coefficients and the associated p-values.

3 Results

3.1 Optimal Temporal Window Length of Recent Tracking History

Figure 2 plots the RMS 3D tracking error for the Calypso study as a function of the length of temporal window for each 2D-3D inference method. The IDC, G-PDF, and A-PDF methods

had significantly smaller errors with >60 s temporal windows. In contrast, the Kalman filter was relatively insensitive to the choice of the window length. The best optimal window length for each method, defined to be the window length with the smallest RMS 3D tracking error for the Calypso study, is shown in Table 1. These values were used in the rest of the study.

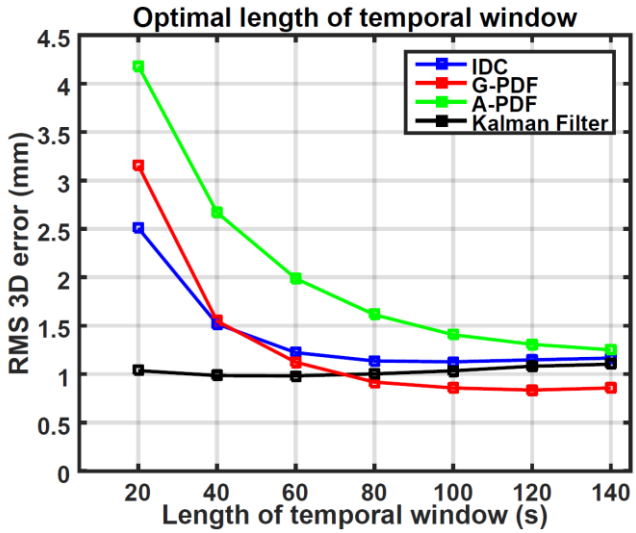


Figure 2. RMS 3D tracking error for the Calypso study as a function of the length of temporal window for each 2D-3D inference method.

Table 1. The optimal length of the temporal window for each method, measured in seconds and number of intrafraction kV frames.

Method	IDC	G-PDF	A-PDF	Kalman Filter
Temporal Window (s)	100	120	140	60
Number of Frames	500	600	700	300

3.2 Tracking errors

Table 2 summarizes the tracking errors of the four 2D-3D inference methods when applied to the Synchrony and Calypso traces. None of the methods showed significant systematic errors in either study. The 3D errors were larger in the Calypso study than in the Synchrony study, indicating that all four methods performed more accurately with inferred traces. The overall RMS and 95th percentile 3D error ranged from 0.45-0.68 mm for the Synchrony study and

0.84-1.25 mm for the Calypso study. The Gaussian PDF method was the overall best performer, followed by the Kalman filter, the IDC method, then the arbitrary-shape PDF method. The tracking trajectories from a typical fraction are shown in Figure 3.

Table 2. A summary of the overall dimensional and 3D tracking errors of the different 2D-3D inference methods applied to the Synchrony and Calypso motion traces.

Synchrony study (inferred traces)				
Errors (mm)	IDC	G-PDF	A-PDF	Kalman Filter
LR – Mean (±std)	0.00 (±0.45)	0.00 (±0.32)	0.02 (±0.52)	-0.02 (±0.43)
AP – Mean (±std)	0.01 (±0.40)	0.00 (±0.32)	-0.01 (±0.44)	0.02 (±0.42)
SI – Mean (±std)	0.00 (±0.01)	0.00 (±0.01)	0.00 (±0.01)	0.00 (±0.09)
3D – RMS	0.60	0.45	0.68	0.60
3D – 95th percentile	1.13	0.74	1.06	1.07
Calypso study (internal traces)				
Errors (mm)	IDC	G-PDF	A-PDF	Kalman Filter
LR – Mean (±std)	-0.04 (±0.72)	0.01 (±0.59)	-0.06 (±0.91)	-0.06 (±0.66)
AP – Mean (±std)	0.00 (±0.87)	-0.01 (±0.59)	0.03 (±0.85)	0.04 (±0.72)
SI – Mean (±std)	0.00 (±0.02)	0.00 (±0.01)	0.00 (±0.01)	0.01 (±0.06)
3D – RMS	1.13	0.84	1.25	0.98
3D – 95th percentile	2.28	1.72	2.64	2.05

Figure 4 shows the Tukey boxplots of the RMS and 95th percentile values of the 3D magnitude of tracking error for each fraction. Similarly to the findings from Table 2, errors were generally larger in the Calypso study than in the Synchrony study. However, there were more outlier fractions in the Synchrony study where large 3D errors were observed. This is likely because there were more fractions with exceptionally large and unpredictable motion in the Synchrony study, as the Synchrony study included 160 6-108 minute fractions as compared to 28 four minute fractions in the Calypso study. As an example, the fraction with the largest tracking errors for the arbitrary-shape PDF method is shown in Figure 5. For this particular fraction the patient had exceptionally large LR and small SI tumor motion. All of

the 2D-3D inference methods yielded larger errors for this fraction. However, the IDC and arbitrary-shape PDF methods were particularly unstable, resulting in RMS 3D tracking errors of 2.94 mm and 3.37 mm, and 95th percentile 3D tracking errors of 6.02 mm and 8.82 mm, respectively. In comparison, the Gaussian PDF and Kalman filter methods had RMS 3D tracking errors of 1.16 mm and 1.57 mm, and 95th percentile 3D tracking errors of 2.58 mm and 3.35 mm, respectively.

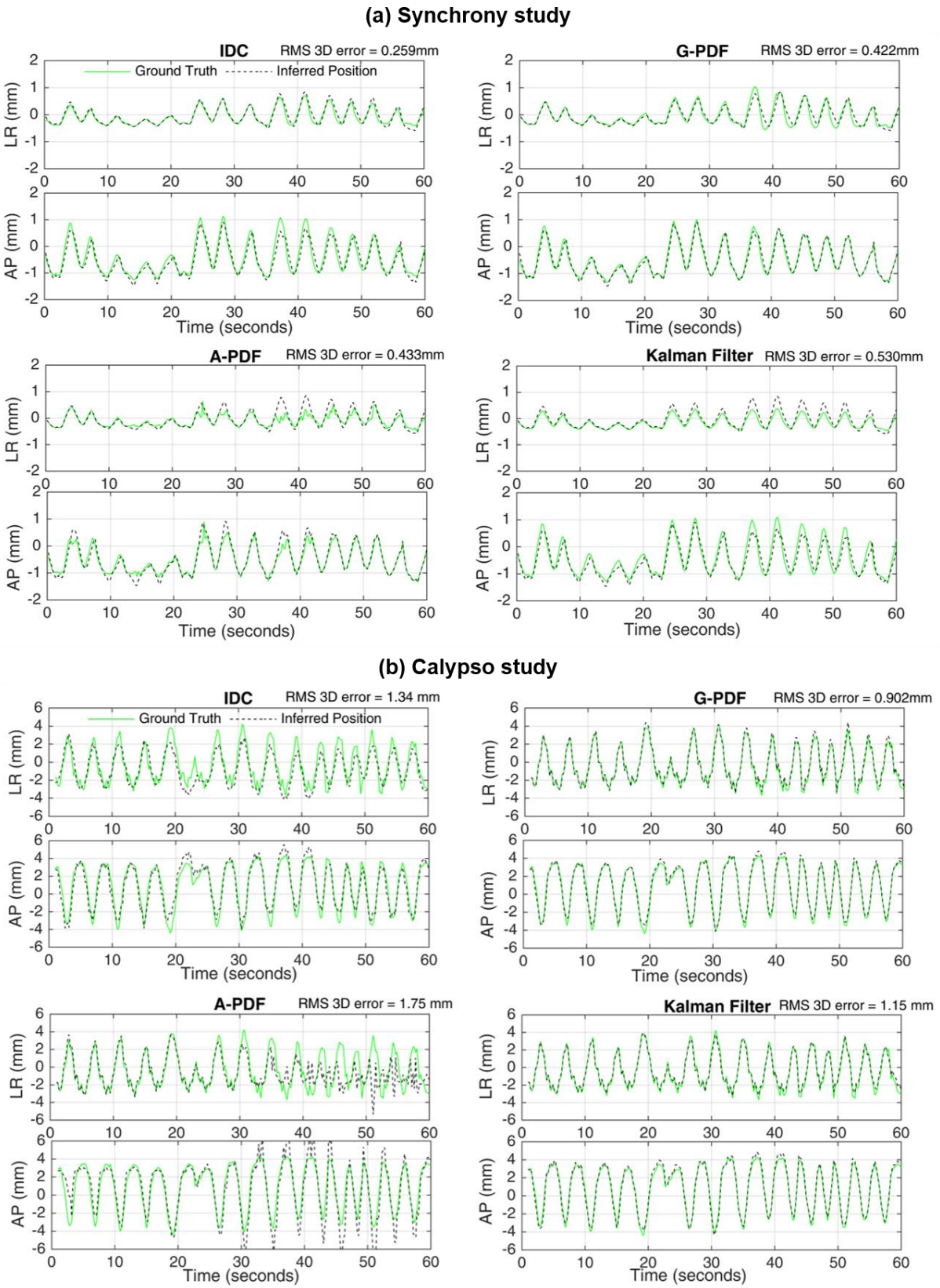


Figure 3. Example tracking trajectories (the beginning one minute) with errors representing the majority of the fractions in the Synchrony and Calypso studies. The RMS 3D error of each particular fraction is also shown. The SI motion is not included as it is always unambiguously resolved.

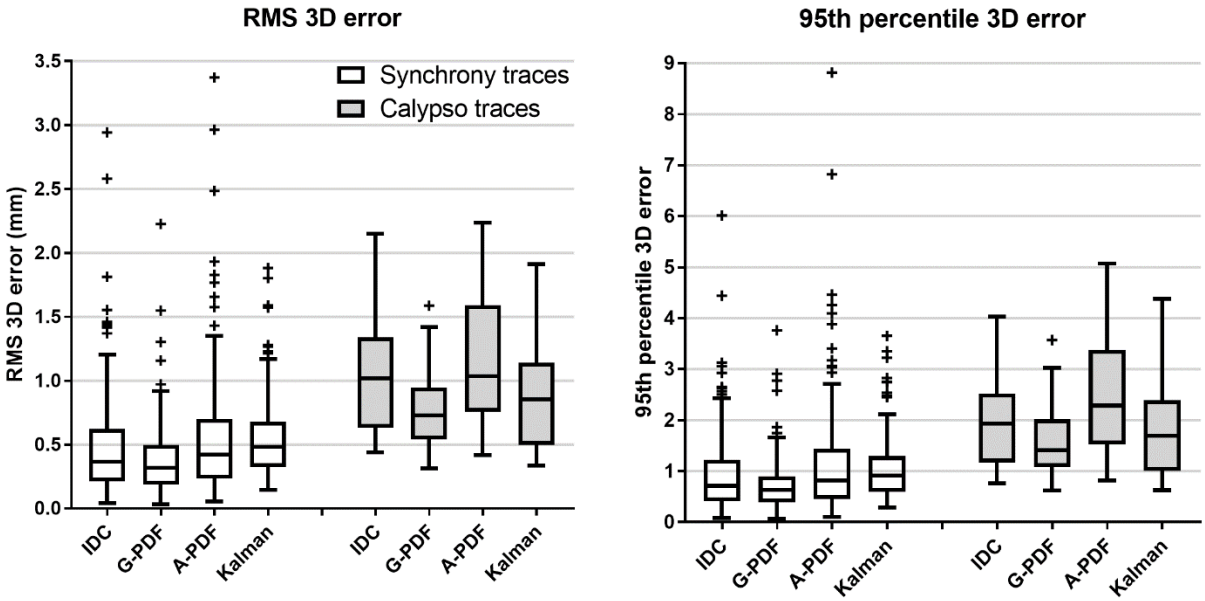


Figure 4. Tukey boxplots of the RMS and 95 percentile values of 3D magnitude of tracking error for each fraction. The box represents the first, second (median), and third quartile. The top and bottom whiskers represent 1.5 interquartile range above the third quartile and below the first quartile, respectively, with the outliers plotted as data points outside of this range.

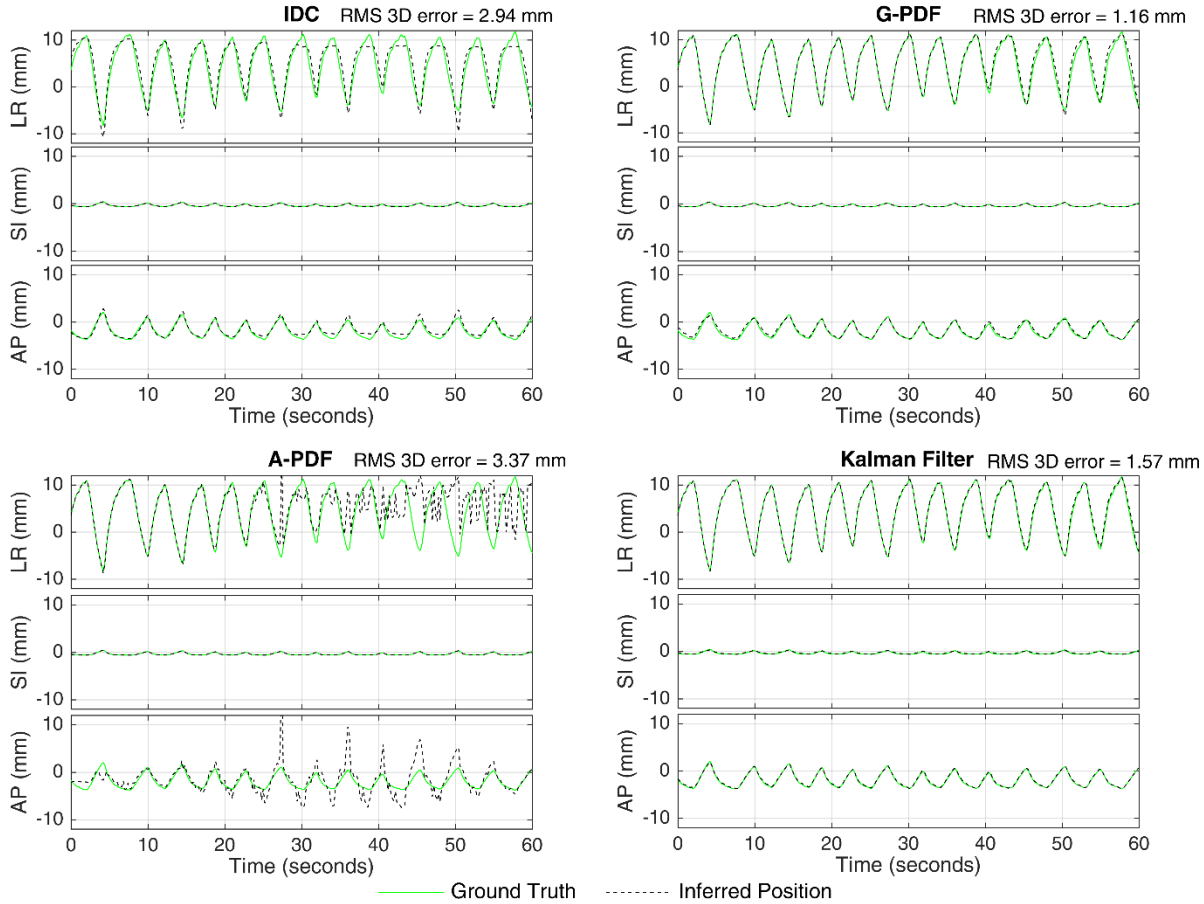


Figure 5. A Synchrony case where large tracking errors were observed for the IDC and A-PDF methods. For this fraction the patient had exceptionally large LR motion and small SI motion.

3.3 Impact of motion parameters

Table 3 summarizes the Spearman correlation coefficients between motion range, interdimensional correlation, BPI and the mean absolute tracking errors in the LR or AP direction. Motion magnitude was found to be the most important factor contributing to tracking error. In both the Synchrony and Calypso studies, larger motion magnitude was strongly correlated with larger tracking error for both LR and AP and for all of the 2D-3D inference methods ($p<0.001$). The impacts of interdimensional correlation and BPI were found to be different in the Synchrony and Calypso studies. In the Synchrony study, smaller interdimensional correlation was found to correlate with larger tracking error in the AP direction for all methods, and in the LR direction for the IDC method. Smaller BPI was found to correlate with larger tracking error in the LR direction for the IDC, Gaussian PDF, and arbitrary-shape PDF methods, and in the AP direction for the IDC and arbitrary-PDF methods. In contrast, the impacts of interdimensional correlation and BPI on tracking errors were much smaller in the Calypso study. Smaller interdimensional correlation was only found to correlate with larger LR error for the IDC method. Correlation between BPI and tracking error was not established for any of the methods.

Table 4 summarizes the range of motion parameters of the Synchrony and Calypso traces. A notable difference between the two is that the Synchrony traces exhibited larger interdimensional correlation than the Calypso traces, especially in the AP direction. This concurs with the expectation that inferred traces have larger interdimensional correlation than internal traces. Since larger interdimensional correlation was found to correlate with smaller tracking errors for the Synchrony traces as shown in Table 3, this could explain the overall smaller errors of the Synchrony study as shown in Table 2 and Figure 4.

Table 3. A summary the Spearman correlation coefficients between tumor motion range, interdimensional correlation, BPI and the mean absolute tracking errors in the LR or AP direction. Cases with p-values smaller than 0.05 are highlighted with bold fonts. The definitions of motion range, interdimensional correlation, and BPI are provided in Section 2.4.

		Synchrony study (inferred traces)			
		IDC	G-PDF	A-PDF	Kalman Filter
Motion range	LR	**0.76****	**0.76****	**0.72****	**0.61****
	AP	**0.66****	**0.60****	**0.68****	**0.55****
Interdimensional correlation	LR	**-0.10***	-0.05	0.09	-0.08
	AP	**-0.33****	**-0.28****	**-0.21***	**-0.17***
BPI	LR	**-0.40****	**-0.31****	**-0.44****	0.04
	AP	**-0.19***	-0.14	**-0.30****	-0.02
		Calypso study (internal traces)			
		IDC	G-PDF	A-PDF	Kalman Filter
Motion range	LR	**0.79****	**0.86****	**0.82****	**0.79****
	AP	**0.88****	**0.76****	**0.86****	**0.71****
Interdimensional correlation	LR	**-0.42***	-0.10	0.13	0.11
	AP	0.16	0.37	0.33	-0.14
BPI	LR	-0.09	-0.07	-0.22	-0.06
	AP	-0.13	-0.14	-0.15	-0.10

*: p-value < 0.05; **: p-value < 0.001

Table 4. A summary of the range of tumor motion parameters of the Synchrony and Calypso traces. The range of each quantity is quoted as the [5^{th}, 50^{th}, 95^{th}] percentile values across all fractions.

	Synchrony	Calypso
LR motion magnitude (mm)	[0.46, 1.79, 5.60]	[1.44, 2.30, 6.98]
AP motion magnitude (mm)	[0.70, 2.22, 6.53]	[0.94, 3.91, 7.92]
LR-SI correlation	[0.09, 0.65, 0.95]	[0.00, 0.54, 0.87]
AP-SI correlation	[0.10, 0.73, 0.95]	[0.19, 0.47, 0.88]
LR BPI	[0.10, 0.28, 0.67]	[0.20, 0.39, 0.61]
AP BPI	[0.09, 0.21, 0.44]	[0.27, 0.40, 0.53]

4 Discussion

kV-based target tracking is a rising field as it utilises the on-board imager which is standard equipment on most modern cancer radiotherapy systems. A number of teams have been working on kV image-guided target tracking systems, translating them into clinical trials[10, 22, 23]. To quantify and compare the accuracy and precision of kV-based tracking errors, the reliability of 2D-3D inference methods must be investigated. A major limitation of the studies to date[6, 12, 13] is that they have used traces inferred from external motion as ground truths. which may not represent true internal target tumor motion accurately, and may underestimate tracking errors.

This study is the first to compare the accuracy of 2D-3D inference methods using both inferred traces and internal traces. Concurring with our hypothesis, the main finding of this study is that internal traces result in larger 2D-3D tracking errors than inferred traces. RMS and 95th percentile 3D errors around 0.84-1.25 mm and 1.72-2.64 mm are to be expected from internal traces, compared to 0.45-0.68 mm and 0.74-1.13 mm from inferred traces. This result suggests that using internal traces acquired from systems such as Calypso as the ground truth is crucial for kV-based tracking studies. The use of inferred traces as the ground truth may underestimate the magnitude of tracking errors, as inferred traces have higher interdimensional correlation (cf. Table 4), which may be biasedly advantageous for most 2D-3D inference methods.

The accuracy of the different 2D-3D inference methods when applied to thoracic/abdominal motion were also directly compared for the first time. The result was consistent between the Synchrony and Calypso studies, with the Gaussian PDF being the overall most accurate method, followed by the Kalman filter, the IDC, and the arbitrary-shape PDF method. Interestingly, the optimal temporal window of recent tracking history can change considerably from method to method. Based on the Calypso study, a larger temporal window (120-140 s) is suggested for PDF based methods to ensure the stability of the estimated PDFs,

while a smaller temporal window (60 s) is advantageous for the Kalman filter to capture the most recent motion pattern. Within a 60 s temporal window, the Kalman filter exploits information within roughly 10-20 breathing cycles.

Figure 5 shows that patient with small SI but large LR/AP motion can be very challenging for kV based tracking. Since all of the methods rely on the interdimensional correlation between SI and the LR/AP components to some extent, small SI motion can make the model extremely sensitive to noise in the SI measurement. This is particularly true for the IDC method, since the LR and AP components are directly estimated from the SI component. The arbitrary-shape PDF method also suffers from this problem as it is more prone to overfitting noise in the motion pattern. By modelling the PDF as a superposition of all the backprojected traces within the temporal window, the number of parameters in the model is essentially the number of frames within the window (700 in this study) multiplied by two (2D position on kV image). In contrast, the Gaussian PDF and the Kalman filter methods have substantially fewer parameters, making them more robust against noise and extreme motion pattern.

Motion magnitude was found to be the largest contributing factor to tracking error in the corresponding direction for all four methods in both the Synchrony and Calypso studies. On the other hand, interdimensional correlation and BPI were found to negatively correlate with tracking errors in the Synchrony study, but not in the Calypso study. This means that motion magnitude should be considered the main factor for estimating kV-based tracking error due to 2D-3D inference.

The errors obtained from our Synchrony study are comparable to previous studies. In the original IDC paper by Chung et al[6], mean RMS 3D error of 0.35 mm was reported, with the 95th percentile per-fraction RMS 3D error being 1.03 mm. The 95th percentile 3D error at any point was 3.59 mm. In the original Gaussian PDF paper by Poulsen et al[12], the mean RMS 3D

error was 0.17 mm but per-fraction RMS 3D error was up to 2.89 mm. Maximum 3D error at any point was 13.6 mm. In the original arbitrary-shape PDF paper by Li et al[13], the RMS 3D error was 0.65 mm for the phantom experiment (lung trace), and 2.5 mm for the simulated elliptic motion trace. The highest 95th percentile 3D error was around 5 mm. However, as only four traces were tested, a direct comparison with our study is difficult. In the original Kalman filter study by Shieh et al.[14], mean 3D error of 1.6-2.9 mm was reported when applied to markerless tumor tracking, in which case tumor segmentation error contributes largely to the overall tracking error due to poor soft tissue visibility on kV images. In contrast, our study assumes zero segmentation error to isolate the impact of 2D-3D inference. Consequently, the errors reported in our study are expected to be much smaller.

There are a couple of limitations in this study. Firstly, the Calypso traces exhibited a small amount of noise due to the intrinsic measurement uncertainty of the system. The tracking accuracy and precision have been reported to be around 0.5 mm and 0.7 mm, respectively[24]. However, this level of noise was not found to be a concern for this study given the motion magnitude of thoracic targets is usually much larger. Secondly, this study investigated 1.6 degrees/s gantry rotation speed during intrafraction tracking, which is relatively slow. A faster gantry rotation speed may improve the accuracy of kV-based tracking, since the temporal window of recent tracking history would contain points measured across a wider angular range. Nevertheless, this limitation does not affect the main findings of this work, as all of the methods were tested with the same gantry rotation speed in both the Synchrony and Calypso studies. Finally, in order to isolate the impact of the 2D-3D inference process, this work has not investigated the effects of marker or tumor segmentation error. Marker segmentation error is usually very small and is expected to have minimal impacts on the results reported in this work. On the other hand, tumor segmentation error can be much larger due to poor soft tissue visibility on kV images. A comparison with the study by Shieh et al.[14]

(markerless tumor tracking using the Kalman filter method), which reported a mean 3D error of 1.6-2.9 mm, suggests that tumor segmentation error contributes 1-2 mm to the overall 3D tracking error.

5 Conclusion

Using inferred traces as the ground truth can underestimate tracking error of kV-based tracking studies. RMS and 95th percentile 3D errors around 0.84-1.25 mm and 1.72-2.64 mm are to be expected from internal traces, compared to 0.45-0.68 mm and 0.74-1.13 mm from inferred traces. The Gaussian PDF method is the most accurate 2D-3D inference method for tracking thoracic/abdominal targets. Larger motion magnitude in the LR or AP direction leads to larger tracking error in the corresponding direction, and should be used to estimate kV-based tracking error due to the 2D-3D inference process.

Acknowledgement

This study was supported by NHMRC Early Career Fellowship APP1120333 and CINSW Early Career Fellowship CS00481. The authors would like to thank A/Prof Per Poulsen for his assistance in the implementation of the Gaussian PDF method and insights to the findings. The authors would like to thank the anonymous reviewers for providing valuable feedback that has greatly improved this work.

References

1. Suh, Y., S. Dieterich, B. Cho, et al., *An analysis of thoracic and abdominal tumour motion for stereotactic body radiotherapy patients.* Physics in medicine and biology, 2008. **53**(13): p. 3623.
2. Ge, J., L. Santanam, C. Noel, et al., *Planning 4-dimensional computed tomography (4DCT) cannot adequately represent daily intrafractional motion of abdominal tumors.* International Journal of Radiation Oncology Biology Physics, 2013. **85**(4): p. 999-1005.
3. Koybasi, O., P. Mishra, S.S. James, et al., *Simulation of dosimetric consequences of 4D-CT-based motion margin estimation for proton radiotherapy using patient tumor motion data.* Physics in medicine and biology, 2014. **59**(4): p. 853.
4. Keall, P.J., G.S. Mageras, J.M. Balter, et al., *The management of respiratory motion in radiation oncology report of AAPM Task Group 76.* Medical Physics, 2006. **33**(10): p. 3874-3900.
5. Caillet, V., J.T. Booth, and P. Keall, *IGRT and motion management during lung SBRT delivery.* Physica Medica, 2017.

6. Chung, H., P.R. Poulsen, P.J. Keall, et al., *Reconstruction of implanted marker trajectories from cone-beam CT projection images using interdimensional correlation modeling.* Medical Physics, 2016. **43**(8): p. 4643-4654.
7. Wiersma, R., W. Mao, and L. Xing, *Combined kV and MV imaging for real-time tracking of implanted fiducial markers.* Medical physics, 2008. **35**(4): p. 1191-1198.
8. Cho, B., P.R. Poulsen, A. Sloutsky, et al., *First demonstration of combined kV/MV image-guided real-time dynamic multileaf-collimator target tracking.* International Journal of Radiation Oncology* Biology* Physics, 2009. **74**(3): p. 859-867.
9. Zhang, P., M. Hunt, H. Pham, et al., *Intrafractional 3D localization using kilovoltage digital tomosynthesis for sliding-window intensity modulated radiation therapy.* Physics in medicine and biology, 2015. **60**(17): p. N335.
10. Keall, P.J., J. Aun Ng, R. O'Brien, et al., *The first clinical treatment with kilovoltage intrafraction monitoring (KIM): A real-time image guidance method.* Medical physics, 2015. **42**(1): p. 354-358.
11. Nguyen, D.T., R. O'Brien, J.-H. Kim, et al., *The first clinical implementation of a real-time six degree of freedom target tracking system during radiation therapy based on Kilovoltage Intrafraction Monitoring (KIM).* Radiotherapy and Oncology, 2017. **123**(1): p. 37-42.
12. Poulsen, P.R., B. Cho, and P.J. Keall, *A method to estimate mean position, motion magnitude, motion correlation, and trajectory of a tumor from cone-beam CT projections for image-guided radiotherapy.* International Journal of Radiation Oncology* Biology* Physics, 2008. **72**(5): p. 1587-1596.
13. Li, R., B.P. Fahimian, and L. Xing, *A Bayesian approach to real-time 3D tumor localization via monoscopic x-ray imaging during treatment delivery.* Medical physics, 2011. **38**(7): p. 4205-4214.
14. Shieh, C.-C., V. Caillet, M. Dunbar, et al., *A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy.* Physics in Medicine and Biology, 2017. **62**(8): p. 3065.
15. Kalman, R.E., *A new approach to linear filtering and prediction problems.* Journal of basic Engineering, 1960. **82**(1): p. 35-45.
16. Kupelian, P., T. Willoughby, A. Mahadevan, et al., *Multi-institutional clinical experience with the Calypso System in localization and continuous, real-time monitoring of the prostate gland during external radiotherapy.* International Journal of Radiation Oncology* Biology* Physics, 2007. **67**(4): p. 1088-1098.
17. Willoughby, T.R., P.A. Kupelian, J. Pouliot, et al., *Target localization and real-time tracking using the Calypso 4D localization system in patients with localized prostate cancer.* International Journal of Radiation Oncology* Biology* Physics, 2006. **65**(2): p. 528-534.
18. Booth, J.T., V. Caillet, N. Hardcastle, et al., *The first patient treatment of electromagnetic-guided real time adaptive radiotherapy using MLC tracking for lung SABR.* Radiotherapy and Oncology, 2016. **121**(1): p. 19-25.
19. Seppenwoolde, Y., H. Shirato, K. Kitamura, et al., *Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy.* International Journal of Radiation Oncology* Biology* Physics, 2002. **53**(4): p. 822-834.
20. Ruan, D., J.A. Fessler, J.M. Balter, et al., *Real-time profiling of respiratory motion: baseline drift, frequency variation and fundamental pattern change.* Physics in Medicine and Biology, 2009. **54**(15): p. 4777-4792.
21. Li, G., M. Caraveo, J. Wei, et al., *Rapid estimation of 4DCT motion-artifact severity based on 1D breathing-surrogate periodicity.* Medical physics, 2014. **41**(11).
22. Cho, B., P.R. Poulsen, and P.J. Keall, *Real-time tumor tracking using sequential kV imaging combined with respiratory monitoring: a general framework applicable to commonly used IGRT systems.* Physics in medicine and biology, 2010. **55**(12): p. 3299.

23. Keall, P.J., J.A. Ng, P. Juneja, et al., *Real-Time 3D Image Guidance Using a Standard LINAC: Measured Motion, Accuracy, and Precision of the First Prospective Clinical Trial of Kilovoltage Intrafraction Monitoring–Guided Gating for Prostate Cancer Radiation Therapy.* International Journal of Radiation Oncology* Biology* Physics, 2016. **94**(5): p. 1015-1021.
24. Muralidhar, K.R., K. Komanduri, B.K. Rout, et al., *Commissioning and quality assurance of Calypso four-dimensional target localization system in linear accelerator facility.* Journal of Medical Physics / Association of Medical Physicists of India, 2013. **38**(3): p. 143-147.

Figure list

Table 1. The optimal length of the temporal window for each method, measured in seconds and number of intrafraction kV frames.

Method	IDC	G-PDF	A-PDF	Kalman Filter
Temporal Window (s)	100	120	140	60
Number of Frames	500	600	700	300

Table 2. A summary of the overall dimensional and 3D tracking errors of the different 2D-3D inference methods applied to the Synchrony and Calypso motion traces.

Synchrony study (inferred traces)				
Errors (mm)	IDC	G-PDF	A-PDF	Kalman Filter
LR – Mean (±std)	0.00 (±0.45)	0.00 (±0.32)	0.02 (±0.52)	-0.02 (±0.43)
AP – Mean (±std)	0.01 (±0.40)	0.00 (±0.32)	-0.01 (±0.44)	0.02 (±0.42)
SI – Mean (±std)	0.00 (±0.01)	0.00 (±0.01)	0.00 (±0.01)	0.00 (±0.09)
3D – RMS	0.60	0.45	0.68	0.60
3D – 95^{th} percentile	1.13	0.74	1.06	1.07
Calypso study (internal traces)				
Errors (mm)	IDC	G-PDF	A-PDF	Kalman Filter
LR – Mean (±std)	-0.04 (±0.72)	0.01 (±0.59)	-0.06 (±0.91)	-0.06 (±0.66)
AP – Mean (±std)	0.00 (±0.87)	-0.01 (±0.59)	0.03 (±0.85)	0.04 (±0.72)
SI – Mean (±std)	0.00 (±0.02)	0.00 (±0.01)	0.00 (±0.01)	0.01 (±0.06)
3D – RMS	1.13	0.84	1.25	0.98
3D – 95^{th} percentile	2.28	1.72	2.64	2.05

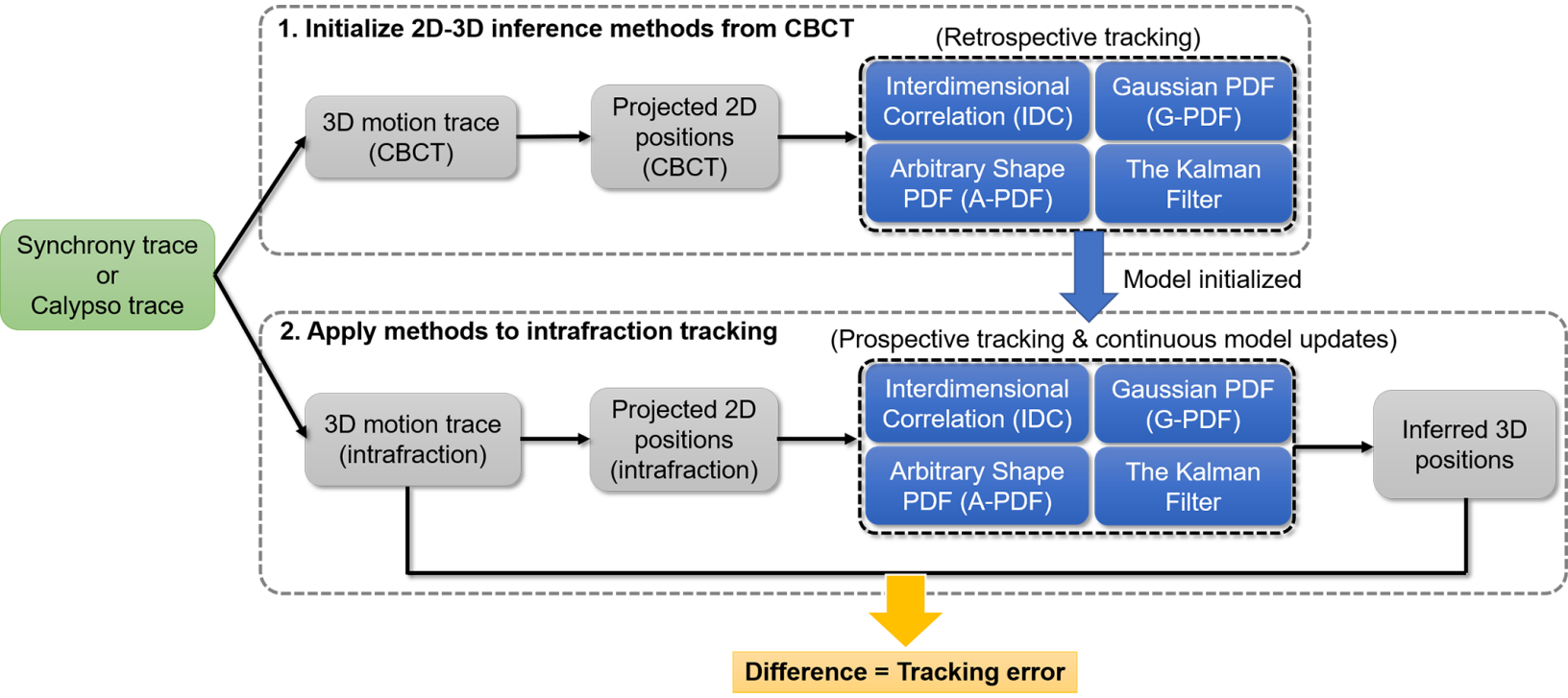
Table 3. A summary the Spearman correlation coefficients between tumor motion range, interdimensional correlation, BPI and the mean absolute tracking errors in the LR or AP direction. Cases with p-values smaller than 0.05 are highlighted with bold fonts. The definitions of motion range, interdimensional correlation, and BPI are provided in Section 2.4.

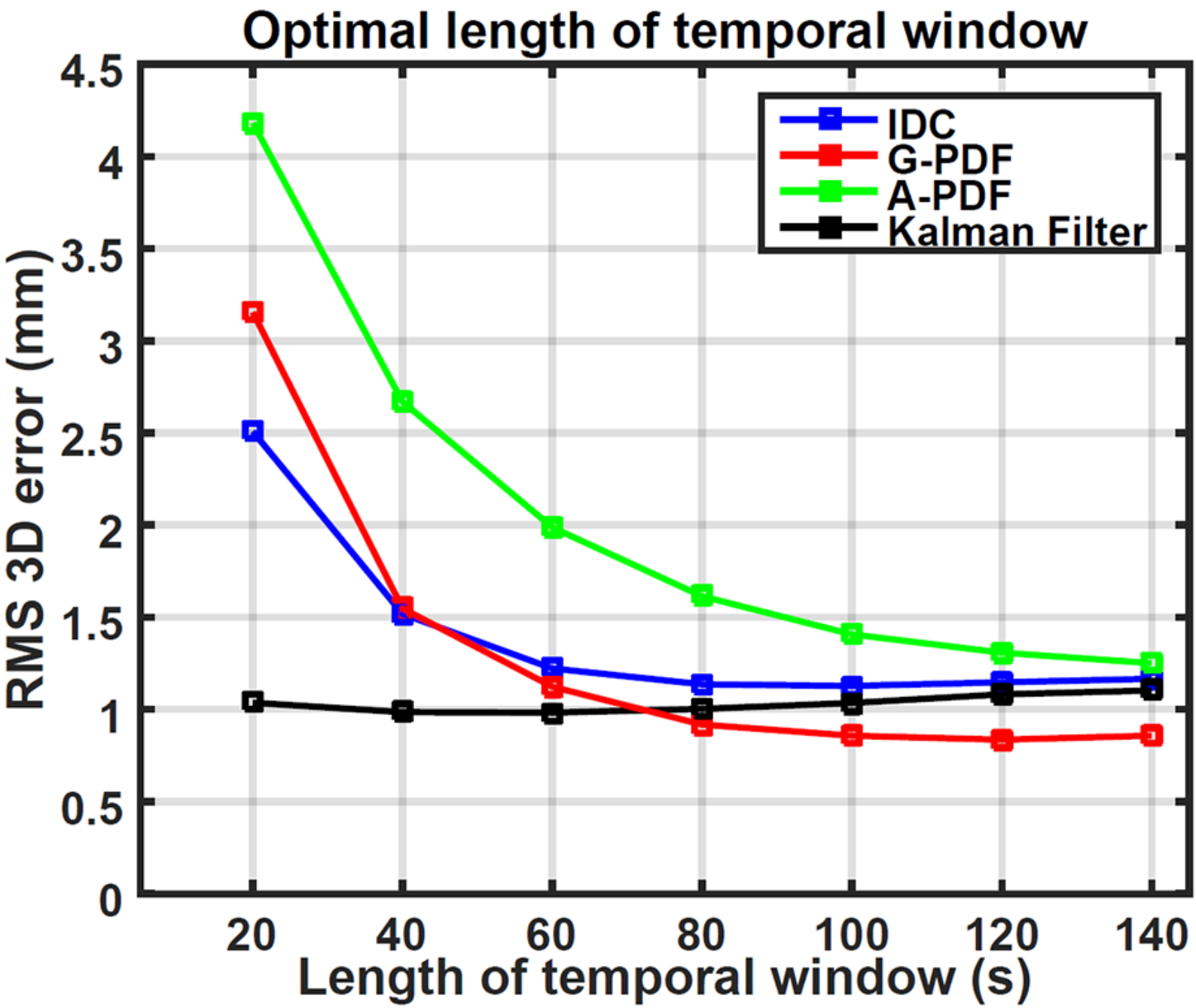
		Synchrony study (inferred traces)			
		IDC	G-PDF	A-PDF	Kalman Filter
Motion range	LR	**0.76****	**0.76****	**0.72****	**0.61****
	AP	**0.66****	**0.60****	**0.68****	**0.55****
Interdimensional correlation	LR	**-0.10***	-0.05	0.09	-0.08
	AP	**-0.33****	**-0.28****	**-0.21***	**-0.17***
BPI	LR	**-0.40****	**-0.31****	**-0.44****	0.04
	AP	**-0.19***	-0.14	**-0.30****	-0.02
		Calypso study (internal traces)			
		IDC	G-PDF	A-PDF	Kalman Filter
Motion range	LR	**0.79****	**0.86****	**0.82****	**0.79****
	AP	**0.88****	**0.76****	**0.86****	**0.71****
Interdimensional correlation	LR	**-0.42***	-0.10	0.13	0.11
	AP	0.16	0.37	0.33	-0.14
BPI	LR	-0.09	-0.07	-0.22	-0.06
	AP	-0.13	-0.14	-0.15	-0.10

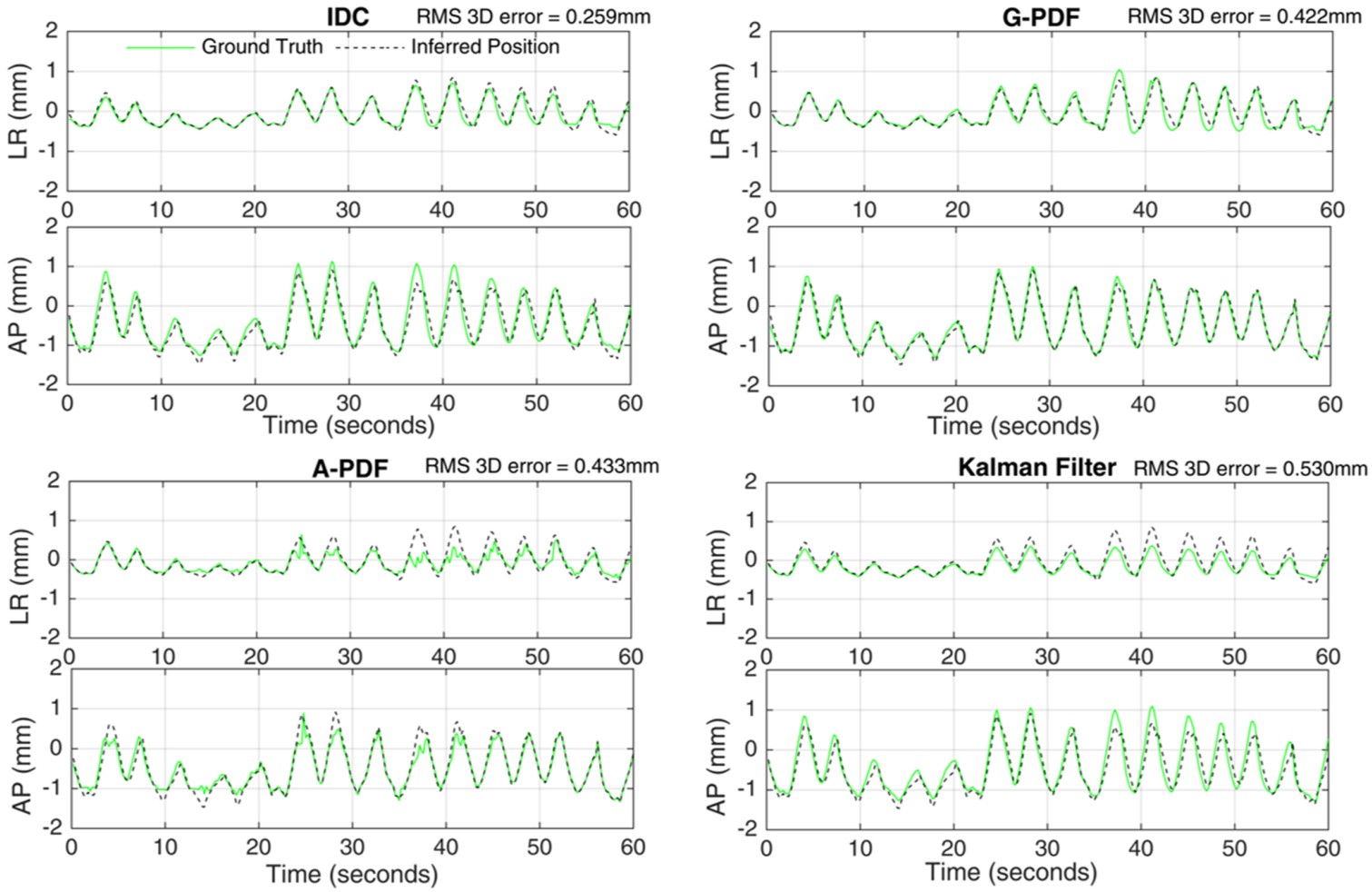
*: p-value < 0.05; **: p-value < 0.001

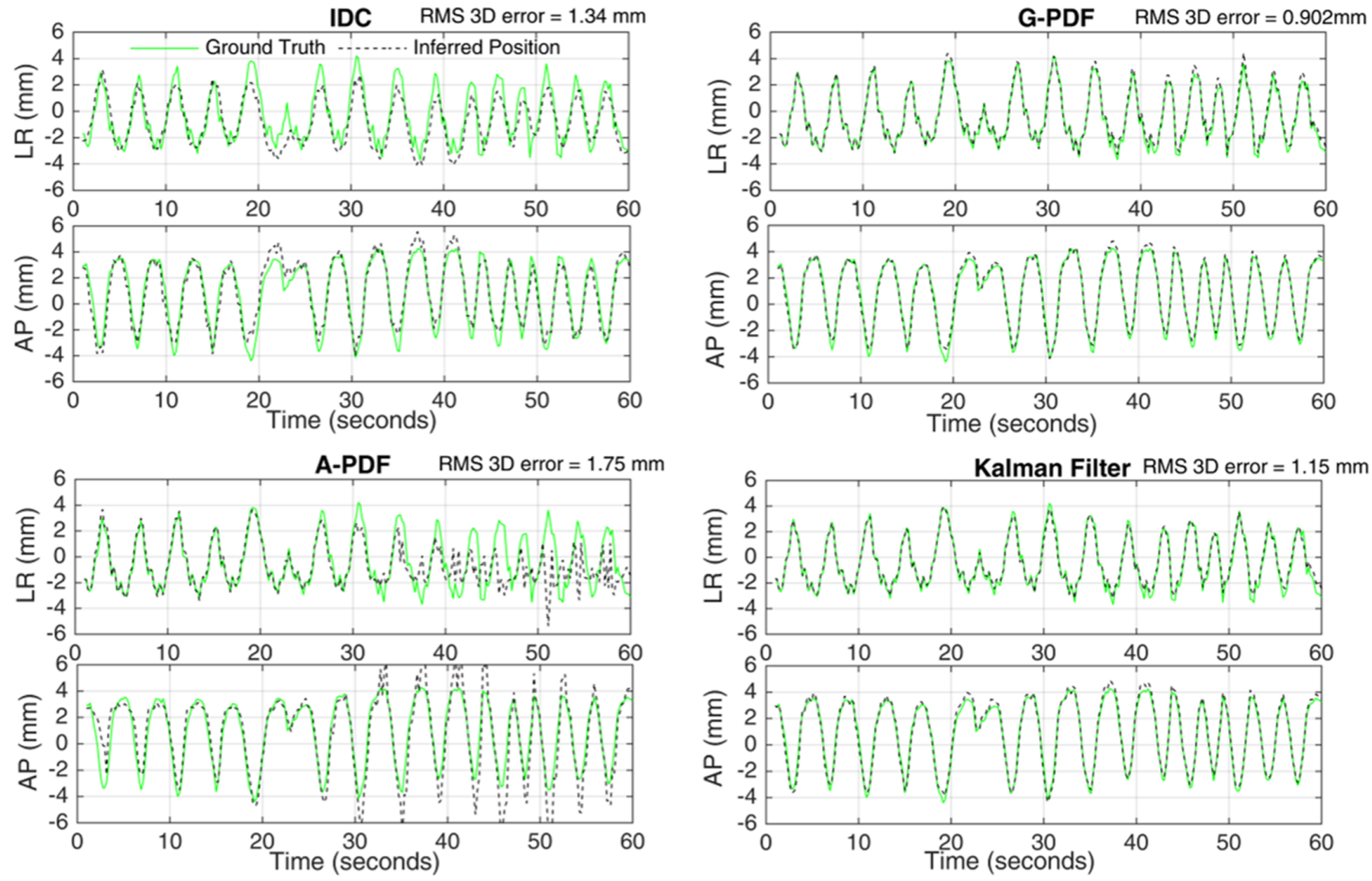
Table 4. A summary of the range of tumor motion parameters of the Synchrony and Calypso traces. The range of each quantity is quoted as the [5^{th}, 50^{th}, 95^{th}] percentile values across all fractions.

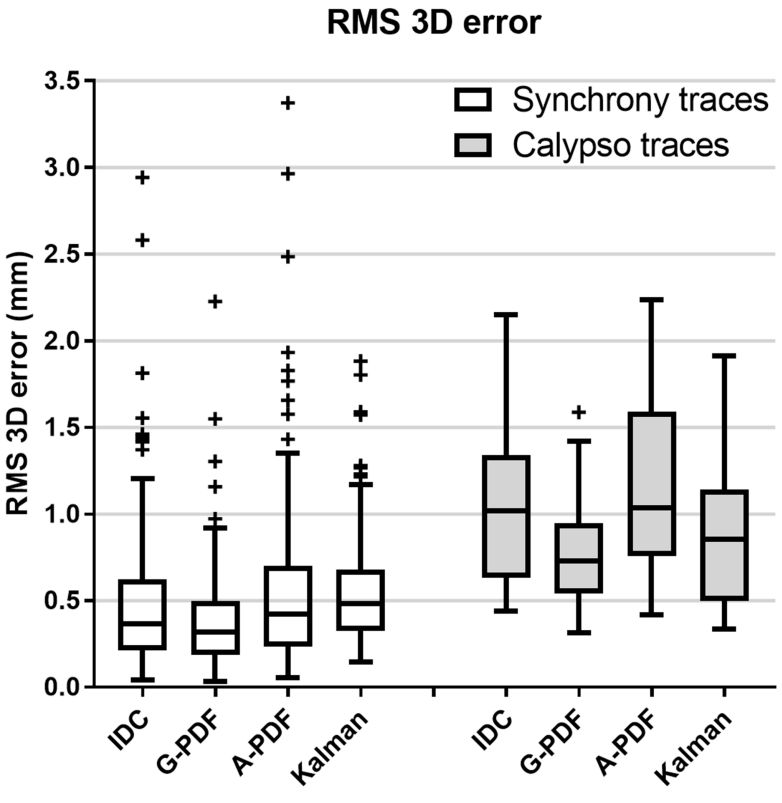
	Synchrony	Calypso
LR motion magnitude (mm)	[0.46, 1.79, 5.60]	[1.44, 2.30, 6.98]
AP motion magnitude (mm)	[0.70, 2.22, 6.53]	[0.94, 3.91, 7.92]
LR-SI correlation	[0.09, 0.65, 0.95]	[0.00, 0.54, 0.87]
AP-SI correlation	[0.10, 0.73, 0.95]	[0.19, 0.47, 0.88]
LR BPI	[0.10, 0.28, 0.67]	[0.20, 0.39, 0.61]
AP BPI	[0.09, 0.21, 0.44]	[0.27, 0.40, 0.53]

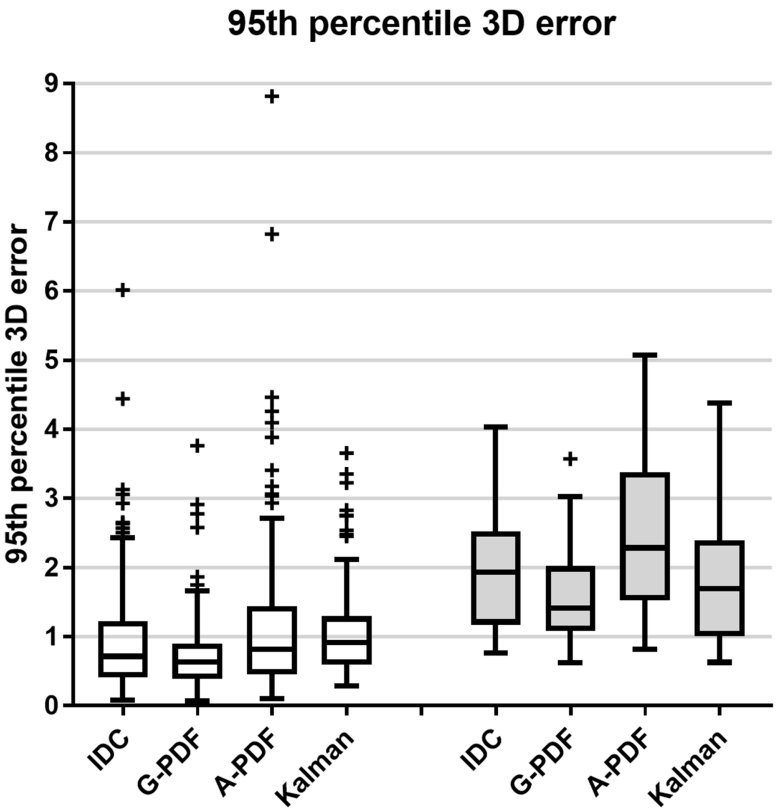

1. Initialize 2D-3D inference methods from CBCT
(Retrospective tracking)
3D motion trace (CBCT)
Projected 2D positions (CBCT)
Interdimensional Correlation (IDC)
Gaussian PDF (G-PDF)
Arbitrary Shape PDF (A-PDF)
The Kalman Filter
Synchrony trace or Calypso trace
Model initialized
2. Apply methods to intrafraction tracking
(Prospective tracking & continuous model updates)
3D motion trace (intrafraction)
Projected 2D positions (intrafraction)
Interdimensional Correlation (IDC)
Gaussian PDF (G-PDF)
Arbitrary Shape PDF (A-PDF)
The Kalman Filter
Inferred 3D positions
Difference = Tracking error

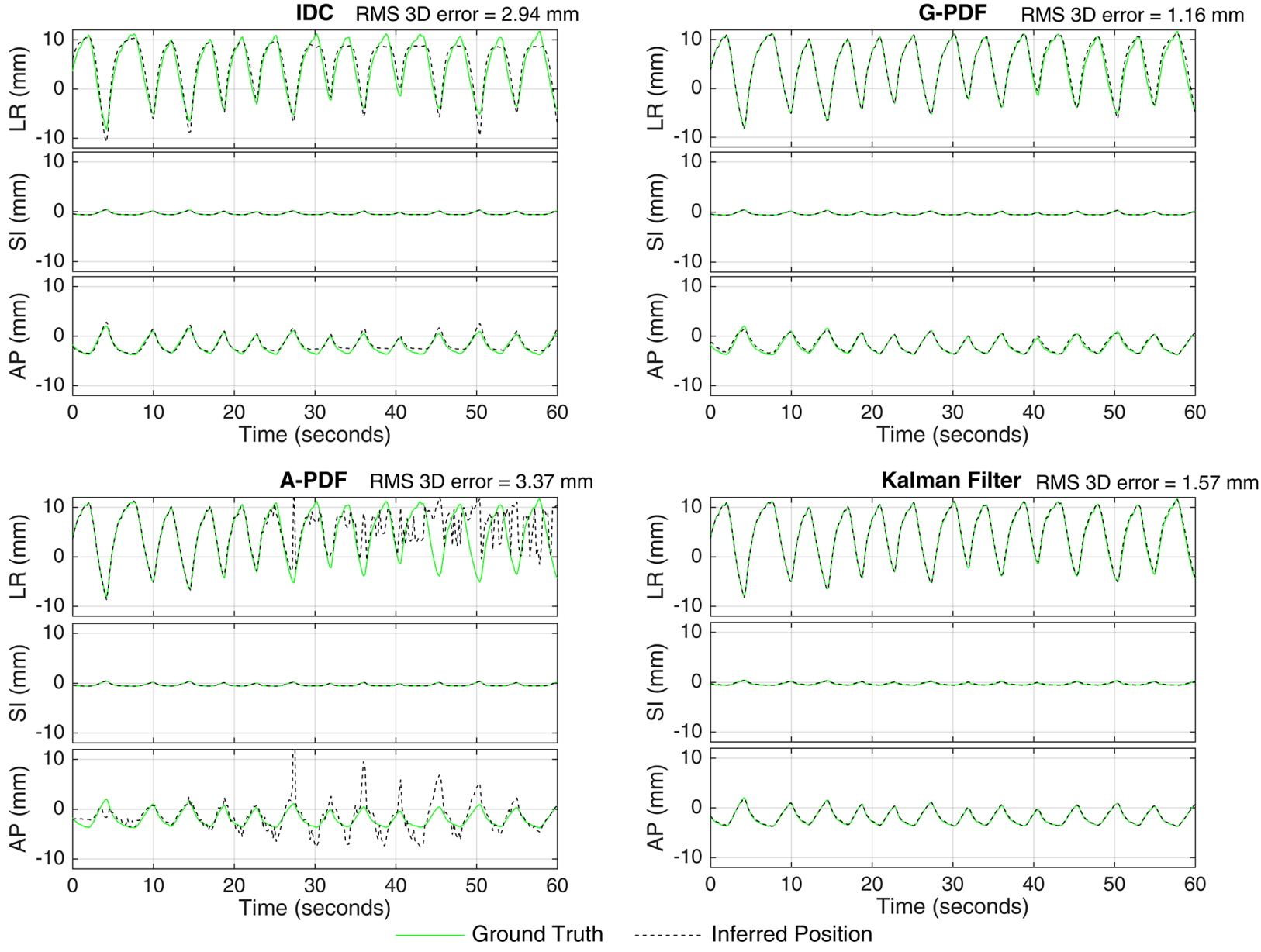

Optimal length of temporal window
RMS 3D error (mm)
Length of temporal window (s)
IDC
G-PDF
A-PDF
Kalman Filter
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
20
40
60
80
100
120
140


(a) Synchrony study
IDC
RMS 3D error = 0.259mm
Ground Truth
Inferred Position
G-PDF
RMS 3D error = 0.422mm
A-PDF
RMS 3D error = 0.433mm
Kalman Filter
RMS 3D error = 0.530mm
LR (mm)
AP (mm)
Time (seconds)


(b) Calypso study
IDC
RMS 3D error = 1.34 mm
Ground Truth
Inferred Position
G-PDF
RMS 3D error = 0.902mm
A-PDF
RMS 3D error = 1.75 mm
Kalman Filter
RMS 3D error = 1.15 mm
LR (mm)
AP (mm)
Time (seconds)


RMS 3D error
Synchrony traces
Calypso traces
RMS 3D error (mm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
IDC
G-PDF
A-PDF
Kalman
IDC
G-PDF
A-PDF
Kalman


95th percentile 3D error
95th percentile 3D error
9
8
7
6
5
4
3
2
1
0
IDC
G-PDF
A-PDF
Kalman
IDC
G-PDF
A-PDF
Kalman


IDC RMS 3D error = 2.94 mm
G-PDF RMS 3D error = 1.16 mm
A-PDF RMS 3D error = 3.37 mm
Kalman Filter RMS 3D error = 1.57 mm
LR (mm)
SI (mm)
AP (mm)
10
0
-10
0
10
20
30
40
50
60
Time (seconds)
Ground Truth
Inferred Position