Physics in Medicine & Biology

ACCEPTED MANUSCRIPT • OPEN ACCESS

Modelling systematic anatomical uncertainties of head and neck cancer patients during fractionated radiotherapy treatment

To cite this article before publication: Poppy Nikou *et al* 2024 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/ad611b

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is **© 2024 The Author(s). Published on behalf of Institute of Physics and Engineering in Medicine by IOP Publishing Ltd**.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 4.0 licence, this Accepted Manuscript is available for reuse under a CC BY 4.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by/4.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 144.82.114.249 on 19/07/2024 at 09:16

Modelling systematic anatomical uncertainties of head and neck cancer patients during fractionated radiotherapy treatment

Poppy Nikou[1,*], **Anna Thompson**[2], **Andrew Nisbet**[1], **Sarah Gulliford**[1,2], **Jamie McClelland**[1]

[1]University College London, WC1E 6AE, London
[2]University College London Hospital, NW1 2BU, London

E-mail: `poppy.nikou.20@ucl.ac.uk`

Abstract.

Objective: Head and neck cancer patients experience systematic as well as random day to day anatomical changes during fractionated radiotherapy treatment. Modelling the expected systematic anatomical changes could aid in creating treatment plans which are more robust against such changes.

Approach: Inter- patient correspondence aligned all patients to a model space. Intra- patient correspondence between each planning CT scan and on treatment cone beam CT scans was obtained using diffeomorphic deformable image registration. The stationary velocity fields were then used to develop B-Spline based patient specific (SM) and population average (AM) models. The models were evaluated geometrically and dosimetrically. A leave-one-out method was used to compare the training and testing accuracy of the models.

Main results: Both SMs and AMs were able to capture systematic changes. The average surface distance between the registration propagated contours and the contours generated by the SM was less than $2mm$, showing that the SM are able to capture the anatomical changes which a patient experiences during the course of radiotherapy. The testing accuracy was lower than the training accuracy of the SM, suggesting that the model overfits to the limited data available and therefore, also captures some of the random day to day changes. For most patients the AMs were a better estimate of the anatomical changes than assuming there were no changes, but the AMs could not capture the variability in the anatomical changes seen in all patients. No difference was seen in the training and testing accuracy of the AMs. These observations were highlighted in both the geometric and dosimetric evaluations and comparisons.

Significance: In this work, a SM and AM are presented which are able to capture the systematic anatomical changes of some head and neck cancer patients over the course of radiotherapy treatment. The AM is able to capture the overall trend of the population, but there is large patient variability which highlights the need for more complex, capable population models.

1. Introduction

Radiotherapy treatment aims to deliver the prescribed radiation dose to the target volume whilst adhering to the tolerance doses of the surrounding organs at risk (OARs) and healthy tissues. Proton therapy is the modality of choice for complex anatomical sites where a large number of OARs are proximal to the tumour site. A complex geometry requires a highly conformal dose distribution which can be created by exploiting the dose-depth profile of protons. Such challenging anatomical sites are common in head and neck (H&N) cancer patients and are therefore often treated with proton radiotherapy. The treatment plan is optimised from the anatomy seen in the planning CT (pCT) and the prescribed dose is delivered over a number of fractions.

Inter-fractional anatomical changes have been observed in H&N cancer patients. These changes can be systematic and progressive, such as weight loss, parotid gland shrinkage and tumour shrinkage (Barker 2004)(Robar 2007)(Ricchetti 2011)(Sanguineti et al. 2013)(Iliadou et al. 2021), or can be random day to day changes such as changes in neck flexion and set-up (Delishaj et al. 2018). Systematic progressive changes present similarly in H&N cancer patients. On the other hand, random day to day changes are patient specific. Anatomical changes can lead to uncertainties in the delivered dose distribution, especially with increasing conformality of the plan. This has the potential to reduce the efficacy of treatment by increasing toxicities to OARs and healthy tissues along with a reduced dose to the tumour (Hunter et al. 2013)(Noble et al. 2019).

Clinical practice in proton beam therapy mitigates for anatomical changes and set-up errors in H&N cancer patients prospectively by optimizing and evaluating the treatment plans against rigid shifts ($3mm$) and range uncertainty (3.5%) scenarios. However, more complex anatomical changes occur in this population of patients. Therefore, patients are monitored during treatment using cone beam CT (CBCT) imaging. Alongside the pCT, the CBCT imaging is used in a dosimetric review to assess whether a patient requires a new treatment plan. Creating a new treatment plan is time consuming and labour intensive. Ideally, a patient would be administered a treatment plan which is robust against all likely anatomical changes. To create anatomically robust treatment plans, a patient's anatomical changes must be anticipated in advance.

Deep learning frameworks have been proposed to estimate the anatomical changes of cancer patients during radiotherapy. These models directly predict either the deformation fields (Lee et al. 2020)(Pastor-Serrano et al. 2023)(Romaguera et al. 2020) or CT images (Lee et al. 2022). These deep learning models however, have a limited field of view and require information from the previous time point to generate a prediction for the next time point. For anatomically robust treatment plans all potential anatomical changes are needed in advance of treatment starting.

Statistical population models have also been proposed to estimate the anatomical changes of cancer patients during radiotherapy (Ehrhardt et al. 2010)(Szeto 2017)(Rigaud 2019)(Budiarto et al. 2011)(Argota-Perez 2022)(Zhang 2022)(Zhang et al. 2023). Most methods do not differentiate between systematic progressive and random day to day changes, so they cannot model the patterns in the systematic changes over treatment. Methods which do differentiate, and are able to model the systematic changes, require regular imaging at fixed time points during treatment e.g. the same day or week (Argota-Perez 2022)(Zhang 2022)(Zhang et al. 2023). Furthermore, this is often of the form of CT imaging which is not routinely acquired during treatment of H&N cancer patients (Zhang 2022)(Zhang et al. 2023). Patients do receive on-treatment CBCT scans, but the frequency of acquiring CBCT imaging varies as it is not often standardized. In this work we present a novel methodology for building patient specific and population average models, which are continuous in time, despite having irregular and sparse imaging data available. The models are built using CBCT images, and focus on capturing the systematic changes. The long term goal of this work is using the model predictions to assist in robust treatment planning for H&N cancer patients.

The models developed in this work are also built using a novel approach based on stationary velocity fields which does not require the use of complex and/or approximate methods to transport the velocity fields from the patient space to the model space. The motivation and formulation of this method is presented in Section 2, their application to H&N patients is presented in Section 3, and evaluated in Section 4.

2. Statistical Population Models

Modelling anatomical changes is a challenging research problem. Statistical population models of anatomical changes have been attempted for thorax (Ehrhardt et al. 2010)(Szeto 2017), cervical (Rigaud 2019) and H&N (Argota-Perez 2022)(Zhang 2022)(Zhang et al. 2023) cancer patients. These studies have a common methodology:

(i) Establish intra- patient correspondence to a patient specific reference frame,
(ii) Establish inter- patient correspondence to a population reference frame,
(iii) Map intra- patient correspondence to model space,
(iv) Build statistical population model,
(v) Map results of statistical population model to new patient space,
(vi) Predict anatomical changes of new patient.

Correspondence between images is found using deformable image registration. The registrations are often performed in a diffeomorphic framework since it guarantees the transformations provide a one-to-one mapping, are always invertable, continuous and preserve topology between images. A popular and efficient approach for ensuring the transformations are diffeomorphic is to use a stationary velocity field parameterisation (Ashburner 2007). Such registrations have been previously validated for various medical imaging applications (Veiga et al. 2015)(Stavropoulou et al. 2021)(Ehrhardt et al. 2010)(Hernandez et al. 2009).

Statistical population models reported in the literature have been built using both deformation fields and stationary velocity fields. In this work, we choose to work with stationary velocity fields to enable the models to be built using the log-Euclidean framework (Arsigny V. 2006). The mathematical motivation for this is discussed in Section 8.1. in the Appendix.

Building statistical population models requires all transformations to be in a common model space. This ensures that the reference images of all intra- patient correspondences are spatially aligned. The transformations should have the same effect in the model space as they do in the patient space. To achieve this the transformations must be transported, i.e. *resampled* and *re-orientated*. For more information on transporting transformations please see Section 8.2. in the Appendix.

A transformation $T_{t\to 0}$ in patient space which maps an image at time t to the reference frame image ($t = 0$) can be transported into model space using

$$\tilde{T}_{t\to 0} = T_{p\to M}(T_{t\to 0}(T_{M\to p})), \tag{1}$$

where $T_{p\to M}$ is the inverse of $T_{M\to p}$, and maps from patient space to model space. See Section 8.2. in the Appendix for a derivation of Equation 1.

If $T_{t\to 0}$ is represented as a stationary velocity field it must be converted to a deformation field, using the scaling and squaring approach (Ashburner 2007), before the composition in Equation 1. In this way, $\tilde{T}_{t\to 0}$ will likewise be represented as a deformation vector field (Ehrhardt et al. 2010). In this work, we aim to work with stationary velocity fields in the model space. It is possible to estimate a stationary velocity field that represents a given deformation field but this is complicated, computationally demanding, and subject to approximation error (Ehrhardt et al. 2010)(Hernandez et al. 2009). Other papers have presented methods to directly approximate the transported velocity fields but such methods are also complicated, computationally demanding, and/or potentially subject to large approximation errors (Zhang 2022)(Lorenzi et al. 2011)(Lorenzi & Pennec 2013).

In contrast, we propose a simple and elegant solution to obtain the stationary velocity fields in the model space. The proposed methodology is as follows:

(i) Ensure all images are rigidly aligned,

(ii) Establish inter- patient correspondence to a population reference frame, $T_{M\to p}$,

(iii) Resample images into model space using $T_{M\to p}$,

(iv) Establish intra- patient correspondence to a patient specific reference frame in model space $\tilde{T}_{t\to 0}$.

This way $\tilde{T}_{t\to 0}$ is directly calculated as a stationary velocity field, $\tilde{v}_{t\to 0}$. This method can potentially introduce small errors due to the resampling of the patient images into model space. However, we expect this error to be negligible for our application. This approach enables us to accurately and efficiently obtain the stationary velocity fields in the model space, and hence build our statistical model from the velocity fields using the log-Euclidean framework (Arsigny V. 2006).

To evaluate the models, the model generated results must first be transported back into patient space. In patient space, there is no need to represent the transformations as velocity fields so they can simply be converted to deformation fields and transported as such

$$T_{mod,t\to 0} = T_{M\to p}(\tilde{T}_{mod,t\to 0}(T_{p\to M})), \tag{2}$$

where $\tilde{T}_{mod,t\to 0}$ is the model generated transformation in model space and $T_{mod,t\to 0}$ is the model generated transformation in patient space. Explicitly therefore, the remaining steps become:

(v) Build statistical population model in model space, $\tilde{T}_{mod,t\to 0}$,

(vi) Transport results of statistical population model to new patient space using Equation 2, $T_{mod,t\to 0}$,

(vii) Use $T_{mod,t\to 0}$ to predict anatomical changes of new patient.

3. Methods

3.1. Patient Data

A cohort of 20 H&N cancer patients were used in this study, which were treated with photon radiotherapy (IMRT and VMAT). The cohort contained patients with a mixture of tumour types, lateralities and progression. The primary tumour volume was treated with 65-70 *Gy*, whilst surrounding neck nodes and proximal regions were prescribed 54-60 *Gy*. Each patient had a pCT and a number of CBCTs. The number of CBCTs varied per patient (n=4 - 10), and were acquired on different fractions for each patient. All patients in the cohort were also replanned for a variety of reasons, including weight loss, set-up difficulties and tumour progression. Therefore, all patients also had a rescan CT (rCT). The anatomical changes seen over the course of treatment of two of the patients were considered to be outliers from the distribution. Therefore, these patients were not used to build or evaluate the models.

3.2. Inter- Patient Correspondence

Inter- patient correspondence was obtained using an iterative groupwise registration algorithm. The algorithm simultaneously generated a population average image which defined the model space. The NiftyReg registration software was used for the registrations, which implements rigid, affine and deformable registrations‡. The rigid and affine registration used a block matching algorithm (Modat et al. 2014). The deformable registrations used a B-spline Free-Form Deformation transformation model defined by a regular grid of control points and a multi-resolution approach (Modat et al. 2010). In this work we used the stationary velocity field parameterisation for the deformable registrations that is available in NiftyReg.

---

‡ https://github.com/KCL-BMEIS/niftyreg

The groupwise registration algorithm was initiated using one patient's pCT as the reference. This patient was selected arbitrarily, since all pCT and CBCT images had similar spatial resolutions. The groupwise registration algorithm was defined as:

(i) Perform registrations between all patient images and the reference patient image,
(ii) Calculate the mean transformation and it's inverse,
(iii) Compose the inverse mean transformation with each registration result,
(iv) Resample each patient image using the composed transformations,
(v) Calculate the average image from the resampled images,
(vi) Set the average image as the reference image.

The above steps were repeated for one or more iterations using rigid, affine, and/or deformable registrations. In this work we first performed one iteration using rigid registrations, one iteration using affine registrations, and then four iterations of deformable registrations. Each iteration of deformable registration increased in the number of resolution levels and decreased the control point spacing. The total number of resolution levels and control point spacing for these registrations were 2, 3, 3, 3 and 20, 15, 10, 10 voxels used for iteration 1, 2, 3, 4 respectively.

Following the approach of Veiga et al. (2021), the deformable registrations were performed using multi-channel images. The images were made up of two channels, one contained the CT image in Hounsfield units and one contained an intensity map of the bony anatomy. The intensity map was obtained by thresholding the CT images between 150 - 3000 Hounsfield units. The two channels were separately weighted between the soft tissue:bony anatomy, as 1:2, 1:2, 1:2, 5:1 in the four registrations, respectively. These weights were used to focus the registrations initially on the bony anatomy and then on the soft tissue. Multi-channel registrations were needed since the anatomical differences seen between patients were large.

The remaining registration parameters were determined empirically to give good results for both the inter- and intra- patient registrations. A five level coarse-to-fine resolution approach was used. The local normalised cross correlation was chosen as a similarity measure for both channels, with a five voxel standard deviation Gaussian window. This similarity measure was chosen since it accounts for intensity inhomogeneities that can be present in CBCT images. To encourage a smooth transformation, linear elastic energy was selected as a regularisation term in the objective function with a weighting of 0.1. A conjugate gradient optimiser was used to minimise the objective function and a free form deformation grid spacing of 10 voxels was used.

The average patient image contained only the anatomical regions which were present in all the pCTs of all the patients. Any visibly large tumour regions were also masked from the pCTs when building the atlas. Masking of this region meant the voxels in the CTV were ignored in the registration between a patient's pCT scan and the atlas. The deformable registration will estimate the transformation in the region of the tumour by smoothly interpolating the deformation from the surrounding unmasked anatomy. This is required when building the average patient image, as different patients have tumours in different places, so they cannot be meaningfully aligned by the inter-patient registrations.

A coronal and sagittal slice of the average patient image are shown in Figure 1. The spatial resolution of the atlas was $1.3{\times}1.3{\times}2.5mm$.

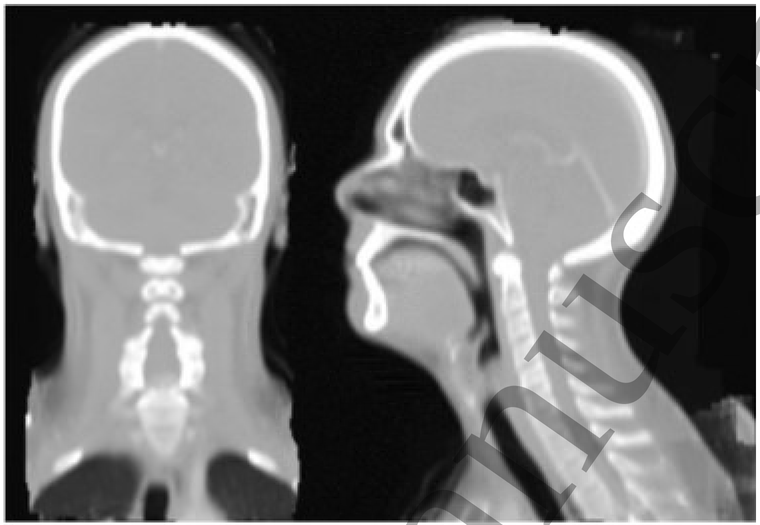

Figure 1: Coronal and sagittal slice of average patient atlas created using a groupwise registration algorithm.

The groupwise registration resulted in the transformations, $T_{M\to p}$, which map from model space to patient space for each patient p, and hence can be used to resample the patient images into model space. All 20 patients were used in building the atlas.

3.3. Intra- Patient Registrations

Intra- patient correspondence was obtained by registering the pCT and CBCTs in model space. Before the CBCT images could be resampled into model space they must first be approximately aligned with the pCT in patient space. Therefore, an initial manual translation was used to align the pCT and CBCT scans. A groupwise rigid (translations and rotations) registration aligned the CBCTs to each other. A rigid registration was then performed between the first CBCT in the series and the pCT. This transformation was applied to all the CBCTs to align them to the pCT. To mimic set-up in clinical practice, these steps were performed using a limited field of view which included only the first two vertebrae (C1 and C2) and the bottom of the skull of the patient. The registrations were performed since the actual shifts used to set up the patients were not available.

For each patient p, the transformation $T_{M\to p}$ was then used to resample the pCT and all CBCT images into the model space.

The resampled pCT was deformably registered to each of the resampled CBCTs in model space using NiftyReg with the same parameters as given in Section 3.2. Multi channel registrations were not required as the anatomical changes which a patient experienced were smaller than the anatomical differences seen in the inter-patient registrations. The intra- patient registrations just used single channel images containing the original image intensities.

The registrations were validated in preliminary experiments by comparing the warped pCT contours to the rCT contours and to contours which were delineated by a consultant oncologist directly on to the CBCTs. The registrations were deemed to have similar uncertainty as the manual contours, and so were deemed to be of sufficient quality.

The intra- patient registrations result in the transformation, $\tilde{T}_{reg,p,t\to0}$, which maps from the CBCT image at time point t to the pCT image for patient p, in model space, and is parameterised by the stationary velocity field $\tilde{v}_{reg,p,t\to0}$.

3.4. Patient Specific Models

Patient specific models (SMs) aimed to estimate $\tilde{v}_{reg,p,t\to0}$ over any time point t during treatment, $t = 1...T$. The variable $\tilde{v}_{reg,p,t\to0}$ is defined by a control point grid, with a vector at each control point. A cubic B-Spline was fit separately to each component of the vector field i.e. in x, y, z component of each vector over time. The cubic B-Spline function is the same as the one commonly used for image registration (Rueckert et al. 1999).

Four control points were placed equidistantly in time so that the B-Spline function was defined between the first and last CBCT scan for that patient. Four control points is the minimum number of control points needed to define a BSpline function. Therefore, this number of control points was chosen to reduce the overfitting of the model, since few CBCT images were available.

The B-Spline was fit using linear least squares. Once the model has been fit it can be used to calculate $\tilde{v}_{SM,p,t\to0}$ the SM velocity field, and $\tilde{T}_{SM,p,t\to0}$, the corresponding SM transformation, for patient p at any time t between the first and last CBCT scans.

3.5. Population Average Models

Population average models (AMs) aimed to capture the average anatomical changes which occured to a population of patients over the course of treatment.

The SMs described in Section 3.4 provided a compact way of describing the anatomical changes for an individual patient during their course of treatment. These models are already spatially aligned as they have all been constructed in the model space, but the control points are not temporally aligned as the last CBCT images were acquired on different days for different patients. Therefore, velocity fields were calculated from the SMs for every day from day 1 (corresponding to the first CBCT image) to day 28 (as the last CBCT for all patients was on or after day 28). Cubic B-Spline functions were then fit to these estimated velocity fields using least squares as before, but this time the four control points were placed equidistantly in time between days 1 and 28 for all patients.

The control points were now placed at the same points in time for all patients, the AM can be constructed by taking the average value of each control point over all patients.

This resulted in a smooth, continuous, and compact model in both space and time of the average anatomical changes which occured in the population. The AMs could then be used to estimate the average velocity fields, $\tilde{v}_{AM,t\rightarrow 0}$, and corresponding transformations $\tilde{T}_{AM,t\rightarrow 0}$, for any timepoint, t, between day 1 and day 28, $t = 1...28$.

3.6. Transporting Deformation Fields

To evaluate the performance of the models and their application of aiding in proton therapy treatment planning of H&N cancer patients, the transformations were transported into the patient space. Transformations $\tilde{T}_{reg,p,t\rightarrow 0}$, $\tilde{T}_{SM,p,t\rightarrow 0}$ and $\tilde{T}_{AM,p,t\rightarrow 0}$ were transported using Equation 2.

4. Evaluation

4.1. Geometric Model Evaluation

A geometric evaluation was performed to test the accuracy of the SM and AM, for modelling anatomical changes. So that the two models can be directly compared, only timepoints considered by all models were included in this evaluation.

The pCT and delineated structures from the pCT were then resampled using $T_{reg,p,t\rightarrow 0}$, $T_{SM,p,t\rightarrow 0}$ and $T_{AM,p,t\rightarrow 0}$. In this evaluation the registration results between the pCT and CBCT, $T_{reg,p,t\rightarrow 0}$, were considered as 'bronze standard' ground truth in the absence of manually delineated CBCTs. The evaluation was conducted at every CBCT time point within a 28 day time-frame, $t = 1..28$.

The structures which were used in this evaluation were the high dose and low dose clinical target volume (CTV), the body, spinal cord, brainstem and left and right parotid glands. These structures were used as they were available for all patients from data collected as part of their routine clinical care. The DicomRTtool was used to convert these structures from their point wise form in the radiotherapy structure file to a binary format (Anderson et al. 2021). These structures were used as they were available for all patients. For each evaluation the DICE, average surface distance and 95th percentile surface distance were calculated. The medpy library was used to calculate these metrics §.

The training and testing accuracy of the SM and AM were calculated by building the models using all the data at first and using a leave-one-out (LOO) method. The SM$_{LOO}$ left out $\tilde{v}_{SM,p,t\rightarrow 0}$ from one of the time points when building the model, excluding $t = 1$ and $t = T$, to avoid testing the extrapolation accuracy of the model. For the AM$_{LOO}$ testing, the models were built leaving out the patient they were evaluated against.

§ https://github.com/loli/medpy/

4.2. Geometric Model Comparison

In order to quantitatively compare the added information which the models would provide in radiotherapy treatment, the geometric model evaluation results were benchmarked against the magnitude of the anatomical changes which occured during treatment. Explicitly, the pCT structures ('Plan') as well as the propagated structures from the SM, SM_{LOO}, AM and AM_{LOO}, were compared to the 'bronze standard' registration propagated structures. The average surface distances for all the structures and comparisons were plotted for comparison.

4.3. Dosimetric Model Evaluation and Comparison

A dosimetric evaluation was performed to test the accuracy of the SM and AM for dosimetric calculations. An intensity modulated proton therapy (IMPT) treatment plan was created for three patients. These patients were chosen as ones which had (1) the least CBCT imaging, (2) weekly CBCT imaging, and (3) the most CBCT imaging. The aim was to capture the full range of imaging scenarios. This sample of patients provided a 17% test set size for the models.

The treatment plans were created using the Eclipse treatment planning software v16.1 (Varian Medical Systems, Palo Alto, USA). For all dose calculations the proton convolution superposition algorithm was used (Shen, 2016). Versions 16.1 were used for both algorithms. A radiobiological effectiveness of the protons of 1.1 was used (Paganetti 2014).

A prescription dose of 65 Gy and 54 Gy in 30 fractions were set to treat the high dose and low dose CTV, respectively. The treatment plan objectives of the CTVs and of the OARs are given in Table 1.

The treatment plans were created with four fields. The arrangement of the fields depended on the geometry of the target tumour. All fields were considered simultaneously in a multi-field optimization scheme. Range shifters of $3cm$ - $5cm$, depending on the tumour depth, were placed in the beam line to modulate the proton fields. An air gap of $15cm$ was set for each patient.

Optimization target volumes (OTVs) were defined using the high and low dose CTVs, with a $3mm$ expansion in all directions. Robustness scenarios were created with $\pm 3mm$ shifts and $\pm 3.5\%$ range uncertainties in combination and individually, totaling 14 uncertainty scenarios. The optimization was conducted robustly on the CTVs, and non-robustly on the OTVs and to the contoured OARs. Each plan was conducted using multifield optimization.

The evaluation of the treatment plan was performed on the nominal (anatomy as seen in the pCT) scenario and the 14 uncertainty scenarios. The clinical goals were calculated on the target volumes and serial OARs for all scenarios. For the parallel OARs, the clinical goals were evaluated only on the nominal case. The results of this analysis for the three proton therapy treatment plans are given in Table 1. The metrics considered on the target volumes were the dose to 99% and 95% of the volume, $D_{99\%}$ and $D_{95\%}$ respectively. The metric considered for the serial OARs was the maximum dose to $0.1cm^3$ of volume, D_{max}. The metric considered for the parallel OARs was the mean dose to the volume, D_{mean}.

Structure	Metric	Clinical Goal / Gy	Nominal / Gy			Worst case / Gy		
			1	2	3	1	2	3
High dose CTV	$D_{99\%}$	> 58.5	60.9	60.0	62.0	57.6	57.2	58.9
High dose CTV	$D_{95\%}$	> 61.8	63.0	62.6	63.2	61.0	61.2	62.3
Low dose CTV	$D_{99\%}$	> 48.6	51.3	51.4	52.3	48.2	48.1	50.2
Low dose CTV	$D_{95\%}$	> 51.3	52.5	52.4	52.9	41.6	50.6	52.3
Brainstem	D_{max}	< 54.0	38.9	42.3	51.0	45.2	47.4	58.8
Spinal Cord	D_{max}	< 48.0	42.7	39.4	39.4	46.7	50	45.3
Ispi Parotid Gland	D_{mean}	< 24.00	21.7	24.7	33.1	-	-	-
Contra Parotid Gland	D_{mean}	< 24.00	23.3	19.2	26.4	-	-	-

Table 1: Proton therapy treatment plan clinical goals and evaluation on nominal and worst case scenario, for three patients.

The pCT and structures were resampled for the three test patients using $T_{reg,p,t\to 0}$, $T_{SM,p,t\to 0}$ and $T_{AM,p,t\to 0}$, at all t where CBCT images were available, except $t = 1$ and $t = 28$ to avoid testing the extrapolation accuracy of the model. The generated CTs and structure sets were imported into the clinical treatment planning system and the IMPT treatment plans were recalculated on the registration and model generated CTs and structure sets. The clinical goals were then reported.

The clinical goals calculated on the images and structure sets which were created using $T_{reg,p,t\to 0}$ were considered as 'bronze standard'. The differences between these and the clinical goals calculated from the Plan, SM, and AM were calculated.

5. Results

5.1. Geometric Model Evaluation

Table 2 and 3 shows the results of the geometric evaluation for the SM and SM$_{LOO}$. For all metrics, the SM results were smaller than the SM$_{LOO}$ results. The average difference between the two model results in median and interquartile range in DICE, average surface distance and 95th percentile surface distance was 0.1 (0.29), 0.5 (0.93) mm and 1.3 (0.03) mm, respectively.

When looking at the surface distance results, the SM$_{LOO}$ was a higher average surface distance than the SM, but the median was $<2mm$ for all structures. This was on the order of the resolution of the CT scans ($\sim 1\times 1\times 2.5mm$). The 95th percentile surface distance showed good correspondence for most structures, with the median below $6.12mm$ for all structures.

A similar trend was shown with the DICE metric. The median DICE for most structures is above 0.8. The spinal cord and the low dose CTV for the SM_{LOO} were a little lower, at 0.69 and 0.79 respectively. This is due to the random day to day changes in neck flexion which are captured by the spinal cord and low dose CTV (since it contains nodal regions which are often in the shoulder region), which the model did not capture well.

Overall, the results of the geometric evaluation showed that the SM had a higher accuracy that the SM_{LOO}. This suggested that the fitting accuracy was higher than the testing accuracy, and therefore the model was overfitting to the data. This could have resulted in the models fitting to both systematic and some random day to day anatomical changes. To improve on this, a larger number of images at different time points are needed. In this dataset, the number of images were on the same order as the number of control points of the models.

Structure	Average Surface Distance (mm)		95^{th} percentile Surface Distance (mm)	
	SM	SM_{LOO}	SM	SM_{LOO}
High Dose CTV	0.58 (0.48 - 0.82)	1.11 (0.84 - 1.43)	2.50 (2.18 - 2.50)	3.08 (2.63 - 4.37)
Low Dose CTV	0.86 (0.66 - 1.02)	1.47 (1.22 - 1.85)	2.50 (2.50 - 2.82)	4.14 (3.32 - 5.16)
Body	0.91 (0.83 - 1.28)	1.90 (1.50 - 2.35)	3.08 (2.65 - 4.82)	6.12 (5.00 - 8.01)
Spinal Cord	0.81 (0.57 - 1.20)	1.43 (1.07 - 2.08)	2.18 (1.80 - 3.01)	3.56 (2.78 - 5.62)
Left Parotid	0.45 (0.37 - 0.56)	0.85 (0.62 - 1.02)	1.80 (1.27 - 2.18)	2.50 (2.18 - 2.76)
Right Parotid	0.36 (0.26 - 0.51)	0.64 (0.49 - 1.03)	1.38 (0.98 - 2.07)	2.50 (1.95 - 2.76)
Brainstem	0.25 (0.16 - 0.45)	0.50 (0.29 - 0.79)	1.17 (0.98 - 1.95)	1.95 (1.07 - 2.50)

Table 2: Average surface distance and 95^{th} percentile surface distance calculated between structures resampled with ground truth registration results and: the patient specific model results (SM) and the leave-one-out patient specific model results (SM_{LOO}). The median and interquartile range is shown over all CBCT time points within 28 days are shown.

Structure	DICE	
	SM	SM_{LOO}
High Dose CTV	0.96 (0.93 - 0.97)	0.93 (0.89 - 0.94)
Low Dose CTV	0.88 (0.84 - 0.90)	0.79 (0.71 - 0.82)
Body	0.98 (0.98 - 0.99)	0.97 (0.96 - 0.97)
Spinal Cord	0.83 (0.75 - 0.88)	0.69 (0.54 - 0.78)
Left Parotid	0.94 (0.91 - 0.95)	0.89 (0.86 - 0.91)
Right Parotid	0.94 (0.91 - 0.96)	0.89 (0.85 - 0.93)
Brainstem	0.96 (0.94 - 0.98)	0.93 (0.89 - 0.96)

Table 3: DICE score calculated between structures resampled with ground truth registration results and: the patient specific model results (SM) and the leave-one-out patient specific model results (SM_{LOO}). The median and interquartile range is shown over all CBCT time points within 28 days are shown.

Table 4 and 5 show the geometric evaluation results for the AM. Small differences were seen between the median AM and AM_{LOO} results. The average difference in median and interquartile range in DICE, average surface distance and 95^{th} percentile surface distance was 0 (0), 0.1 (0.1) mm and 0.2 (0.2) mm, respectively. This shows that the training and testing accuracy of the AM are very similar.

The median average surface distance of all structures is $<2.5mm$. This was on the order of the resolution of the CT scans ($\sim 1 \times 1 \times 2.5mm$). The median 95^{th} percentile surface distance was bellow $8.1mm$ for all structures. The DICE scores were above 0.8 for all structures except for the low dose CTV and spinal cord. Both structures were found in the neck and shoulder region, in which both systematic and random day to day changes in neck flexion can occur. These results suggested that the models do not capture all the possible anatomical changes which occured to patients in this region.

Structure	Average Surface Distance (mm)		95^{th} percentile Surface Distance (mm)	
	AM	AM_{LOO}	AM	AM_{LOO}
High Dose CTV	1.44 (1.13 - 2.04)	1.49 (1.19 - 2.12)	4.84 (3.41 - 6.70)	5.00 (3.54 - 6.98)
Low Dose CTV	2.09 (1.51 - 2.94)	2.19 (1.59 - 3.10)	6.06 (4.34 - 8.38)	6.37 (4.50 - 8.98)
Body	2.44 (1.93 - 3.90)	2.50 (2.07 - 4.09)	8.11 (6.24 - 12.91)	8.61 (6.65 - 13.42)
Spinal Cord	2.15 (1.42 - 2.85)	2.24 (1.46 - 3.07)	5.49 (3.85 - 7.83)	5.82 (3.88 - 8.42)
Left Parotid	1.26 (0.89 - 1.62)	1.31 (0.92 - 1.71)	3.09 (2.54 - 4.65)	3.17 (2.58 - 4.85)
Right Parotid	0.91 (0.65 - 1.18)	0.95 (0.67 - 1.23)	2.62 (2.50 - 3.13)	2.68 (2.50 - 3.17)
Brainstem	0.60 (0.41 - 0.94)	0.61 (0.42 - 0.98)	1.95 (1.27- 2.50)	1.95 (1.27 - 2.50)

Table 4: Average surface distance and 95^{th} percentile surface distance calculated between structures resampled with ground truth registration results and: the population average model results (AM) and the leave-one-out population average model results (AM_{LOO}). The median and interquartile range is shown over all CBCT time points within 28 days are shown.

Structure	DICE	
	AM	AM_{LOO}
High Dose CTV	0.90 (0.86 - 0.92)	0.89 (0.85 - 0.92)
Low Dose CTV	0.69 (0.57 - 0.78)	0.67 (0.55 - 0.77)
Body	0.96 (0.93 - 0.97)	0.95 (0.93 - 0.97)
Spinal Cord	0.54 (0.36 - 0.71)	0.53 (0.35 - 0.70)
Left Parotid	0.82 (0.77 - 0.87)	0.82 (0.76 - 0.86)
Right Parotid	0.87 (0.82 - 0.90)	0.87 (0.81 - 0.89)
Brainstem	0.91 (0.89 - 0.94)	0.91 (0.88 - 0.94)

Table 5: DICE score calculated between structures resampled with ground truth registration results and: the population average model results (AM) and the leave-one-out population average model results (AM_{LOO}). The median and interquartile range is shown over all CBCT time points within 28 days are shown.

In order to visually evaluate the SM, an animation is included. To view the animation please view this figure on the journal web page. Figure 2 shows the pCT of a patient which has been resampled over 35 days using the SM model results. A progressive weight loss was observed in this patient.

In the animation, the patients support system and bolus are also seen to deform. These regions were masked from the registrations. The CT imaging bed was masked since it is not the same as the treatment bed seen in the CBCT images. The bolus was masked as, often, after the patient receives a replan they have a new or no bolus in the CBCT images. In these regions, the deformation field is simply extrapolated and is not constrained.

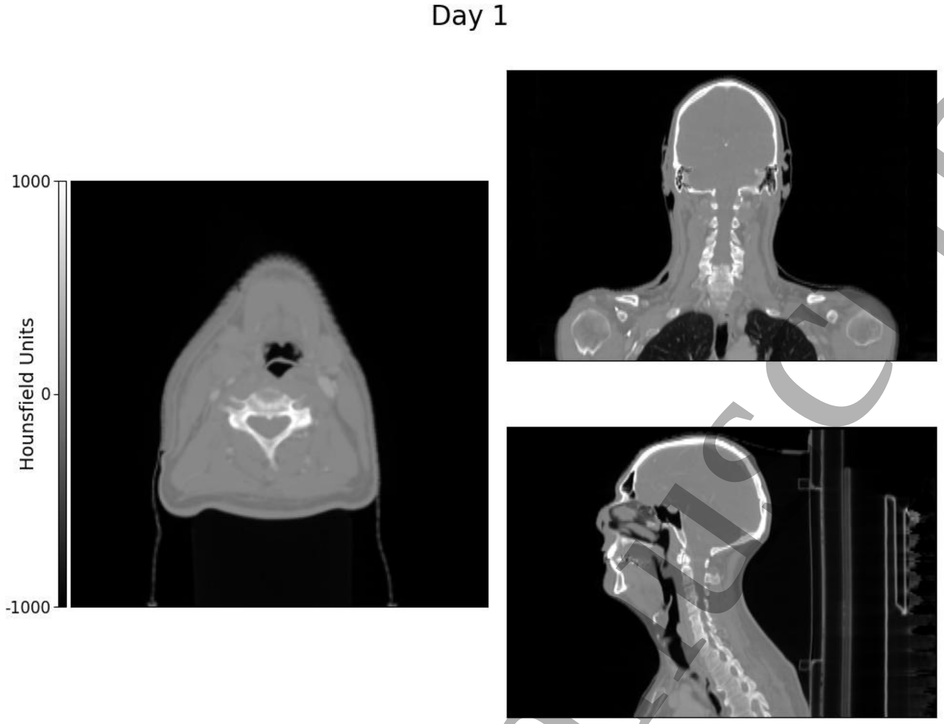

Figure 2: Thumbnail capture of animation of anatomical changes of a patient produced by resampling the patient's pCT using the patient specific model results. To see animation, please view this figure on the journal web page.

In order to visually evaluate the AM, two animations are included. To view the animations please view the figures on the journal web page. Figure 3 and 4 show the pCT of two patients which were resampled daily using the SM (left) and AM (centre) model results, and the difference (right) between the two. In the difference image, the green shows no difference between the two images, the yellow shaded regions highlight regions of difference. This evaluation assumed that the SM results were able to capture the anatomical changes of the patient well, and were therefore provided as a reference.

In Figure 3 a good correspondence was seen between the two models. However, Figure 4 showed that the AM did not capture the anatomical changes which occurred for this particular patient. In the animation, the displayed patient lost weight over the course of treatment which lead to changes in the set-up of the patient. This is not captured by the AM. This suggested that for some patients the anatomical changes were more varied.

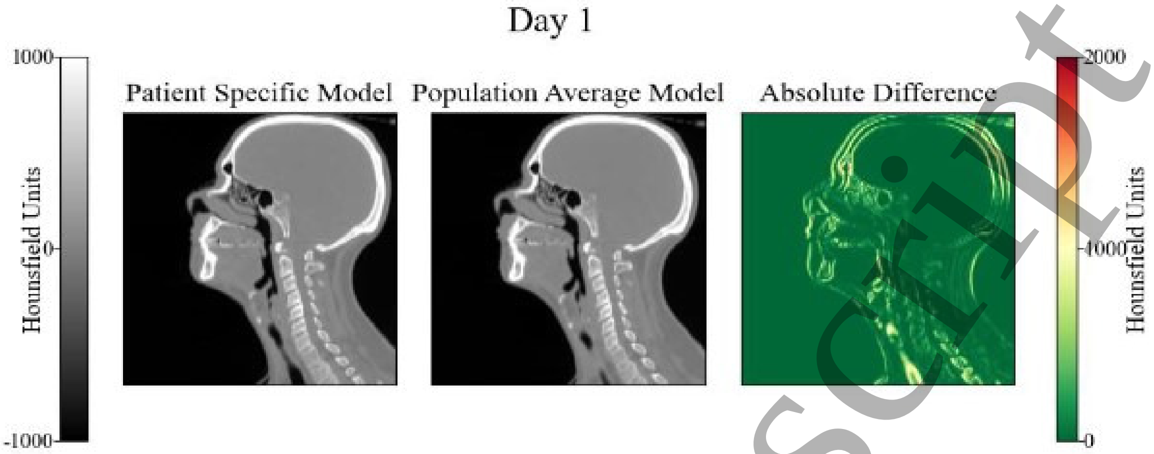

Figure 3: Thumbnail capture of animation of anatomical changes of a patient produced by resampling the patient's pCT using the patient specific model results (left), population average model results (centre) and the difference between the two model results. The green regions shows no difference between the two images, the yellow shaded regions highlight regions of difference. To see animation, please view this figure on the journal web page.

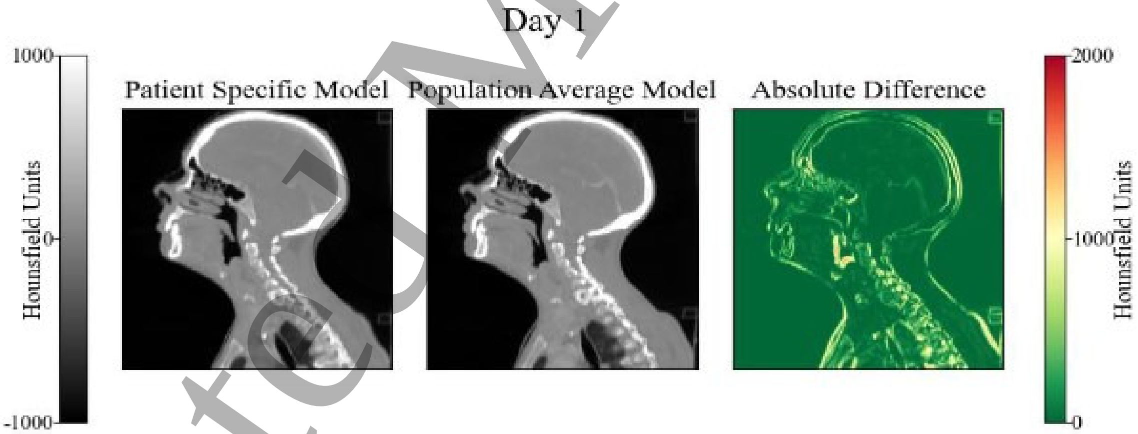

Figure 4: Thumbnail capture of animation of anatomical changes of a patient produced by resampling the patient's pCT using the patient specific model results (left), population average model results (centre) and the difference between the two model results. The green regions shows no difference between the two images, the yellow shaded regions highlight regions of difference. To see animation, please view this figure on the journal web page.

5.2. Geometric Model Comparison

Figures 5 shows the geometric model comparisons for the Plan, AM, AM_{LOO}, SM and SM_{LOO}. All structures were compared to the 'bronze standard' ground truth registration structures.

From the LOO models, the SM_{LOO} had the biggest improvement over the Plan across all metrics and structures. The median average surface distance and outliers were the smallest. The average difference between the Plan and SM_{LOO} models resulted in median DICE, average surface distance and 95^{th} percentile surface distance was 0, $0.2mm$, and $0.7mm$, respectively. This suggested that despite the model overfitting, it was still able to pick out systematic anatomical changes which occurred during treatment.

The AM_{LOO} outperformed or had similar performance to the Plan for all metrics and structures. The average difference between the Plan and AM_{LOO} models resulted in median DICE, average surface distance and 95^{th} percentile surface distance was 0.1, $0.7mm$, and $2.1mm$, respectively. This suggested that, overall, the AMs did provide some additional information on the anatomical changes despite their simplicity. The number of outliers for the AM_{LOO} were larger than for the Plan, indicating that for some patients it was better to assume no change, than the average change. This suggested that for the majority of patients the anatomical changes were similar to the average changes, but for some patients the changes were more varied. This is an inherent limitation stemming from the model's complexity.

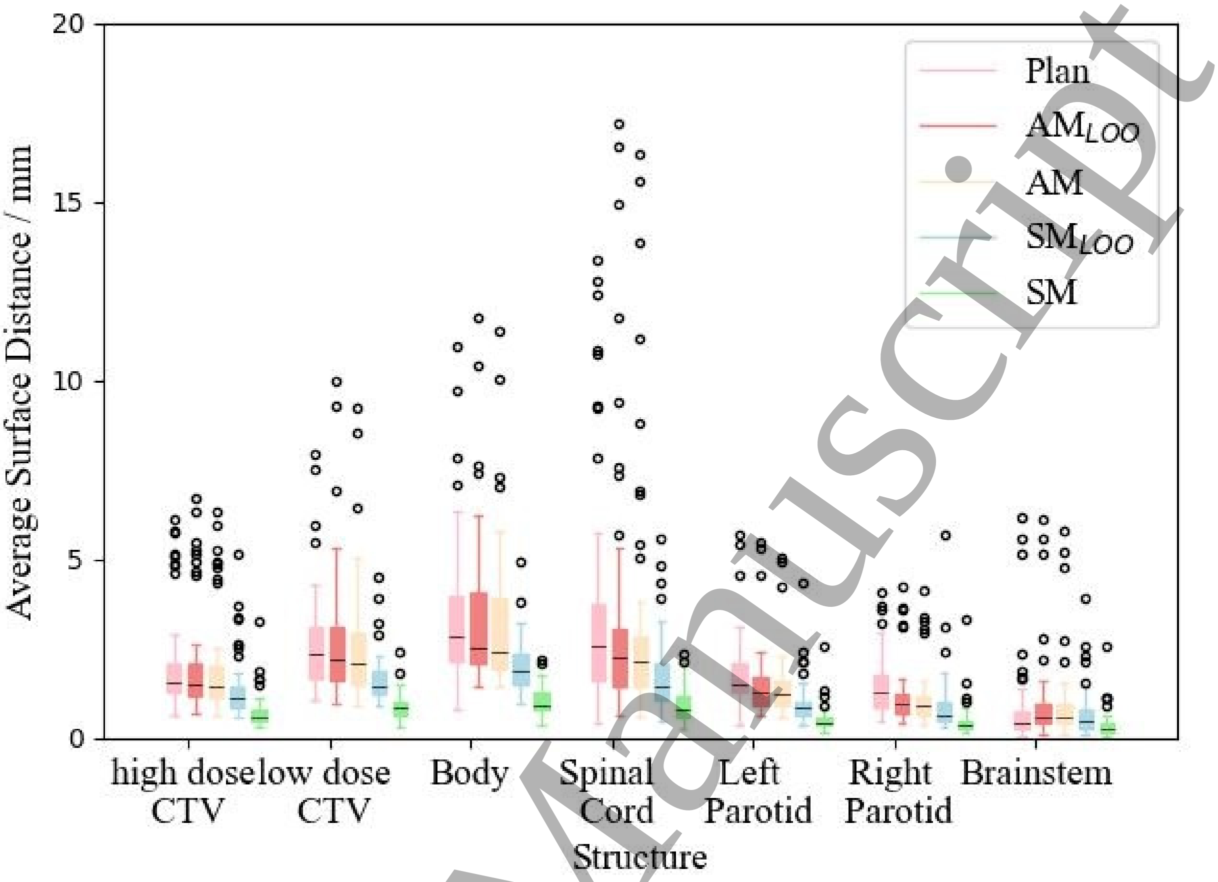

Figure 5: Average surface distance calculated between structures resampled with ground truth registration results and: the planning CT structures simulating no anatomical change (Plan), the leave-one-out population average model results (AM_{LOO}), the population average model results (AM), the leave-one-out patient specific model results (SM_{LOO}) and the patient specific model results (SM). The median, range and interquartile range over all CBCT time points within 28 days are shown.

5.3. Dosimetric Model Evaluation and Comparison

Figures 6, 7 and 8 show the results of the dosimetric evaluation and comparison. These figures correspond respectively to the patients with the least CBCT imaging, weekly CBCT imaging and the most CBCT imaging. The axis have been scaled equally on all three figures for comparison. All dose were metrics calculated on the registration and model generated CTs and structure sets. These were generated by resampling the pCT and delineated structures from the pCT using $T_{reg,p,t\to0}$, $T_{SM,p,t\to0}$ and $T_{AM,p,t\to0}$. The registrations were considered as 'bronze standard' ground truths in this evaluation and comparison.

For the first patient shown in Figure 6, it was observed that the anatomical changes had a large dosimetric impact on the OARs. The median absolute dose difference for the contralateral parotid gland was 11.2 Gy. For the remaining OARs the median absolute dose difference was below 5.5 Gy. In contrast, the median absolute dose difference for the CTVs was below 2.2 Gy.

Figure 6 shows that the SM and SM_{LOO} outperformed or had similar performance to the Plan on all structures. The average improvement, overall structures, in median absolute dose difference offered by the SM and SM_{LOO} was 1.9 Gy and 0.6 Gy, respectively. The AM and AM_{LOO} outperformed or had similar performance to the Plan for all structures, except for the ipsilateral parotid gland. The AM and AM_{LOO} had similar performance to each other, and consistently outperformed by the SM and SM_{LOO}.

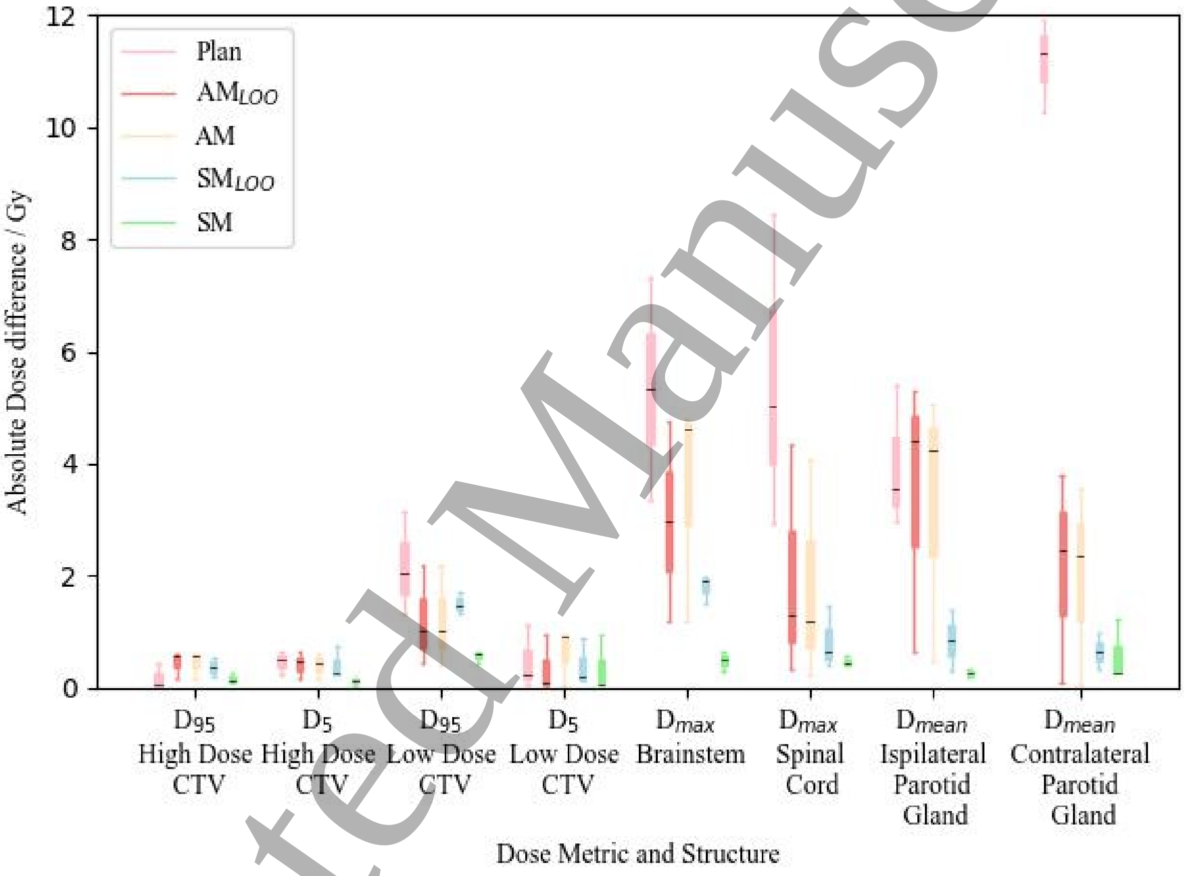

Figure 6: Absolute dose difference within structures resampled with ground truth registration results and: the planning CT structures from the nominal scenario simulating no anatomical change (Plan), the leave-one-out population average model results (AM_{LOO}), the population average model results (AM), the leave-one-out patient specific model results (SM_{LOO}) and the patient specific model results (SM). The median, range and interquartile range over all CBCT time points within 28 days are shown. These are the results for one patient in the population, which had the least CBCT imaging.

The results of the second patient are shown in Figure 7. The anatomical changes had a smaller dosimetric impact to those observed in the first patient. The largest dosimetric impact was on the spinal cord and contralateral parotid gland, with a median absolute dose difference of 4.3 *Gy* and 4.0 *Gy*, respectively. The remaining OARs and CTVs had an average absolute dose difference below 2.5 *Gy* and 2.6 *Gy*, respectively.

Figure 7 shows that all the models outperformed or had similar performance to the Plan for all structures. The average improvement, overall structures, in median absolute dose difference offered by the SM and SM_{LOO} was 1.9 *Gy* and 0.6 *Gy*, respectively. The AM and AM_{LOO} had similar performance to each other, and are generally outperformed by the SM and SM_{LOO} . The AM and AM_{LOO} improved on the Plan with the largest improvements seen in the spinal cord with an absolute dose difference of 2.2 *Gy* between each model and the Plan.

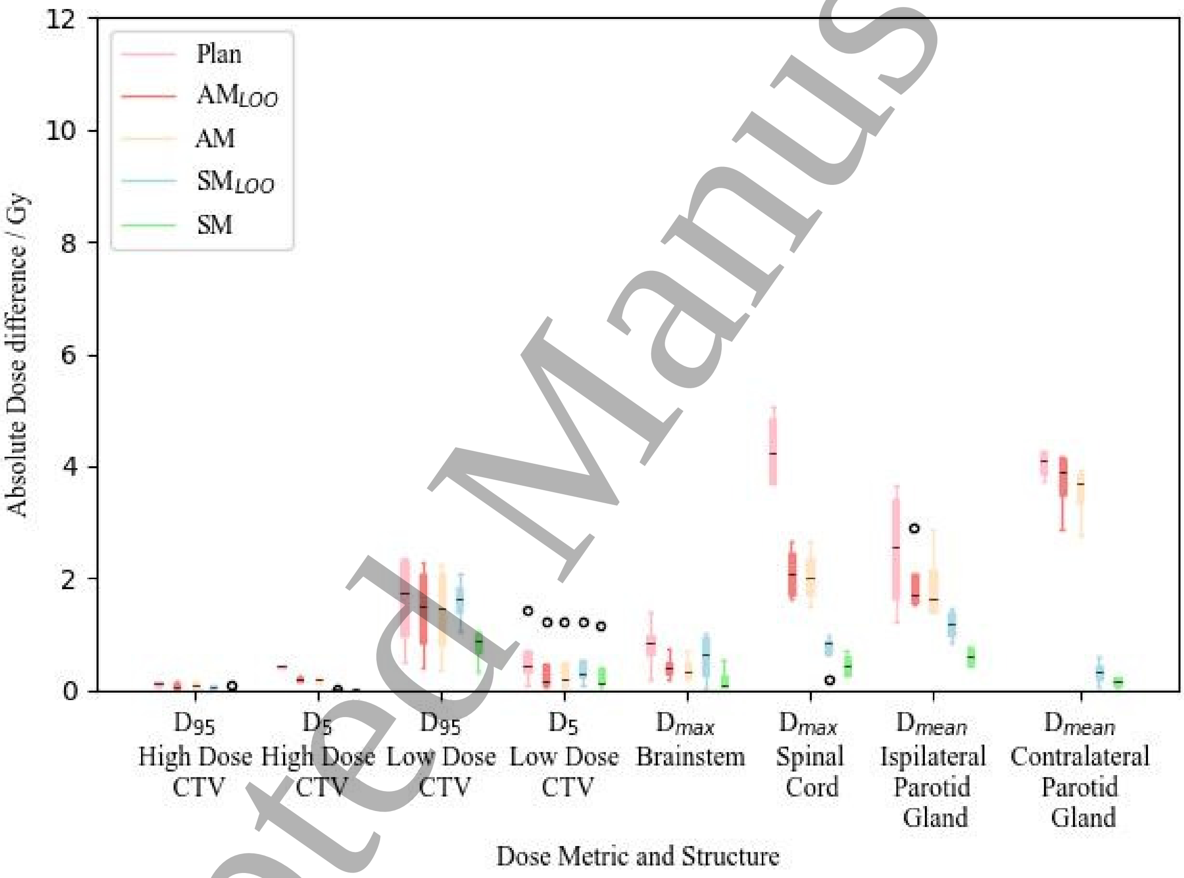

Figure 7: Absolute dose difference within structures resampled with ground truth registration results and: the planning CT structures from the nominal scenario simulating no anatomical change (Plan), the leave-one-out population average model results (AM_{LOO}), the population average model results (AM), the leave-one-out patient specific model results (SM_{LOO}) and the patient specific model results (SM). The median, range and interquartile range over all CBCT time points within 28 days are shown. These are the results for one patient in the population, which had weekly CBCT imaging.

The results for the third patient are shown in Figure 8. The anatomical changes for this patient caused the largest absolute dose difference in the brainstem and ipsilateral parotid gland. The median absolute dose difference for these structures was 3.4 Gy and 3.3 Gy, respectively. For the remaining OARs the median absolute dose difference was below 2.5 Gy. The anatomical changes also caused an absolute dose difference in the CTVs which was higher than the other two patients. In the D_{95} of the high dose CTV, a median absolute dose difference of 3.4 Gy was observed.

The model results for the third patient were more varied. The SM outperformed the Plan for all structures. The SM_{LOO} outperformed the Plan or had similar performance for all structures, except for the spinal cord. This showed that the models were overfitting for the spinal cord. This patient had the most CBCT images. Since CBCT images are acquired for set-up purposes, it is likely that this patient had difficulties with set-up due to their anatomical changes. The results suggested that there exists some random day to day anatomical changes in the structure which were captured by the SM but not the SM_{LOO}.

The AM and AM_{LOO} had a similar performance to each other, and outperformed or had similar performance to the Plan on all structures, except the brainstem and spinal cord. This suggested that the anatomical changes seen in the spinal cord were not captured by the AM and AM_{LOO}. These changes could include a mixture of systematic changes and random day to day changes, which were not all captured by the AM and AM_{LOO}.

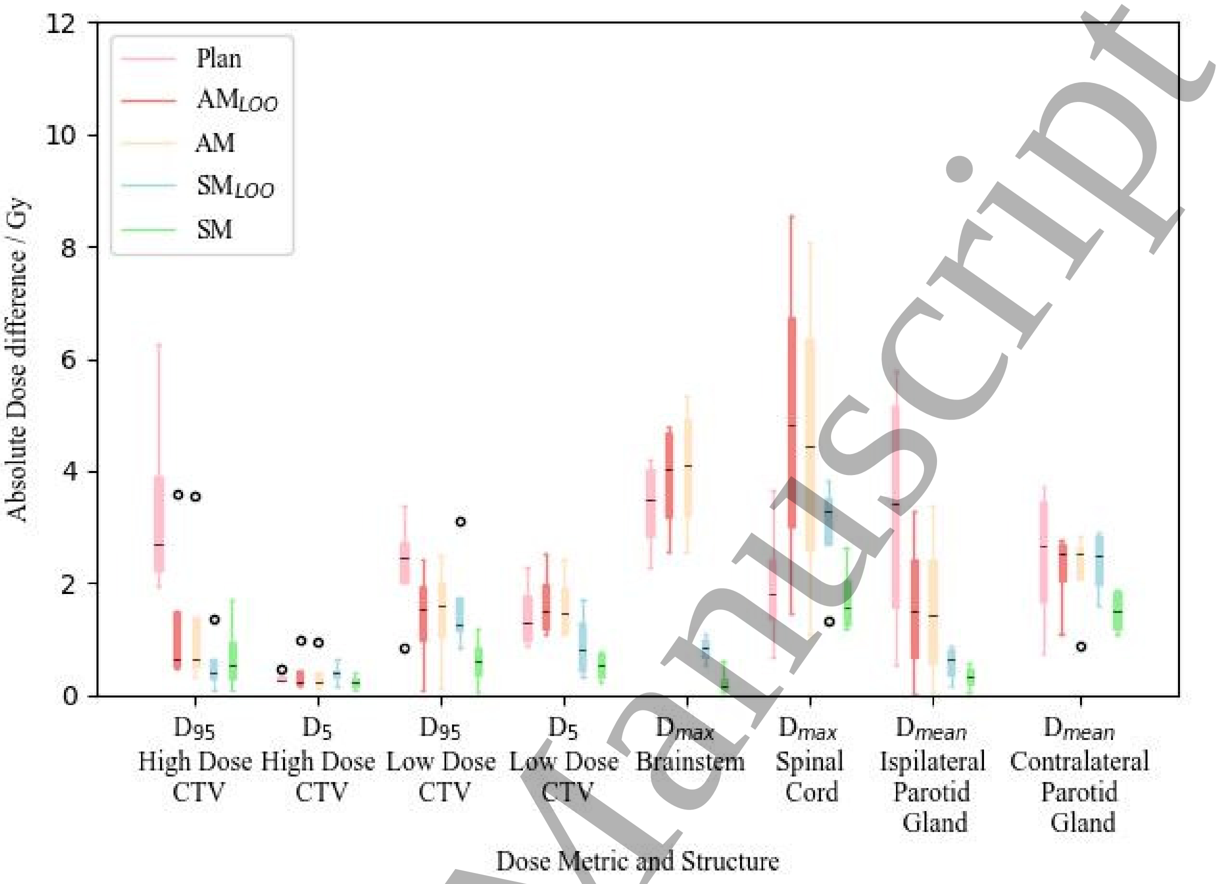

Figure 8: Absolute dose difference within structures resampled with ground truth registration results and: the planning CT structures from the nominal scenario simulating no anatomical change (Plan), the leave-one-out population average model results (AM$_{LOO}$), the population average model results (AM), the leave-one-out patient specific model results (SM$_{LOO}$) and the patient specific model results (SM). The median, range and interquartile range over all CBCT time points within 28 days are shown. These are the results for one patient in the population, which had the most CBCT imaging in the population.

When comparing the D$_{95}$ of the low dose CTV, an improvement or similar performance to the Plan was seen for all patients across all models. For the D$_{95}$ of the high dose CTV the models and Plan had a similar performance for two of the patients. For the final patient the models provided significant improvement, on average, of 2.1 Gy over the Plan. In general however, the differences between the Plan and model results were small across all patients. The average difference in D_{95} between the Plan and the models was 0.9 Gy for the low dose CTV and for the high dose CTV. This small variability suggested that the method of robust planning ensured good coverage to the CTVs throughout treatment. However, the models in this work were also built with a dataset of patients which were used in this study included a wide range of tumour types, lateralities and progression. These models therefore, do not explicitly model a specific CTV. This could also account for the small improvements provided by the models in the CTV dose.

For the OARs, absolute dose differences were higher. This was expected as all plans were only robustly optimized to the CTVs. The differences in the ipsilateral and contralateral parotid glands were due to the gland locations within the dose distribution. The ipsilateral parotid gland was often located within the high dose region, where the dose distribution was quite robust to the changes. The contralateral parotid gland however, was often located within a dose gradient between the high and low dose CTV, therefore changes in anatomy can lead to large differences in dose.

The SM and SM_{LOO} improved on the plan for all OAR structures and all patients. The AM and AM_{LOO} improved on the Plan for all patients for both parotid glands, except the ipsilateral parotid gland of the first patient. A similar pattern was seen for the brainstem and spinal cord when comparing the three patients. The AM and AM_{LOO}, for these structures, improved on the Plan for two out of three of the patients. This suggested that the spinal cord does undergo some systematic changes which were seen in the first two patients, but more variability existed within the population.

Overall, the dosimetric evaluation and comparison highlighted the robustness of the treatment plans to deliver the dose to the high and low dose CTVs. Larger differences in dose metrics could accumulate in the OARs. The SM results suggested it was able to model the changes in the OARs. The AM showed improvement on assuming no change for some structures and patients but not all. This motivates the need for more complex models which could capture these more complex changes.

When comparing the results of the three patients, no particular trends were observed, suggesting that the number of CBCTs had little effect on this evaluation and comparison.

6. Discussion

In this work we present the methodology for developing a patient specific model (SM) and population average model (AM) to model the anatomical changes in head and neck (H&N) cancer patients over the course of radiotherapy treatment. Inter- patient registrations were used to resample patient imaging into an average model space. Intra- patient registrations, parameterised by stationary velocity fields, were used to track longitudinal anatomical changes. A B-Spline function was then fit to the stationary velocity fields, allowing for the stationary velocity fields of a particular patient to be approximated at all points in time. This defines the basis of the SM. The parameters of the B-Spline functions were then averaged, to calculate an AM. Both models were then evaluated and compared geometrically and dosimetrically.

The development of the AM relies heavily on the SM accurately estimating the anatomy in between time points. Any errors in the SM results could result in errors in the AM. The SMs were built using cubic B-Splines which require a minimum of four temporal control points. In Section 5 it was shown that the SMs were overfitting to the data, suggesting that the SMs could be fitting to both the systematic changes and the random day to day changes. In the future, this will be addressed by collecting data from patients with more CBCTs. Although more CBCT imaging will be included in the criteria for the new patient cohort, the frequency of the imaging may still be irregular. The number of temporal control points was selected, in this work, as the minimum since few CBCTs were available per patient. In future work, the dataset with more CBCT imaging will be used to optimize the number of temporal control points with the aim of extracting systematic changes without overfitting to the random day to day changes.

The models presented in this work were built using a small population of patients which were treated with photon radiotherapy. The evaluation and application of these models however, is focused on proton radiotherapy. It is expected that the systematic changes, such as weight loss and tumour shrinkage, should present similarly in the two patient cohorts since some of these observations are independent of treatment type (Cacicedo et al. 2013) (Wang et al. 2023). Further validation of the model is needed on patients which were treated with IMPT.

The dataset of patients which were used in this study included a wide range of tumour types and lateralities. The AM therefore, did not model one CTV. Sub-groups of patients could be created based on CTV characteristics, since they are likely to also impact the progressive anatomical changes which are presented. In developing population models, there exists a tradeoff between building separate models for sub-groups of patients and ensuring that each model contains enough patient data. In this work all patients were included in a singular AM. Despite this, the results show that the AM is able to predict the anatomical changes which occurred during treatment better than assuming no anatomical change. This is especially highlighted in the OARs.

The AM in this work is built on a relatively small dataset of 18 patients. The small dataset enabled us to model the average change with the AM, but meant that the variability within the population could not be modelled. It also limited the time over which these models could be created (28 days). In clinical practice most H&N cancer patients are treated over 6-7 weeks (30-35 days). In this cohort, for most patients, CBCT imaging was not available after day 28. As seen in the results, the AM provides a reasonable approximation of the anatomical changes seen in most individual patients. However, there was also considerable variation in the changes observed suggesting the AM did not provide a good approximation to the anatomical changes for all patients. This is due to simplicity of the model. To tackle these issues a model which incorporates variability, such as principal component analysis could be used (Zhang et al. 2023). To build such models, a larger dataset of patients is also needed to capture the variability within the population. Both of these tasks are encompassed in the future work of this study.

The long-term goal of this work is to use the model predictions to assist in robust treatment planning for H&N cancer patients. Currently, in creating a robust treatment plan in proton beam therapy uncertainty scenarios are considered based on rigid shifts and proton range uncertainty scenarios. These scenarios are often used to optimize and evaluate the plan on. The AM could be used to generate a scenario of the average anatomical change, on which the treatment plan could also be optimized and/or evaluated. This clinical methodology could be extended to use multiple likely scenarios of anatomical changes, generated by a more capable statistical population model. Further research is needed to determine the optimal number of scenarios for both optimization and evaluation for this particular application.

The application of the models presented in this work is focused on robust treatment planning in proton beam therapy. However, these models could also be extended and incorporated into adaptive radiotherapy workflows e.g. by updating the predictions using CBCTs acquired during treatment. In proton beam therapy, the adaptive radiotherapy workflow is currently performed offline and requires a rescan CT (Oud et al. 2024). The model predictions could be used to prospectively predict the point of replanning and be used to create a new proton therapy treatment plan. Such workflows and their benefits have been shown in the literature but are not addressed in this work (Zhang 2022).

The dosimetric evaluation and comparison presented in this work involved three different patients, each with a different CTV. Although the treatment plans were created with as much similarity as possible, due to the differences in CTVs, the treatment plans result in dose gradients being placed in different regions of anatomy. Therefore, even if the patients experienced similar anatomical changes the dosimetric impact of these changes may be different. Therefore, the dosimetric evaluation should not be used to draw firm conclusions about how well the models predict the dosimetric impact of the anatomical changes, since small differences in the predictions may lead to large dosimetric differences from the ground truth. It is used to demonstrate the dosimetric impact of the anatomical changes and how the model outcomes can be used to perform dosimetric calculations on. Further studies using larger patient number are required to fully assess the ability of the models to estimate the dosimetric impact of the anatomical changes which occur.

7. Conclusion

Anatomical uncertainties in radiotherapy can lead to a difference in the dose which is delivered to the patient from the dose which was planned. Head and neck cancer patients have been seen to experience large anatomical changes over the course of treatment. The anatomical changes can be both systematic progressive or random day to day changes. Ideally, a treatment plan would be administered which is robust against all anatomical changes.

In this work we present a novel methodology of developing patient specific and population average models to extract the systematic anatomical changes seen in patients over the course of treatment. The models are based on few imaging data taken at irregular frequencies during treatment. A geometric and dosimetric evaluation and comparison showed that these models are able to capture the systematic anatomical changes but not random day to day changes for some patients. For most patients, the models are better than assuming no anatomical change. The large variability within the population however, also suggests the need for more complex models.

8. Appendix

8.1. Statistics on Stationary Velocity Fields

This Section describes a simple thought experiment which highlights the pitfalls of using deformation fields for building population models, and motivates the use of stationary velocity fields.

Consider a 90° rotation clockwise and anticlockwise of four points on a square, as shown in Figure 9. The vectors represent the change in position of each point following each transformation. A deformation field stores the final location of these points.

Averaging the location of the final points would result in a point at the centre of the square, for all examples. The resulting average vectors therefore would point from the edge of the square towards the centre. The average deformation field would cause the square to shrink to a single point. Figure 10 shows this schematically. This goes against intuition in which the average of two opposing 90° rotations would result in the identity transform, with all points being in their original location.

This simple example highlights the pitfalls of calculations on deformation fields. Stationary velocity fields do not have this property. Explicitly, if a transformation is represented as a velocity field, then the vector for a 90 degree clockwise rotation and 90 degree anticlockwise rotation will average to zero. Therefore, stationary velocity fields are used in this work.

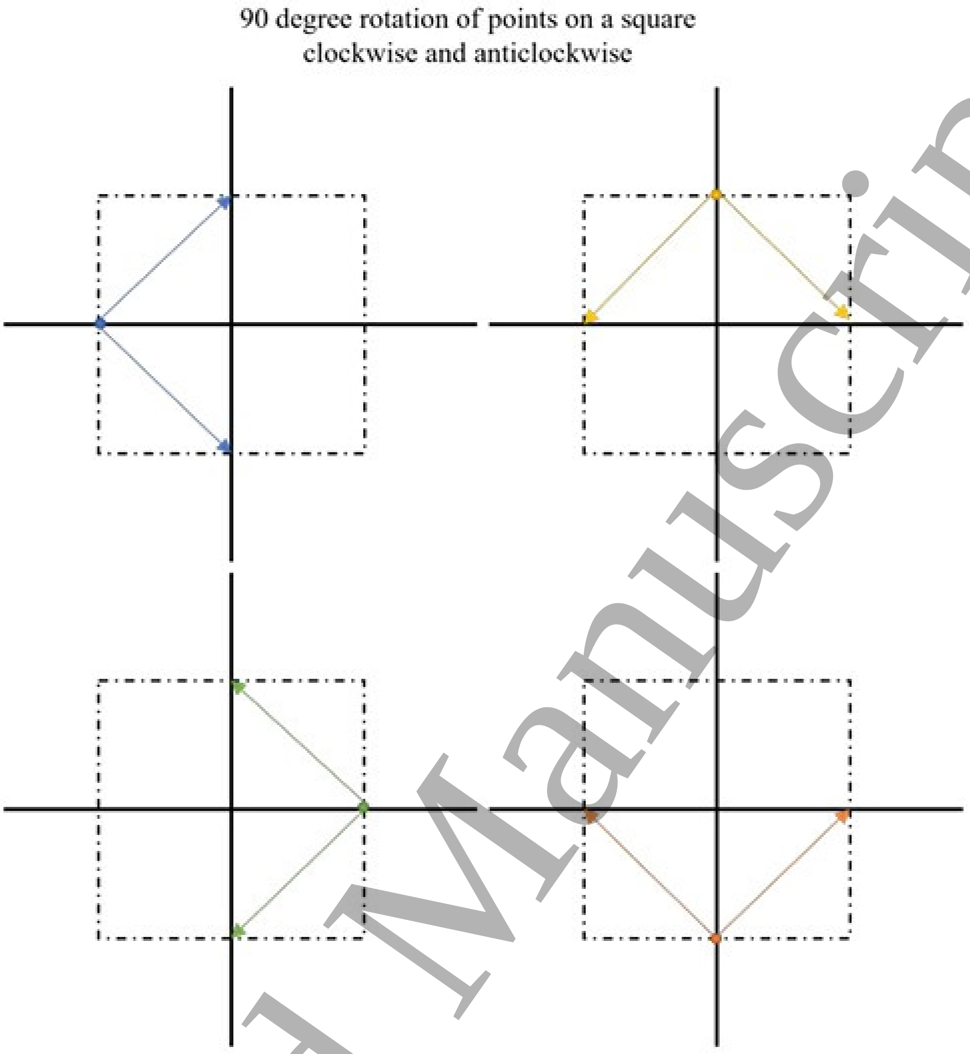

Figure 9: Schematic diagram of a 90° rotation of points on a square, both clockwise and anticlockwise. The tip of the arrows represent the final positions of the points.

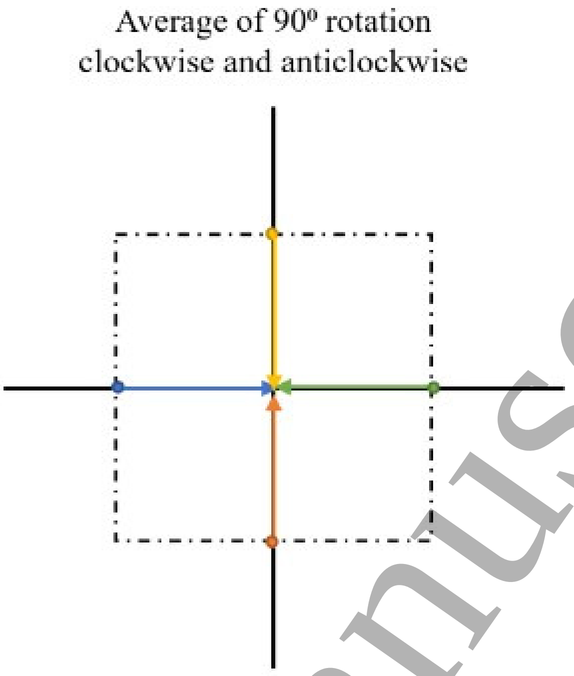

Figure 10: Schematic diagram showing the average of the two vectors which represented a clockwise and anticlockwise rotation of the square for a particular point. The vectors all point to the centre of the square.

8.2. Transporting Vector Fields

A transported vector field should result in a transformation which has the same effect in the model space as it does in the patient space. In this Section we motivate the need to both *resample* and *re-orientate* vectors when transporting them.

Figure 11 represents a patient space in the left image and a model space in the right image, related by a 180° rotation. This example is used to highlight the impact of incorrectly transporting transformations when rotations exist between the two spaces. The example chosen is an extreme example, as it is likely that the patient space will be more closely aligned to the model space. However, in the alignment between the spaces there may exist local rotations, which will be impacted by incorrect transporting of vector fields.

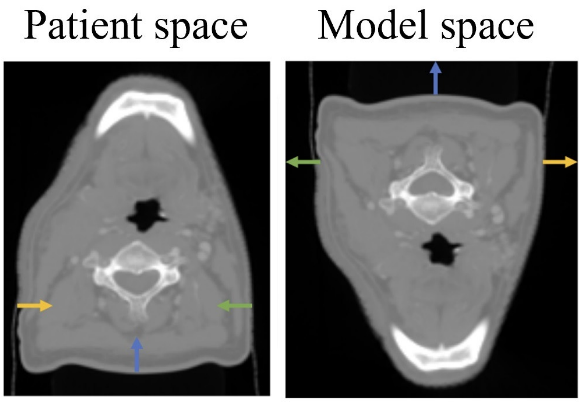

Figure 11: Schematic diagram showing the resampling of a transformation from a patient space to a model space. The two spaces are related by a 180° rotation.

Transporting of vector fields requires the vectors to be both *resampled* and *re-orientated*. If the transformations are mapped from one space to another by simply resampling the vector fields, this can lead to the transformations having a different effect in model space than in patient space. This is demonstrated in Figure 11.

Three deformation vectors are shown as coloured arrows in Figure 11. In patient space the vectors are pointing inwards. This represents a patient getting smaller e.g. due to weight loss. If the vectors are simply resampled but not re-orientated into model space, the vectors will now be pointing outwards. In model space, the deformation field now represents the patient getting larger as shown on the right of Figure 11. Although the vectors have been resampled into the correct spatial location, they have not been re-orientated.

Mathematically, the vectors shown in patient space are (1,0,0), (0,1,0), (-1,0,0). A clockwise rotation of θ applied to a point (x,y), results in a point (x', y') calculated using

$$x' = x\cos(\theta) - y\sin(\theta), y' = x\sin(\theta) + y\cos(\theta). \tag{3}$$

In the example in Figure 11, $\theta = 180°$. Therefore, $x' = x \times -1$ and $y' = y \times -1$. A correct re-orientation of these vectors would result in (-1,0,0), (0,-1,0), (1,0,0). However, in Figure 11 it is shown that by simply resampling the vectors into model space, we result in vectors (1,0,0), (0,1,0), (-1,0,0), which have not been re-orientated. So again, although the vectors have been resampled into the correct spatial location, they have not been re-orientated.

In order to correctly transport vectors from patient space to model space, the vectors must be both *resampled* and *re-orientated*. To define the correct mathematical formulation for transporting vector fields, we use Figure 12 as a reference.

Consider a transformation $T_{t\to 0}$ which maps the image at time t to the reference frame image (time 0) in patient space. This transformation could be used to resample the reference frame image to represent the image at time t. The transformation $T_{M\to p}$ maps from model space to the patient space, and hence can be used to resample the patient images from patient space into model space. The transported transformation $\tilde{T}_{t\to 0}$ is defined such that

$$T_{M\to p}(\tilde{T}_{t\to 0}) = T_{t\to 0}(T_{M\to p}), \tag{4}$$

and can therefore be calculated directly by

$$\tilde{T}_{t\to 0} = T_{p\to M}(T_{t\to 0}(T_{M\to p})), \tag{5}$$

where $T_{p\to M}$ is the inverse of $T_{M\to p}$. This can be seen schematically in Figure 12.

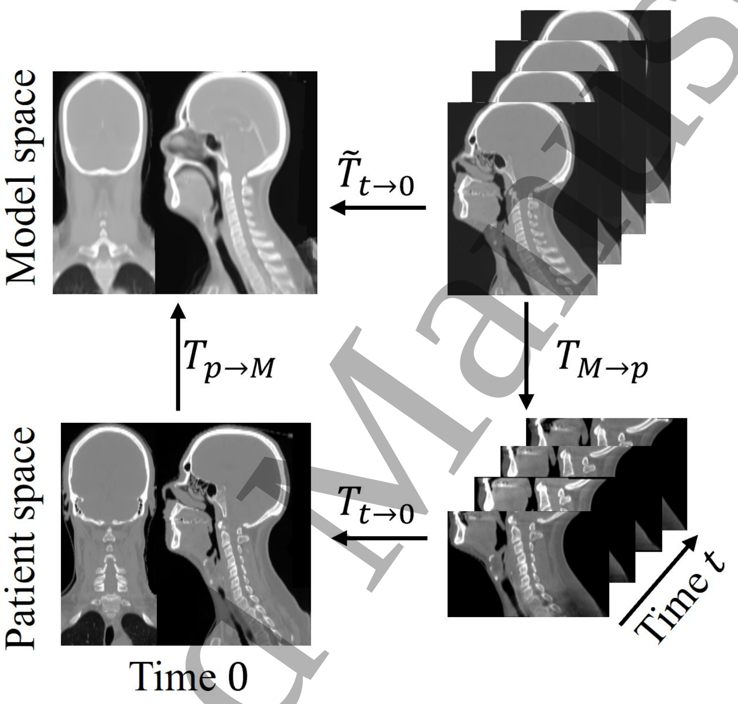

Figure 12: Schematic representation of inter- patient correspondence, $T_{M\to p}$ and it's inverse $T_{p\to M}$ as well as intra- patient correspondence in both a patient space, $T_{t\to 0}$, and model space, $\tilde{T}_{t\to 0}$.

To realise Equation 5 in Figure 12 consider calculating the location of a point ($\tilde{x}_0$, $\tilde{y}_0$, $\tilde{z}_0$) defined in model space at $t = 0$ from the location $(\tilde{x}_t, \tilde{y}_t, \tilde{z}_t)$ defined in model space at t. This could be achieved using:

$$(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0) = \tilde{T}_{t\to 0}(\tilde{x}_t, \tilde{y}_t, \tilde{z}_t). \tag{6}$$

Alternatively, the location of a point ($\tilde{x}_0$, $\tilde{y}_0$, $\tilde{z}_0$) can be calculated using the following steps:

$$(x_t, y_t, z_t) = T_{M\to p}(\tilde{x}_t, \tilde{y}_t, \tilde{z}_t), \tag{7}$$

$$(x_0, y_0, z_0) = T_{t\to 0}(x_t, y_t, z_t), \tag{8}$$

$$(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0) = T_{p\to M}(x_0, y_0, z_0). \tag{9}$$

Equations 7-9 can be simplified to

$$(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0) = T_{p\to M}(T_{t\to 0}(T_{M\to p}(x_0, y_0, z_0))). \tag{10}$$

Equating the two methods in Equation 6 and 10, we can realise

$$\tilde{T}_{t\to 0} = T_{p\to M}(T_{t\to 0}(T_{M\to p})) \tag{11}$$

as defined in Equation 5.

References

Anderson, B. M., Wahid, K. A. & Brock, K. K. (2021), 'Simple python module for conversions between dicom images and radiation therapy structures, masks, and prediction arrays', *Practical Radiation Oncology* **11**, 226–229.

Argota-Perez, R. (2022), 'Evaluating principal component analysis models for representing anatomical changes in head and neck radiotherapy', *Physics and Imaging in Radiation Oncology* **22**, 13–19.

Arsigny V., Commowick O., P. X. . A. N. (2006), A log-euclidean framework for statistics on diffeomorphisms, Springer Berlin Heidelberg, pp. 924–931.

Ashburner, J. (2007), 'A fast diffeomorphic image registration algorithm', *NeuroImage* **38**, 95–113.

Barker, J. (2004), 'Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated ct/linear accelerator system', *International journal of radiation oncology, biology, physics* **59**, 960–970.

Budiarto, E., Keijzer, M., Storchi, P. R., Hoogeman, M. S., Bondar, L., Mutanga, T. F., de Boer, H. C. J. & Heemink, A. W. (2011), 'A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases', *Physics in Medicine and Biology* **56**, 1045–1061.

Cacicedo, J., Casquero, F., Martinez-Indart, L., Hoyo, O. d., Iturriaga, A. G. d., Navarro, A. & Bilbao, P. (2013), 'A prospective analysis of factors that influence weight loss in patients undergoing radiotherapy', *Chinese Journal of Cancer* .

Delishaj, D., Ursino, S., Pasqualetti, F., Matteucci, F., Cristaudo, A., Soatti, C. P., Barcellini, A. & Paiar, F. (2018), 'Set-up errors in head and neck cancer treated with imrt technique assessed by cone-beam computed tomography: a feasible protocol', *Radiation Oncology Journal* **36**, 54–62.

Ehrhardt, J., Werner, R., Schmidt-Richberg, A. & Handels, H. (2010), 'Statistical modeling of 4d respiratory lung motion using diffeomorphic image registration.', *IEEE transactions on medical imaging* **30**, 251–265.

Hernandez, M., Bossa, M. & Olmos, S. (2009), 'Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows', *International Journal of Computer Vision 2009 85:3* **85**, 291–306.

Hunter, K. U., Fernandes, L. L., Vineberg, K. A., McShan, D., Antonuk, A. E., Cornwall, C., Feng, M., Schipper, M. J., Balter, J. M. & Eisbruch, A. (2013), 'Parotid glands dose–effect relationships based on their actually delivered doses: Implications for adaptive replanning in radiation therapy of head-and-neck cancer', *International Journal of Radiation Oncology Biology Physics* **87**, 676–682.

Iliadou, V., Economopoulos, T. L., Karaiskos, P., Kouloulias, V., Platoni, K. & Matsopoulos, G. K. (2021), 'Deformable image registration to assist clinical decision for radiotherapy treatment adaptation for head and neck cancer patients', *Biomedical Physics & Engineering Express* **7**, 055012.

Lee, D., Hu, Y.-c., Kuo, L., Alam, S., Yorke, E., Li, A., Rimner, A. & Zhang, P. (2022), 'Deep learning driven predictive treatment planning for adaptive radiotherapy of lung cancer', *Radiotherapy and Oncology* **169**, 57–63.

Lee, D., Zhang, P., Nadeem, S., Alam, S., Jiang, J., Caringi, A., Allgood, N., Aristophanous, M., Mechalakos, J. & Hu, Y.-C. (2020), 'Predictive dose accumulation for hn adaptive radiotherapy', *Physics in Medicine Biology* **65**, 235011.

Lorenzi, M., Ayache, N. & X., P. (2011), 'Schild's ladder for the parallel transport of deformations in time series of images', *Lecture Notes in Computer Science* .

Lorenzi, M. & Pennec, X. (2013), 'Efficient parallel transport of deformations in time series of images: From schild's to pole ladder', *Journal of Mathematical Imaging and Vision* **50**, 5–17.

Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S. & Ourselin, S. (2014), 'Global image registration using a symmetric block-matching approach', *Journal of Medical Imaging* **1**, 024003.

Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C. & Ourselin, S. (2010), 'Fast free-form deformation using graphics processing units', *Computer Methods and Programs in Biomedicine* **98**, 278–284.

Noble, D. J., Yeap, P.-L., Seah, S. Y., Harrison, K., Shelley, L. E., Romanchikova, M., Bates, A. M., Zheng, Y., Barnett, G. C., Benson, R. J., Jefferies, S. J., Thomas, S. J., Jena, R. & Burnet, N. G. (2019), 'Anatomical change during radiotherapy for head and neck cancer, and its effect on delivered dose to the spinal cord', *Radiotherapy and Oncology* **130**, 32–38.

Oud, M., Breedveld, S., Rojo-Santiago, J., Giżyńska, M. K., Kroesen, M., Habraken, S., Perkó, Z., Heijmen, B. J. M. & Hoogeman, M. S. (2024), 'A fast and robust constraint-based online re-optimization approach for automated online adaptive intensity modulated proton therapy in head and neck cancer', *Physics in Medicine and Biology* .

Paganetti, H. (2014), 'Relative biological effectiveness (rbe) values for proton beam therapy. variations as a function of biological endpoint, dose, and linear energy transfer', *Physics in Medicine and Biology* **59**, R419–R472.

Pastor-Serrano, O., Habraken, S., Hoogeman, M. S., Lathouwers, D., Schaart, D. R., Nomura, Y., Xing, L. & Perkó, Z. (2023), 'A probabilistic deep learning model of inter-fraction anatomical variations in radiotherapy', *Physics in medicine biology/Physics in medicine and biology* **68**, 085018–085018.

Ricchetti, F. (2011), 'Volumetric change of selected organs at risk during imrt for oropharyngeal cancer', *Int J Radiat Oncol Biol Phys* **80**, 161–168.

Rigaud, B. (2019), 'Statistical shape model to generate a planning library for cervical adaptive radiotherapy', *IEEE Transactions on Medical Imaging* **38**, 406–416.

Robar, J. (2007), 'Spatial and dosimetric variability of organs at risk in head-and-neck intensity-modulated radiotherapy', *International journal of radiation oncology, biology, physics* **68**, 1121–1130.

Romaguera, L. V., Plantefève, R., Romero, F. P., Hébert, F., Carrier, J.-F. & Kadoury, S. (2020), 'Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks', *Medical Image Analysis* **64**, 101754.

Rueckert, D., Sonoda, L., Hayes, C., Hill, D., M.O, L. & D.J, H. (1999), 'Nonrigid registration using free-form deformations: Applications to breast mr images', *NeuroImage* **18**, 712–721.

Sanguineti, G., Ricchetti, F., Thomas, O. & McNutt, T. (2013), 'Pattern and predictors of volumetric change of parotid glands during intensity modulated radiotherapy', *British Journal of Radiology* **86**.

Stavropoulou, A., Szmul, A., Chandy, E., Veiga, C., Landau, D. & McClelland, J. R. (2021), 'A multichannel feature-based approach for longitudinal lung ct registration in the presence of radiation induced lung damage', *Physics in Medicine Biology* **66**, 175020.

Szeto, Y. (2017), 'Geometrical uncertainties a population based statistical model for daily geometric variations in the thorax', *Radiotherapy and Oncology* **123**, 99–105.

Veiga, C., Lim, P., Anaya, V. M., Chandy, E., Ahmad, R., D'Souza, D., Gaze, M., Moinuddin, S. & Gains, J. (2021), 'Atlas construction and spatial normalisation to facilitate radiation-induced late effects research in childhood cancer', *Physics in Medicine Biology* **66**, 105005.

Veiga, C., Lourenço, A., Mouinuddin, S., van Herk, M., Modat, M., Ourselin, S., Royle, G. & McClelland, J. R. (2015), 'Toward adaptive radiotherapy for head and neck patients: Uncertainties in dose warping due to the choice of deformable registration algorithm', *Medical Physics* **42**, 760–769.

Wang, M., Zhu, Y., Ju, Y., Long, T., Jin, L., Zhang, J., Zhao, Y., Wang, H. & Wu, R. (2023), 'Changes in nutritional status of cancer patients undergoing proton radiation therapy based on real-world data', *Journal of Healthcare Engineering* **2023**.

Zhang, Y. (2022), 'Dir-based models to predict weekly anatomical changes in head and neck cancer proton therapy', *Physics in Medicine & Biology* **67**(9), 095001.

Zhang, Y., Alshaikhi, J., Tan, W., Royle, G. & Bär, E. (2023), 'A probability model for anatomical robust optimisation in head and neck cancer proton therapy', *Physics in Medicine Biology* **68**(1), 015014.