A probability model for anatomical robust optimisation in head and neck cancer proton therapy

Ying Zhang[1], Jailan Alshaikhi[2], Wenyong Tan[3], Gary Royle[1], Esther Bär[1,4]

1. Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom
2. Saudi Proton Therapy Center, King Fahad Medical City, Riyadh, Saudi Arabia
3. Department of Oncology, Shenzhen Hospital of Southern Medical University Shenzhen 518101, China
4. University College London Hospitals NHS Foundation Trust, Radiotherapy Physics, 250 Euston Road, London NW1 2PG, United Kingdom

E-mail: ying.zhang.18@ucl.ac.uk

May 2021

Abstract.

Objective: Develop an anatomical model based on the statistics of the population data and evaluate the model for anatomical robust optimisation in head and neck (H&N) cancer proton therapy.

Approach: Deformable Image Registration (DIR) was used to build the probability model (PM) that captured the major deformation from patient population data and quantified the probability of each deformation. A cohort of 20 nasopharynx patients was included in this retrospective study. Each patient had a planning CT and 6 weekly CTs during radiotherapy. We applied the model to 5 test patients. Each test patient used the remaining 19 training patients to build the PM and estimate the likelihood of a certain anatomical deformation to happen. For each test patient, a spot scanning proton plan was created. The PM was evaluated using proton spot location deviation and dose distribution.

Main results: Using the proton spot range, the PM can simulate small non-rigid variations in the first treatment week within 0.21 $\pm$0.13 mm. For overall anatomical uncertainty prediction, the PM can reduce anatomical uncertainty from 4.47$\pm$1.23 mm (no model) to 1.49$\pm$1.08 mm at week 6. The 95% confidence interval (CI) of dose metric variations caused by actual anatomical deformations in the first week is -0.59 $\sim$ -0.31 % for low-risk CT D_{95}, and 0.84 $\sim$ 3.04 Gy for parotid D_{mean}. On the other hand, the 95% CI of dose metric variations simulated by the PM at the first week is -0.52 $\sim$ -0.34% for low-risk CTV D_{95}, and 0.58 $\sim$ 2.22 Gy for parotid D_{mean}.

Significance: The PM improves the estimation accuracy of anatomical uncertainty compared to the previous models and does not depend on the acquisition of the weekly CTs during the treatment. We also provided a solution to quantify the probability of an anatomical deformation. The potential of the model for anatomical robust optimisation is discussed.

Keywords: Anatomical model, uncertainty evaluation, proton therapy

Submitted to: *Phys. Med. Biol.*

1. Introduction

Intensity modulated proton therapy (IMPT) has advantages in delivering a conformal dose distribution to the target while minimizing the dose to the adjacent normal tissue [1, 2, 3, 4], exploiting the steep falloff of the Bragg peak. However, this precise delivery technique has inherent sensitivity to uncertainties, which are especially common in the head and neck (H&N) cancer treatment. Uncertainties may degrade the quality of treatment. To help design the congruous mitigation methods, uncertainty needs to be systematically explored.

Current research in H&N proton therapy delivery focuses on the mitigation of anatomical changes[5, 6]. For H&N cancer patients, the anatomical changes can be divided into two types: small non-rigid variations (sNRVs) and progression. For sNRVs, nasal filling, jaw movement, neck folds, spine flexion and shoulder position changes are common during treatment [7, 5]. sNRVs occur randomly. The dosimetric influence of these changes was revealed by Zhang *et al.*[8]. For progression changes, Tan *et al.* [9] reported an average tumor volume shrinkage of 36.5% (20% $\sim$ 60%.) in a cohort of 20 nasopharynx cancer patients. Bhide *et al.* [10] showed that for 20 H&N patients, the parotid volume decreased with a reduction rate between 21.3% and 42%, with an average medial shift of the parotid centre of mass of 2.3 mm (as measured at week 4)

Anatomical robust optimisation (aRO) is a recently proposed technique to design robust plans of resilience to anatomical changes. In aRO, multiple CTs acquired during the treatment are included in the optimisation[11, 12]. However, it either requires multiple scanning before planning[11] or needs the images acquired during the treatment[12]. While multiple scanning gives the extra imaging dose to the H&N patient and affects the efficiency of a busy proton therapy practice, the dependence of images acquired during the treatment also compromises the benefits of reducing the rate of replanning. Anatomical models that simulate the possible geometric variations from a population of patient data remove the requirement of multiple scanning and the dependence of CT images acquired during the treatment by taking the predicted images into aRO. Several mathematical models have been proposed to account for anatomical changes[13, 14, 15, 16]. Yu *et al.* used an anatomical model for deformable image registration (DIR) evaluation. Zhang *et al.* used a predictive model in head and neck patients in offline adaptive proton therapy [15, 16].

The previous models proposed by Yu *et al.*[14] and Zhang *et al.*[15] mainly predict the patient-specific progressive changes during the treatment. They required weekly CTs during the treatment to build or update the model, which limited the benefits of creating a robust plan against anatomical changes at the planning stage. In reality, a CT is only a snapshot of the anatomy, it cannot fully represent the anatomical changes that happened to the patients during the treatment. Small non-rigid anatomical changes always exist. Apart from that, the influences of acute toxicities from cancer treatment on H&N patients' eating during the treatment also determine the weekly anatomical changes. Considering that accurately predicting anatomical changes during

the treatment at the beginning is challenging, an alternative way is to build a probability model that generates the predicted images based on the statistics of the population data. To be more specific, the model generates plausible deformations instead of one specific deformation. To capture the major deformations in a population, we investigated the use of principal component analysis (PCA). PCA finds the best orthogonal basis, the principal components (PCs), whose variance of the projections of the data are ranked from the greatest to the smallest. Thus, it is possible to restore information using a limited number of PCs that describe the majority of anatomical deformations. The basics of PCA were detailed in Lever *et al.* [17]. However, the assumption that generated deformations would happen with equal probability is not reasonable. To date, it is still challenging to quantify the probability of a certain type of anatomical deformation to arise during the treatment course. In this paper, we exploited the orthogonality of PC to do the quantification. The calculated probability can assist to assess the anatomical uncertainty for a patient cohort. Also, the predicted images with higher possibilities can be chosen for aRO and the corresponding possibility can be used to design robust optimisation objectives in aRO.

In this work, we aim to: 1) Develop a probability model (PM) based on PCA to model major deformations in patients. 2) Quantify the probability of each type of anatomical deformation based on population data. 3) Validate the feasibility of the PM to measure anatomical uncertainty based on range and dose distribution.

2. Method and material

2.1. Patient data

Twenty nasopharyngeal carcinoma patients were recruited retrospectively. Each patient underwent a planning CT (pCT) and a weekly repeat CT (rCT_t), where t (t=0,1,2,3,...) represents the week of CT scanning. CT images were all acquired using a Brilliance Big Bore CT simulator (Philips, Inc, Cleveland, OH, USA). The details are listed in Table 1. Contours in the planning CT and weekly CTs were manually delineated by an oncologist. The leave-one-out strategy was applied for 5 test patients. Each test patient used the 19 training patients to build the model.

Table 1: CT image acquisition details

Tube voltage	Reconstruction diameter	Slince thickness	Pixel Spacing	Data Collection Diameter
120kVp	500 mm	3 mm	0.98 mm	600 mm

For all 5 test patients, an original (nominal) IMPT treatment plan with three beam fields (60°, 180°, 300°) was generated using the Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA). All plans generated throughout this study were robustly optimised with ±3 mm setup and ±3.5% range uncertainty for CTVs and critical organs at risk (OARs). A relative biological effectiveness (RBE) of 1.1 for proton beams was

used. The dosimetric goals and priorities for all plans in this study are summarised in Table 2. A plan was deemed acceptable if the goals set for the CTV and serial organs are fulfilled in the nominal scenario (the error free distribution) as well as all 12 dose distributions (3 mm orthogonal shifts combined with the ±3.5% range error) in a robust evaluation. More clinical characteristics of the patients can be found in the paper of Tan *et al.* [9, 18].

Table 2: Dosimetric goals of the treatment plans created in this study

.

Structure	Goal under uncertainty
High-risk-CTV	D_{95}(The minimum dose to 95% of target volume) > 95% of prescription dose (72.6 Gy(RBE), 33 fractions)
Low-risk-CTV	D_{95} > 95% of prescription dose (63 Gy(RBE), 33 fractions)
CTV	D_2 (The minimum dose to the hottest 2% volume) < 107% of prescription dose
Spinal cord	D_{max} (The maximum dose in the volume) < 45 Gy(RBE)
Brainstem	D_{max} <55 Gy(RBE)
Chiasm	D_{max} <55 Gy(RBE)
Structure	Goal in Nominal
Parotid glands	D_{mean} (The mean dose in the volume) <26 Gy(RBE)
Oral cavity	D_{mean} <40 Gy(RBE)
Larynx	D_{mean} <40 Gy(RBE)
Proton planning information: MFO planning; spot spacing size: 5mm; energy range: 70 MeV – 250MeV; range shifter: 5cm dose calculation algorithm: Pencil beam scanning (PBS); optimisation algorithm: Nonlinear Universal Proton Optimiser.	

2.2. Methods

2.2.1. Deformable image registration We chose to use the diffeomorphic image registration implemented in NiftyReg [19] to identify anatomical changes between two images. It is invertible, differentiable and whose inverse is also differentiable [20, 21]. The deformation vector fields (DVFs) ($\boldsymbol{\phi}$) are expressed as:

$$\boldsymbol{\phi} = \exp(\boldsymbol{v}), \tag{1}$$

where $\boldsymbol{v}$ is the stationary velocity field (SVFs) of the diffeomorphic image registration [21] used in this project.

The inverse transformation $\boldsymbol{\phi}^{-1}$ can be calculated as

$$\boldsymbol{\phi} = \exp(\boldsymbol{v}) \quad \Rightarrow \quad \boldsymbol{\phi}^{-1}(x) = \exp(-\boldsymbol{v}). \tag{2}$$

To obtain inter-fraction DVFs of patients, rigid registration was done before DIR using Niftyreg. In DIR, pCT was the reference geometry, each rCT_t was registered to its respective pCT to produce transformation $\boldsymbol{v}_{p\to t}$, where p stands for pCT.

The generated $\boldsymbol{v}_{p\to t}$ is applied to the rCTs and its contours to generate dCTs and propagated contours. The nominal plan is applied to the dCTs to calculate the dose distributions respectively. In an ideal algorithm, dCTs should be the same as the pCT. Therefore, dCTs should have the same dose distribution as the planning dose distribution. The DIR algorithm was fully evaluated by the 3-dimensional mean surface distance (MSD) between propagated contours and planning contours, and the 3D gamma

index (2%/2mm) between accumulated dose distribution of dCTs and the planning dose distribution.

The maximum value of MSD was below 3 mm (slice thickness) for all structures. The gamma index between accumulated dose using dCTs and planning dose is above 97.73%. An illustration of the DIR based on the contour of the spinal cord is shown in Appendix A.

2.2.2. Probability model In the clinic, the magnitude of uncertainty is estimated from population data [22, 23, 24, 25]. In this paper, we used the PM to statistically model the anatomical changes of population based on PCA. The procedure of building the PM was divided into the following steps and repeated for each treatment week t.

(i) To ensure that the inter-fraction DVFs of patients were in the same space and had the same resolution, we projected the DVFs into the atlas space using:

$$\boldsymbol{v}_{a,p\to t} = \boldsymbol{R}_{a\to p}^{-1} \circ \boldsymbol{v}_{p\to t} \circ \boldsymbol{R}_{a\to p}. \tag{3}$$

where $\boldsymbol{R}_{a\to p}$ presents the SVF from pCT to altas CT.

(ii) The average SVF for week t in the atlas space was calculated as the expectation value E of the deformation $\boldsymbol{v}_{a,p\to t}$ of the training dataset

$$\mathbf{E}(\boldsymbol{v}_{a,p\to t}) = \frac{1}{N_p} \sum_{pi} \boldsymbol{v}_{a,p\to t}^{pi}, \tag{4}$$

where N_p is the number of patients used in this model and pi is the patient index.

(iii) The random deformation of each patient at week t in the atlas space can be calculated as following:

$$\boldsymbol{v}_{a,p\to t}^{\mathrm{rand},pi} = \boldsymbol{v}_{a,p\to t}^{pi} - E(\boldsymbol{v}_{a,p\to t}). \tag{5}$$

(iv) The random deformations of all training patients at week t were composed to a random deformation matrix in the atlas space, referred to as $\boldsymbol{v}_{a,p\to t}^{\mathrm{rand}}$, which was represented approximately using

$$\boldsymbol{v}_{a,p\to t}^{\mathrm{rand}} = (\boldsymbol{v}_{a,p\to t}^{\mathrm{rand},1}, \boldsymbol{v}_{a,p\to t}^{\mathrm{rand},2}, \cdots, \boldsymbol{v}_{a,p\to t}^{\mathrm{rand},N_p}) \approx \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,L} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N_p,1} & \alpha_{N_p,2} & \cdots & \alpha_{N_p,L} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \vdots \\ \boldsymbol{V}_L \end{bmatrix}. \tag{6}$$

$\boldsymbol{V}_l$ is the PC vector, also called eigenvector. $\alpha_{i,l}$ is the coefficient of the l-th eigenvector belonging to the i-th training set. L is the number of eigenvectors used to build the model. L was chosen to be able to represent 90% of population variations. Each column of the α matrix represents the coefficients of one eigenvector.

(v) The probability density function (PDF) of α of an eigenvector was annealed using kernel density estimation [26]. The estimation used

$$p_l(\alpha) = \frac{1}{N_p \cdot \sigma_l \sqrt{2\pi}} \sum_{i=1}^{N_p} \exp^{-(\alpha-\alpha_{i,l})^2/2\sigma_l^2}, \tag{7}$$

$$\sigma_l = \frac{1.06}{{N_p}^{0.2}} \sqrt{\frac{\sum_i^{N_p} (\alpha_{i,l} - \frac{1}{N_p}\sum_i^{N_p} \alpha_{i,l})^2}{N_p - 1}}, \tag{8}$$

where p_l is the probability distribution of the coefficients of the l-th eigenvector. The comparison between real distribution of a column of α and the annealed distribution is shown in figure 1.

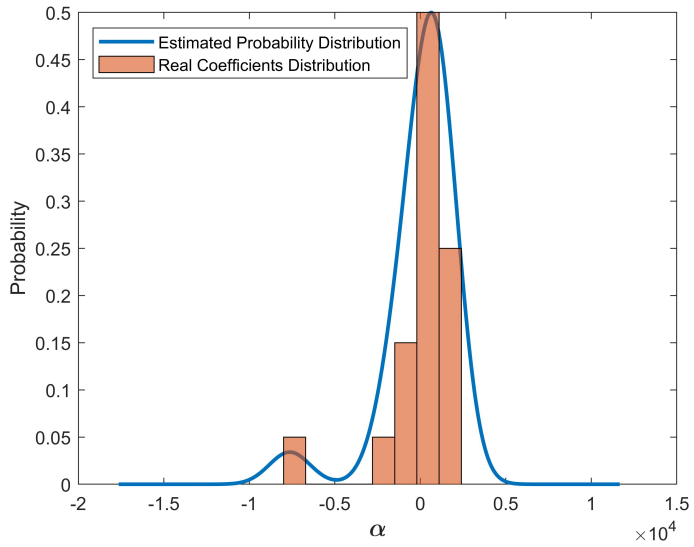


Figure 1: The comparison between real distribution of a column of α and the annealed distribution.

(vi) The sampled numbers from p_l formed the l-th column of the $\tilde{\alpha}$ matrix. Because the distribution was estimated from a limited training dataset, sampling extended the coefficients to capture all the possible random anatomical changes resulting in

$$\tilde{\boldsymbol{v}}^{\text{rand}}_{a,p\to t} = (\tilde{\boldsymbol{v}}^{\text{rand},1}_{a,p\to t}, \tilde{\boldsymbol{v}}^{\text{rand},2}_{a,p\to t}, \cdots, \tilde{\boldsymbol{v}}^{\text{rand},N_s}_{a,p\to t}) \approx \begin{bmatrix} \tilde{\alpha}_{1,1} & \tilde{\alpha}_{1,2} & \cdots & \tilde{\alpha}_{1,L} \\ \tilde{\alpha}_{2,1} & \tilde{\alpha}_{2,2} & \cdots & \tilde{\alpha}_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\alpha}_{N_s,1} & \tilde{\alpha}_{N_s,2} & \cdots & \tilde{\alpha}_{N_s,L} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \vdots \\ \boldsymbol{V}_L \end{bmatrix}. \tag{9}$$

Each row of the predictive matrix $\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}} = (\tilde{\alpha}_{i,1}, \tilde{\alpha}_{i,2}, \cdots, \tilde{\alpha}_{i,L})$ in equation 9 was multiplied with $\boldsymbol{V}$ to form a predicted random deformation for week t, represented by $\tilde{\boldsymbol{v}}^{\text{rand},i}_{a,p\to t}$. i is the index of the predicted random deformations. N_s is the number of samples.

(vii) A deformation of the PM for week t is:

$$\boldsymbol{v}_{a,p\to t}^{\mathrm{PM},i} = E(\boldsymbol{v}_{a,p\to t}) + \tilde{\boldsymbol{v}}_{a,p\to t}^{\mathrm{rand},i}, \quad i\forall(1 \sim N_s). \tag{10}$$

(viii) Each deformation $\boldsymbol{v}_{a,p\to t}^{\mathrm{PM},i}$ was transformed into the space of an individual patient using

$$\boldsymbol{v}_{p\to t}^{\mathrm{PM},i} = \boldsymbol{R}_{a\to p}^{-1} \circ \boldsymbol{v}_{a,p\to t}^{\mathrm{PM},i} \circ \boldsymbol{R}_{a\to p}. \tag{11}$$

(ix) To warp the pCT, the transformation must be directed from the predicted anatomy to the pCT. The deformation $\boldsymbol{v}_{t\to p}^{\mathrm{PM},i}$ needs to be reversed using:

$$\boldsymbol{v}_{t\to p}^{\mathrm{PM},i} = -\boldsymbol{v}_{p\to t}^{\mathrm{PM},i}. \tag{12}$$

(x) The warped images $\mathrm{CT}_t^{\mathrm{PM},i}$ were acquired using:

$$\boldsymbol{\phi}_{t\to p}^{\mathrm{PM},i} = \exp(\boldsymbol{v}_{t\to p}^{\mathrm{PM},i}), \tag{13}$$

$$\mathrm{CT}_t^{\mathrm{PM},i} = \boldsymbol{\phi}_{t\to p}^{\mathrm{PM},i}(\mathrm{pCT}), \tag{14}$$

where $\mathrm{CT}_t^{\mathrm{PM},i}$ is the i-th predicted image of week t. $\boldsymbol{\phi}_{t\to p}^{\mathrm{PM},i}$ is also applied to contours of planning CT to get the deformed contours on $\mathrm{CT}_t^{\mathrm{PM},i}$.

We can now obtain N_s predicted images for week t. Considering that eigenvectors are orthogonal, the probability distribution of their coefficients are independent. Therefore, the probability of predicted images with specific $\boldsymbol{\alpha_i}$ can be calculated by the joint probability

$$P(\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}}) = p_1(\tilde{\alpha}_{i,1}) \cdot p_2(\tilde{\alpha}_{i,2}) \cdot \ldots \cdot p_L(\tilde{\alpha}_{i,L}), \qquad \sum_i^{N_s} P(\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}}) = 1. \tag{15}$$

2.3. Model evaluation

In this section, we evaluated the PM in terms of estimating anatomical uncertainty based on weighted spot location deviation (WSLD) and dose distribution.

2.3.1. Model evaluation based on weighted spot location deviation. WSLD measured the uncertainty based on range changes. We used the spot positions and weights, derived from the treatment plan file as both the spot positions and weights [27] affect the dose distribution. The method to determine the spot location is described in Appendix B. The WSLD is presented in equation 16:

$$\mathrm{WSLD} = \sum_r |\boldsymbol{r}_{\mathrm{uncertainty}} - \boldsymbol{r}_{\mathrm{reference}}| \cdot w_r, \ \sum w_r = 1, \tag{16}$$

where $\boldsymbol{r}$ is a spot position in the CT. $\boldsymbol{r}_{\mathrm{reference}}$ is the spot location in the reference frame (pCT). $\boldsymbol{r}_{\mathrm{uncertainty}}$ is the spot location under an anatomical variation (predicted $\mathrm{CT}^{\mathrm{PM}}$). The spot weight w_r is normalized to 1.

The WSLD is calculated for each of the $\mathrm{CT}^{\mathrm{PM}}$s. The total WSLD that combines the WSLD of $\mathrm{CT}_t^{\mathrm{PM},i}$ and its estimated probability from the extended population was calculated using

$$\mathrm{WSLD}_t^{\mathrm{PM}} = \sum_{i=1}^{N_s} (\mathrm{WSLD}^{\mathrm{CT}_t^{\mathrm{PM},i}} \cdot P(\boldsymbol{\alpha_i})). \tag{17}$$

$\mathrm{WSLD}^{\mathrm{CT}_t^{\mathrm{PM},i}}$ is the WSLD from the predicted image $\mathrm{CT}_t^{\mathrm{PM},i}$ at week t. $P(\boldsymbol{\alpha^i})$ is the probability of the predicted image $\mathrm{CT}_t^{\mathrm{PM},i}$. N_s is the number of samples produced by the PM.

The $\mathrm{WSLD}^{\mathrm{PM}}$ is used to estimate the anatomical uncertainty from three aspects: small non-rigid variations (sNRVs), total anatomical uncertainty, and residual anatomical uncertainty from the PM, as follows:

- The small non-rigid positioning anatomical uncertainty and total anatomical uncertainty simulated by the PM. The PM statistically summarized the probability of anatomical changes for each week. Progressive changes induced by the radiation in the first week are not significant [9]. Therefore, the WSLD estimated by the PM in the first week of treatment represented the influence from sNRVs such as tongue movement, shoulder positioning or small rotations. The WSLD from later fractions was the combined influence of sNRVs and progressive changes (total anatomical uncertainty).
- The residual anatomical uncertainty from the PM - The difference between the estimated anatomical uncertainty from the PM and actual anatomical uncertainty was used to evaluate the accuracy of the PM. We referred to it as the residual anatomical uncertainty ($\Delta\mathrm{WSLD}_{\mathrm{res}}$), see equation 18,

$$\Delta\mathrm{WSLD}_t^{\mathrm{res}} = \mathrm{WSLD}_t^{\mathrm{real}} - \mathrm{WSLD}_t^{\mathrm{PM}}, \tag{18}$$

  where $\mathrm{WSLD}_t^{\mathrm{real}}$ is the actual anatomical uncertainty calculated by the WSLD between rCT_t and pCT, which is also corresponding to the residual anatomical uncertainty of no model. $\mathrm{WSLD}_t^{\mathrm{PM}}$ is the anatomical uncertainty at week t estimated by the PM. Ideally, the model should approach a ΔWSLD of 0 for each treatment week t.

2.3.2. Model evaluation based on dose distribution. Anatomical deformations lead to dose variations. We recalculated the dose on the deformed images using the original IMPT plan. Then we compared 1) the actual dose variations from the training dataset, and 2) the dose variations simulated by the PM. The deformations of an exemplary patient in the first week were chosen to demonstrate the dosimetric influence from small non-rigid variations.

1) To obtain the actual dose variations, we applied the actual deformations of 20 patients in the first week to the planning CT of the exemplary patient to obtain 20

actual sNRVs. The original IMPT plan of this exemplary patient was applied to the 20 actual sNRVs to calculate the dose variations. These variations can be illustrated as dose volume histogram (DVH) bands in the nominal DVHs of organs, referred to as actual DVH bands. We also calculated perturbed dose metrics for each considered dose metric Dx (e.g. D_{95}). The perturbed dose metrics subtract the nominal dose metrics to obtain the dose metric discrepancy ΔDx.

2) To obtain the dose variations simulated by the PM, 20 CT_1^{PM} (CT of the PM in the first week) were selected following the joint probability distribution of the PM for the exemplary patient. The same IMPT plan was also applied to the 20 CT_1^{PM} to create DVH bands, referred to as the simulated DVH bands from the PM. The dose metric discrepancy simulated by the PM was referred to as ΔDx'.

3. Results

The exemplary patient's slice images from the rCT_1, rCT_6 and 2 predicted CTs of the PM in the first week and the sixth week are shown in figure 2.

3.1. Model evaluation based on weighted spot location deviation.

The WSLD of the anatomical uncertainty is estimated from the PM. The result is shown in average WSLD with 95% confidence interval (CI) (see figure 3). The estimated anatomical uncertainty from small non-rigid anatomical changes accounted for a range uncertainty of 2.18±0.19 mm. The estimated total anatomical uncertainty (from sNRVs and progressive changes) can reach 3.09±0.26 mm at week 6.

The residual anatomical uncertainty from no model and the PM($N_s = 100$) were compared in figure 4. In no model, predicted images were replaced by the planning CT. When the anatomical uncertainty estimated from the predicted images of the PM was considered, the residual anatomical uncertainty was reduced from 4.47±1.23 mm (no model) to 1.49±1.08 mm (PM) at week 6, achieving a significant improvement as compared to no model.

The comparison of individual cases between the actual WSLD (using rCTs) and the estimated WSLD from the PM are listed in Table 3.

Table 3: WSLD caused by actual anatomical deformations (using rCTs) and WSLD estimated by the PM in each test patient and week.

Week	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5	
	rCT(mm)	PM(mm)	rCT(mm)	PM(mm)	rCT(mm)	PM(mm)	rCT(mm)	PM(mm)	rCT(mm)	PM(mm)
1	1.72	2.12	1.96	1.88	2.57	2.53	2.32	2.14	1.89	2.27
2	2.52	2.42	2.07	2.18	1.99	2.71	2.30	3.11	2.54	2.55
3	3.43	2.75	2.15	2.30	2.59	3.23	2.57	3.20	3.49	2.72
4	4.93	2.65	2.78	2.65	2.94	3.62	3.44	3.14	4.69	3.10
5	5.62	2.97	2.73	2.30	4.02	3.53	2.91	3.15	5.57	3.20
6	5.23	2.78	3.12	3.07	5.13	3.63	2.62	2.89	6.27	3.10

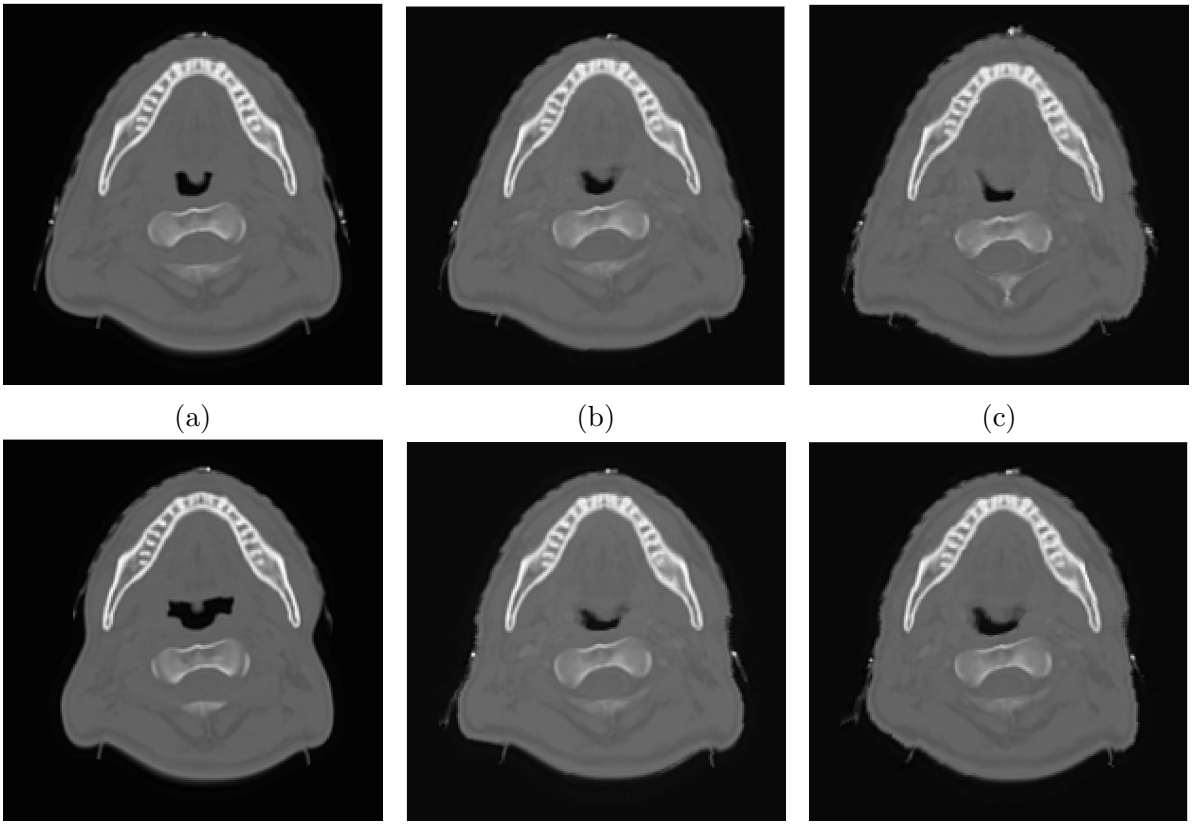


Figure 2: The exemplary patient's slice images from the rCT_1, rCT_6 and 2 predicted CT of the PM in the first week and the sixth week. (a) shows a slice image from the rCT_1 of the exemplary patient. (b)-(c) show slice images from 2 predicted CTs of the PM in the first week. (d) shows the same slice image from the rCT_6 of the exemplary patient. (e)-(f) show slice images from 2 predicted CTs of the PM in the sixth week.

3.2. Model evaluation based on dose distribution.

For the exemplary patient, the actual DVH bands in the first week and the simulated DVH bands from the PM in the first week is shown in figure 5. Supporting the rationality of the PM, the simulated DVH bands of the PM demonstrate similar variations as the actual DVH bands.

The DVH bands are intuitive but only demonstrate the worst cases. The statistics of dose metrics need to be summarised to reveal the true ability of the PM. We listed the maximum, minimum, mean value (μ), and margin of error (E) (used to form 95% CI) of the ΔDx from the actual sNRVs and the ΔDx' from the simulated sNRVs of the PM in Table 4.

The 95% CI of dose metric variations caused by actual anatomical deformations in the first week is from -0.59 % to -0.31 % (low-risk CTV D_{95}), from 0.84 Gy to 3.04 Gy (parotid glands D_{mean}) and from -0.96 Gy to 1.90 Gy (spinal cord D_2) for low-risk CTV D_{95}, parotid glands D_{mean} and spinal cord D_2, respectively. While the range of dose

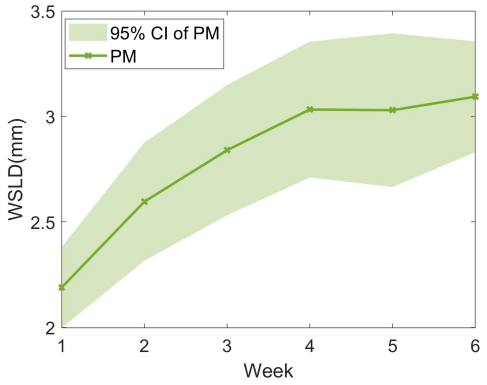


Figure 3: The anatomical uncertainty estimated from the PM in WSLD for each week. The result is estimated in average WSLD with 95% CI across the 5 test cases. The WSLD in the first week presents the uncertainty from non-rigid positioning. The WSLD in the following weeks evaluates the combined effect of anatomical uncertainty from non-rigid positioning and progressive anatomical changes (total anatomical uncertainty).

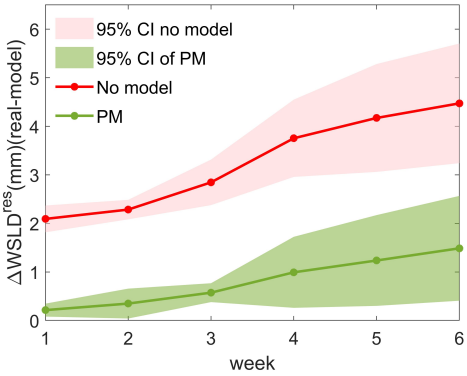
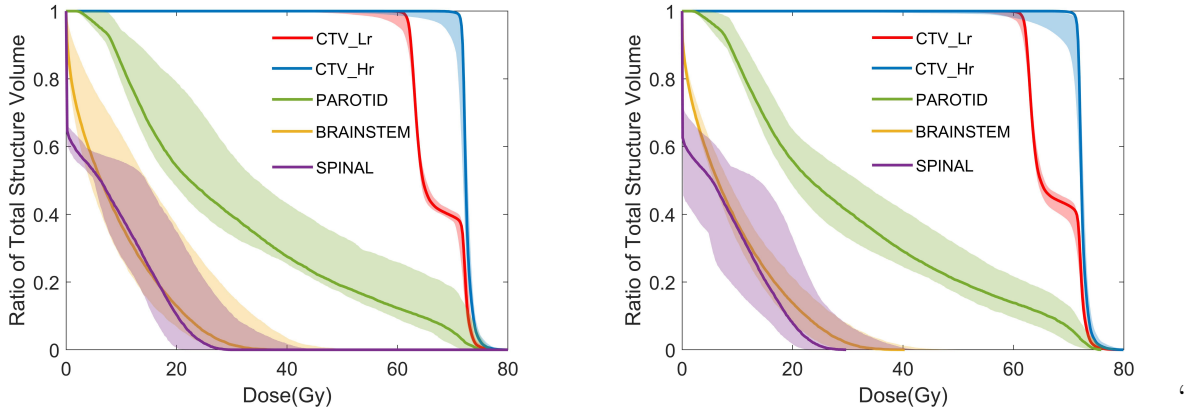


Figure 4: The residual anatomical uncertainty in WSLD. The residual anatomical uncertainty comes from no model,in which predicted images were replaced by planning CT, and the PM(N_s = 100) were compared. The result indicates the average difference with 95% CI between the estimated WSLD from the PM and the actual WSLD across the 5 test dataset.



(a) (b)

Figure 5: Evaluation of the PM based on dose distribution. a) shows the bandwidth from actual DVH band in the first week. b) shows the bandwidth simulated from the PM in the first week.

metric variations simulated by the PM is from -0.52% to -0.34% (low-risk CTV D_{95}), from 0.58 Gy to 2.22 Gy (parotid glands D_{mean}) and from -1.65 Gy to 1.35 Gy (spinal

Table 4: Dose metrics discrepancy (perturbed dose metrics - nominal dose metric) from the actual sNRVs (ΔDx) and the simulated sNRVs from the PM (ΔDx') are listed in the maximum, minimum, mean value (μ) and margin of error (E) of 95% confidence interval.

	Low-risk CTV (ΔD_{95}(%))		High-risk CTV(ΔD_{95}(%))		Parotid Glands(ΔD_{mean}(Gy))		Brainstem(ΔD_2(Gy))		Spinal(ΔD_2(Gy))	
	Actual	Simulated	Actual	Simulated	Actual	Simulated	Actual	Simulated	Actual	Simulated
Maximum	-0.05	0	-0.19	0	6.83	5.81	8.91	4.98	12.59	12.03
Minimum	-1.46	-0.97	-3.12	-5.83	-2.15	-2.35	-5.08	-5.12	-3.98	-3.04
μ	-0.45	-0.43	-1.02	-0.88	1.95	1.4	-0.08	0.04	0.47	-0.15
E	0.14	0.09	0.40	0.38	1.10	0.82	1.09	1.1	1.43	1.50

cord D_2) for low-risk CTV D_{95}, parotid glands D_{mean} and spinal cord D_2, respectively.

4. Discussion

In this paper, an anatomical model was developed based on the statistics of the population data. We quantified the probability of an anatomical deformation to arise during the treatment. The model accuracy was evaluated based on WSLD and dose distribution.

Zhang *et al.* [15] showed that the effect of anatomical progressions in the first week of treatment was not significant. Therefore, we can use the PM of the first treatment week to evaluate the uncertainty from small non-rigid anatomical changes. The PM can simulate the small non-rigid variations in the first treatment week within 0.21 $\pm$0.13 mm. For overall anatomical uncertainty prediction, the PM can reduce anatomical uncertainty from 4.47$\pm$1.23 mm (no model) to 1.49$\pm$1.08 mm at week 6 (see figure 4). Compared to the anatomical uncertainty estimated by the previous AM/RIM model [15], the PM further improves the estimation accuracy by 1.24 mm at week 3 and 0.4 mm at week 6 on average. Besides, the PM model does not rely on the acquisition of the weekly CT during the treatment, making it more suitable to estimate anatomical uncertainty.

For the dose metrics in Table 4, the 95% CI of the simulated dose metric of the PM in the first week is basically within the 95% CI of actual anatomical deformations, supporting that the PM is feasible to simulate the anatomical variations.

The inclusion of more scenarios in the training dataset can improve the probability estimation. In probabilistic treatment planning or robust optimisation, uncertainty scenarios are often described using uniform distribution [28] or normal distribution [29] in the cost function. However, it is difficult to correspond one uncertainty scenario with a probability. This work exploited the independence between PCs to calculate the probability for each predicted CT. This can be used to design the cost functions of aRO.

The possibility of using cone-beam CT (CBCT) to build the model has been discussed in [15]. In a very recent study, plan robustness against anatomical changes was investigated by aRO. Mesías *et al.* [11] and Yang *et al.* [12] both concluded that this method improved plan robustness toward anatomical changes and reduced the number

of plan adaptations for H&N patients. However, Mesías *et al.* [11] required multiple scanning to produce extra CT images before treatment for robust optimisation. It will add burden to a busy clinic. Yang *et al.* [12] used the images from the first plan adaption to include progressive anatomical changes into the second adaptive plan. However, it limits the creation of a robust plan at the early planning stage. To overcome these limitations, the PM was developed based on the population data to capture systematic progressions and comprehensive random deformations of H&N patients, making it possible to include anatomical changes before treatment without extra burden[28]. Online adaption is an aspirational technique intended for same-day application. However, their results were based on static images, acquired several minutes (median reported adaption time: 12 minutes) before treatment application[30]. To consider the possibility of small patient movements during the waiting time, non-rigid positional changes can in the future be included in robust optimisation for current online adaption techniques. Such changes can, for example, be inferred from the here suggested PM in the first treatment week.

aRO often comes at the cost of increasing the integral dose[30]. Since the PM is developed at each time-point, aRO using the PM can be explored to find the best way of balancing plan robustness and integral dose increase. The potential strategies are using: 1) the PM of different weeks, 2) the systematic progressive changes of different weeks plus the small non-rigid variation of the first week. Future studies using different strategies for aRO are underway.

5. Conclusion

The novelty of the paper is present in two aspects: 1) Instead of producing one predicted CT at each time point, we produce predicted CTs based on the statistics of the population each week, which is more close to reality. 2) we give a solution to calculate the probability of a certain type of anatomical change. The probability improves the accuracy of estimating the anatomical uncertainty and can be used to design the cost function of aRO. Future studies validating the potential clinical application, such as aRO, are underway.

Conflict of interest: None

Funding statement: Ying Zhang is supported by the scholarship from the China Scholarship Council Nos. 201809150003. Esther Bär is supported by the CRUK RadNet City of London Award C7893/A28990. Wenyong Tan is supported by the National Natural Science Funding of China (81974462).

Data share statement:

Restricted access. The authors cannot make patient data publicly available due to data use agreement. Programming code is available on the GitHub

https://github.com/YingClara92/Anatomical_Models.git.

References

[1] Marloes Steneker, Antony Lomax, and Uwe Schneider. Intensity modulated photon and proton therapy for the treatment of head and neck tumors. *Radiotherapy and Oncology*, 80(2):263–267, 2006.

[2] Tara A Van De Water, Antony J Lomax, Hendrik P Bijl, Marije E De Jong, Cornelis Schilstra, Eugen B Hug, and Johannes A Langendijk. Potential benefits of scanned intensity-modulated proton therapy versus advanced photon therapy with regard to sparing of the salivary glands in oropharyngeal cancer. *International Journal of Radiation Oncology* Biology* Physics*, 79(4):1216–1224, 2011.

[3] S. E. McGowan, N. G. Burnet, and A. J. Lomax. Treatment planning optimisation in proton therapy. 86(1021):20120288, 2013.

[4] Hideki Minatogawa, Koichi Yasuda, Yasuhiro Dekura, Seishin Takao, Taeko Matsuura andTakaaki Yoshimura, Ryusuke Suzuki, and et al. Potential benefits of adaptive intensity-modulated proton therapy in nasopharyngeal carcinomas. *Journal of applied clinical medical physics*, 22(1):174–183, 2021.

[5] Tom Bruijnen, Bjorn Stemkens, Chris H J Terhaard, Jan J W Lagendijk, Cornelis P J Raaijmakers, and Rob H N Tijssen. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. *Radiotherapy and Oncology*, 130:82–88, 2019.

[6] Mislav Bobić, Arthur Lalonde, Gregory C Sharp, Clemens Grassberger, Joost M Verburg, Brian A Winey, Antony John Lomax, and Harald Paganetti. Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy. *Physics in Medicine & Biology*, 66:055023, 2021.

[7] Eric K Hansen, M Kara Bucci, Jeanne M Quivey, Vivian Weinberg, and Ping Xia. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics*, 64(2):355–362, 2006.

[8] Ying Zhang, Jailan Alshaikhi, Richard A. Amos, Wenyong Tan, Virginia Marin Anaya, Yaru Pang, Gary Royle, and Esther Bär. Pre-treatment analysis of non-rigid variations can assist robust IMPT plan selection for head and neck patients. *Medical Physics*, pages 1–11, 9 2022.

[9] Wenyong Tan, Yanping Li, and Guang Han. Target volume and position variations during intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma. *OncoTargets and therapy*, 6:1719, 2013.

[10] Shreerang A. Bhide, Mark Davies, Kevin Burke, Helen A. McNair, Vibeke Hansen, Y. Barbachano, I. A. El-Hariry, Kate Newbold, Kevin J. Harrington, and Christopher M. Nutting. Weekly Volume and Dosimetric Changes During Chemoradiotherapy With Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Observational Study. *International Journal of Radiation Oncology Biology Physics*, pages 1360–1368, 2010.

[11] Macarena Cubillos-Mesías, Esther G.C. Troost, Fabian Lohaus, Linda Agolli, Maximilian Rehm, Christian Richter, and Kristin Stützer. Including anatomical variations in robust optimization for head and neck proton therapy can reduce the need of adaptation. *Radiotherapy and Oncology*, 131:127–134, 2019.

[12] Zhiyong Yang, Xiaodong Zhang, Xianliang Wang, X. Ronald Zhu, Brandon Gunn, Steven J. Frank, Yu Chang, Qin Li, Kunyu Yang, Gang Wu, Li Liao, Yupeng Li, Mei Chen, and Heng Li. Multiple-CT optimization: An adaptive optimization method to account for anatomical changes in intensity-modulated proton therapy for head and neck cancers. *Radiotherapy and Oncology*, 142:124–132, 2020.

[13] Panagiotis Tsiamas, Hassan Bagher-Ebadian, Farzan Siddiqui, Chang Liu Christian A Hvid, Joshua P Kim, Stephen L Brown, Benjamin Movsas, and Indrin J Chetty. Principal component

analysis modeling of Head-and-Neck anatomy using daily Cone Beam-CT images. *Medical physics*, 45(12):5366–5375, 2018.

[14] Z. Henry Yu, Rajat Kudchadker, Lei Dong, Yongbin Zhang, Laurence E. Court, Firas Mourtada, Adam Yock, Susan L. Tucker, and Jinzhong Yang. Learning anatomy changes from patient populations to create artificial CT images for voxel-level validation of deformable image registration. *Journal of Applied Clinical Medical Physics*, pages 246–258, 2016.

[15] Ying Zhang, Stacey McGowan Holloway, Megan Zoë Wilson, Jailan Alshaikhi, Wenyong Tan, Gary Royle, and Esther Bär. DIR-based models to predict weekly anatomical changes in head and neck cancer proton therapy. *Physics in Medicine & Biology*, 67(9), 2022.

[16] Ying Zhang, Jailan Alshaikhi, Richard A. Amos, Matthew Lowe, Wenyong Tan, Esther Bär, and Gary Royle. Improving workflow for adaptive proton therapy with predictive anatomical modelling: A proof of concept. *Radiotherapy and Oncology*, 173:93–131, 2022.

[17] Jake Lever, Martin Krzywinski, and Naomi Altman. Points of significance: Principal component analysis. *Nat Methods*, 14:641–643, 2017.

[18] Wenyong Tan, Jianzeng Ye, Ruilian Xu, Xianming Li, Wan He, Xiaohong Wang, Yanping Li, and Desheng Hu. The tumor shape changes of nasopharyngeal cancer during chemoradiotherapy: The estimated margin to cover the geometrical variation. *Quantitative Imaging in Medicine and Surgery*, 6(2):115, 2016.

[19] Marc Modat, Gerard R. Ridgway, Zeike A. Taylor, Manja Lehmann, Josephine Barnes, David J. Hawkes, Nick C. Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98(3):278–284, 2010.

[20] Brian C Hall. *Lie algebras and the exponential mapping.* Lie Groups, Lie Algebras, and Representations. Springer, New York, NY, 2003.

[21] Jan Ehrhardt, René Werner, Alexander Schmidt-Richberg, and Heinz Handels. Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE transactions on medical imaging*, 30(2):251–265, 2010.

[22] Simon Van Kranen, Angelo Mencarelli, Coen Rasch Suzanne Van Beek, Marcel Van Herk, and Jan Jakob Sonke. Adaptive radiotherapy with an average anatomy model: Evaluation and quantification of residual deformations in head and neck cancer patients. *Radiotherapy and Oncology*, 109(3):463–468, 2013.

[23] J Unkelbach and U Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Physics in Medicine & Biology*, 49(17):4005, 2004.

[24] M Van Herk, P Remeijer, C Rasch, and J V Lebesque. The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology Biology Physics*, pages 1121–1135, 2000.

[25] Alessandra Bolsi, Antony J Lomax, Eros Pedroni, Gudrun Goitein, and Eugen Hug. Experiences at the Paul Scherrer Institute with a remote patient positioning procedure for high-throughput proton radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 71(5):1581–1590, 2008.

[26] Martin J Murphy, Francisco J Salguero, Jeffrey V Siebers, David Staub, and Constantin Vaman. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. *Medical physics*, 39(2):573–580, 2012.

[27] Heng Li, Xiaodong Zhang, Peter Park, Wei Liu, Joe Chang, Zhong xing Liao, Steve Frank, Yupeng Li, Falk Poenisch, Radhe Mohan, Michael Gillin, and Ronald Zhu. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. *Radiotherapy and Oncology*, 114(3):367–372, 2015.

[28] Albin Fredriksson, Anders Forsgren, and Björn Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical Physics*, 38(3):1672–1684, 2011.

[29] Rom´an Bohoslavsky, Marnix G Witte, Tomas M Janssen, and Marcel van Herk. Probabilistic objective functions for margin-less IMRT planning. *Physics in medicine and biology*, 58(11):3563, 2013.

[30] Arthur Lalonde, Mislav Bobić, Brian Winey, Joost Verburg, Gregory C.Sharp, and Harald Paganetti. Anatomic changes in head and neck intensity-modulated proton therapy: comparison between robust optimization and online adaptation. *Radiotherapy and Oncology*, 159:39–47, 2021.

[31] Cheuk Fai Lui. Density heterogeneity tool for optimising beam angle selection in proton therapy, 2018.

Appendix A. An illustration of the deformable image registration

The contours of rCTs are propagated to the planning CT using DVFs. In an ideal situation (ideal DIR algorithm and no delineation bias between different images). The propagated contours should be exactly the same as the planning contour. An example of the propagated contour of the spinal cord of a rCT_3 is shown on the planning CT with its corresponding planning contour.

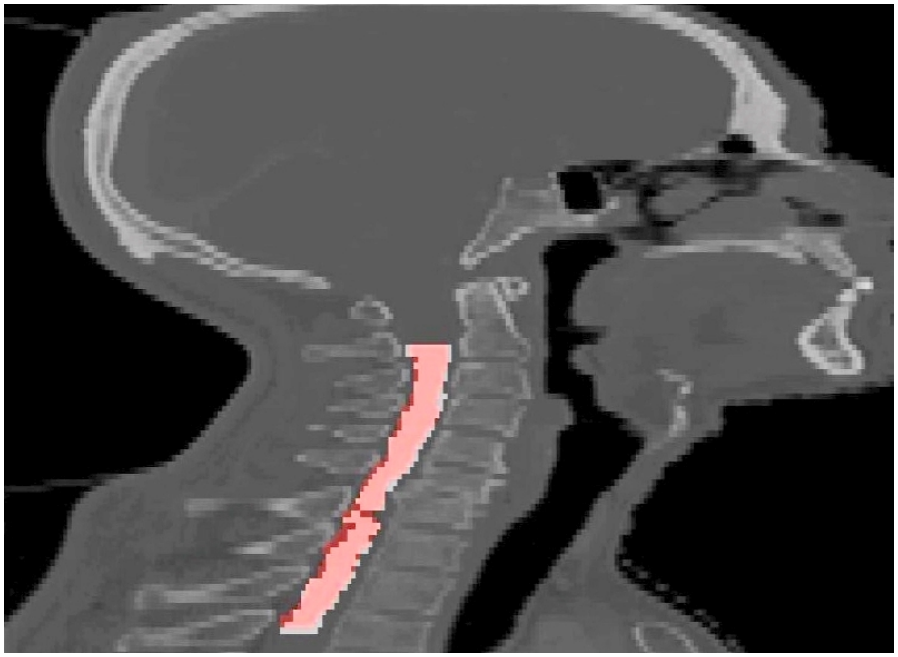


Figure A1: The propagated contour of the spinal cord of a rCT_3 is compared with its original planning contour on planning CT. The pink area represents the propagated contour of the spinal cord. The blue area represents the planning contour of the spinal cord.

Appendix B. Acquiring spot location

For each of the 5 test patients, the treatment plans were exported from the Varian treatment planning system. Information of the spot positions (X,Y) and energy/layer (Z) were extracted from the plan files. (X, Y) are recorded relative to the isocenter (the center of the target). (X, Y) with the beam angle can specify the beam's path. The beam energy (Z) determines the depth of the spot along the path by calculating the WEPL using equation B.1.

$$WEPL = \sum_{i,j,k \in S} RSP_{i,j,k} \cdot d_{i,j,k}, \tag{B.1}$$

where S is a set of voxels which contain the beam path, $d_{i,j,k}$ is the path length of the beam inside the voxel (i, j, k). $\mathrm{RSP}_{i,j,k}$ is the voxel-wise relative stopping power

estimated from CT numbers using a clinical calibration curve. We assumed parallel beams. A ray tracing algorithm [31] was used to calculate the length in each voxel passed by the beam.