Forecasting longitudinal changes in oropharyngeal tumor morphology throughout the course of head and neck radiation therapy

Adam D. Yock

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Arvind Rao

Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Lei Dong

Scripps Proton Therapy Center, San Diego, California 92121 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Beth M. Beadle and Adam S. Garden

Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030

Rajat J. Kudchadker

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Laurence E. Court^{a)}

Department of Radiation Physics and Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

(Received 6 February 2014; revised 24 June 2014; accepted for publication 24 June 2014; published 17 July 2014)

Purpose: To create models that forecast longitudinal trends in changing tumor morphology and to evaluate and compare their predictive potential throughout the course of radiation therapy.

Methods: Two morphology feature vectors were used to describe 35 gross tumor volumes (GTVs) throughout the course of intensity-modulated radiation therapy for oropharyngeal tumors. The feature vectors comprised the coordinates of the GTV centroids and a description of GTV shape using either interlandmark distances or a spherical harmonic decomposition of these distances. The change in the morphology feature vector observed at 33 time points throughout the course of treatment was described using static, linear, and mean models. Models were adjusted at 0, 1, 2, 3, or 5 different time points (adjustment points) to improve prediction accuracy. The potential of these models to forecast GTV morphology was evaluated using leave-one-out cross-validation, and the accuracy of the models was compared using Wilcoxon signed-rank tests.

Results: Adding a single adjustment point to the static model without any adjustment points decreased the median error in forecasting the position of GTV surface landmarks by the largest amount (1.2 mm). Additional adjustment points further decreased the forecast error by about 0.4 mm each. Selection of the linear model decreased the forecast error for both the distance-based and spherical harmonic morphology descriptors (0.2 mm), while the mean model decreased the forecast error for the distance-based descriptor only (0.2 mm). The magnitude and statistical significance of these improvements decreased with each additional adjustment point, and the effect from model selection was not as large as that from adding the initial points.

Conclusions: The authors present models that anticipate longitudinal changes in tumor morphology using various models and model adjustment schemes. The accuracy of these models depended on their form, and the utility of these models includes the characterization of patient-specific response with implications for treatment management and research study design. © 2014 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4887815>]

Key words: morphology, shape analysis, modeling, prediction, adaptive radiation therapy

1. INTRODUCTION

Consideration of changes in patient anatomy is critical in the administration of radiation therapy. The radiation therapy treatment plan and other patient management decisions are based on the patient's anatomy as depicted in a single set of pretreatment computed tomography (CT) images. If the configuration of the anatomy at the time of treatment departs from that depicted in the pretreatment CT images, the integrity of the optimized dose distribution may be compromised. As a result, clinicians expend considerable effort to ensure precise patient positioning and to minimize the effects of changes in patient anatomy.

Throughout the course of radiation therapy, different types of anatomic changes may occur, each with a different effect on the dose distribution. Random motions that occur during or between treatment fractions wash out and blur the dose distribution, while systematic motions common to all treatment fractions shift it. In addition, there are longitudinal anatomic changes that trend over time as the patient responds to therapy.

Because of their progressive nature, longitudinal changes have a dosimetric effect that may be greater than the effect due to intra- or interfractional motion. The discrepancy between the planned and delivered dose distribution grows as the patient anatomy departs further and further from that observed in the planning CT images. However, the progressive nature of longitudinal changes also suggests that these trends may not only be measured, but also predicted.

Prediction has been achieved using very simple descriptions of the tumor, such as volume, position, and surface area.^{1–4} Other measures used to describe radiation therapy targets and organs at risk include contour slice area and curvature,⁵ central axis position,⁶ and mathematical notions of sphericity⁷ and eccentricity.⁸ These descriptors are generally intuitive, easy to calculate, and can be readily incorporated into quantitative analysis. However, they are also reductive—irretrievably discarding a large amount of spatial information. Because they no longer retain a complete description of morphology (size, shape, and position), reductive descriptors are nongenerative—they cannot be used to reconstruct the original configuration of that which they describe. This compromises the specificity and utility of these otherwise convenient descriptors.

Generative descriptors of morphology are high-dimensional, making them more challenging to acquire and analyze and requiring larger data sample sizes. Despite the challenges, many techniques have been developed to quantify morphology with generative descriptors, as surveyed by Iyer *et al.*⁹ and Tangelder *et al.*¹⁰ Such descriptors have found utility in medicine—predominantly in computational neuroanatomy¹¹ and neurosurgery,¹² as well as in computer-aided diagnosis of breast cancer.¹³

In radiation therapy, generative descriptors of morphology are typically either finite element models based on biomechanical assumptions,^{4,14,15} vector models derived from deformable image registration,¹⁶ or statistical models representing a large number of landmarks.^{17–21} Their applications have

primarily been limited to modeling random tissue deformation, based on a set of patient images acquired over time. These images are considered to depict random instances of anatomy; however, they also include the effects of longitudinal changes. Without explicit consideration of longitudinal effects, models not only misrepresent random anatomic changes but also forfeit their potential to forecast longitudinal ones.

The prediction of longitudinal anatomic changes that occur during radiation therapy is of interest as it could facilitate necessary adjustments during treatment. Currently, such adjustments are reactionary. For example, clinicians may adjust treatment having noticed that a patient has lost a large amount of weight. Dramatic weight loss suggests the patient's anatomy, and thus the delivered dose distribution, has also changed. A desirable solution would facilitate identification and adjustment of the treatment in real time. However, the speed with which adjustments to a radiation therapy treatment plan can be made is limited. As a result, only a few adjustments are feasible, and they lag behind the anatomic changes. Accurate predictive models of changes in tumor morphology, if developed and validated, might compensate for this lag and influence treatment management decisions.

The purpose of this work was to create generative models to forecast longitudinal changes in tumor morphology and to evaluate and compare their predictive potential throughout the course of radiation therapy.

2. METHODS

Morphologies of oropharyngeal gross tumor volumes (GTVs) were represented by a set of landmarks positioned at the GTVs' surfaces. The relative positions of these landmarks were described in two ways: (1) by the distance from the surface landmarks to the GTV centroid (see details in Sec. 2.C.1) and (2) by a spherical harmonic decomposition of this set of distances (see details in Sec. 2.C.2). Three models (see details in Sec. 2.D), each with multiple model adjustment schemes (see details in Sec. 2.E), were used to forecast the change in GTV morphology throughout the course of treatment. The forecast accuracy of these models was evaluated and compared (see details in Sec. 2.F).

2.A. Instances of tumor morphology

The morphology of 35 GTVs (17 primary and 18 nodal) was retrospectively analyzed over the course of radiation therapy. These GTVs were acquired from the CT images of 19 patients (12 with multiple GTVs) who had undergone intensity-modulated radiation therapy of the oropharynx. These patients were randomly selected from an adaptive radiation therapy protocol described by Schwartz *et al.*²² with selection criteria including stage III, IVa, or IVb squamous cell carcinoma of the oropharynx as defined by the American Joint Committee on Cancer and an Eastern Cooperative Oncology Group performance status of 0–2. All patients were treated to 69.96 or 70.00 Gy with 2.12 or 2.00 Gy per fraction. Patient information is presented in Table I. As part of a clinical protocol, these patients received “daily” CT-on-rails imaging

TABLE I. Patient information.

Patient	Age (years)	Disease site	Disease stage ^a	Concurrent systemic therapy
1	62	Base of tongue	T1 N2A	Cisplatin
2	41	Tonsil	T3 N2B	Cisplatin and carboplatin
3	53	Tonsil	T2 N2C	Cetuximab
4	56	Base of tongue	T3 N2C	Cisplatin, carboplatin, and paclitaxel
5	57	Base of tongue	T3 N2A	Cisplatin
6	58	Base of tongue	T4 N2B	Cisplatin and cetuximab
7	50	Base of tongue	T2 N2B	Cetuximab
8	49	Tonsil	T1 N2B	Cisplatin
9	42	Base of tongue	T2 N2A	Cisplatin
10	54	Base of tongue	T4 N0	Cisplatin
11	62	Tonsil	T4 N0	Cetuximab
12	56	Base of tongue	T2 N2B	Cisplatin
13	50	Base of tongue	T2 N2A	Cisplatin
14	51	Tonsil	T3 N2B	Cisplatin
15	50	Base of tongue	T4 N2B	Cisplatin
16	38	Tonsil	T1 N2B	Cisplatin
17	68	Tonsil	T2 N2B	Cisplatin
18	69	Tonsil	T3 N1	Cisplatin
19	44	Base of tongue	T2 N2B	Cetuximab

^aNo patient had metastatic disease.

before each treatment fraction for image-guided alignment. To identify the GTVs in the daily CTs, the GTVs delineated by a physician on the pretreatment planning CT images were deformed to the daily CT images using in-house software. The deformable registration software is described by Wang *et al.* and, when applied to mathematically deformed head and neck anatomy, resulted in a mean error of 0.2 mm with 99% of errors being less than 2 mm.²³ After deformation, each GTV contour was reviewed individually and residual inaccuracies were considered acceptable for our analysis. For consistency, although some patients received as many as 35 treatment fractions, only the first 32 treatment fractions were considered as these CTs were available for all 19 patients. A 20th patient had been considered initially but was excluded due to a greater number of unavailable CTs.

2.B. Positioning surface landmarks

The position of the GTV centroid was determined on the images of each CT. A computer algorithm was then used to position 614 surface landmarks at the maximum radial extent of the GTV from the centroid at 10° polar and azimuthal angular intervals. These landmarks were points on the smooth surface of the GTV and did not necessarily correspond to noticeable features of the GTV or the anatomy. The angular sampling of the GTV surface was oriented relative to the patient's bony anatomy in order to remove the influence of variations in CT coordinates and patient setup. This was accomplished by first identifying four bony anatomic landmarks on the images of each CT and then using these landmarks to establish an internal coordinate system. The four landmarks were (1) the basion, at the anterior boundary of the foramen magnum, (2) the incisive foramen, in the anterior hard palate, (3) the left cochlea, and (4) the right cochlea. These land-

marks were chosen because they were easily and reproducibly identified on the pretreatment and daily CT images (Fig. 1). The landmarks were placed by a single observer to eliminate interobserver variability. The lateral dimension of the internal bony anatomy coordinate system was determined by the line between the left and right cochlea landmarks. An initial anterior–posterior dimension was determined by the line between the incisive foramen and the basion landmarks. The cross-product of unit vectors aligned with the lateral and initial anterior–posterior dimensions determined the superior–inferior dimension. Then, the cross-product of the lateral and superior–inferior dimensions was used to adjust the initial anterior–posterior dimension to ensure orthogonality. The coordinate system was translated to place its origin at the midpoint between the left and right cochlea landmarks. Placing surface landmarks at regular angular intervals with respect to an internal, bony anatomy coordinate system maintained correspondence between the landmarks across different sets of CT images and different patients.

2.C. Morphology descriptions

Given the sets of surface landmarks, we used the two methods below to describe the shape and size of each GTV. These descriptions were separately combined with each GTV's position to yield a morphology feature vector.

2.C.1. Radial extent morphology descriptor

The first of the two morphology feature vectors was based on a radial extent morphology descriptor. The distances between the GTV centroid and the GTV surface landmarks were listed in a row vector according to the angle of the surface landmark relative to the centroid. This vector was

FIG. 1. Locations of bony anatomy landmarks. (a) Sagittal plane depicting the basion (Ba) and the incisive foramen (IF) landmarks. Inset: Close up of incisive foramen landmark in axial plane. (b) Coronal plane depicting the right cochlea (RC) and the left cochlea (LC) landmarks.

concatenated with a vector of the GTV centroid coordinates. The result was a feature vector that could specifically reconstruct the position of the surface landmarks to approximate the original GTV morphology. This feature vector served as the basis for quantitative comparison and generation of morphology models. This approach closely resembles Lele and Richtsmeier's Euclidean distance matrix analysis.²⁴

2.C.2. Spherical harmonic morphology descriptor

The second morphology feature vector was generated by a spherical harmonic decomposition of the distances between

each GTV's centroid and surface landmarks. Spherical harmonics are the solutions to the angular portion of Laplace's equation in spherical coordinates. Their real form is presented in Eq. (1), where $Y_l^m(\theta, \varphi)$ denotes the spherical harmonic with degree l (a whole number) and order m (an integer from $-l$ to l) at polar and azimuthal angles θ and φ , respectively. $P_l^m(\cdot)$ denotes the associated Legendre function. Sets of spherical harmonic terms are frequently employed as an orthonormal basis to describe phenomena as diverse as atomic orbitals,^{25,26} molecular binding sites,²⁷ computer graphics lighting,^{28,29} planetary gravitational and magnetic fields,^{30–33} and the cosmic microwave background.³⁴

$$Y_l^m(\theta, \varphi) = \begin{cases} \sqrt{\frac{2l+1}{2\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) \cos(m\varphi) & \text{if } m > 0 \\ \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) & \text{if } m = 0 \\ \sqrt{\frac{2l+1}{2\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) \sin(m\varphi) & \text{if } m < 0 \end{cases} \quad (1)$$

Spherical harmonics are compact and efficient descriptors for shapes that are approximately spherical. They are single-valued, radial functions—an attribute sometimes referred to as being “star-shaped.” As such, they cannot describe concavities that obscure a direct line-of-sight from the spherical harmonic's centroid to its surface. However, because the surface landmarks were placed at the maximum extension of the GTV at regular angular intervals, any such concavities—none of which appeared to be of considerable consequence when approximating the shape of the GTV—had already been removed from the description and posed no challenge to the spherical harmonic decomposition.

For each GTV, the set of distances from the centroid to the surface landmarks was projected onto a basis composed of spherical harmonic terms. The result was a series of coefficients that, when combined with the spherical harmonic basis, returned the observed distances between the GTV centroid and the GTV surface landmarks. Determination of the spherical harmonic coefficients was achieved by fitting them to the observed surface landmark radial distances using ordinary least squares optimization. The coefficients were then compiled into a row vector concatenated with the GTV centroid coordinates, which produced a morphology feature vector similar to the radial extent feature vector described above. Previous applications of spherical harmonics sometimes combine harmonics of the same degree to achieve a rotationally invariant descriptor. In this work, however, we maintain the independence of each harmonic to retain a descriptor capable of identifying the particular orientation of the GTV. The result is a descriptor that can uniquely reconstruct the GTV morphology. Any residual discrepancies are limited to those represented by the higher order terms of diminishing effect that were disregarded via termination of the spherical harmonic series.

2.D. Morphology models

Each element of the two morphology feature vectors was considered a sample from a continuous time series. To reinforce this notion and to reduce noise, we smoothed the feature vector elements in the time dimension using cubic smoothing splines. The regularized time series of the feature vector elements were then used to generate three different morphology models. These models were based on information available from a set of GTVs used to establish model parameters (the training GTVs) and/or from the initial instance of the GTV to be predicted (the test GTV). Identification of the test and training GTVs was by the iterative process of leave-one-out cross-validation described below in Sec. 2.F.

The first model was referred to as the static model because it assumed no change from the initial value of the feature vector. The static model is described in Eq. (2), where $V_{0,\text{test}}$ denotes the initial feature vector of the test GTV, and $V_{t,\text{test}}$ denotes its value at treatment fraction t :

$$V_{t,\text{test}} = V_{0,\text{test}}. \quad (2)$$

The second model was a linear model in which the feature vector changed at a constant rate determined to yield a total change equal to the median total change of the training dataset. It is described in Eqs. (3) and (4), where β reflects the constant rate of change determined by feature vectors of GTVs in the training set ($V_{t,\text{train}}$):

$$V_{t,\text{test}} = V_{0,\text{test}} \times (1 + \beta \cdot t), \quad (3)$$

$$\beta = \text{median} \left(\frac{V_{32,\text{train}} - V_{0,\text{train}}}{V_{0,\text{train}}} \right) \times \left(\frac{1}{32} \right). \quad (4)$$

The third model [Eq. (5)] was a mean model that represented the average relative change in the feature vector based on a training dataset:

$$V_{t,\text{test}} = V_{0,\text{test}} \times \text{mean} \left(\frac{V_{t,\text{train}}}{V_{0,\text{train}}} \right). \quad (5)$$

For the three feature vector elements that designated the position of the GTV centroid in these models, the absolute change in value was considered rather than the relative change because of the large variation in initial GTV centroid position. Changes in the remaining feature vector elements were relative, having been normalized to their initial values. In addition, a maximum daily relative change of 25% (a value based on preliminary modeling experience) was set for all elements of the spherical harmonic feature vectors to regularize the model forecast.

2.E. Model adjustment schemes

The effect of adjusting each model during the course of treatment to improve accuracy was also considered. We tested these effects by adding 0, 1, 2, 3, or 5 points at which the model parameters were adjusted to match the true value of the morphology being forecast (adjustment points). This process simulated a clinical scenario similar to adaptive radiation therapy, in which a patient may be reimaged midtreatment,

FIG. 2. Diagram of model adjustment schemes. Time point 0 represents the time of the pretreatment CT. Subsequent time points mark treatment fractions. Model parameters were determined from the pretreatment CT images and images taken at the adjustment points (closed circles) and applied to subsequent treatment fractions (open circles) until the next adjustment point or the end of treatment. For readability, adjustment schemes are grouped vertically by number of adjustment points.

leading to a corresponding adjustment in treatment management. The effort invested into adjusting each model was considered dependent on the number of adjustment points. Multiple schemes with 1 and 2 adjustment points at different times during treatment were considered. The timing of the adjustments is depicted in Fig. 2.

The models with adjustment points were generated as described above, but with model parameters specific to each interadjustment point interval. Versions of Eqs. (2)–(5) that describe model predictions between adjustment points at treatment fractions p and q are presented in Eqs. (6)–(9). These equations also apply prior to the first adjustment point or subsequent to the last adjustment point when $p = 0$ and $q = 32$, respectively. To elaborate on these equations, when adjustment points were included in the static model [Eq. (6)], the value of the feature vector observed at p was considered the new value. This value was held constant until q or the end of treatment. Similarly, the rate of change for the linear model was determined by the median change of the training dataset at q or the end of treatment [Eqs. (7) and (8)]. The forecast of the mean model was shifted at p so that the average change in the feature vector was applied to the morphology observed rather than to the average training set morphology [Eq. (9)]. This process of adjusting the model parameters was repeated at each adjustment point:

$$V_{p \leq t < q, \text{test}} = V_{p, \text{test}}, \quad (6)$$

$$V_{p \leq t < q, \text{test}} = V_{p, \text{test}} \times (1 + \beta_p \times t), \quad (7)$$

$$\beta_p = \text{median} \left(\frac{V_{q, \text{train}} - V_{p, \text{train}}}{V_{p, \text{train}}} \right) \times \left(\frac{1}{q - p} \right), \quad (8)$$

$$V_{p \leq t < q, \text{test}} = V_{p, \text{test}} \times \text{mean} \left(\frac{V_{t, \text{train}}}{V_{p, \text{train}}} \right). \quad (9)$$

2.F. Evaluation of model forecast accuracy

The accuracy with which each model forecasts the position of the surface landmarks of a test GTV was evaluated using leave-one-out cross-validation. For leave-one-out cross-validation, the parameters of the models were derived on all but one of the GTVs. The models were then applied to the remaining GTV, predicting the feature vector, and thus the positions of the surface landmarks, at each treatment fraction. The accuracy of each model was assessed by measuring the root mean squared error (RMSE) in millimeters between the physical position of each surface landmark as represented in the forecast and true feature vectors. This process is repeated for a total of 35 times so that each GTV serves as the test GTV.

Separately for the two morphology descriptors, we identified the static, linear, and mean models that resulted in the lowest median RMSE for adjustment schemes with an equal number of adjustment points. These three models were considered to provide the best forecast accuracy for a certain investment in adjusting the models. In each case, the static model with the best prediction accuracy was compared to that with 1 fewer adjustment point using a Wilcoxon signed-rank test (SPSS Statistics 19, IBM, Armonk, NY) to determine if a significant improvement in forecast accuracy was achieved for the investment of adding an adjustment point. In addition, the linear and mean models were compared to the static model in order to determine if a significant improvement in forecast accuracy was achieved for the investment of using a nonstatic model, again with a Wilcoxon signed-rank test.

3. RESULTS

The radial extent feature vectors comprised 617 elements. For the spherical harmonic feature vectors, the first 100 terms up through spherical harmonics of degree 9 were sufficient to reduce the error between the approximated GTV shape and the observed set of landmarks to within about 0.5 mm. Figure 3 presents the error in approximating the observed GTV shape using spherical harmonics of different maximum degrees. Each distribution represents the median Euclidean distances between the approximated and observed position of each surface landmark for each GTV. Equation (10) elucidates the calculation of this average error for a particular

FIG. 3. Median error in approximating GTV shape with spherical harmonics. Each distribution represents the median Euclidean distances between the approximated and observed surface landmark positions for the population of 35 GTVs. The box range is the 25th to the 75th percentile error, and the whisker range is the minimum to the maximum error.

FIG. 4. Forecast error for radial extent morphology descriptor models. Each set of three distributions depicts the static (left), linear (center), and mean (right) morphology models featuring an equal number of adjustment points. Each individual distribution represents the test GTV errors from the 35 iterations of leave-one-out cross-validation. The box range is the 25th to the 75th percentile of RMSE, and the whisker range is the 10th to the 90th percentile.

GTV i (ε_i) based on the approximated and observed positions of each landmark l (Approximated $_{i,l}$ and Observed $_{i,l}$, respectively). Additional terms beyond those of degree 9 further decreased the error but with diminishing returns. The combination of these terms with the GTV centroid coordinates resulted in a 103-element spherical harmonic feature vector:

$$\varepsilon_i = \text{median}_{\text{over } l} (\text{Approximated}_{i,l} - \text{Observed}_{i,l}). \quad (10)$$

Figures 4 and 5 are box-and-whisker plots depicting the forecast error of the three models acquired with the radial extent and spherical harmonic morphology descriptors, respectively. For both descriptors, the forecast error was on the order of millimeters, and both the median error and the range of errors decreased as the number of adjustment points increased.

FIG. 5. Forecast error for spherical harmonic morphology descriptor models. Each set of three distributions depicts the static (left), linear (center), and mean (right) morphology models featuring an equal number of adjustment points. Each individual distribution represents the test GTV errors from the 35 iterations of leave-one-out cross-validation. The box range is the 25th to the 75th percentile of RMSE, and the whisker range is the 10th to the 90th percentile.

TABLE II. Difference in forecast error of static models with various numbers of adjustment points for the radial extent morphology descriptor.^a

Model A	Model B	Maximum Δ RMSE (mm)	Median Δ RMSE (mm)	Minimum Δ RMSE (mm)	<i>p</i> -Value	% Δ RMSE ≤ 0
Static (0) ^b	Static (1)	0.2	−1.2	−4.5	<0.001	97.1
Static (1)	Static (2)	0.6	−0.4	−3.0	<0.001	97.1
Static (2)	Static (3)	0.2	−0.3	−1.5	<0.001	88.6
Static (3)	Static (5)	−0.1	−0.4	−1.8	<0.001	100.0

^aDifference calculated as Model B − Model A.^bParentheticals denote the number of adjustment points included in the model.

Tables II–V tabulate the differences in the models depicted in Figs. 4 and 5. Table II presents the effect on forecast error of incorporating additional adjustment points into the static model for the radial extent morphology descriptor. Adding each adjustment point decreased the RMSE by a statistically significant amount ($p < 0.001$). The magnitude of the decrease was largest (median Δ RMSE = −1.2 mm) with the addition of the first adjustment point and tended to decrease with additional points. Furthermore, for each comparison, the minimum Δ RMSE was of greater magnitude than the maximum Δ RMSE, suggesting the benefit achieved in the best-case scenario outweighed the cost incurred in the worst-case scenario. Finally, a decrease in the RMSE was observed in the overwhelming majority (approximately 89%–100%) of GTVs.

Table III compares the linear and mean models with the lowest median RMSE for a certain investment in model adjustment to that of the static model. These models were derived from the radial extent morphology descriptor. Statistically significant decreases in the RMSE were observed in the linear and mean models compared to the static models for adjustment schemes up to the first and second adjustment points, respectively. For adjustment schemes with a greater number of adjustment points, the ranges in Δ RMSE were smaller and more symmetric around zero. This reflected less drastic, more comparable tradeoffs between the best- and worst-case scenarios. The proportion of patients for whom the linear and mean models decreased the RMSE was about 77% for the

scheme lacking any adjustment and tended to decrease with the successive addition of adjustment points.

The results for the spherical harmonic morphology descriptor in Table IV are very similar to those of the radial extent morphology descriptor in Table II. The RMSE decreased significantly with the addition of each adjustment point ($p < 0.001$), and the largest change occurred with the addition of the first point (median Δ RMSE = −1.2 mm). The asymmetric ranges of Δ RMSE around zero again suggested that the benefit of the best-case scenarios outweighed the cost of the worst-case scenarios. As was the case for the radial extent morphology descriptor, the majority of GTVs demonstrated a decrease in the RMSE with the addition of each adjustment point.

Table V depicts the same information as Table III but for the spherical harmonic morphology descriptor. For the comparison of the linear and static models, the patterns of Table V resemble those of Table III. The linear models provided a statistically significant decrease in the RMSE compared to the static models up to the addition of a second adjustment point; however, further decreases with additional adjustment points were not statistically significant. Conversely, the decrease in the RMSE provided by the mean model without any adjustment points only approached statistical significance ($p = 0.058$), while including 2 or more adjustment points resulted in statistically significant increases in the RMSE ($p \leq 0.003$). For the linear models, with the exception

TABLE III. Difference in forecast error when comparing linear and mean models with static models for the radial extent morphology descriptor.^a

Model A	Model B	Maximum Δ RMSE (mm)	Median Δ RMSE (mm)	Minimum Δ RMSE (mm)	<i>p</i> -Value	% Δ RMSE ≤ 0
Static (0) ^b	Linear (0)	0.6	−0.2	−1.3	0.001	77.1
Static (0)	Mean (0)	0.9	−0.2	−1.6	0.005	74.3
Static (1)	Linear (1)	0.4	−0.1	−0.6	0.054	62.9
Static (1)	Mean (1)	0.3	−0.1	−0.6	0.036	65.7
Static (2)	Linear (2)	1.9	0.0	−1.5	0.968	48.6
Static (2)	Mean (2)	1.1	0.1	−1.3	0.097	37.1
Static (3)	Linear (3)	0.1	−0.0	−0.2	0.725	54.3
Static (3)	Mean (3)	0.2	−0.1	−0.3	0.126	62.9
Static (5)	Linear (5)	0.2	−0.0	−0.1	0.907	57.1
Static (5)	Mean (5)	0.1	−0.0	−0.1	0.990	51.4

^aDifference calculated as Model B − Model A.^bParentheticals denote the number of adjustment points included in the model.

TABLE IV. Difference in forecast error of static models with various numbers of adjustment points for the spherical harmonic morphology descriptor.^a

Model A	Model B	Maximum Δ RMSE (mm)	Median Δ RMSE (mm)	Minimum Δ RMSE (mm)	<i>p</i> -Value	% Δ RMSE ≤ 0
Static (0) ^b	Static (1)	0.3	−1.2	−4.7	<0.001	97.1
Static (1)	Static (2)	0.4	−0.4	−2.8	<0.001	97.1
Static (2)	Static (3)	0.2	−0.3	−1.5	<0.001	88.6
Static (3)	Static (5)	−0.1	−0.4	−1.8	<0.001	100.0

^aDifference calculated as Model B − Model A.^bParentheticals denote the number of adjustment points included in the model.

of that lacking any adjustment points, the ranges of Δ RMSE were reasonably symmetric around zero. Those for the mean models became increasingly skewed in the positive direction with the addition of adjustment points. Here, in contrast to the mean models of the radial extent morphology descriptor in Table III, the cost of the worst-case scenario outweighed the benefit of the best-case scenario.

Figures 6 and 7 depict contours of forecast morphology of an example GTV overlaid on the patient's anatomy of the 32nd treatment fraction. Figure 6 compares static models with varying numbers of adjustment points while Fig. 7 compares the linear and static models with no adjustment points. In Fig. 6, the contours of the static models with additional adjustment points more closely conform to the observed GTV morphology. In Fig. 7, the GTV morphology predicted by the linear model approximates the observed GTV morphology, and does so noticeably better than that predicted by the static model. Both the linear and static model predictions were based exclusively on the pretreatment CT images acquired several weeks prior to the realization of the depicted anatomy.

4. DISCUSSION

Our data show that the accuracy of forecasting tumor morphology throughout the course of radiation therapy can be improved by using nonstatic, morphology models and by adding points during treatment at which the models are adjusted. Our models and adjustment schemes, applied to two different tu-

mor morphology descriptors, retained information regarding tumor size, shape, and position, and could construct the forecast tumor morphology.

The effect on forecast accuracy of adding adjustment points was greater than that of selecting linear or mean models over static models. However, for a particular number of adjustment points, incorporating the use of a linear or mean model provided a further improvement in accuracy (with the exception of the mean models for the spherical harmonic morphology descriptor). As adjustment points were added, the improvement in forecast accuracy resulting from the additional points or from the use of the nonstatic models diminished. As a result, there may be a point in the application of models like these at which the improved accuracy is considered not worth the additional investment in model selection or model adjustment. In the case of the mean models for the spherical harmonic morphology descriptor, the decline in forecast accuracy was exacerbated by the addition of adjustment points.

That the use of nonstatic models and the addition of model adjustment points would improve the forecast accuracy is not surprising. The linear and mean models included prior information regarding the morphological changes observed in similar tumors, so it is sensible that such a forecast would perform better than a naïve one. Similarly, models with adjustment points incorporated prior information specific to the morphology being predicted. Therefore, an even greater improvement is to be expected. The decrease in forecast accuracy

TABLE V. Difference in forecast error when comparing linear and mean models with static models for the spherical harmonic morphology descriptor.^a

Model A	Model B	Maximum Δ RMSE (mm)	Median Δ RMSE (mm)	Minimum Δ RMSE (mm)	<i>p</i> -Value	% Δ RMSE ≤ 0
Static (0) ^b	Linear (0)	0.5	−0.2	−1.5	0.003	68.6
Static (0)	Mean (0)	0.9	−0.2	−1.6	0.058	57.1
Static (1)	Linear (1)	0.4	−0.1	−0.6	0.035	65.7
Static (1)	Mean (1)	0.5	0.0	−0.4	0.577	48.6
Static (2)	Linear (2)	0.4	−0.0	−0.3	0.349	54.3
Static (2)	Mean (2)	1.3	0.3	−1.3	0.003	25.7
Static (3)	Linear (3)	0.2	−0.0	−0.3	0.849	54.3
Static (3)	Mean (3)	0.8	0.1	−0.2	0.003	22.9
Static (5)	Linear (5)	0.2	−0.0	−0.2	0.913	60.0
Static (5)	Mean (5)	0.9	0.1	−0.1	<0.001	8.6

^aDifference calculated as Model B − Model A.^bParentheticals denote the number of adjustment points included in the model.

FIG. 6. Forecast contours of an example GTV as depicted on the anatomy of treatment fraction 32. (a) Axial image, (b) sagittal image, and (c) coronal image. Sky blue color wash: original GTV contour deformed to treatment fraction 32, considered ground truth. Red contour: forecast morphology according to the static model with no adjustment points. Yellow contour: forecast morphology according to the static model with 1 adjustment point. Dark blue contour: forecast morphology according to the static model with 5 adjustment points.

FIG. 7. Forecast contours of an example GTV as depicted on the anatomy of treatment fraction 32. (a) Axial image, (b) sagittal image, and (c) coronal image. Sky blue color wash: original GTV contour deformed to treatment fraction 32, considered ground truth. Red contour: forecast morphology according to the static model with no adjustment points. Dark blue contour: forecast morphology according to the linear model with no adjustment points.

improvement with the increase in number of adjustment points is also intuitive—the models approached the true morphological values, and there were fewer fractions between adjustments for their forecasts to diverge. Similar abilities to forecast morphology was to be expected from the radial extent and spherical harmonic morphology descriptors, as they were both able to closely approximate tumor morphology.

The magnitude of the changes in forecast error and the relative performance of the nonstatic models were not apparent prior to this work. We expected the mean models might provide improvements in forecast accuracy greater than the linear models because the mean models were more flexible, lacking the additional constraint of a constant rate of change between adjustment points. We did not anticipate that mean models with adjustment points would provide forecast accuracy inferior to that of the static and linear models for the spherical harmonic morphology descriptors.

It is unclear why the mean model performed poorly for the spherical harmonic morphology descriptor. Perhaps, while they provided an effective way to describe morphology, spherical harmonics were less suited to describe *changes in* morphology. To transfer information regarding the change in a feature vector element from one instance of morphology to another is more intuitive when the feature vector element represents the radial extent in a particular direction rather than the weight attributed to a particular geometrical component. The values and relative changes in the former are also more consistent across tumors. As a result, population-driven models, such as the linear and mean models, may be more effective for the radial extent morphology descriptor. In addition, the constraint of the linear models to change at a constant rate might have made them more robust to noisy forecasts that arose from the selection of descriptor.

Although reductive, nongenerative descriptors have been used to describe and even forecast shape,^{1–4} the potential to forecast longitudinal changes in tumor morphology directly has not been investigated. We are therefore unable to compare our forecast accuracy with that achieved by others. Takao *et al.*¹⁴ did use finite element models to measure longitudinal changes in the shape of cervical lymph nodes, though they did not attempt to evaluate the potential of their models to forecast these changes. They suggested, however, that changes in lymph node shape can be predicted indirectly by predicting changes in tumor volume. Comparisons of the performance of our models with those of nongenerative descriptors or with those of random anatomic changes that may or may not include longitudinal changes are not appropriate.

The work presented here need not be limited to analyzing trends in GTV morphology. While the models are designed to forecast longitudinal changes, the conceptual basis behind them could be applied to organs at risk as well as to configurations of landmarks that comprise multiple organs or that retain meaning beyond direct correspondence to any specific organ. Furthermore, the radial extent and spherical harmonic morphology descriptors could also be utilized to measure random or systematic variations.

Models of morphology descriptors are also not limited to being used in the oropharynx. However, the methods pre-

sented here may require modifications to address the types of anatomic variations characteristic of alternative disease sites. For example, the prostate, subject to the deformations of the nearby bladder and rectum, may exhibit random motions of larger magnitude than those of the oropharyngeal tumors we studied. And a random motion that occurs during the acquisition of the pretreatment CT images may propagate throughout the treatment as a systematic error. These variations must be considered when generating models of morphology descriptors for the prostate. As a second example, lung tumors exhibit considerable variations in morphology (primarily position) due to respiratory motion. This large, periodic, intrafractional motion cannot be ignored when generating descriptors and models of lung tumors as it was in our work describing tumors of the oropharynx.

Patients included in the work presented here had similar diseases and treatment regimens. Response of a patient's disease and normal anatomy might not only vary due to disease site as described above, but also due to clinical considerations or treatment strategies not reflected in our patient sample. For instance, our models, based on patients treated with chemotherapy concurrent with their radiation, may not be appropriate for an individual who has received neoadjuvant chemotherapy prior to radiation. Our models may be further limited by ignoring potentially relevant clinical or histological variables such as human papillomavirus infection status. Further investigation is required to fully characterize the scope of these particular methods.

Although we were successful in improving the accuracy of forecasting changes in tumor morphology, we encourage caution. The morphology descriptors we presented are both approximations of surfaces and may not accurately reflect surfaces that are particularly convoluted. In addition, we used a relatively small set of sample tumors to determine the model parameters. There may be additional modes of morphological variation that this set did not account for. Thus, the forecast trends may simply not apply to a particular patient. In addition, subgroups of tumors that share characteristic changes in morphology might exist and provide further improvement in forecast accuracy as case solutions, if new tumors can be classified correctly.

Because a forecast will always remain, at best, an educated guess, and because of the risk posed to a patient were he or she to exhibit trends in anatomic changes dissimilar to those observed during model generation, models should not be considered a direct replacement for imaging during radiation therapy. However, anticipating anatomic changes with more complete morphology models like those presented here has a number of potential applications. First, with respect to a specific patient's treatment, a generative model that constructs the forecast morphology may provide insight as to how the tumor will likely move and deform relative to organs at risk and other neighboring anatomy. These models may better inform clinicians regarding decisions such as the appropriate number and timing of subsequent imaging exams and interventions. In addition, comparing a specific patient's response during treatment to population model predictions could determine if the patient is demonstrating an uncharacteristic

response. This information has implications for the management of the remainder of that patient's treatment, as well as for the design of future research studies. Finally, studies such as those aimed to reconstruct the dose delivered to a patient may also benefit from interpolating between imaging exams using models that more accurately reflect the changing tumor morphology.

5. CONCLUSION

We have presented models that anticipate longitudinal changes in tumor morphology using various models and model adjustment schemes. The accuracy of these models was evaluated and depended on their form. Their utility includes the characterization of patient-specific response with implications for treatment management and research study design. We encourage further investigation into the potential of integrating comprehensive, dynamic measures of patient response and anatomic variability into radiation therapy.

ACKNOWLEDGMENTS

The authors thank Jinzhong Yang, Ph.D., and Joy Zhang, Ph.D., for computational support of this work, and Kathryn Carnes for editing this manuscript. The authors also acknowledge financial support from the American Legion Auxiliary and The University of Texas Graduate School of Biomedical Sciences at Houston. No author has a conflict of interest to report regarding any facet of this study.

^{a)}Electronic mail: LECourt@MDAnderson.org

- ¹R. M. Seibert, C. R. Ramsey, J. W. Hines, P. A. Kupelian, K. M. Langen, S. L. Meeks, and D. D. Scaperth, "A model for predicting lung cancer response to therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **67**, 601–609 (2007).
- ²M. Azimi, "Predictive modelling of head and neck cancer tumour based on statistical analysis," *Int. J. Collaborative Enterprise* **3**, 60–71 (2012).
- ³A. K. Kamrani and M. Azimi, "Geometrical modeling of H&N cancer tumor," *Rapid Prototyping J.* **17**, 55–63 (2011).
- ⁴S. Takao, S. Tadano, H. Taguchi, and H. Shirato, "Computer simulation of radiotherapy for malignant tumor: A mechanical analogy method," *J. Biomech. Sci. Eng.* **4**, 576–588 (2009).
- ⁵M. S. Hoogeman, M. Van Herk, D. Yan, L. J. Boersma, P. C. M. Koper, and J. V. Lebesque, "A model to simulate day-to-day variations in rectum shape," *Int. J. Radiat. Oncol., Biol., Phys.* **54**, 615–625 (2002).
- ⁶G. G. Zeng, S. L. Breen, A. Bayley, E. White, H. Keller, L. Dawson, and D. A. Jaffray, "A method to analyze the cord geometrical uncertainties during head and neck radiation therapy using cone beam CT," *Radiother. Oncol.* **90**, 228–230 (2009).
- ⁷U. Bagci, J. H. Yao, K. Miller-Jaster, X. J. Chen, and D. J. Mollura, "Predicting future morphological changes of lesions from radiotracer uptake in 18F-FDG-PET images," *PLoS One* **8**, 1–14. (2013).
- ⁸L. H. Schwartz, J. A. C. Colville, M. S. Ginsberg, L. Wang, M. Mazumdar, J. Kalaigian, H. Hricak, D. Ilson, and G. K. Schwartz, "Measuring tumor response and shape change on CT: Esophageal cancer as a paradigm," *Ann. Oncol.* **17**, 1018–1023 (2006).
- ⁹N. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani, "Three-dimensional shape searching: State-of-the-art review and future trends," *Comput. Aided Des.* **37**, 509–530 (2005).
- ¹⁰J. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools Appl.* **39**, 441–471 (2008).
- ¹¹M. Styner, J. A. Lieberman, R. K. McClure, D. R. Weinberger, D. W. Jones, and G. Gerig, "Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4872–4877 (2005).
- ¹²C. Davatzikos, D. G. Shen, A. Mohamed, and S. K. Kyriacou, "A framework for predictive modeling of anatomical deformations," *IEEE Trans. Med. Imaging* **20**, 836–843 (2001).
- ¹³A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of benign and malignant masses based on Zernike moments," *Comput. Biol. Med.* **41**, 726–735 (2011).
- ¹⁴S. Takao, S. Tadano, H. Taguchi, K. Yasuda, R. Onimaru, M. Ishikawa, G. Bengua, R. Suzuki, and H. Shirato, "Accurate analysis of the change in volume, location, and shape of metastatic cervical lymph nodes during radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **81**, 871–879 (2011).
- ¹⁵D. Yan, D. A. Jaffray, and J. W. Wong, "A model to accumulate fractionated dose in a deforming organ," *Int. J. Radiat. Oncol., Biol., Phys.* **44**, 665–675 (1999).
- ¹⁶G. van der Wielen, T. Mutanga, L. Incrocci, W. Kirkels, E. Vasquez Osorio, M. Hoogeman, B. Heijmen, and H. de Boer, "Deformation of prostate and seminal vesicles relative to intraprostatic fiducial markers," *Int. J. Radiat. Oncol., Biol., Phys.* **72**, 1604–1611 (2008).
- ¹⁷Y. Jeong, R. J. Radke, and D. M. Lovelock, "Bilinear models for inter- and intra-patient variation of the prostate," *Phys. Med. Biol.* **55**, 3725–3739 (2010).
- ¹⁸E. M. Vázquez Osorio, M. S. Hoogeman, A. Al-Mamgani, D. N. Teguh, P. C. Levendag, and B. J. M. Heijmen, "Local anatomic changes in parotid and submandibular glands during radiotherapy for oropharynx cancer and correlation with dose, studied in detail with nonrigid registration," *Int. J. Radiat. Oncol., Biol., Phys.* **70**, 875–882 (2008).
- ¹⁹M. Sohn, M. Birkner, D. Yan, and M. Alber, "Modelling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and evaluation," *Phys. Med. Biol.* **50**, 5893–5908 (2005).
- ²⁰M. Sohn, B. Sobotta, and M. Alber, "Dosimetric treatment course simulation based on a statistical model of deformable organ motion," *Phys. Med. Biol.* **57**, 3693–3709 (2012).
- ²¹E. Fontenla, C. A. Pelizzari, J. C. Roeske, and G. T. Y. Chen, "Using serial imaging data to model variabilities in organ position and shape during radiotherapy," *Phys. Med. Biol.* **46**, 2317–2336 (2001).
- ²²D. L. Schwartz, A. S. Garden, J. Thomas, Y. P. Chen, Y. B. Zhang, J. Lewin, M. S. Chambers, and L. Dong, "Adaptive radiotherapy for head-and-neck cancer: Initial clinical outcomes from a prospective trial," *Int. J. Radiat. Oncol., Biol., Phys.* **83**, 986–993 (2012).
- ²³H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.* **50**, 2887–2905 (2005).
- ²⁴S. Lele and J. Richtsmeier, *An Invariant Approach to Statistical Analysis of Shapes* (Chapman and Hall, Boca Raton, FL, 2001).
- ²⁵M. Bander and C. Itzykson, "Group theory and hydrogen atom 1," *Rev. Mod. Phys.* **38**, 330–345 (1966).
- ²⁶M. Bander and C. Itzykson, "Group theory and hydrogen atom 2," *Rev. Mod. Phys.* **38**, 346–358 (1966).
- ²⁷R. J. Morris, R. J. Najmanovich, A. Kahraman, and J. M. Thornton, "Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons," *Bioinformatics* **21**, 2347–2355 (2005).
- ²⁸B. Cabral, N. Max, and R. Springmeyer, "Bidirectional reflection functions from surface bump maps," *Comput. Graph.* **21**, 273–281 (1987).
- ²⁹F. X. Sillion, J. R. Arvo, S. H. Westin, and D. P. Greenberg, "A global illumination solution for general reflectance distributions," *Comput. Graph.* **25**, 187–196 (1991).
- ³⁰A. S. Konopliv, W. B. Banerdt, and W. L. Sjogren, "Venus gravity: 180th degree and order model," *Icarus* **139**, 3–18 (1999).
- ³¹F. G. Lemoine, D. E. Smith, D. D. Rowlands, M. T. Zuber, G. A. Neumann, D. S. Chinn, and D. E. Pavlis, "An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor," *J. Geophys. Res., [Planets]* **106**, 23359–23376, doi:10.1029/2000JE001426 (2001).
- ³²B. Tapley, J. Ries, S. Bettadpur, D. Chambers, M. Cheng, F. Condi, B. Gunter, Z. Kang, P. Nagel, R. Pastor, T. Pekker, S. Poole, and F. Wang, "GGM02—An improved Earth gravity field model from GRACE," *J. Geod.* **79**, 467–478 (2005).
- ³³N. Olsen, H. Luhr, T. J. Sabaka, M. Manda, M. Rother, L. Tofner-Clausen, and S. Choi, "CHAOS—a model of the Earth's magnetic field derived from CHAMP, Orsted, and SAC-C magnetic satellite data," *Geophys. J. Int.* **166**, 67–75 (2006).
- ³⁴M. Koamionkowski, A. Kosowsky, and A. Stebbins, "Statistics of cosmic microwave background polarization," *Phys. Rev. D* **55**, 7368–7388 (1997).