

Advanced Radiotherapy Planning Based on Probabilistic Concepts:

Understanding and modelling geometrical
uncertainties in head and neck radiotherapy to
evaluate plan robustness

A thesis submitted to the University of Manchester for the degree of Doctor
of Philosophy in the Faculty of Biology, Medicine, and Health

2022

Jennifer S Robbins
School of Medical Sciences
Division of Cancer Sciences

Contents

Abstract	8
Declaration	9
Copyright Statement	10
Acknowledgements	11
The Author	11
1 Introduction	13
1.1 Uncertainties.	15
1.1.1 Set-up uncertainties and organ motion.	17
1.1.2 Anatomical deformations.	18
1.1.2.1 Deformable image registration.	18
1.1.2.2 Principal component analysis.	20
1.2 Margin-based planning.	23
1.2.1 Limitations to margin-based planning.	24
1.3 Robust planning.	25
1.4 Probabilistic methods.	26
1.5 Evaluating with uncertainties.	27
1.6 Critical evaluation of literature.	29
1.7 Aims and structure of thesis.	37
2 Evaluating principal component analysis models for representing anatomical changes in head and neck radiotherapy	39
2.1 Introduction.	41
2.2 Materials and methods.	42
2.2.1 Evaluation strategy for PCA-based deformation models.	42
2.2.2 Evaluation toolbox.	43
2.2.3 PCA models examples.	44
2.2.4 Model Evaluation.	46
2.3 Results.	47
2.3.1 Model robustness (LOOCV)	47
2.3.2 Model generalisability.	48
2.4 Discussion.	50
2.5 Conclusion.	53
2.6 Acknowledgements.	53
Appendices	54
2.A Supplementary material.	54
2.A.1 Z-score.	54
2.A.2 Model sensitivity.	54
2.A.3 Patient data.	55

2.A.4 Cumulative variance ratio.	56
2.A.5 Increased number of input scans.	58
2.A.6 LOOCV vs generalisability.	60
2.A.7 Validation patient 2.	60
3 Probabilistic evaluation of plan quality for time-dependent anatomical deformations in head and neck cancer patients	63
3.1 Introduction.	64
3.2 Materials and methods.	65
3.2.1 Patient data.	66
3.2.2 PCA-based deformation models.	67
3.2.2.1 Pre-processing.	68
3.2.2.2 Average geometry using group-wise registration.	68
3.2.2.3 Generate DVFs from CBCTs.	68
3.2.2.4 PCA model creation.	69
3.2.2.5 Creating simulated deformations.	70
3.2.2.6 PCA model evaluation.	70
3.2.3 Treatment simulations.	71
3.2.3.1 Anatomical deformations.	71
3.2.3.2 Set-up uncertainties.	71
3.2.3.3 Anatomical deformations and set-up uncertainties.	72
3.2.3.4 Dose accumulation of actual treatment.	72
3.2.3.5 Volume change.	72
3.3 Results.	73
3.3.1 PCA-based deformation models.	73
3.3.1.1 Model evaluation.	73
3.3.2 Set-up uncertainties.	74
3.3.3 Treatment simulations.	74
3.4 Discussion.	77
3.5 Conclusion.	81
3.6 Acknowledgements.	81
Appendices	82
3.A Supplementary material.	82
3.A.1 Non-rigid registrations.	82
3.A.2 Different patient numbers for training.	82
3.A.3 Volume change of OARs.	83
3.A.4 Validation patient 2.	84
4 Can anatomical deformations in head and neck cancer be accounted for as set-up uncertainties?	85
4.1 Introduction.	87
4.2 Materials and methods.	88
4.2.1 Patient data.	88
4.2.2 Plan creation.	89

4.2.3 Treatment simulations.	90
4.3 Results.	91
4.4 Discussion.	95
4.5 Conclusion.	98
4.6 Acknowledgements.	98
Appendices	99
4.A Supplementary material.	99
4.A.1 Probabilistic planning algorithm.	99
5 The impact of target volume shape on plan robustness with different planning approaches	101
5.1 Introduction.	102
5.2 Materials and methods.	104
5.3 Results.	107
5.4 Discussion.	111
5.5 Conclusion.	116
5.6 Acknowledgements.	116
Appendices	117
5.A Supplementary material.	117
5.A.1 High, medium and low dose spread.	117
5.A.2 Isodose contours.	118
6 Discussion	125
6.1 Comparison to literature.	126
6.2 Limitations and future work.	130
6.3 Clinical relevance.	139
7 Conclusions	141
8 Publications and presentations	143
8.1 Publications.	143
8.1.1 Peer-reviewed.	143
8.1.2 Non peer-reviewed.	143
8.1.3 In preparation.	143
8.2 Presentations.	144

Word count ~ 30,000

List of Tables

2.1	Summary of patient datasets.	44
3.1	Patient characteristics.	66
3.2	Percentage of Z-score with values less than.	74
3.3	M , Σ and σ values from the 30 training patients.	74
4.1	Patient characteristics.	88
4.2	OAR planning objectives.	89
4.3	Uncertainty settings required to reach the desired robustness.	94
5.1	Shape, volume, surface area and sphericity of each target shape.	105
5.2	Planning cost functions and their weights.	106

List of Figures

1.1	Effect of systematic and random uncertainties.	16
1.2	Example of DIR with different bending energy parameters.	20
1.3	2-dimensional example of PCA.	21
2.1	Flow chart for generating PCA anatomical deformation models.	45
2.2	M_{res}^{90} for the LOOCV.	47
2.3	M_{res}^{90} for each week of treatment.	48
2.4	M_{res} maps for patient-specific models.	49
2.5	M_{res} maps for the population-based model.	49
2.A.1	M_{res}^{90} per component for simulated DVFs with added Gaussian noise.	55
2.A.2	Cumulative variance ratio for components in a PCA model from dataset 1.	57
2.A.3	Cumulative variance ratio for components in a PCA model from dataset 2.	57
2.A.4	Cumulative variance ratio for components in the population-based model.	58
2.A.5	M_{res}^{90} for models created from 1 or 2 weeks of data for patient 1, dataset 2.	58
2.A.6	M_{res}^{90} for models created from 1 or 2 weeks of data for patient 10, dataset 2.	59
2.A.7	M_{res} maps for models created from 1 or 2 weeks of input data for patient 1.	59
2.A.8	M_{res} maps for models created from 1 or 2 weeks of input data for patient 10.	59
2.A.9	Scatter plot of M_{res}^{90} for the LOOCV vs the model generalisability.	60
2.A.10	Average pCT and the pCT of validation patient 2.	61
3.1	Flow chart for generating time-varying PCA anatomical deformation models.	67
3.2	Cumulative variance ration of the different PCA models.	73
3.3	M_{res} maps and M_{res}^{90} values for the different PCA models.	74
3.4	Difference between planned and simulated dose.	76
3.A.1	M_{res}^{90} for different numbers of training patients.	83
3.A.2	Real and simulated volume change throughout the treatment course.	84
4.1	Simulated D98 values and the difference from the planned D98.	91
4.2	Simulated mean OAR doses and the difference from the planned doses.	92
4.3	Simulated maximum OAR doses and the difference from the planned doses.	93
4.4	Interpolated OAR doses for desired level of robustness.	94
5.1	Violin plots of CTV_{min}	107
5.2	Violin plots of D99.	108
5.3	Scatter plot of 90 th percentile of CTV_{min} vs minimum distance between the CTV and 95% isodose surface.	109
5.4	Conformity index of the different plans.	109
5.5	Scatter plot of sphericity vs a) CTV_{min} and b) conformity index.	110
5.6	Dose distributions for the sphere and cube plans.	111
5.A.1	Mean distance between CTV and isodose surfaces.	117
5.A.2	Isodose contours for the different plans.	118
6.1	Effect of registration uncertainties.	131

6.2	Example of intra-patient registrations.	132
6.3	Example of inter-patient registrations.	133
6.4	Difference between methods of mapping DVFs to the average geometry.	134

The University of Manchester

Jennifer S Robbins

Doctor of Philosophy

Advanced radiotherapy planning based on probabilistic concepts

June 27, 2022

Abstract

Geometrical uncertainties in radiotherapy such as differences in patient set-up positioning, and organ motion/deformations can cause the target to be underdosed or organs at risk to be overdosed. The aim of this thesis was to understand geometrical uncertainties in head and neck (H&N) radiotherapy; how to model them, how to evaluate their impact and how to make plans that are robust against them.

One way of modelling anatomical deformations is by using principal component analysis (PCA). In this thesis, a method for evaluating how well PCA models represent unseen deformations within a patient or population was first developed. This evaluation scheme was demonstrated in H&N cancer patients for both patient specific and population-based models. In the studied cohorts, the largest residual errors were found around the oropharynx.

Next, we developed a population-based time-dependent model for anatomical deformations in H&N using data from 30 H&N patients. This involved creating a PCA model for the systematic components of the deformations and weekly models for the random components. These models were then used to simulate many treatments and the effect of deformations on the delivered dose to the patient was evaluated alone and in combination with set-up uncertainties. The effect of anatomical deformations was found to be similar to or smaller than that of set-up uncertainties for all organs considered except the larynx and the primary clinical target volume (CTV).

Considering this finding, we then investigated whether plans could be created using tools to account for set-up uncertainties that were robust to anatomical deformations. We compared plans created using margin, robust and probabilistic approaches for different uncertainty settings. Our results show that margin-based plans were the most robust to anatomical deformations using only methods to account for set-up uncertainties.

Finally, we investigated whether the shape of the target affects the robustness of plans for set-up uncertainties. We compared margin, robust and probabilistic plans for different target shapes and found that as the CTV was less spherical, the plan robustness decreased. Margin-based plans were seen to be over-robust to set-up uncertainties and both robust and probabilistic planning approaches were seen to underdose voxels on the 'corners' of more complex CTV shapes.

Collectively, these results show that while in many cases plans can be created that are robust to geometrical uncertainties, including deformations, robust and probabilistic planning approaches should be used with care to ensure adequate target coverage.

Declaration

A version of the work presented in Chapter 2 of this thesis was included in a thesis submitted to Aarhus University for the degree of Doctor of Philosophy by co-author Raul Argota-Perez. No other portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <https://www.library.manchester.ac.uk/about/regulations/>) and in The University’s Policy on Presentation of Theses.

Acknowledgements

Firstly, I would like to thank my supervisors Marcel van Herk, Eliana Vasquez Osorio and Andrew Green for their support and encouragement over the last four years. In particular, I would like to thank Eliana for dealing with my frustration and getting me through the 'Summer of Registrations' in one piece.

Thank you to everyone in the RRR department for providing a supportive working environment filled with advice, celebrations and snacks – even through the lockdowns.

I would like to say a big thank you to my family for supporting me through everything, not just in the past four years but also the 20+ years before that. In particular, I would like to thank my parents for always being there, and not complaining when I announce last minute that I am coming to stay to 'help me write better'.

I would also like to thank my friends, in particular Shetland Pony and the POD, for putting up with me and providing much needed distractions, entertainment and a listening ear for my rants. I greatly appreciate all the cloak swishing, online scribble sessions and general insanity that comes so naturally to us.

And of course, I would like to thank my wife, Bella, for everything; for coping with me on my bad days as well as the good, for making me laugh even when I really want to be grumpy, and most importantly, for cooking me food and doing the washing up on the many occasions I decided that was the only time I could possibly do any work.

Last, but not least, I would like to thank Koda Bear for providing all the snuggles and ensuring I had enough naps.

The Author

The author graduated from The University of Manchester in July 2018 with a First Class Honours degree (MPhys) in Physics, and has been working part-time as a Science Explainer at Jodrell Bank since July 2017.

Blank page

Chapter 1

Introduction

Cancer is one of the leading causes of premature death (deaths occurring between the ages of 30-70) in the world [1], with almost 10.0 million deaths and an estimated 19.3 million new cases in 2020 [2]. The number of new cases is predicted to almost double to 28.4 million in 2040 [2]. Radiotherapy is an important modality for treating cancer, with over half of all patients receiving it as part of their treatment [3], increasing to roughly 80% of patients with head and neck (H&N) cancer [4].

The intention of radiotherapy is to deliver high energy radiation to the cancerous cells in order to kill them by causing damage to their DNA. Generally, the higher the radiation dose delivered to the tumour, the higher the chance of killing all its cells. However, the radiation causes damage to the DNA of all cells it passes through and not just the cancerous cells. Damage to the surrounding normal tissue and organs at risk (OARs) can cause side effects for the patient, some of which can be quite severe. For example, possible side effects in H&N radiotherapy can include xerostomia (dry mouth) and dysphagia (difficulty swallowing), both of which can have a large impact on the quality of life for the patient [5]–[7]. This means that one of the main challenges of radiotherapy is getting the balance between depositing a large enough dose to the tumour to kill all the cancer cells while simultaneously limiting the damage to the healthy tissue and keeping the risk of side effects as low as possible. Because of this, radiotherapy is normally delivered over multiple sessions, called fractions, over the course of several days or weeks as cancer cells are less able to repair their DNA than healthy tissue. This provides a better therapeutic ratio, i.e., killing more cancer cells than healthy cells for the same dose.

The planning process typically begins with a computed tomography scan (CT) where the anatomy of the patient can be visualised. There, the tumour and other surrounding OARs are delineated. The visible tumour volume on the scan is extended to a clinical target volume (CTV) to account for surrounding microscopic tumour deposits not visible on the CT. The CTV is then considered the target for the high radiation dose. Depending on the site and stage of the tumour, a set of clinical goals are decided upon, including minimum

levels of dose to the target and maximum levels of dose to the different OARs and healthy tissue. These clinical goals are weighted to determine the priority of the different objectives and then input into a planning system which optimises the dose distribution to ensure it hits all the goals, or as many as possible if they cannot all be achieved simultaneously. This plan is then used to treat the patient.

However, there are numerous uncertainties associated with the radiotherapy planning and treatment process, including geometrical uncertainties. Geometrical uncertainties arise from things such as differences in patient set-up positioning or movement/deformation of organs within the patient. For example, H&N patients often experience weight loss throughout the treatment [8], the parotids can shrink [9]–[18] and the neck and shoulders can change position and shape on a day-to-day basis [19]–[21]. These changes in the patient's anatomy cause the actual dose delivered to the patient to be different from the planned dose [22], [23]. This can make the treatment less effective by underdosing the target and can also overdose OARs, increasing the risk of side-effects. Currently there are limited options to predict such errors and their impact for an individual patient prior to treatment. These uncertainties can be categorised as inter-fraction, meaning the changes happen between one fraction and the next, and intra-fraction where the changes happen within a given fraction. This thesis focuses on inter-fraction uncertainties.

In order to limit the effect of geometrical changes, plans need to be created that are robust to the uncertainties present in the radiotherapy process. In practical terms, the uncertainty in question needs to first be measured, and then modelled such that the effect of the uncertainty on the delivered dose can be investigated. Ideally, treatment plans should take the uncertainty in question into account. Conventionally, uncertainties are accounted for by using margins, where the CTV is expanded to create a larger planning target volume (PTV) which is then treated with a high dose. However, other planning approaches have been developed where the planning system enforces robustness algorithmically rather than by using margins. In robust planning, the optimiser samples a number of scenarios with different errors and the worst-case out of the sampled scenarios is optimised. This method is commonly used in proton therapy [24]–[31] but has also been considered for photon therapy [32]–[35]. Another method, probabilistic planning, directly takes the probability of many sampled scenarios into account when optimising the plan. While

probabilistic planning is not yet used clinically, there have been a few studies investigating its effectiveness in different situations [36]–[41]. Many of these studies have shown that robust and probabilistic planning can improve OAR sparing when compared to margin-based plans for the same or improved level of target coverage. However, since multiple scenarios are sampled at each iteration of the optimisation, robust and probabilistic planning are more computationally expensive and take longer to run than conventional margin-based planning.

Most treatment planning systems have built-in methods for robust planning and some also have a research interface allowing for the implementation of new methods that have been used to provide prototypes of probabilistic planning [42].

1.1 Uncertainties

Some uncertainties arise from physical limitations such as the accuracy of the dose calculation algorithms or the quality of the scan used for planning. The quality of the scan will affect the ability of the clinicians to correctly delineate the target and OARs, called delineation uncertainties. Scan quality can also affect the estimate of tissue density which affects how much dose is absorbed and so adds uncertainty to the delivered dose. This latter effect is considered to be small in photons [43] but is a larger concern in proton therapy [39], [44], [45].

However, a large cause of uncertainty is through differences in the positioning of patient anatomy between planning and actually delivering the treatment. Planning is generally done on a single scan, the planning CT (pCT), days or sometimes weeks before the actual treatment starts. This pCT is just a ‘snapshot’ of the patient’s anatomy at a specific moment in time and will not represent exactly what the patient’s anatomy will look like ‘on average’ during treatment.

These uncertainties can be split into systematic uncertainties and random uncertainties (Figure 1.1). Random uncertainties are day-to-day variations and are different for each treatment fraction. For standard fractionations, they can be considered to cause an overall blurring of the dose distribution as seen by the CTV. Systematic errors are the same throughout the whole treatment course and typically arise from a difference in the pCT with the average position throughout treatment delivery. They generally manifest as an

overall shift or deformation of the delivered dose compared to the planned dose. The effect of systematic uncertainties tends to be larger than that of random uncertainties.

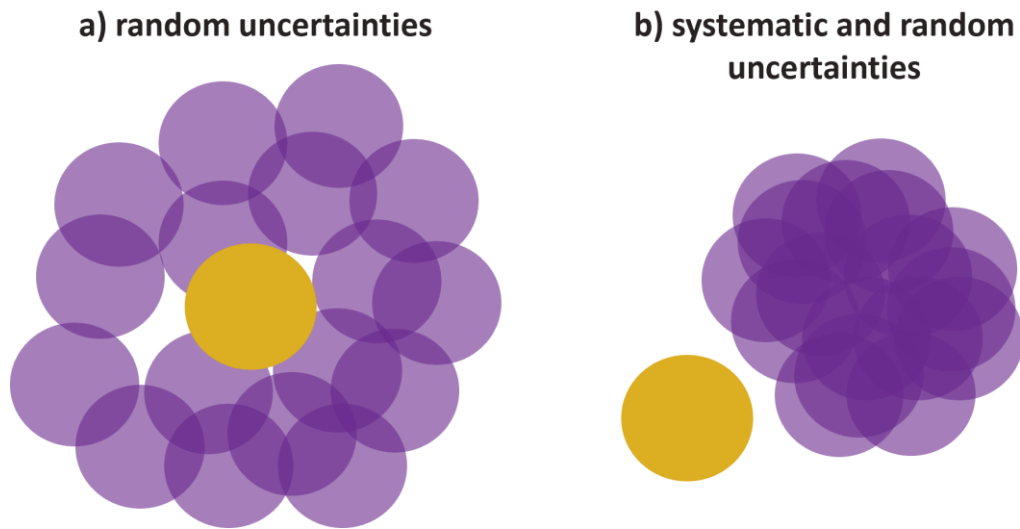


Figure 1.1: An example of a) random uncertainties only and b) systematic and random uncertainties. The yellow circle shows the organ position at planning and the purple circles show the organ position during the different treatment fractions.

Ideally, uncertainties should be reduced as much as possible. Immobilisation devices such as masks can help limit movement and reduce differences in patient set-up position. However, these masks do not affect motion or deformation of organs within the patient and can become 'loose' if the patient loses weight.

Currently, image guided radiotherapy is used for most patients. This is where the patient is imaged before treatment and adjustments to the couch positioning are used to line up the target as much as possible with the position on the pCT. This approach cannot account for any deformations of the body and its organs.

If during treatment it is seen that the patient anatomy has changed too much from the planning anatomy, it is possible to re-plan the patient based on a new scan taken part way through the treatment. However, this only partly reduces the systematic errors, and there will still be day-to-day variations that are not accounted for.

Online planning could be used where a patient is imaged before treatment and a full plan is then created and delivered while the patient is still on the treatment bed. This assumes that the patient remains still while they are waiting for the plan to be created, and can only be used to account for inter-fraction motion. This is very time consuming for both patient

and clinicians so is less practical for regular use in clinic. However, it is being pioneered on MR-Linacs [46], [47].

The next sections will describe the two main sources of geometrical uncertainty in more detail.

1.1.1 Set-up uncertainties and organ motion

Uncertainties due to patient set-up position and inter-fraction organ motion are often treated in the same way. Set-up uncertainties occur when the patient is positioned differently on the couch and organ motion occurs when the organ has been displaced within the body, relative to the bony anatomy. Both manifest as a difference in target or OAR position with respect to their planned position. These uncertainties are generally treated as rigid body translations, although some studies have also considered rotations. The effect of random rotations can be approximated as a local translation around the centre of mass [36].

Models for set-up uncertainties and organ motion generally assume dose-shift invariance, which means that it is assumed the dose distribution will remain unchanged within the patient in the face of positional errors. This assumption is not valid in proton therapy as the dose distribution strongly depends on the density of the tissue the beam travels through, meaning a small set-up displacement could alter the dose distribution considerably. As such, for proton treatments, the dose distribution should be recalculated every time a set-up error is simulated. However, for photon radiotherapy, the dose-shift invariance assumption has shown to be fairly accurate in most cases [43]. This assumption mainly breaks down when there are large changes in tissue density, or at air-tissue interfaces. If the organ is close to the surface of the patient, the dose distribution must be extended outside the patient as the dose is only calculated within the patient body and not in the air [41], [48].

Set-up uncertainties and organ motion can be modelled for the patient as a whole, or for each organ individually. They are the simplest uncertainties to model as they are assumed to follow a Gaussian distribution. The standard deviation of the Gaussian distribution can be different in each of the cardinal directions, creating an ellipsoidal error kernel. Random and systematic uncertainties have to be considered separately and may have different

standard deviations, σ and Σ respectively, although they are often similar [49]. The effect of random set-up uncertainties and inter-fraction organ motion can be approximated by convolving the dose distribution with the random error kernel [50]–[52]. In most cases, this is a good approximation, however it relies on the assumption of a large number of fractions, and so is less accurate for treatments with a small number of fractions [53]. Systematic uncertainties are modelled by translating the blurred dose distribution with shifts sampled from the systematic error kernel.

The size of these uncertainties will vary from site to site. Van Kranen *et al.* [54] measured an overall Σ of 1.1-1.2 mm in the different cardinal directions and σ to be 1.4-1.5 mm in their cohort of H&N cancer patients. However, when considering different OARs separately, they found quite different error distributions for the different OARs, which is evidence of deformations.

1.1.2 Anatomical deformations

In addition to rigid translations, organs can also change their shape, i.e., deform. In H&N, this includes both progressive changes throughout the treatment course and day-to-day variations. Weight loss and parotid shrinking are some of the most common examples of progressive changes, with an average parotid volume decrease of $26 \pm 11\%$ [55] seen throughout treatment. Day-to-day variations include changes to the flexion of the neck and spine, with an average angular change in the yaw (rotation around the anterior-posterior axis) of $0.97 \pm 5.57^\circ$ at the C2 vertebra [19].

Note that even though planning systems can deal with translational errors, e.g., by using robust planning, there are currently no planning systems explicitly dealing with anatomical deformations.

1.1.2.1 Deformable image registration

In order to create a model for the deformations present, the actual deformations seen in the patient or population must first be extracted. This is generally done by using deformable image registration (DIR) to map each point on a repeat scan to the corresponding point on the reference scan. Generally, the reference scan, R , is the scan used for planning, the pCT, and the repeat scan is one of the scans taken at the time of treatment, known as the floating image, F . The DIR outputs the transformation, T , that best

transforms F into R . This transformation describes a vector for each voxel showing how it deforms from one scan to the other. This set of vectors is known as a displacement vector field (DVF).

There are three main components to DIR [56]: the transformation model, the objective function and the optimisation. The transformation model describes how the floating image can be changed to match it to the reference image. It defines the degrees of freedom and the parameters which can be altered to optimise the similarity between the transformed floating image, F' , and the reference image. Throughout this thesis, the transformation model that has been used for DIR is a type of free-form deformation (meaning each voxel can move independently), specifically a B-spline model [57]. This method fits B-spline (a type of polynomial function used for curve fitting) curves at each point in a grid of control points, and optimises the coefficients of each B-spline to ensure the best possible matching between the images [58].

The objective function is optimised to ensure the best match between the transformed floating image and the reference image. It consists of two parts – the similarity metric which gives a measure of how well the two images match, and the penalty terms which can be used to ensure smoother or perhaps more realistic deformations. These two parts are combined to make a single objective function as shown in the following equation:

$$\text{Objective function} = (1 - \alpha) \times SM - \alpha \times PT, \quad (1.1)$$

where SM is the similarity metric, PT is the penalty term and $0 \leq \alpha \leq 1$ is the penalty term parameter.

Several similarity measures have been proposed [59]. The similarity measure used in this thesis is the local normalised cross correlation (LNCC), which aims to maximise the product of the intensities in the two images, over a local region. This product is normalised to account for the fact that the intensities in the two images may not be the same, and just assumes a linear relationship between the intensities in the two images [59]. To calculate the LNCC, for each voxel, the normalised cross correlation (NCC) is calculated within a small Gaussian window surrounding the voxel. The NCC value within the window is then summed over each voxel in the reference scan [60]. The NCC is calculated using the following equation:

$$NCC = \frac{\sum_x (R(x) - \bar{R}) \cdot (F'(x) - \bar{F}')}{U \sqrt{\sum_x (R(x) - \bar{R})^2 \sum_x (F'(x) - \bar{F}')^2}}, \quad (1.2)$$

where U is the number of voxels in the reference scan.

The bending energy, which is the sum of the second derivatives of the transform [61], is used in all DIR in this thesis as a penalty term. Having a low α allows the optimiser more freedom to match the intensities of the images, however if this parameter is too low it can cause un realistic deformations (see Figure 1.2).

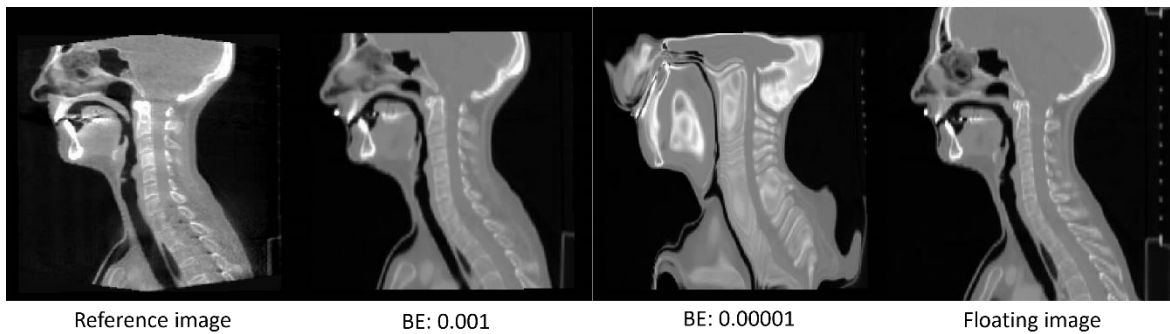


Figure 1.2: An example of DIR with different penalty term parameters for the bending energy

1.1.2.2 Principal component analysis

There have been various studies looking into modelling anatomical deformations. For example, Mageras *et al.* [62] mapped rectum contours from a population database onto a new patient's pCT. Hoogeman *et al.* [63] varied the shape of the rectum slice-by-slice using probability distributions from a population of similar patients, considering only the area, and difference in area and position compared to neighbouring slices. However, the most common way of modelling anatomical deformations is by using a statistical method called principal component analysis (PCA) [64]–[72], which isolates the main modes of deformation present in the input data [73].

PCA is a statistical method that makes use of the correlation between the different dimensions to reduce high dimensional data to a set of principal modes of variation, called components. The output of PCA contains the mean of the input data points and a set of orthogonal principal components, consisting of eigenvectors and the corresponding eigenvalue. Each eigenvector is made up of a linear combination of each of the original dimensions and each data point can then be described by the set of principal components, instead of the original dimensions [73]. This is the mathematical equivalent of transforming

the coordinate system to a new basis. The eigenvalue gives the variance of the data along that eigenvector, and so gives an indication of the size of the variations in that direction. The principal components are ordered by variance, from largest to smallest, so the first component is the one with the most variance. This statistical method operates under the assumption that the data in each dimension is normally distributed about a mean point and that any point in the dataset can be created by summing linear combinations of the different components. A simple example of PCA in 2-dimensions is shown in Figure 1.3.

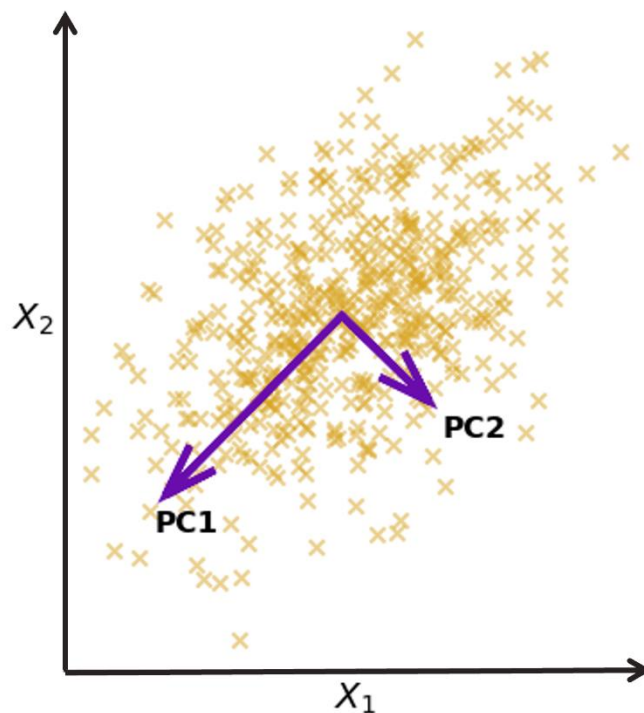


Figure 1.3: A simple 2-dimensional example of PCA. The 'x's represent the data points and the arrows show the principal components. The direction of the arrow shows the eigenvector, and the length of the arrow is scaled by the eigenvalue, or variance, of that component. Note that PC1 describes the axis of largest variation on the dataset, and the two components are orthogonal. After PCA, each data point can be described in terms of (PC1, PC2) instead of (x_1, x_2) .

PCA is useful for modelling anatomical deformations from DVFs. Each DVF contains the displacement in each of the 3 spatial dimensions for each point within the patient and indirectly describes their correlations. This means each DVF has $3U$ dimensions, where 3 corresponds to the spatial directions and the U is the number of voxels. PCA then reduces this large amount of information to give the main modes of deformation present within the set of input DVFs. This is commonly done in one of two ways: looking at each voxel within a scan or each point on the surface of a specific organ or set of organs. As is typically the case when modelling anatomical deformations, if the number of dimensions, N , is much

larger than the number of input scans, n , $n-1$ components are produced from the model [65].

PCA models can be used to investigate the main modes of deformation present in the set of input scans. One can just look at the eigenvectors of the first few components to investigate how each voxel in the model is displaced, and the eigenvalue will give an estimate of the size of the variation in that direction. Alternatively, simulated DVFs can be produced from a PCA model by taking a linear combination of the components, each multiplied by a scaling factor sampled from a Gaussian distribution, using the following equation:

$$\mathbf{v}_{sim} = \bar{\mathbf{v}} + \sum_l u_l \mathbf{e}_l \quad (1.3)$$

where \mathbf{v}_{sim} is a simulated DVF, $\bar{\mathbf{v}}$ is the mean from the model and \mathbf{e}_l is the eigenvector for component l . u_l is a scaling factor, randomly sampled from a Gaussian distribution with a variance equal to the eigenvalue of component l . These simulated DVFs can then be used to run treatment simulations or create coverage probability maps.

PCA is commonly run using multiple input DVFs produced from repeat scans of the same patient. However, this can only be done once the patient has a large number of repeat scans, which is not the case at the beginning of treatment where modelling anatomical deformations is likely to be most useful. One way to get around this and be able to use these models *a priori* is by running PCA on DVFs from multiple patients to get a population-based model of deformation. Budiarto *et al.* [69], Vile [70] and Tsiamas *et al.* [71] all created population-based PCA models where the deformations within each patient were mapped to a 'standard' or average organ shape in order to get a common reference point for all the patients. DVFs in this average geometry from multiple patients were then used as inputs for the PCA models. Vile [70] used a population-based model in order to explicitly split out the systematic and random components to the anatomical deformations, which would not be possible for a single patient model.

Most PCA-based studies assume that anatomical deformations remain the same throughout treatment and do not account for time trends that may affect organ shape. Li *et al.* [74] incorporated a time-varying element into their lung based PCA models, but this

was just based on the periodic lung motion and not on any longer term changes that may occur as the treatment progresses. Currently, there are no studies that have explicitly incorporated time trends into their PCA models for anatomical deformations, something that is particularly important in H&N.

1.2 Margin-based planning

Once the uncertainties have been identified and modelled, they should ideally be accounted for in the planning process. Conventionally, this is done by expanding the CTV to a PTV following a set margin recipe, with the aim of delivering the high dose homogeneously to the PTV. Margin recipes are designed to account for all the uncertainties in the radiotherapy process and ensure that there is a certain probability of achieving correct target coverage across all possible scenarios. They are calculated based on the standard deviations of the systematic and random uncertainties, Σ and σ respectively. Σ and σ can contain components from various sources of uncertainty, e.g., set-up uncertainties and organ motion, added in quadrature.

Stroom *et al.* [75] used the criteria that at least 99% of the CTV received at least 95% of the prescribed dose, and estimated that the required margin, m , would be

$$m = 2\Sigma + 0.7\sigma. \quad (1.4)$$

Alternatively, van Herk *et al.* [76] used the aim of ensuring that at least 90% of patients received a minimum dose to the CTV of at least 95% of the prescribed dose. They calculated that the required margin would be

$$m = 2.5\Sigma + 0.7\sigma, \quad (1.5)$$

assuming a standard beam width of 3.2 mm. These recipes are very similar, with the difference being what was defined as an acceptable level of coverage in each case. Both margin recipes are based on a number of assumptions, such as: a large number of fractions, dose-shift invariance, homogeneous tissue density and perfect conformity, i.e., the 95% isodose contour perfectly follows the PTV.

While margins are mainly used for the CTV to PTV expansion, similar margin recipes have been derived for OARs to produce a planning organ at risk volume (PRV) [77].

1.2.1 Limitations to margin-based planning

There are several limitations to the margin concept. In some cases, OARs may be very close to, or inside, the PTV margin. Generally, the optimisation process is designed to prioritise PTV coverage, so this leads to a high dose being delivered to the nearby OAR. While it would be possible to increase the priority of limiting the dose to that OAR, this can cause an underdose to the edge of the PTV. Using this margin concept, it is challenging to find the optimal balance between overdosing the OAR and underdosing the PTV. It may be that the CTV is unlikely to be found in the volume of the overlap of the PTV and OAR, and so underdosing that edge of the PTV could be preferable to overdosing the OAR in that instance. This information is lost, however, when using PTV margins.

In some situations, the CTV is close to the skin and extending it with a margin could cause the PTV to extend outside of the patient's body. This is clearly not physical and causes problems for the plan optimiser as most systems cannot compute dose delivered to the air.

The size of the PTV necessary can also depend on how steep the dose gradient is around the CTV. If the dose gradient is very steep, the delivered dose to the CTV will be much more sensitive to variations due to uncertainties than a shallower dose gradient. This means that in cases where there is a shallow dose gradient surrounding the CTV, the necessary PTV margins to ensure adequate CTV coverage will be smaller. This is generally not taken into account in the generation of the PTV margins.

PTV margins are derived assuming dose-shift invariance and a large number of fractions, which is not always the case, as explained in Section 1.1.1. This means that in some circumstances, where these assumptions do not hold true, these margin recipes may not be accurate.

Margin recipes were derived on spherical, homogeneous PTVs, much larger than the set-up uncertainties, meaning they have low curvature. However, this is not true of all CTVs in actual patients. Witte *et al.* [78] showed that the density and size of the CTV would actually affect the size of margin required to account for the random uncertainties. Zheng *et al.* [79] also considered cylindrical CTVs and CTVs with concave regions of varying sizes, and concluded that along with different sizes, different shape CTVs would also need different margins to account for the random component of the uncertainties.

Margin recipes were also derived for ideal dose distributions, assuming that the 95% isodose contour would perfectly follow the PTV, without considering whether this was realistic or deliverable. In actuality, this assumption is not met, and the dose distribution is not conformal, and there is effectively an additional margin beyond the PTV [80]. This means that margin-based plans are actually often more robust to uncertainties than expected [49], [81] and are more conservative than intended. While this can be seen as a good thing in terms of target coverage, it does mean that there may be a higher than necessary dose to the OARs, increasing the risk of possible side effects.

Some radiotherapy schedules include features such as boosts, which means a higher dose is prescribed to a subset of the tumour volume [82], which create non-homogenous prescriptions. Non-homogenous dose distributions go against the assumptions made in the original margin recipe derivation and so margin recipes would need to be adjusted in these cases [36].

1.3 Robust planning

Robust planning, sometimes called worst-case optimisation, samples a set of discrete error scenarios and optimises based on the worst-case of these scenarios. The most common implementation of robust planning, known as composite worst-case optimisation, just looks at the total cost for each error scenario, and optimises on the worst-case for the scenario as a whole. An alternative approach is to determine the worst-case scenario on a voxel-by-voxel basis, called voxel-wise worst-case optimisation. For this implementation, the worst-case dose distribution is calculated by taking the minimum dose value across the sampled scenarios for voxels inside the target and the maximum for any voxel outside the target. It is then this worst-case dose distribution that is optimised. Consequently, the plan may be overly conservative and is a lower bound for the absolute worst-case situation as the worst-case for voxel i may occur for a different scenario than the worst-case for voxel j [83]. An objective-wise worst-case optimisation approach has also been used, where the worst-case scenario for each individual objective is considered and optimised [24].

Fredriksson and Bokrantz [84] compared composite-wise, objective-wise and voxel-wise worst-case methods in one prostate case. They found that the voxel-wise worst-case method is much more conservative than the other two as it takes into account situations

that are not going to happen and so compromises on unfeasible situations. This can result in higher target doses but also higher OAR doses than are strictly necessary. In this thesis, the composite worst-case implementation is used for robust planning.

For set-up uncertainties in photon radiotherapy, there are generally seven scenarios sampled in robust planning: the nominal scenario (assuming zero error) and the maximum error in each of the cardinal directions. This assumes that any smaller set-up errors in a given direction will have smaller dosimetric errors [85] and will not contribute to the worst-case scenario. However, as the expected range of the uncertainty increases, it may be necessary to include more scenarios [45]. In RayStation for example, as the robustness setting increases above 5 mm, more scenarios are sampled, including some which combine the three spatial directions (i.e., along the diagonals). While this may lead to a more accurate worst-case situation, it increases the computational time, so a balance has to be struck.

1.4 Probabilistic methods

As well as just taking different error scenarios into account during the optimisation process, probabilistic planning considers the probability distribution of the error scenarios. This means that error scenarios that are more likely to occur are given more consideration, while situations that only occasionally happen have a much lower impact on the overall solution. There are multiple ways of doing this, including coverage probability optimisation (CP) [86], conditional value at risk optimization (CVaR) [87] and stochastic optimisation [36]. In CP, coverage probability maps for the target and OARs are created and, for each voxel, the different objective functions are weighted by the probability of the corresponding target or OAR occurring in that voxel. This approach is particularly useful for areas where an OAR overlaps the PTV. CVaR is a method initially used in investment managing [88], and optimises the expected value of a specified percentage, e.g., 10%, of the worst-case scenarios. It is essentially a mix between worst-case optimisation and stochastic optimisation. Stochastic optimisation looks at optimising the expectation value of the objective function by combining the value of the cost function for each scenario with the probability of that scenario occurring. At each iteration of the optimisation, the dose distribution is first blurred by convolving it with the random error kernel. Then, the

different systematic errors are sampled by shifting the dose distribution and assuming dose-shift invariance. Some implementations also make use of a confidence level, and only optimise based on scenarios up to a certain level of probability. In these implementations, the objective function for each of S error scenarios is calculated and then sorted into ascending order of cost. The weighted sum only sums the first s of the sampled scenarios, until the cumulative probability of those scenarios is greater than or equal to the specified confidence level [36].

Probabilistic methods heavily depend on the model and parameters of the uncertainties used in the optimisation process. If the uncertainties are underestimated in the planning process, or the model used is not very realistic, this can have implications on the effectiveness of the probabilistic approach. It can also become challenging to incorporate all possible geometrical variations into one probabilistic model [36].

This thesis focuses on the stochastic optimisation approach, henceforth simply referred to as ‘probabilistic planning’.

1.5 Evaluating with uncertainties

Once a plan has been created, it should be assessed to see how well it performs and how robust it is to the considered uncertainties. Generally, in clinical settings, radiotherapy plans are just assessed by looking at the planned dose distribution and specific dose metrics for the targets and OARs. However, this just gives an idea of how good the plan is in the nominal scenario and does not explicitly take uncertainties into account. Just because a plan looks good in the nominal scenario, does not mean it will remain a good plan under uncertainties. Generally PTV coverage is still used to report plan quality and comparisons [89] as this is assumed to account for the uncertainties present. However, the PTV is not an exact representation of the dose delivered to the CTV under uncertainties. Both robust and probabilistic planning do not use the PTV and optimise directly on the CTV, and so PTV coverage is often lower than on margin-based plans which were optimised on the PTV. This does not necessarily mean the plan is worse, just that the uncertainties have been included differently and PTV coverage is not always a good measure for CTV coverage under uncertainties. Ideally, target coverage should be assessed on the CTV and uncertainties should be explicitly considered in the plan evaluation, to see how robust the plan is.

The most common way to evaluate and compare how well different plans account for uncertainties is by simulating a set of treatments. These generally simulate each individual fraction of a treatment course and accumulate the dose to each voxel across all fractions. For the simulations, only one systematic error is sampled throughout the whole treatment, and different random errors are sampled for each fraction. This is then repeated multiple times to simulate, for example, 1000 treatments. For each treatment, specific dose metrics can be calculated and the distribution of these across the different simulated treatments can be used to evaluate the plan.

There are two main things to consider when evaluating the plan under uncertainties: the 'specified' value and the variation on that value. The mean or a specific percentile (e.g., 90th percentile) is commonly used as the 'specified' value, to give an idea of how much dose would be delivered in a general treatment. The range or standard deviation of the dose values for the simulated treatments can also be used to give an idea of how much the delivered dose could change from this 'specified' value. However, running 1000 treatment simulations is computationally expensive and can take a long time.

There is no standard way of defining whether a plan is robust or not. Albertini et al. [90] define a plan to be robust if "*even in the case of uncertainties (e.g. range and spatial uncertainties), the agreement between the calculated and the delivered dose is good in the planned clinical volume*". This statement is rather vague, and so different groups have devised different ways of categorising the robustness, or lack of, for a plan. For their robustness evaluator tool, Loebner et al. [91] defined the Robustness Index of a plan to be the fraction (either weighted or un-weighted, depending on the situation) of the sampled scenarios that achieve a specific dose criteria. Similarly, other groups have specified a particular percentage of simulated treatments that must achieve a given dose objective for a plan to be considered robust. For example, Van der Voort et al. [92] considered a plan robust if 98% of the CTV received at least 95% of the prescribed dose in at least 98% of the simulated treatments.

In this thesis, plan robustness is quantified by considering the 90th percentile of the distribution of the minimum (or near minimum) dose to the target for a set of simulated treatments. A plan is considered 'robust' if 90% of the simulated treatments reach at least

95% of the prescribed target dose. If the 90th percentile is much higher than this dose threshold of 95% of the prescribed dose, a plan is considered over-robust.

Currently, no commercial treatment planning systems include simple methods for directly comparing and evaluating plan robustness under uncertainties [93].

1.6 Critical evaluation of literature

A major part of the work presented in this thesis is creating anatomical deformation models for H&N cancer patients using PCA. PCA anatomical deformation models has previously been used in H&N cancer patients by Tsiamas *et al.* [71] and Chetvertkov *et al.* [72].

Tsiamas *et al.* [71] used data from 18 H&N cancer patients to create a set of different PCA models. They created both individual patient and population-based models, and compared models made focussing on just single organs to a models made using all organs. For each model, they used only the first ten components and looked at the percentage predicted spatial displacement, calculated for each component from the RMS of the eigenvector.

For each patient, individual models were created for each organ, and one model using all organs. For the population-based model, only the first 30 fractions were used as different patients were treated with a different number of fractions. If there was no repeat scan available for a given fraction, the previous scan was used. For each organ, the DVFs from each of the first 30 fractions for each patient was input. Similar to the individual patient models, one model was created for each individual organ, and then a multi-organ model was also created.

They found that between 3 and 4 principal components were needed for each of the individual patient models to reach a cumulative 95% predicted spatial displacement, however in the population-based model, 4 components only reached between 80-89% predicted spatial displacement. For the individual patient models, on average the parotids had the largest predicted spatial displacement in the first 4 components out of all the different organ models, but for the population-based model it was the pharyngeal constrictor muscles.

The work in this study compares the percentage predicted spatial displacement for a number of different models, looking at the difference between individual patient models

and population-based models and also between models focussing on specific organs with those considering all organs at once. In this thesis, we followed a similar approach for modelling anatomical deformations in H&N. However, no models were created that focus on just one specific organ as the aim of the models was to simulate anatomical deformation across the whole head and neck region and see what effect this has on the delivered dose.

Tsiamas *et al.* make no mention of a 'reference' or 'average' patient geometry for their population-based models and there is no mention of mapping the individual patient DVFs into a common reference frame. One can assume that this means that the DVFs remained in the individual patient reference frame when used as the inputs for the population-based model. In this thesis, we propose to use a common reference frame, as it is needed for population-based models to accommodate differences in patient geometries.

Chetvertkov *et al.* [72] used data from 10 H&N cancer patients to create models of anatomical deformation using standard PCA and regularised PCA which includes a smoothness term and a sparseness term. For the regularised PCA, the smoothness and sparseness parameters were chosen empirically through experimentation, under the assumption that DVFs describing anatomical deformation should be intrinsically smooth. Models were created from DVFs from both synthetic CTs and the clinical CBCTs. For both the synthetic and the clinical DVFs, models were created using just the first 5 DVFs, roughly half the DVFs, and all but the last DVF.

For the clinical CBCTs, the DVFs for each fraction were extracted using deformable image registration. Additionally, a differential DVF was also computed, which was the difference between the DVF for that fraction with the DVF from the previous fraction.

The synthetic CTs were created by considering just a single mode of progressive change as seen in the real patient scans, e.g., parotid shrinking. This synthetic CT was considered the final fraction treatment and the progressive changes for the other fractions were calculated by scaling this final fraction DVF. This was either done linearly, or by modelling early or late response by using a quadratic scaling. These scaled DVFs described the progressive changes for each fraction, and day-to-day changes were added to each fraction by summing linear combinations of the differential DVFs from all clinical CBCTs from previous fractions. The

day-to-day portion of the synthetic DVF was scaled to model small, medium and large random variations.

For each model, the weighting factor of the first couple of components was calculated by taking the inner product of the eigenvector with the input DVF with the PCA mean subtracted. This was done for each input DVF to the model, and a linear regression was done between these weights, and the line of best fit was extrapolated beyond the fractions input into the model. This regression line was then compared to the regression line calculated when the weighting factor for all available DVFs for that patient in that treatment course (either clinical or synthetic), to see how well the model can predict the weighting factor of later treatments.

They found that the regularised PCA performed better than standard PCA as standard PCA seemed to produce more noise in each component. As such, regularised PCA was better at extracting the progressive changes in just the first couple of components. In both the synthetic models and the clinical models, regularised PCA was seen to get a good linear prediction of the progressive changes when half/almost all DVFs were used for model creation, but struggled when just using the first 5 fractions as the input.

Chetverkov *et al.* found that using regularised PCA produced DVFs that were better able to capture the progressive changes in the synthetic DVFs in the first few components of the model. Standard PCA was used throughout this thesis as regularised PCA is more complex as it also includes a smoothness and sparseness term. Each of these terms needs a parameter and the choice of parameter value will affect the output of the PCA. The challenge then comes in determining what the optimum parameter values should be and indeed whether there is an overall 'best' set of parameters. Rather than trying to optimise these parameters and enforce a certain level of smoothness on the DVFs, I chose to use standard PCA for the models in this thesis.

This study looks at whether PCA models can be used to predict the progressive changes in a patient later on in the treatment. However, their method requires some kind of fitting for the weighting factors. They demonstrate a linear fitting method in their results, but suggest that a quadratic fit to describe either early or late response could also be used. This means that the prediction of the long-term progressive changes will heavily depend on the model

chosen. To counter this need for fitting to predict progressive changes, the anatomical deformation models in this thesis actually include separate weekly models to capture the long-term time trends, so fitting from just the early fractions is not required. However, this method only works as the models presented in this thesis are population-based and so data throughout the whole treatment course for a new patient being predicted is not required. In their work, Chetvertkov *et al.* only look at individual patient models where this would not be feasible.

To the best of the author's knowledge, there have not been any studies in H&N that use anatomical deformation models created from PCA to run treatment simulations or account for the uncertainties directly in the planning process. However, this has been done in other sites.

Vile [70] used data from 19 prostate cancer patients to create a set of PCA models and then used them to run a virtual clinical trial, comparing conventional margin-based plans with plans created using the PCA models. In their Ph.D. thesis, they created both individual patient and a population-based PCA model and explicitly split out the systematic and random components of the anatomical deformation by creating a separate model for each. For each patient, a systematic DVF was created by taking a mean of all the DVFs produced from the repeat scans of that patient. Random error DVFs were calculated by taking the residual DVF for each scan – the DVF from the deformable image registrations with the patient mean DVF subtracted.

For the individual patient models, only a model of the random DVFs was produced as there was just one systematic DVF per patient so this would be constant. For the population-based models, all DVFs were mapped into an average patient anatomy and then a systematic PCA model was created using the systematic DVF from each patient and a separate random model was created using the random DVFs from all patients. For their models, they only kept the first L components that made up 95% of the total variance of the model.

To evaluate their models, they looked at the difference between the input DVFs and the reconstruction of those DVFs (calculated using the inner product of the components with the input DVF) using just the first L components from the model. They also ran a leave one

out analysis, where they recreated the population-based model using only the DVFs from 18 patients and calculated the residual error between the DVFs from the remaining patient and the reconstruction of those DVFs from the model. This was repeated, leaving each patient out in turn.

They then used their models to run a virtual clinical trial, comparing plans made using conventional margins to plans created using patient-specific margins derived using the population-based model. The patient-specific margins were derived by sampling 1000 DVFs from the models, with each DVF being a sum of a DVF from the systematic model and one from the random model. These were then mapped back into the patient geometry and used to create coverage probability maps for the organs. The patient-specific margins were then derived by using the 95% iso-probability contour of the CTV. Plans were created using these patient-specific margins and compared to plans created using the clinical margins/margins derived using the van Herk margin formula. In order to compare the plans, the individual patient models were used to sample a random DVF for each fraction of treatment which was summed to the systematic DVF of that patient. For each plan, the dose was accumulated across each of these fractions sampled from the individual patient model, and specific end points were compared.

They found that the margins derived from the coverage probability maps were smaller than the clinical margins/those from the van Herk margin formula, but the plans still gave adequate target coverage while reducing OAR dose in over 75% of the patients.

In this work, Vile explicitly split out the random and systematic components of the anatomical deformation, and this was also done for the models in this thesis. However, in this thesis, the random component of the deformations were further split into different treatment weeks to account for any long-term time trends. This is important in H&N as there are known to be progressive changes throughout treatment. However, there are no specific long-term trends in prostate patients and so this further splitting would likely have little effect on the work done by Vile.

Vile used the inner product between the eigenvectors of the components and the systematic/random DVFs to evaluate their model. They performed this evaluation on DVFs used to create the model, to see what the residual error when only using the first L

components of their model would be. They also used a leave one out analysis to evaluate the models against DVFs from a patient not included in the model. In this thesis, this evaluation method was extended to not only evaluate against input DVFs, but also other DVFs from the same patient that had not previously seen by the model. This extension allowed us to evaluate how well the model can represent unseen changes that also occur in the patient.

When running their virtual clinical trial, Vile mapped the simulated DVFs from the population-based model back into the patient geometry to create the coverage probability maps. In this thesis, this extra step was not done for sake of simplicity and to avoid any possible additional uncertainties that may arise from mapping the DVFs back into the patient geometry. As such, all simulated treatment courses, analysis and evaluations for the population-based model in this thesis were performed in the average patient space, rather than the individual patient geometry.

When sampling from the PCA models to simulated DVFs used to create the coverage probability maps, Vile sampled one DVF from the systematic model and one from the random model and summed them together. However, for each simulated DVF, a different systematic DVF was sampled. This effectively assumes that each treatment was a single fraction. In this thesis, full treatment courses were sampled where the systematic DVF remained constant for all fractions in the treatment and just the random component of the DVF changed.

Another major part of the work presented in this thesis is to create H&N plans that directly account for the uncertainties present by using robust or probabilistic planning. Wagenaar *et al.* [34] and Fontanarosa *et al.* [41] used robust and probabilistic planning respectively to account for set-up uncertainties in their cohorts of H&N patients.

Wagenaar *et al.* [34] compared margin-based plans with robustly-optimised plans for 10 H&N cancer patients, accounting for set-up uncertainties. They used a 5 mm PTV margin and a robustness setting of 5 mm, sampling 7 scenarios at each iteration. Once optimised, both the margin and the robust plans for each patient were renormalised to ensure that the D99 of the CTV reached 95% of the prescribed dose. For each plan, the dose was recalculated on the CBCTs available for that patient and then the CBCT dose was mapped

onto the pCT and accumulated. This dose accumulation was also done using the weekly verification CTs rather than CBCTs to see if there was any difference in recalculating the dose on the two types of scan.

If required, plan adaptation was performed after week 3, reoptimizing both robust and margin plans on the verification CT from that week. The dose accumulation for the adapted plans was then done assuming the original plan for the first 15 fractions, and the adapted plan for all remaining fractions.

They found there was little difference between the dose accumulation done using the CTs and the CBCTs, suggesting that recalculating on the CBCTs is good enough if no CTs are available. The robust plans were seen to have a higher target dose and lower OAR doses, especially in regions such as near the bones or the oesophagus where there were high density gradients.

For their dose accumulation, Wagenaar *et al.* recalculated the dose at each fraction on the CT or CBCT before accumulating it. This was not done in this thesis for simplicity reasons as the static dose cloud approximation was assumed, meaning that changes in the patient anatomy were assumed to have minimal difference to the dose distribution.

In this study, Wagenaar *et al.* renormalised all their plans to ensure a specific level of target coverage. This was not done for any plans created in this thesis. The plan optimiser aims to minimise the objective function for the plan. While renormalising the plan will ensure that the desired level of target coverage is reached, the renormalised plan will no longer be the optimum plan based on the set of specified objectives. Perhaps a better way to ensure a specific level of coverage while still having the optimum plan for the specified objectives would be to adjust the weight of each objective. However, this means that when comparing the robust and margin plans, either the weights of the objectives will be different, or they will not both reach the specified level of target coverage.

Fontanarosa *et al.* [41] created probabilistic plans for 20 H&N cancer patients accounting for set-up uncertainties and compared them to the clinical margin-based plans, using stochastic probabilistic planning. The values of Σ and σ used in the probabilistic plans were taken from values found at their institution but rescaled to ensure the equivalent margin was as close to 5 mm as possible, which was the margin used in the clinical PTV plans. To

create the probabilistic plan, the clinical plan was copied and any ring structures on the PTV were discarded, and the minimum dose objective for the PTV was changed to the CTV. Just one target objective was kept to avoid possible correlation issues stemming from the fact that each objective function is evaluated separately by the planning algorithm and so different sampled scenarios may be included in different objectives.

At each run of the optimisation, 10,000 35 fraction treatments were simulated using the Σ and σ values and the dose to each point was calculated for that treatment. The 90th percentile of the distribution of the specific CTV endpoints (D99, D98, D2 and mean dose) were calculated. If these endpoints matched that of the clinical plan, the optimisation was stopped. If not, the weights of the objectives were adjusted, and this was repeated until the CTV end points of the two plans were roughly the same. The planned dose distribution was first expanded outside of the body to account for any potential shifts of the target that would take it outside the skin, where the dose is not routinely calculated in planning systems. The dose-shift approximation was assumed and so the dose was not recalculated for each simulated geometry.

They found that in general, the probabilistic plans saw increased OAR sparing when compared to the clinical plan for the same level of target coverage. However, the difference between the plans was larger for some patients rather than others, with the biggest differences occurring in case where an OAR overlapped the PTV but not the CTV.

The probabilistic treatment planning algorithm used in this work by Fontanarosa *et al.* is the same one as used throughout this thesis. In this thesis, the dose-cloud approximation was also used, and the planned dose distribution was first extended outside the body of the patient to account for potential movement of the CTV outside the skin. However, only 1,000 simulated treatment courses were run in this thesis as opposed to the 10,000 run by Fontanarosa *et al.*, simply to save on time and computational power.

Fontanarosa *et al.* adjusted the weights of the probabilistic plan to ensure the same level of target coverage, whereas this was not done in this thesis in order to compare the two plans directly without using different weights between the two. Also, Fontanarosa *et al.* used just one probabilistic objective to avoid possible correlation problems. However, the

probabilistic plans created in this thesis used multiple probabilistic target objectives, to keep them in line with the clinical plan.

1.7 Aims and structure of thesis

In summary, geometrical uncertainties play a large part in radiotherapy planning which, when not accounted for, can cause differences between the planned and delivered dose. This in turn can lead to a decreased efficacy of the treatment and increased risk of side-effects for the patient. The size of both set-up uncertainties and organ motion have been studied in H&N cancer patients [9], [54] and conventional margin plans have been shown to be more conservative than expected to these errors [49]. The use of both robust planning and probabilistic planning have been studied as an alternative way to account for these uncertainties in H&N [34], [41]. Both methods have been shown to decrease OAR dose in H&N while maintaining or improving target coverage when compared to conventional margin plans. However, to the best of the author's knowledge, the efficacy of robust vs probabilistic planning for H&N has not been directly compared. In addition, anatomical deformations have been seen to have an impact on the dose delivered to H&N patients [11]–[15], and models of these deformations have been created [71], [72]. However, the robustness of plans to these anatomical deformations has not been assessed. Additionally, uncertainties due to time trends in the anatomical deformations such as tumour shrinkage, weight loss or parotid shrinking have not as of yet been explicitly incorporated into plan optimisations [94].

The general aim of this thesis was to understand geometrical uncertainties in H&N radiotherapy; how to model them, how to evaluate their impact and how to make plans that are robust against them. This thesis is presented in journal format and consists of a number of studies, presented as publications, that link together to achieve this overall aim. Each chapter addresses one or more of the following specific aims:

1. Create and evaluate a time-dependent H&N anatomical deformation model including random and systematic components.
2. Evaluate the impact of set-up uncertainties and anatomical deformations on the target and OAR doses.

3. Compare different planning approaches to deal with set-up uncertainties and anatomical deformations.
4. Evaluate the impact of target shape on plan robustness.

Chapter 2 presents a method for evaluating how well anatomical deformation models created using PCA can represent real unseen anatomical changes in a patient, working towards addressing aim 1. The use of this methodology is demonstrated for both patient-specific and a population-based model for H&N cancer patients.

Chapter 3 develops a time-varying anatomical deformation model for H&N cancer patients and uses it to investigate the effect of these deformations on the clinical plans. This effect is then compared to that of set-up uncertainties and a combination of the two uncertainties, addressing aims 1 and 2.

Chapter 4 investigates whether anatomical deformations in H&N cancer patients can be accounted for by treating them as set-up uncertainties. This is done by creating margin, robust and probabilistic plans accounting for different sizes of set-up uncertainties and evaluating the effect of anatomical deformations on each plan. This addresses aims 2 and 3 of this thesis.

Chapter 5 evaluates the robustness of margin, robust and probabilistic plans to set-up uncertainties for different target shapes. This is used to evaluate the impact that target shape has on plan robustness, addressing aims 3 and 4.

Finally, Chapter 6 presents a general discussion, showing how the previous chapters link together to address the overall aim, as well as identifying future work to continue improving radiotherapy planning for H&N cancer patients.

Chapter 2

Evaluating principal component analysis models for representing anatomical changes in head and neck radiotherapy

This chapter has been published in *Physics and Imaging in Radiation Oncology*, 2022, Volume 22, Pages 13-19, <https://doi.org/10.1016/j.phro.2022.04.002>. The paper is presented here with modifications to the text, formatting and figures to be consistent with the rest of this thesis.

Authors

*Jennifer Robbins¹, *Raul Argota-Perez², Andrew Green¹, Marcel van Herk¹, +Stine Korreman^{2,3,4}, +Eliana Vásquez-Osorio¹

Affiliations

1. Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
2. Department of Oncology, Aarhus University Hospital, Aarhus, Denmark
3. Danish Centre for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark
4. Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

*Both authors contributed equally to this paper

+ Both authors contributed equally to this paper

Author contributions

I, with help from EVO, performed all the deformable registrations for the population-based model and evaluation. I wrote the code used to create the population-based PCA model which was then adapted by RAP for the patient-specific models. RAP, MvH, SK, EVO and I devised the evaluation strategy which was reviewed by AG. I produced the methods figure. I performed all the analysis and generated the figures and text related to the population-based model. Together with RAP, I wrote and revised the introduction, materials and

methods section, discussion and supplementary materials. The whole manuscript was reviewed by all co-authors.

Abstract

Background: Anatomical changes during radiotherapy pose a challenge to robustness of plans. Principal component analysis (PCA) is commonly used to model such changes. We propose a toolbox to evaluate how closely a given PCA model can represent actual deformations seen in the patient and highlight regions where the model struggles to capture these changes.

Methods: We propose to calculate a residual error map from the difference between an actual displacement vector field (DVF) and the closest DVF that the PCA model can produce. This was done by taking the inner product of the DVF with the PCA components from the model. As a global measure of error, the 90th percentile of the residual errors (M_{res}^{90}) across the whole scan was used. As proof of principle, we demonstrated this approach on both patient-specific cases and a population-based PCA in head and neck (H&N) cancer patients. These models were created using deformation data from deformable registrations between the planning computed tomography and cone-beam computed tomography (CBCTs) and were evaluated against DVFs from registrations of CBCTs not used to create the model.

Results: For our example cases, the oropharyngeal and the nasal cavity regions showed the largest local residual error, indicating the PCA models struggle to predict deformations seen in these regions. M_{res}^{90} ranged from 0.4 mm to 6.3 mm across the different models.

Conclusions: A method to quantitatively evaluate how well PCA models represent observed anatomical changes was proposed. We demonstrated our approach on H&N PCA models, but it can be applied to other sites.

2.1 Introduction

Throughout radiotherapy, geometrical uncertainties such as set-up errors or anatomical changes may lead to underdosing the target or overdosing organs at risk. Margins or robust optimisation techniques can deal with these uncertainties, but their applicability is limited for complex changes. Therefore, adaptive strategies are often used for larger anatomical changes [95], [96]. However, current adaptation strategies represent a significant workload and can create bottlenecks in workflows [97], [98].

An alternative is to predict anatomical changes using mathematical models to increase plan robustness. Such models could be used to predict which patients may benefit from plan adaptation, or to identify regions where large anatomical changes are expected. Deformation models could also simulate anatomical changes during planning to increase robustness, e.g. using robust or probabilistic planning [99]. Principal component analysis (PCA) is widely used for creating such models, for instance in lung [68], [100]–[102], prostate [65]–[67], [69], [103], cervix [99] and head and neck (H&N) [71], [72]. The usefulness of such models depends on their ability to accurately simulate future changes in the patient.

In this study, we propose a toolbox to evaluate how well a given PCA model can represent anatomical changes that were not used to generate the model. These tools can be used to evaluate the model robustness, estimate global residual errors within the boundaries of the scan and highlight regions where the model struggles to capture anatomical changes. The aim of this study was to present a proof-of-principle for this method in H&N cancer using both patient-specific and population-based deformation models.

2.2 Materials and methods

PCA models can be created from a set of displacement vector fields (DVF) to simulate anatomical changes. These DVFs use non-rigid registrations to describe the deformations between two images, e.g., the planning CT (pCT) and a cone-beam CT (CBCT).

The resulting model consists of the eigenvectors for each of the principal components of deformation, e_l , the corresponding variance for each component, λ_l , and the mean of the input DVFs, \bar{v} . These can then be used to simulate plausible DVFs (v_{sim}), following Equation 2.1, where the weights, u_l , are randomly selected from a Gaussian distribution centred at zero with variance λ_l .

$$v_{sim} = \bar{v} + \sum_l u_l e_l \quad (2.1)$$

2.2.1 Evaluation strategy for PCA-based deformation models

The proposed method presented here can be used to determine to what degree a PCA model can represent a DVF that was not used to generate the model. We refer to this DVF

as the ‘reference DVF’, \mathbf{v}_{ref} . The closest vector field to the reference DVF, $\mathbf{v}_{closest}$, was generated using the model with an optimal set of weights, w_l , found using Equation 2.2.

$$w_l = \mathbf{e}_l \cdot (\mathbf{v}_{ref} - \bar{\mathbf{v}})^T \quad (2.2)$$

The closest simulated DVF the model can produce, $\mathbf{v}_{closest}$, was found by substituting w_l from Equation 2.2 into Equation 2.1. A measure of the likelihood of this closest DVF being produced by the model, the Z-score, is presented in Supplement 2.A.1.

To quantify how close the closest DVF is to the reference DVF, we defined the residual DVF, \mathbf{v}_{res} , as

$$\mathbf{v}_{res} = \mathbf{v}_{ref} - \mathbf{v}_{closest} \quad (2.3)$$

By taking the magnitude of each vector within \mathbf{v}_{res} , M_{res} , we can identify local regions with larger errors. As a measure of the global model performance, we calculated the 90th percentile of M_{res} , M_{res}^{90} . The 90th percentile is commonly used in the literature [9], but the mean or a different percentile could be selected depending on the intended application.

2.2.2 Evaluation toolbox

Our evaluation scheme can be used in different ways to evaluate how well PCA-models perform. E.g., we can evaluate the robustness and stability of the model, the sensitivity of the model to random noise, and the general ability of the model to simulate anatomical changes within the patient.

Model robustness: To determine how robust a PCA model is to the input DVFs, a leave-one-out cross validation (LOOCV) analysis can be performed by running PCA with one of the input DVFs left out. The DVF that is left out can then be used as \mathbf{v}_{ref} for that iteration.

Model sensitivity: The sensitivity of the model to random noise and the number of components used for the evaluation can be assessed using simulated DVFs, to which Gaussian noise was added (Supplement 2.A.2).

Model generalisability: To evaluate whether a deformation model describes real anatomical changes, the PCA model can be evaluated against unseen DVFs from the patient.

2.2.3 PCA models examples

To test our proposed method in different settings, we created both patient-specific models (datasets 1 and 2) and a population-based model (dataset 3). The datasets are summarized in Table 2.1 and detailed in Supplement 2.A.3.

Data for	Patient-specific PCA models		Population-based PCA models	
			Training	Validation
Dataset	1	2	3a	3b
Nr of patients	24	20	20	20
Site	Sinonasal	Oropharyngeal	Oropharyngeal	Oropharyngeal
Treatment technique	All IMRT	10 IMRT, 10 VMAT	All VMAT	All VMAT
Treatment period	2009-2017	2008-2018	2016-2018	2016-2018
Dose prescription	66-68 Gy, (n=16), 60-66 Gy (n=8)	66-68 Gy (n=17), 76 Gy (n=3).	60-66 Gy	60-66 Gy
Number of CBCTs per patient	30 - 34	33-34 or 56	8 - 31	9 - 24

Table 2.1: Summary of patient datasets.

Dataset 1 includes 24 sinonasal cancer patients collected from the DAHANCA database with 30-34 daily CBCTs per patient, under approval by the Danish Data Protection Agency (1-16-02-676-18).

Dataset 2 includes 20 oropharyngeal cancer patients collected from a single institution with 33-56 CBCTs per patient, under internal approval (in accordance with Danish guidelines).

Dataset 3 includes 40 oropharyngeal patients collected from a single institution with 8-31 CBCTs per patients, under ethics approval from the UK North West - Haydock Research Ethics Committee, (17/NW/0060, local ref. 2018-018). 20 patients were selected as training patients, dataset 3a. The remaining 20 patients were used for validation, dataset 3b.

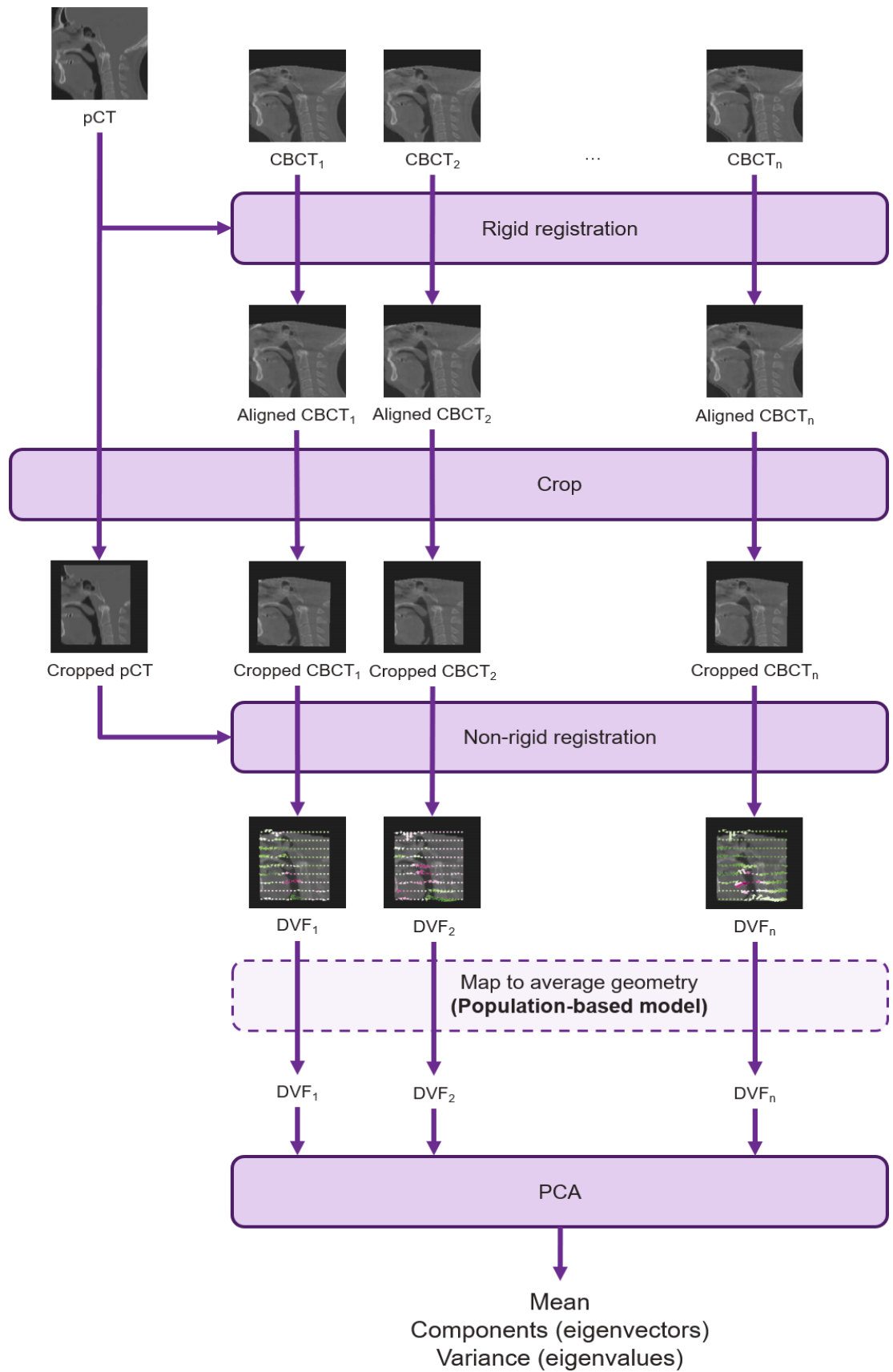


Figure 2.1: Flow chart showing the method for generating PCA models. For the patient-specific models, all CBCTs are from the same patient and are registered to the same pCT for that patient. For the population-based model, each CBCT is from a different patient and is registered to the corresponding pCT.

To generate the DVFs for PCA creation, each CBCT was first rigidly registered to the pCT based on bony anatomy, obtaining the starting point for deformation. The pCT and aligned CBCTs were then cropped to a bounding box to reduce computation time. Then, deformable registration was performed from each CBCT to the pCT, using either NiftyReg [61] (datasets 2 and 3) or Anaconda in RayStation [104] (dataset 1). A scheme summarising our approach is presented in Figure 2.1.

For the patient-specific models, DVFs from the first five days of treatment were used to create each PCA model. The bounding box covered an area including the brainstem, parotids, primary CTV and the nodal CTV (if present) with a 5 cm margin in each direction excluding the shoulders. To investigate the effect of the number of input scans, models were created for two of the patients (dataset 2) using the ten DVFs corresponding to the first two weeks of treatment.

For the population-based model, DVFs mapping one CBCT from each of the six weeks of treatment from each of the 20 training patients were used. To standardise the DVFs between different patients, an average patient geometry was created using a groupwise registration [105] of the pCTs of these patients. The bounding box covered an area including the brainstem, parotids, oral cavity, larynx and primary CTV with a 1.5 cm margin in each direction. All DVFs were then mapped to the average geometry using SimpleITK [106] and were used to create a single PCA model.

The authors visually inspected all deformed CTs and fine-tuned sub-optimal registrations.

We generated all PCA models using Scikit-learn in Python [107]. We report the cumulative variance for each component for the PCA models in Supplement 2.A.4.

2.2.4 Model Evaluation

For all models, we evaluated the robustness with the LOOCV method. We ran an investigation into the model sensitivity for the population-based model and one of the patient-specific models (Supplement 2.A.2).

As test data for the model generalisability method, we used the DVFs from the first CBCT of each of the subsequent five weeks of treatment for patient-specific cases. For the population-based case, we used all remaining DVFs of the training patients (dataset 3a) and

DVFs from all CBCTs of the validation patients (dataset 3b). In each case, we created heat maps of M_{res} to evaluate the local model quality and M_{res}^{90} to evaluate the global model quality.

A pass/fail threshold for M_{res}^{90} can be set depending on the specific application. As an example, we selected 4 mm.

We used all components available for evaluating the PCA models, but in practice one could restrict the number of components, e.g., many PCA models only include components that cover the largest 90% of the total variance of the model.

2.3 Results

2.3.1 Model robustness (LOOCV)

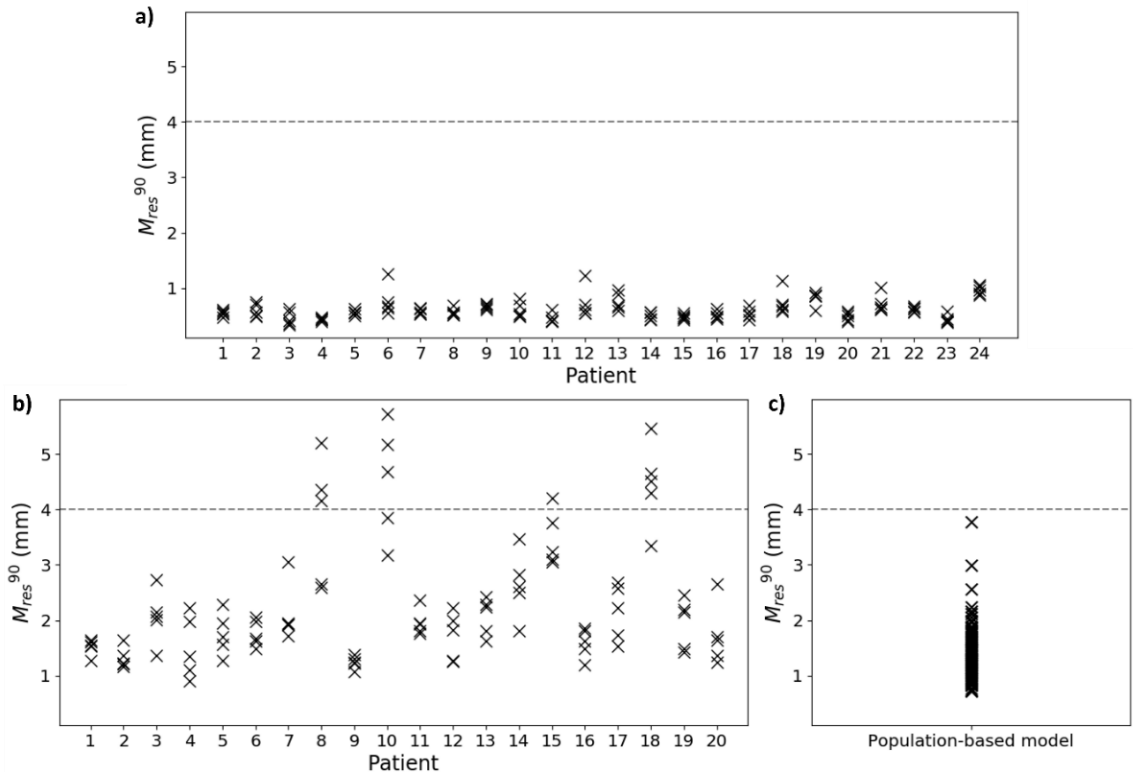


Figure 2.2: M_{res}^{90} for the LOOCV for each of patient-specific model in a) dataset 1 and b) dataset 2, and for c) the population-based model (dataset 3a).

Figure 2.2 shows M_{res}^{90} for the LOOCV for all models. For the patient-specific study, all M_{res}^{90} were well below the 4 mm threshold for dataset 1 (sinonasal cancer), while for dataset 2 (oropharyngeal cancer), models for 4 out of 20 patients had $M_{res}^{90} > 4$ mm. In the population-based study (dataset 3a), all M_{res}^{90} were below 4 mm.

The largest M_{res} were observed in the regions around the oropharynx for all datasets. Additionally, for dataset 1, the region around the nasal cavities had high M_{res} .

2.3.2 Model generalisability

The ability of the patient-specific models created with the DVFs from week 1 to adequately describe deformations present later during treatment is shown in Figure 2.3. Most of the patients passed the 4 mm threshold in subsequent weeks. As expected, a tendency for M_{res}^{90} to increase over time during the treatment course was observed. Using DVFs from the first two weeks resulted in slightly decreased numbers, but with the same trend over time (see data in Supplement 2.A.5). It should be noted that two of the patients who failed the generalisability evaluation also failed the LOOCV (patients 10 and 18). A heat map of the mean M_{res} (across all DVFs used for investigating the model generalisability) for patient 1 of dataset 1 is shown in Figure 2.4. The largest M_{res} values were found in the oropharynx region for all patients, and additionally in the nasal cavities for dataset 1 patients.

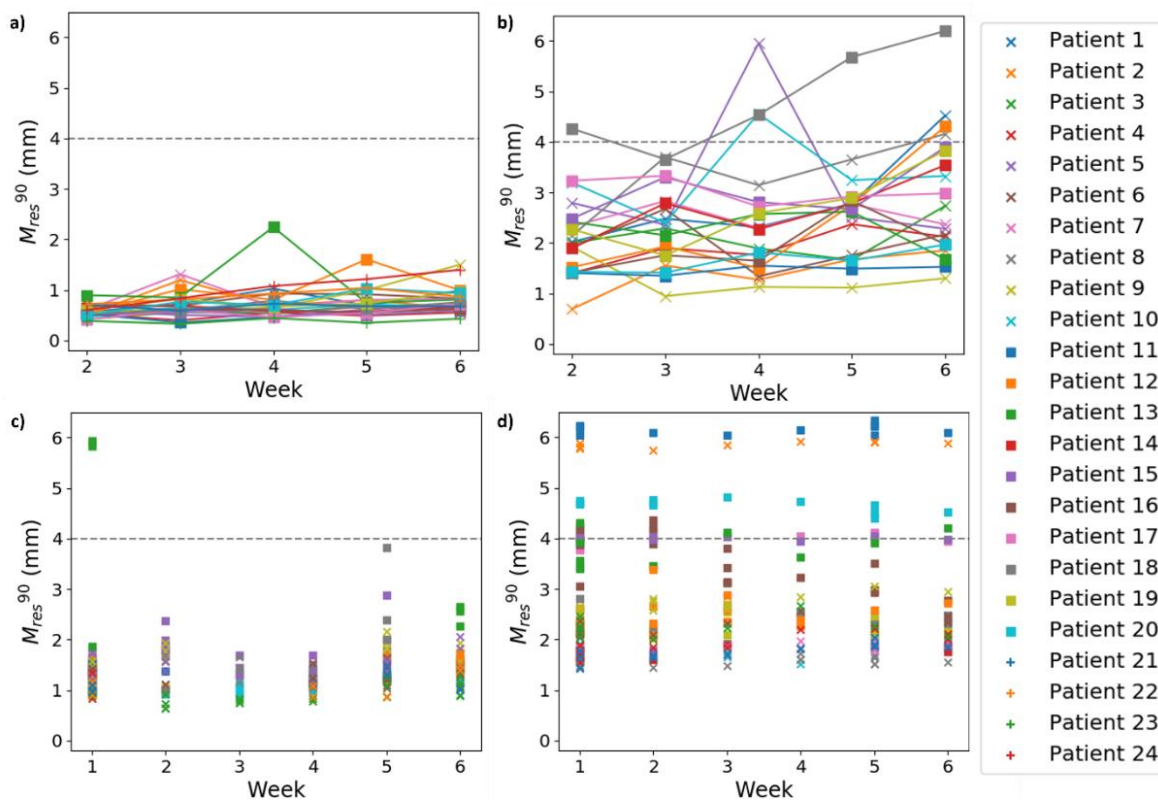


Figure 2.3: The 90th percentile of M_{res} calculated on one DVF from each of weeks 2-6 of treatment for each patient for the patient-specific models created using a) dataset 1, b) dataset 2, and for the population-based model calculated on all DVFs not used for model creation from each of weeks 1-6 using the c) training patients (dataset 3a), d) validation patients (dataset 3b).

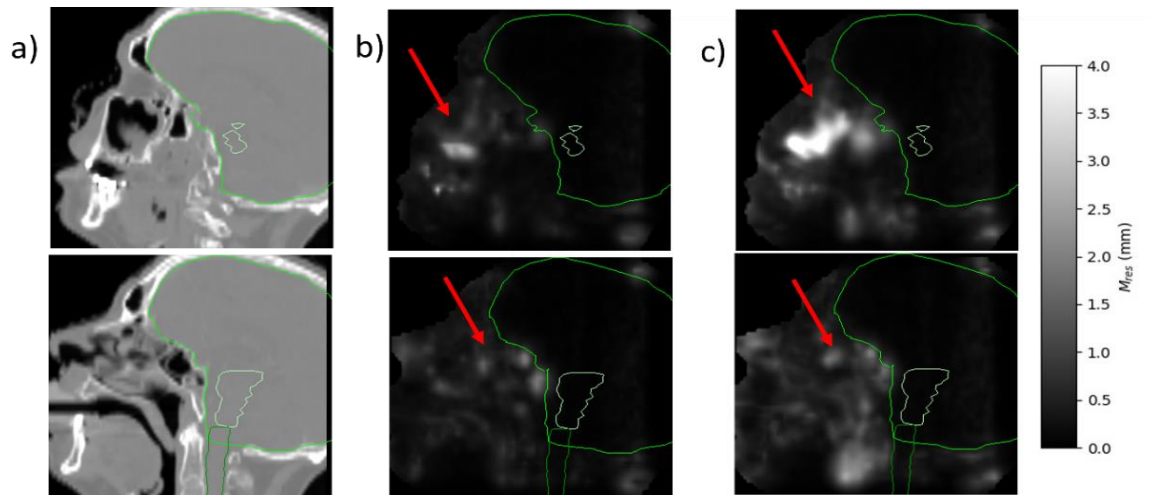


Figure 2.4: Identifying areas with a high M_{res} for the patient-specific model (patient 1, dataset 1). The panels show a) the pCT for a patient in two different slices (top showing nasal cavity region and bottom showing the oropharyngeal region), b) the mean of M_{res} from LOOCV and c) the mean of M_{res} from the following weeks. Contours of the brain, brainstem and spinal cord are shown in green.

For the population-based model, there was little difference in performance between different treatment weeks, but the model performed worse for validation patients than training patients. Three of the validation patients were above the 4 mm threshold for all DVFs evaluated (Figure 2.3). For the training patients, two DVFs from the first week of treatment for a single patient failed to reach the threshold (who also had the second highest value for the LOOCV in week 1). M_{res} was largest around the oropharynx, consistent with the patient-specific results (Figure 2.5). The mean M_{res} values (across all available DVFs) were higher for the validation patients in this and surrounding regions. Around the brainstem and within the skull, M_{res} was consistently low for both the training and validation patients.

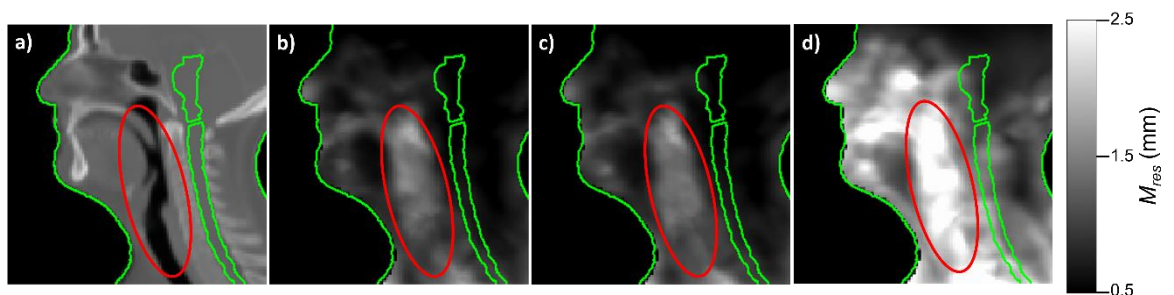


Figure 2.5: Identifying trends in areas with a high M_{res} for the population-based model for both training and validation patients. The panels show a) the average pCT, b) the mean of M_{res} from the LOOCV (120 DVFs), c) the mean of M_{res} from all the DVFs not used to create the PCA model of the training patients (217 DVFs) and d) the mean of M_{res} from all the DVFs of the validation patients (254 DVFs). The region with a consistently high M_{res} is outlined in red. The external contour and the brainstem and spinal cord are shown in green for context.

A scatter plot linking the mean M_{res}^{90} for the LOOCV with the mean M_{res}^{90} for the generalisability is presented in Supplement 2.A.6.

2.4 Discussion

We have presented a toolbox to evaluate how well anatomical changes in unseen data can be described by PCA models and tested it for H&N cancer patients in both patient-specific and population-based contexts. Our evaluation strategy can be used to easily determine patients where PCA modelling could be used (e.g., by applying LOOCV and setting up a threshold of acceptable error) and highlight anatomical regions where the PCA model struggles.

To the best of our knowledge, no other study has proposed such a method. Similar evaluation strategies have focussed on either determining the required number of components to use in a PCA model [65], [68], [69], [100] or the optimal number of input scans needed to create the model [68]. Generally, these strategies tend to evaluate the PCA model against input DVFs, whereas our method compares the model to unseen DVFs. However, Badawi *et al.* [68] evaluated the residual error on unseen scans for PCA models created using a subset of the input scans, but this was mainly used to assess how many weeks' worth of input scans was needed for the residual errors to stabilise. Budiarto *et al.* [69] also evaluated the residual error on three patients not previously seen by their population-based model, but did not evaluate on any unseen DVFs from their training patients. For our population-based model, the validation patients performed worse in the evaluation than the training patients (see Supplement 2.A.7).

Some studies using PCA looked at deformations to the surface contour of specific organs as inputs to the model [65], [68], [69] while others considered each image voxel [67], [74], [101]. Here we have used the voxel-based approach as this allows us to model changes on the entire scan and include multiple organs and their interactions, as well as other surrounding tissue.

Our scheme focuses on computing the magnitude difference between the reference and closest DVF and does not differentiate between differences in magnitude and direction for each vector. Of course, one could also evaluate each of the x , y and z directions separately.

While the method only evaluates individual instantiations of a DVF, by combining multiple results, the toolbox can be used to assess the quality of the model. E.g., by computing M_{res} for multiple known DVFs, this can give an idea of what modes of deformations are/are not captured by the model. In addition, our method also calculates the Z-score for each component, providing an idea of the magnitude of the deformations captured. Providing the Z-scores across all sampled DVFs follow a normal distribution, the model accurately captures the size of the deformations sampled and will correctly describe expected changes.

This work relies on PCA, so all limitations related to this method will apply. A vital part of PCA modelling and the evaluation toolbox are the deformable registrations. Although the registrations were visually inspected to ensure quality, they will still introduce uncertainties into both the PCA model and residual DVFs. For this study, a quantitative assessment of registration uncertainties was not done, but we would recommend this for specific applications depending on the required accuracy. To give an idea of the possible size of these uncertainties, Veiga *et al.* [108] found the mean distance transform between manual and deformed surfaces for H&N registrations using NiftyReg to be 0.3 ± 0.4 mm. We assumed that any uncertainties associated with imperfect registrations would be smaller than the anatomical changes we aim to capture in the model – e.g., Barker *et al.* [9] found the median medial shift of the centre of mass of the parotid throughout treatment to be 3.1 mm.

Previous studies have investigated the generation of PCA models in H&N cancer patients and their evaluation. For example, Tsiamas *et al.* [71], assessed the number of components needed to model the spatial displacements for specific organs, using data from 18 patients to create both individual and population-based models. They focused on comparing the relative variance of the different PCA components between models. They found three to four principal components were sufficient to achieve spatial displacement prediction at the 95% confidence level for normal organs. In another study, Chetvertkov *et al.* [72] generated PCA models considering variations occurring in the whole scan in 10 patients. They focused on whether the weighting factors from the inner product could be used to predict systematic changes throughout the treatment course, and dismissed the errors not

expressed by linear combinations of the eigenvectors. Our results show that these errors can be considerable and should not be dismissed.

The patient-specific models were created from scans from the first week of treatment and could for instance be used either for informing a patient-specific threshold for adaptation, or for including variations seen in the PCA model in a patient-specific robust optimization. In the first case, a set of possible patient geometries could be simulated from the model, dosimetric evaluation could be performed, and an assessment of whether adaptation is required be made. In the second case, a robust plan could be made to include the variations seen in the PCA model, such that within this envelope of variations no adaptation would be required for the rest of the treatment. Our population-based model was created using treatment scans across six weeks from 20 patients, providing the opportunity to use it to simulate possible anatomical changes on new patients, without having to wait for multiple scans. This means population-based models could be used at the planning stage to evaluate the robustness of treatment plans against expected anatomical changes or to directly account for them by using robust or probabilistic planning. Depending on the specific application, PCA models may be tuned to focus either on systematic variations (seen in the lower order components) and/or random variations (in higher order components).

We observed that M_{res}^{90} varied between datasets. One possibility is that the oropharynx tumours in datasets 2 and 3 may affect the model performance as it changes the anatomy and the tumour may shrink during treatment. Another possibility is the difference in bounding boxes between patient datasets. The bounding box for dataset 1 was based on a region surrounding the skull, meaning there was likely to be better anatomical consistency than for datasets 2 and 3, where the bounding boxes generally extended further down the neck which is prone to larger deformations [54]. This shows that the residual error depends on the region of interest being considered, which should be carefully selected to cover clinically relevant organs. One way to eliminate the bias of the bounding box could be to report M_{res}^{90} for the target volume and the relevant organs individually.

By using our evaluation toolbox, regions where a PCA model struggles to represent deformations can easily be identified. For example, the oropharynx presented a challenge for both the patient-specific models (using 5 or 10 input DVFs) and the population-based

model (using 120 input DVFs). This suggests that it is not a limit in the number of scans that is causing this challenge, but rather a limitation of the PCA method itself. PCA assumes variations are normally distributed around a non-zero mean. This assumption is commonly seen in the literature for models for day-to-day changes in organ shape [66], [67], but it has not been validated in H&N. Additionally, for the sinonasal cancer patients, the nasal cavity was highlighted as a challenging region. This is due to the cavity being filled/emptied, which violates the assumption of the variation of the position of each voxel being normally distributed.

2.5 Conclusion

We have proposed and tested a toolbox to evaluate how well PCA models can predict anatomical deformations. We showed how regions were identified where models created from the first week of radiotherapy in H&N cancer patients struggle to represent anatomical changes occurring later during treatment. All tested models had difficulties capturing deformations in the oropharyngeal region, and the nasal cavity for models created on sinonasal cancer patients. Our methods could potentially be used in a variety of scenarios to evaluate and validate PCA models and facilitate incorporation of deformation modelling in various clinical applications.

2.6 Acknowledgements

This work was supported by The Danish Cancer Society (grant no R167-A11003), DCCC-RT - The Danish National Research Centre for Radiotherapy, Danish Cancer Society (grant no. R191-A11526) and Danish Comprehensive Cancer Centre, and by Aarhus University Research Foundation (grant no AUFF-F-2016-FLS-8-4 HD). This work was supported by Cancer Research UK via funding to the Cancer Research Manchester Centre [C147/A25254]. Marcel van Herk was supported by NIHR Manchester Biomedical Research Centre. We kindly acknowledge Dr. Kenneth Jensen, DCPT, for his help with collecting the patient data. We would like to thank Rebecca Holley for helping reduce the word count in the final manuscript.

Appendix 2A

2.A Supplementary material

2.A.1 Z-score

To get an indication of how likely it is for $v_{closest}$ to occur from the PCA model, we calculated the Z-score (the weight divided by $\sqrt{\lambda_l}$) for the weight of each component, and report the maximum Z-score, Z_{max} , for each DVF. Following the empirical rule, if the Z-score is less than 3, there is more than 99.7% chance that it could have been selected from the model. For models with a large number of components the percentage of Z-scores under 3 can also be reported. If this number is significantly less than 99.7%, it is unlikely that $v_{closest}$ could be randomly produced from the model.

Robustness (LOOCV):

Z_{max} ranged from between 0.1-1.6 for dataset 1 and between 0.1-1.8 for dataset 2. For the population-based study (dataset 3a), Z_{max} ranged from 1.8-5.7. 99.2% of all the Z-scores were below 3.

Generalisability:

For the DVFs not used to train the PCA models, Z_{max} ranged between 0.2-1.8 for dataset 1 and 0.2-2 for dataset 2. For the population-based datasets, Z_{max} ranged from 1.7-13.0 for the training patients with 99.0% of the Z-scores below 3 and for the validation patients, Z_{max} ranged from 2.0-17.5 with 91.6% of the Z-scores below 3. The DVFs with the highest Z_{max} also tended to be those with high M_{res}^{90} in the population-based model.

2.A.2 Model sensitivity

We assessed the sensitivity of the PCA models by quantifying the impact of adding differing levels of Gaussian noise to the reference DVF. We randomly produced 50 simulated DVFs using Equation 2.1 and added spatially independent Gaussian noise (standard deviation, SD, of 0, 1, and 2 mm) to each vector. For each simulated DVF (including noise), we calculated M_{res}^{90} using the first l components from the PCA model, where $1 \leq l \leq n-1$

(where n is the number of input DVFs for the model and $n-1$ is the total number of components from the model). These M_{res}^{90} values were plotted against l , along with the mean for all 50 DVFs and are shown in Figure 2.A.1.

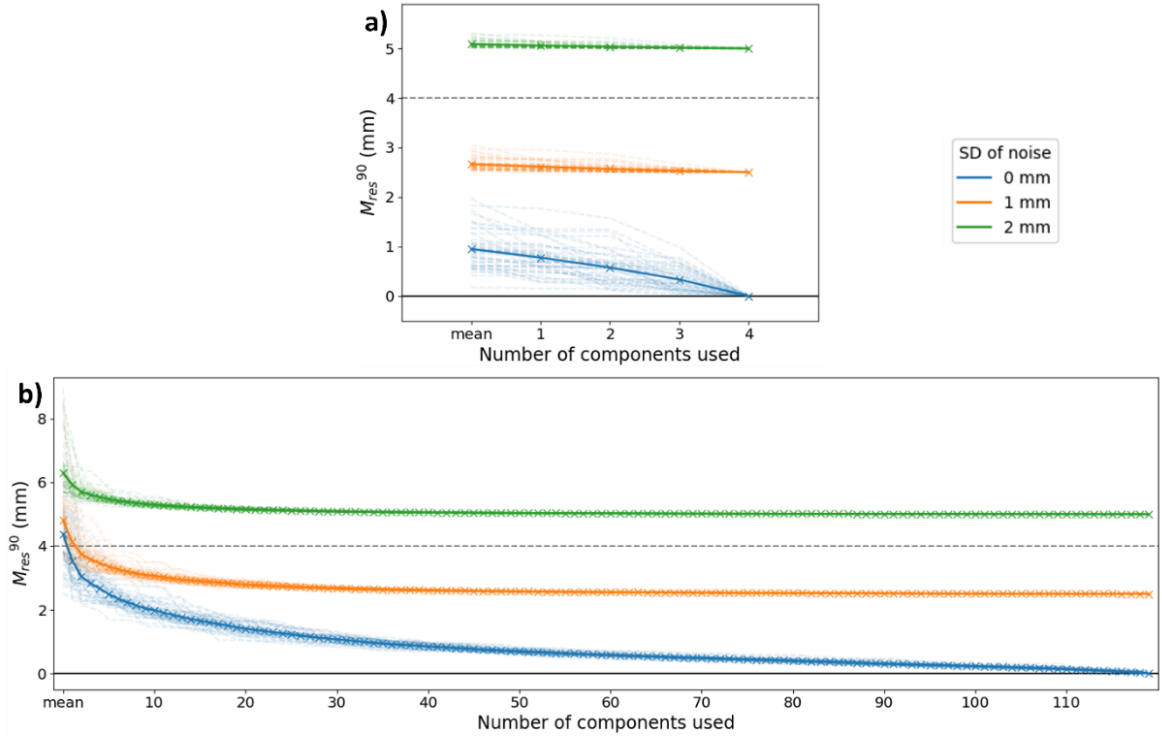


Figure 2.A.1: Comparing M_{res}^{90} for DVFs created from the PCA model with differing levels of noise, analysed using a different number of components from the model. The dashed lines show the results from each of 50 randomly generated DVFs and the solid line shows the mean of those 50. Plot a) shows the patient-specific model (patient 1, dataset 2) and plot b) shows the population-based model.

The tests with the simulated data (Figure 2.A.1) show that for all noise-free simulated DVFs, M_{res}^{90} (and actually all M_{res}) reached zero after all components were accounted for, validating our evaluation approach. We also established that adding Gaussian noise with SD of 2 mm pushed M_{res}^{90} above our chosen threshold of 4 mm. Regardless of the level of Gaussian noise, the Z_{max} values were the same and ranged between 0.3-2.6 for all patient-specific models and 1.9-3.9 for the population-based model. 99.7% of all the Z-scores were below 3 for the population-based model.

2.A.3 Patient data

Patient datasets are detailed below:

Dataset 1: Data from 24 sinonasal cancer patients treated either with primary or postoperative photon therapy (IMRT) with curative intent from 2009 to 2017 were included in this dataset. The data was collected from the DAHANCA database (approval by the Danish Data Protection Agency (1-16-02-676-18)). These patients were treated to a prescribed dose to the primary CTV of either 66-68 Gy for primary radiotherapy or 60-66 Gy for post-operative radiotherapy. Patients had between 30-34 CBCTs.

Dataset 2: Data from 20 oropharyngeal cancer patients treated in the period of 2008-2018 with radical radiotherapy in a single institution were arbitrarily selected. The data was fully anonymised and collected from a clinical database after internal study approval at the clinical department, in accordance with Danish guidelines. 10 patients were treated with IMRT and 10 with volumetric-modulated arc therapy (VMAT), 17 to a prescribed dose of 66-68 Gy (33-34 fractions, 5-6 per week) and 3 to a prescribed dose of 76 Gy (56 fractions, 10 per week). Patients had between 33-56 CBCTs.

Dataset 3: Data from 40 oropharyngeal patients treated between 2016 and 2018 with radical radiotherapy in a single institution were selected. These data were collected from the ukCAT distributed learning database (ethics approval from the UK North West - Haydock Research Ethics Committee, (17/NW/0060), local consent ref. 2018-018). All patients were treated in 30 fractions using VMAT to a prescribed dose of 60-66 Gy to the primary CTV. 20 patients were selected as the training patients, denoted dataset 3a. The remaining 20 patients were used for validation and denoted dataset 3b. Patients had between 8-31 CBCTs.

2.A.4 Cumulative variance ratio

Here we report the cumulative variance for each component in the PCA models for the first patient-specific model in each of datasets 1 and 2 and for the population-based model.

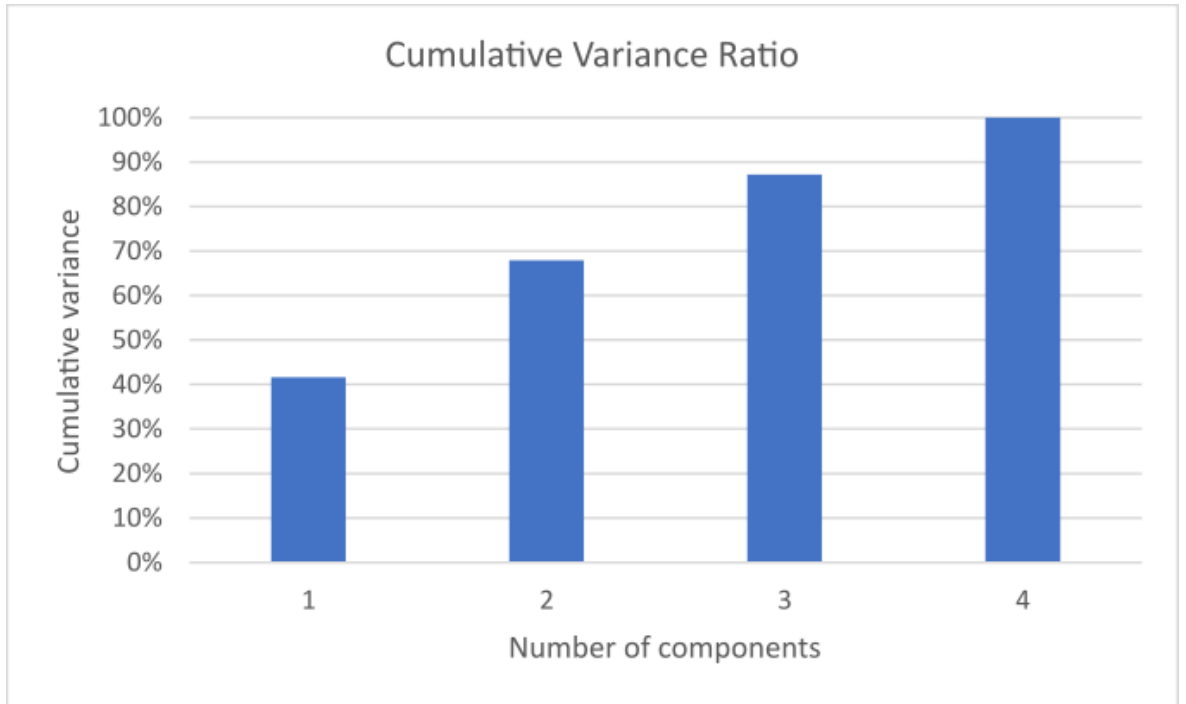


Figure 2.A.2: Cumulative variance ratio of each component for patient 1 of dataset 1.

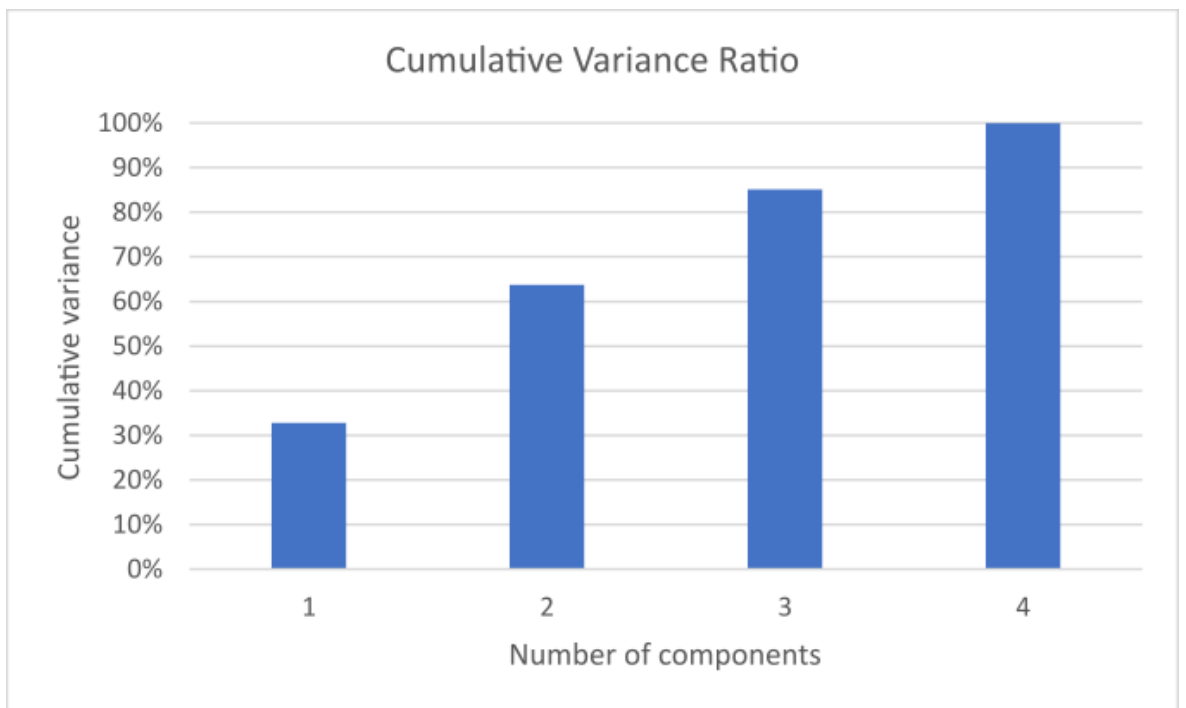


Figure 2.A.3: Cumulative variance ratio of each component for patient 1 of dataset 2.

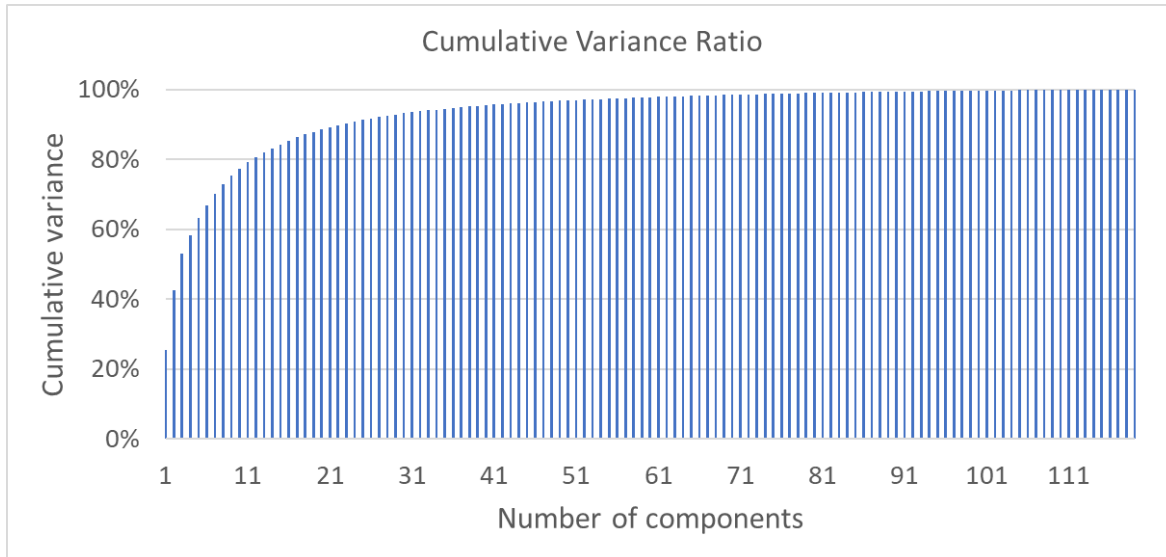


Figure 2.A.4: Cumulative variance ratio of each component for dataset 3a.

2.A.5 Increased number of input scans

As shown in Figures 2.A.5-2.A.8, increasing the number of input scans improved the model and decreased the M_{res} values, but the same region was still highlighted as posing a challenge for the models. This suggests that it is perhaps a fundamental limitation of the PCA method that is causing the difficulties in modelling the changes within the oropharynx.

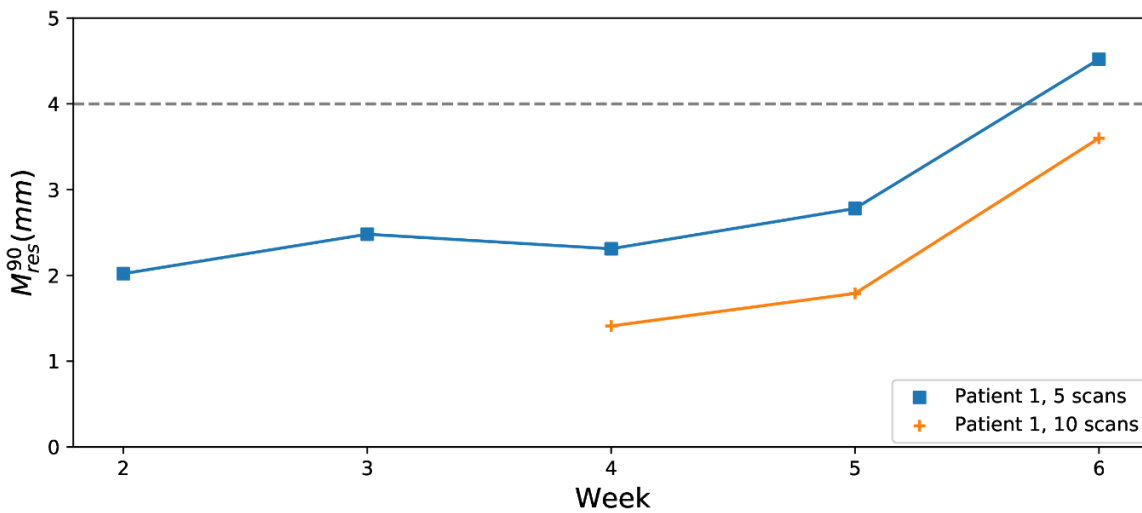


Figure 2.A.5: M_{res}^{90} calculated on the later weeks for patient 1 (dataset 2) with the model created: using the DVFs from first week (blue) and using the DVFs from first two weeks (orange).

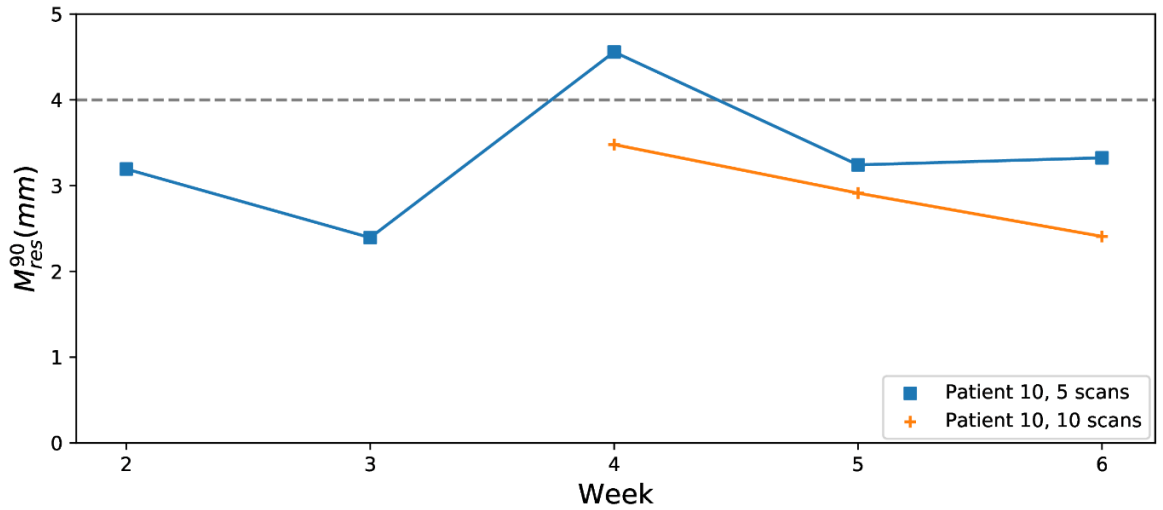


Figure 2.A.6: M_{res}^{90} calculated on the later weeks for patient 10 (dataset 2) with the model created: using the DVFs from first week (blue) and using the DVFs from first two weeks (orange).

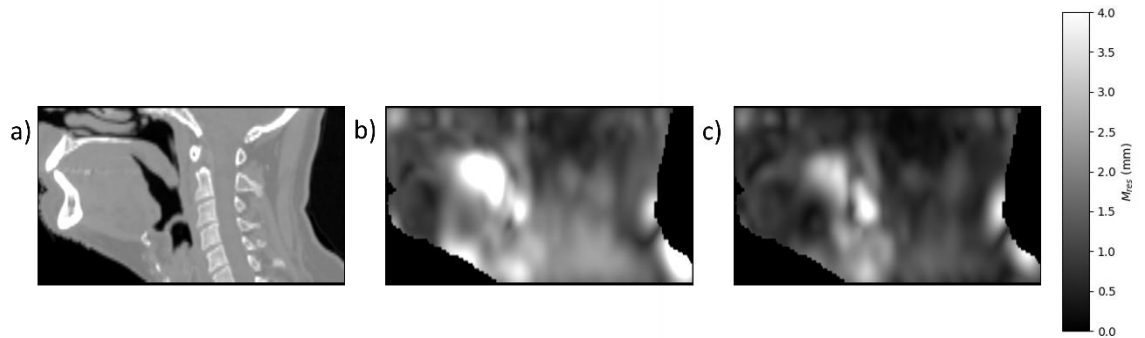


Figure 2.A.7: Identifying areas with high M_{res} for patient 1 (dataset 2). The panels show a) the pCT b) the mean of M_{res} from the later weeks with model created with the DVFs from first week c) the mean of M_{res} from the later weeks with model created with the DVFs from first two weeks.

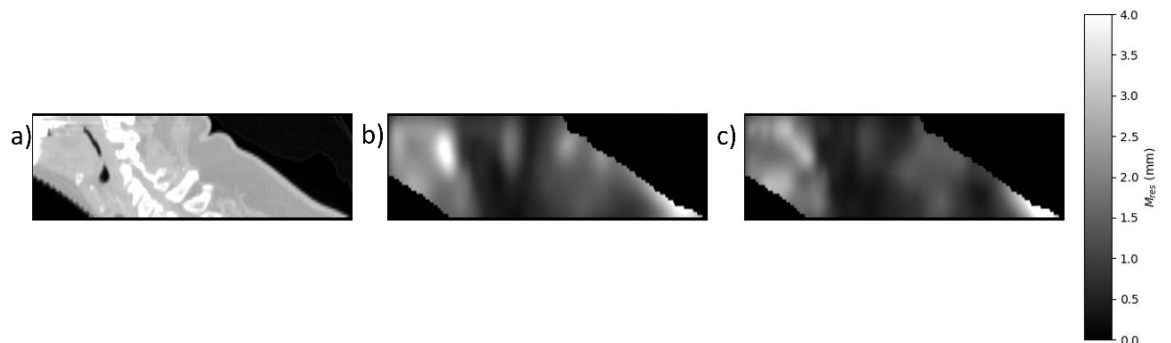


Figure 2.A.8: Identifying areas with high M_{res} for patient 10 (dataset 2). The panels show a) the pCT b) the mean of M_{res} from the later weeks with model created with the DVFs from first week c) the mean of M_{res} from the later weeks with model created with the DVFs from first two weeks.

2.A.6 LOOCV vs generalisability

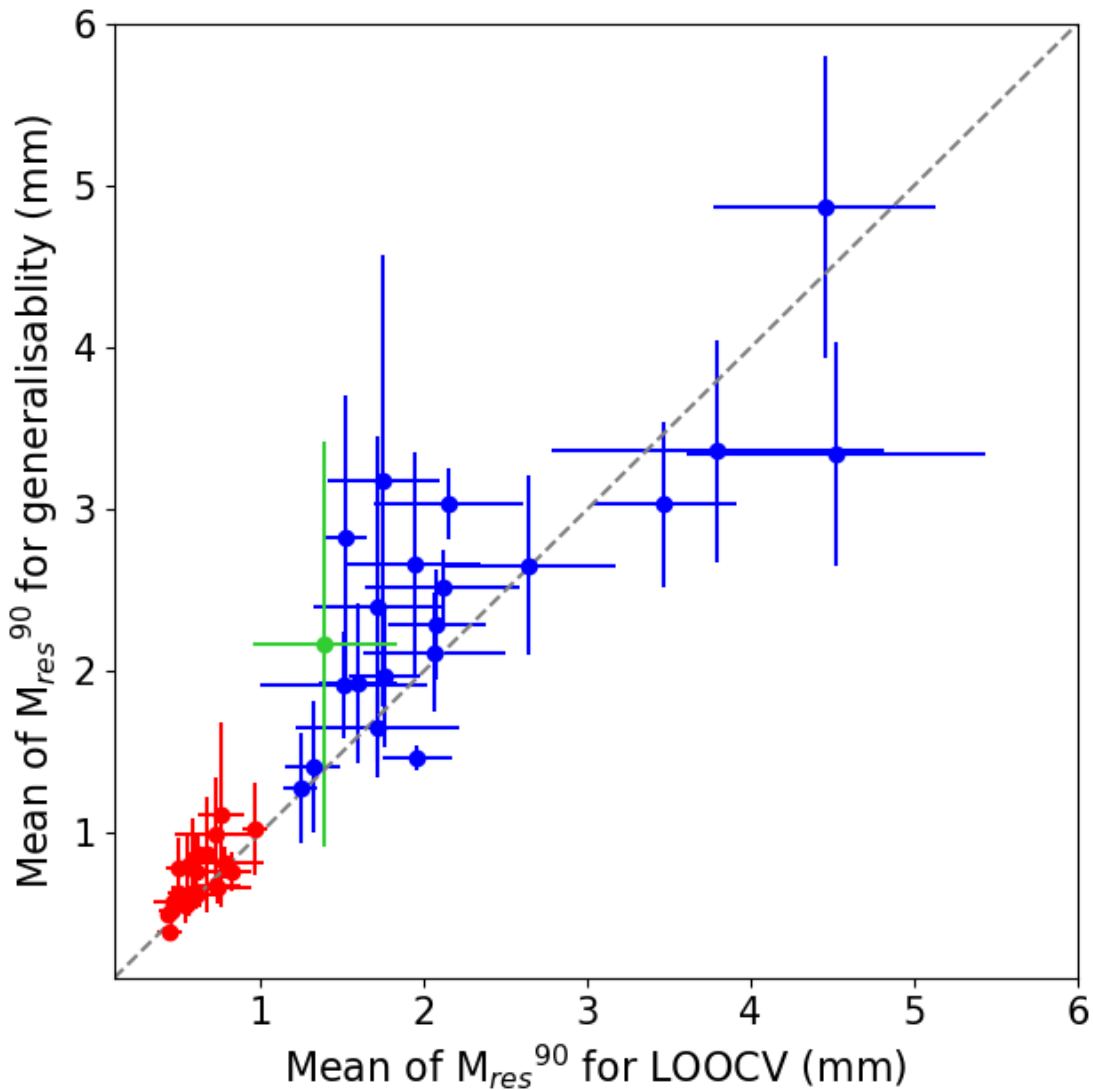


Figure 2.A.9: Scatter plot of the mean M_{res}^{90} for LOOCV (with bars extending ± 1 SD) with the mean M_{res}^{90} for generalisability (with bars extending ± 1 SD) for dataset 1 (red), dataset 2 (blue) and dataset 3a (green).

2.A.7 Validation patient 2

For the population-based model, the validation patients performed worse in the evaluation than the training patients, likely due to systematic differences between the patients. For example, the pCT of validation patient 2 (which had some of the highest M_{res}^{90} values) shows that this patient was particularly large compared to the average pCT (see Figure 2.A.10 below). This suggests that this kind of model works best on patients that are similar to the average patient geometry.

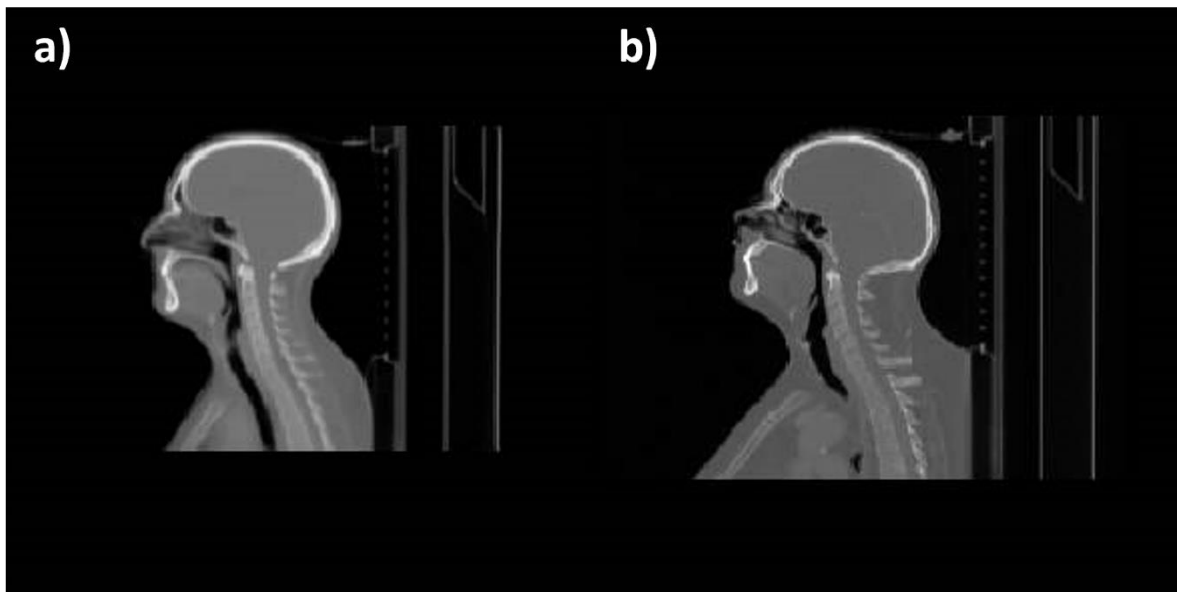


Figure 2.A.10: a) the average pCT of the training patients and b) the pCT of validation patient 2 which showed consistently high M_{res}^{90} values for all scans evaluated against.

Blank page

Chapter 3

Probabilistic evaluation of plan quality for time-dependent anatomical deformations in head and neck cancer patients

Authors

Jennifer Robbins¹, Marcel van Herk¹, Björn Eiben^{2,3}, Andrew Green⁴ and Eliana Vásquez Osorio¹

Affiliations

1. Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
2. Joint Department of Physics, Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, UK
3. Centre for Medical Image Computing, Radiotherapy Image Computing Group, Department of Medical Physics and Biomedical Engineering University College London, London, UK
4. European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom

Author contributions

I designed the experiments together with the co-authors. I pre-processed all the data, including patient selection and prepping the CBCTs for deformable registrations. With help from EVO, we performed and checked all the deformable registrations. I wrote the python code used to create the anatomical deformation models. I wrote and optimised the lua code used to run treatment simulations. I performed all the analysis and produced all the figures. I wrote the manuscript which was reviewed by all the co-authors.

Abstract

Background: In addition to patient set-up uncertainties, anatomical deformations, e.g., weight loss, lead to time-dependent differences between the planned and delivered dose in a radiotherapy course that currently cannot easily be predicted. The aim of this study was to create time-varying prediction models to describe both systematic and random anatomical deformations in head and neck (H&N) cancer.

Methods: Weekly population-based principal component analysis models were generated from on-treatment cone-beam CT scans (CBCTs) of 30 H&N cancer patients, with additional data of 35 patients used as a validation cohort. We simulated treatment courses accounting for a) anatomical deformations, b) set-up uncertainties and c) a combination of both. The dosimetric effects of the simulated deformations were compared to a direct dose accumulation based on deformable registration of the CBCT data.

Results: Set-up uncertainties were seen to have a larger effect on the organ at risk (OAR) doses than anatomical deformations for all organs except the larynx and the primary CTV. Distributions from simulation results were in good agreement with those of the accumulated dose.

Conclusions: We present a novel method of modelling time-varying anatomical deformations in H&N cancer. The effect on the OAR doses from these deformations are smaller than the effect of set-up uncertainties for most OARs. These models can, for instance, be used to predict which patients could benefit from adaptive radiotherapy, prior to commencing treatment.

3.1 Introduction

For head and neck (H&N) cancer, radiotherapy is an important treatment modality with around 80% of patients receiving it [4]. Radiotherapy plans are commonly optimised on the patient anatomy determined in a single planning CT scan (pCT). The pCT, however, only shows a snapshot of the patient anatomy at one point in time. Motion and deformation of the patient's internal anatomy and changes in patient set-up positioning will cause uncertainties in the absorbed dose [109]–[111]. In particular for H&N cancer patients, changes in weight are often observed [8]. Moreover, there are many reports of parotid

gland shrinkage [9]–[17], which is more pronounced in regions treated with high doses [18]. Regarding set-up, the literature reports flexion of the neck and rotations and translations of the patient under the mask as the main sources of variation [19]. These changes can potentially lead to overdosing organs at risk (OARs) or underdosing the target [10], [13], [22], [23].

Various mathematical techniques have been proposed to model and account for anatomical deformations [63], [65]–[70], [112]–[114], with principal component analysis (PCA) showing promise for H&N patients [71], [72]. With respect to set-up uncertainty, patient misalignments are commonly modelled using Gaussian distributions. However, up to now, modelling and evaluating the effect of random and systematic uncertainties in both anatomical deformations and set-up positioning in a time-varying manner has not been done using population models, i.e., to obtain models that are applicable to a new patient with only a single pCT, prior to commencing the actual treatment.

The aim of this study was to create a method to predict the impact of anatomical deformations and set-up uncertainties on the dose delivered to a patient. We built on PCA methods used previously [65], [69], [70] and in Chapter 2, to obtain a population-based statistical representation of anatomical deformations and introduced time-variation to account for time trends such as weight loss. Simulated anatomical deformations from the PCA models were combined with simulated patient misalignments to obtain a set of plausible treatment scenarios taking both the random and systematic effects of these uncertainties into consideration. The estimate of these uncertainties in the delivered dose could be used to “add error bars” in planning dose estimations. Additionally, these models have potential applications in the early prediction of which patients may benefit from adaptive radiotherapy, facilitating resource scheduling.

3.2 Materials and methods

In short, anatomical deformations were modelled at the population level using PCA on the displacement vector fields (DVF) from the registrations of weekly cone-beam CT scans (CBCTs) to the pCTs. To account for time trends, independent weekly PCA models were created, using one CBCT and its corresponding DVF per week of all training patients. Systematic and random set-up uncertainties were modelled as Gaussian distributions. For

each patient, 1000 30 fraction treatment scenarios were simulated (assuming dose-shift invariance [41]) using the clinical dose distributions, accounting for: a) only anatomical deformations, b) only set-up errors, and c) both anatomical deformations and set-up errors. Both the anatomical deformation and the set-up uncertainty simulations included a systematic component for the whole treatment, and a random component for each fraction. To test the method, we compared the distributions of simulated and observed cumulative doses in the primary CTV and the OARs used for planning. The following sections expand this approach and the data used in our study.

3.2.1 Patient data

Variable	Value	Number of training patients	Number of validation patients
T stage	1	7	8
	2	17	17
	3	4	4
	4	1	6
	unknown	1	0
N stage	0	7	9
	1	3	3
	2	19	21
	3	0	2
	unknown	1	0
Sex	Male	23	25
	Female	7	10
Tumour size	<100cc	18	15
	100-200cc	10	17
	>200cc	2	3

Table 3.1: Patient characteristics

Data (including the pCT, on-treatment CBCTs, dose distributions and delineations) from 65 unilateral oropharyngeal patients treated at our institution were used for this study. These data were collected from the ukCAT distributed learning database (ethics approval from the UK North West - Haydock Research Ethics Committee, reference number 17/NW/0060,

local consent ref. 2018-018). All patients had at least one CBCT for each of six weeks of treatment and the field of view of each of the scans included the whole brainstem and extended down to at least the top of the shoulders. All patients were treated in 30 fractions using VMAT. The first CBCT was taken on the first day of treatment and this was considered as day 0. The CBCTs were then split into treatment weeks. Using the available CBCTs, there were 30 patients with CBCTs taken within one day of the second day of each week (i.e., day 1, 8, 15 etc. ± 1 day). This ensured that the selected training CBCTs were taken at the same point in treatment for each patient and as close to seven days apart as possible within this cohort. These 30 patients were selected as training patients, using the weekly CBCTs for generating the PCA models. The remaining 35 patients were used as a validation cohort. The unused CBCTs from the training patients (318 CBCTs) and all CBCTs from the validation patient dataset (466 CBCTs) were used for model evaluation purposes. A summary of patient characteristics is shown in Table 3.1.

3.2.2 PCA-based deformation models

We used PCA to generate deformation models based on the 30 training patients. The general process is illustrated in Figure 3.1.

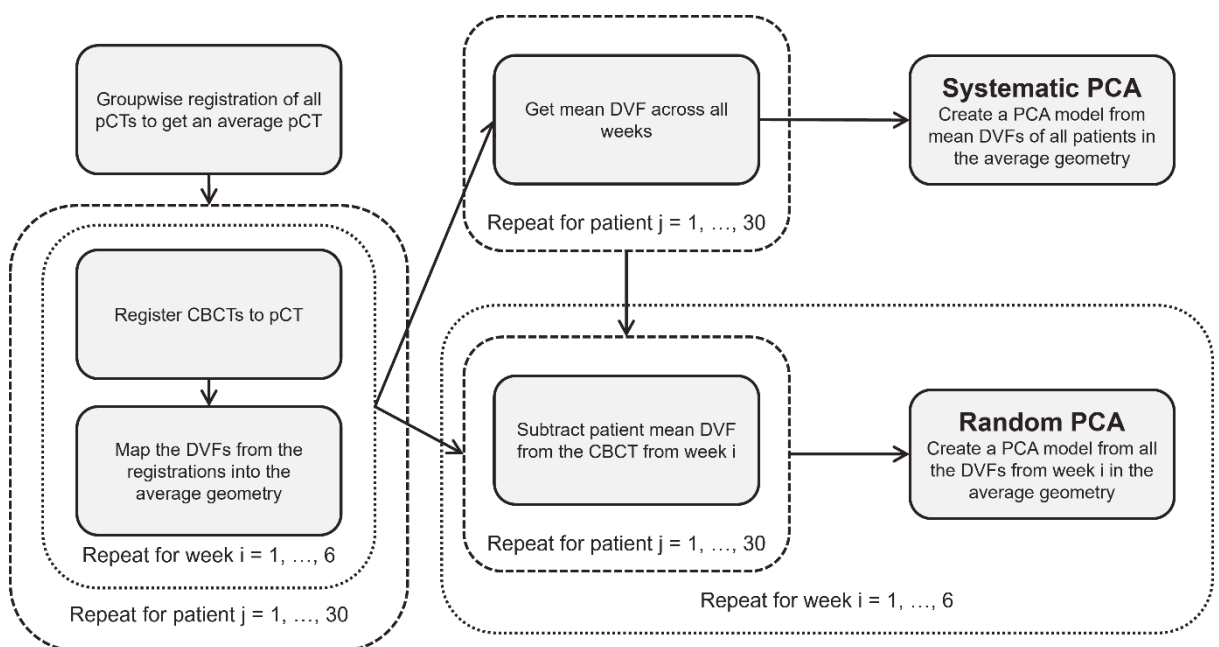


Figure 3.1: Flow chart showing the process for creating the PCA models for systematic and random deformations. First, registrations were performed to get the weekly DVFs from each patient in the average geometry. Patient mean DVFs (across the six treatment weeks) were used as the input for the systematic PCA and the weekly DVFs with the patient mean subtracted were used as input for the random PCA models.

3.2.2.1 Pre-processing

All pCT scans were centred using the centre of mass of the brainstem and, to reduce bias due to target laterality, patients were mirrored in the left-right direction as needed to ensure that the centre of mass of the primary tumour was on the right side for all patients. Next, a shading correction was applied to all CBCTs [115] to improve their quality. For each patient, all CBCTs were then rigidly registered to the pCT based on the bony anatomy in the skull, and the top of the skull from the pCT was patched onto the CBCT for the region outside of the field of view of the CBCT. This registered scan was then used as the input for deformable registration.

3.2.2.2 Average geometry using group-wise registration

To model anatomical deformations across a population of patients, the geometry of each patient must be brought to a common reference. To accomplish this, a group-wise registration of the pCTs of all the training patients was performed using an arbitrary patient's pCT (in our case, patient 1) as the reference geometry. We performed 2 rigid registration iterations followed by 20 non-rigid registration iterations using NiftyReg [61], [105] with a bending energy parameter of 0.0005. After each iteration the reference anatomy was replaced by the newly calculated group average, thus reducing the effect of selecting an initial reference image to a minimum. This produced the average patient geometry with the relevant non-rigid transformations for each individual patient's pCT to this geometry. The pCTs of the validation patients were then non-rigidly registered to this same average patient geometry, also using NiftyReg with a bending energy parameter of 0.0005.

3.2.2.3 Generate DVFs from CBCTs

To describe the deformations between the CBCT and the pCT, non-rigid registrations were performed using NiftyReg [61]. Due to the limited field of view often present in the CBCTs, the registrations were performed with the CBCT as the reference image. Full details of the registration parameters used are shown in Supplement 3.A.1. The resulting DVF was then inverted to get the deformations from the CBCT to the pCT as required for model building. Each DVF was then mapped into the average patient geometry via a resampling filter using SimpleITK [106]. This step brought all DVFs into a common frame of reference, allowing voxel-wise processing.

All DVFs were then cropped to a bounding box, covering an area including the brainstem, parotids, oral cavity, larynx and primary CTV with a 1.5 cm margin in each direction, to reduce computation time.

These DVFs were in the direction from the pCT to the CBCT. We used a direct sampling approach [116] where each point inside an organ was transformed by the vector (point + vector = transformed point). This meant that the points inside the pCT geometry could be transformed into the CBCT (or simulated) geometry, and then the pCT dose per fraction to each point could be accumulated across all fractions in the treatment simulations (see Section 3.2.3 below).

3.2.2.4 PCA model creation

PCA takes a dataset with n data-points, each with N dimensions (that can be correlated with each other), and reduces it down to a smaller set of principal components that are orthogonal and uncorrelated [73]. In our case, we had $N=3U$ dimensions where 3 corresponds to the number of spatial dimensions (x , y , and z) and U was the number of voxels present in the DVFs in the average patient geometry, cropped to the bounding box (583,968 voxels in our case). n is the number of scans input to the model which in our case was 30 for each model.

Using the DVFs obtained from the registrations and mapping of the weekly CBCTs, we created seven PCA models, one for systematic deformations and six (one per week) for random deformations. The systematic PCA model was created using the mean of the six weekly DVFs of each patient as the inputs. The random PCA models were created by subtracting the mean DVF for each patient from their weekly DVFs, using these differences as inputs to build the models.

As $N \gg n$, the maximum number of components generated by PCA is $n-1$ [65], therefore we had 29 available components for each model. Each PCA model contains eigenvectors of the 29 principal components, e_l , their corresponding variances, λ_l , and the mean of the input DVFs, \bar{v} . We used the first L components that covered the largest 90% of the total variance from each model.

3.2.2.5 Creating simulated deformations

For each model, a set of simulated deformations were created. Each simulated DVF, \mathbf{v}_{sim} , was created using the equation:

$$\mathbf{v}_{sim} = \bar{\mathbf{v}} + \sum_{l=1}^L u_l \mathbf{e}_l, \quad (3.1)$$

where u_l is a scaling factor for the l^{th} component, randomly drawn from a Gaussian distribution with a variance equal to λ_l . The overall simulated DVF for a given fraction was created by first simulating a DVF from the systematic model and then one from the relevant random weekly model. The sum of these two DVFs provided the total deformations for that fraction.

3.2.2.6 PCA model evaluation

The PCA models were evaluated against DVFs derived from the CBCTs that were not used in model creation following the scheme detailed in Chapter 2. In short, given a reference DVF, the closest DVF that the model can represent was generated. The differences between the reference DVF and the closest DVF indicated the residual deformations that the model had not been able to replicate.

For the systematic PCA model, for each patient, the mean of all DVFs available for that patient was used as the reference DVF. This total mean for each patient was then subtracted from each of the DVFs for that patient. For each of the weekly random PCA models, each DVF from CBCTs taken within the same week as the model with the patient mean subtracted was used as the reference DVF.

Heat maps of the magnitude of the residual deformations for each voxel, M_{res} , were plotted to indicate the size of the local residual error for the PCA models at different points in the average patient geometry. The 90th percentile of M_{res} , M_{res}^{90} , was calculated to give an estimate of the global residual error for each DVF. The Z-score for each component of each DVF the model evaluated against was calculated.

An investigation into the effect of training the PCA models on a different number of patients is presented in Supplement 3.A.2.

3.2.3 Treatment simulations

For treatment simulations, we started with the planned dose distribution for the patient and mapped it into the average patient geometry. For the sake of simplicity, we assumed dose-shift invariance meaning the dose distribution was not recalculated for each new geometry. This assumption had been shown to work well [43], [117]. The dose distribution was then extended slightly outside the patient's body to improve dose-shift invariance [41], [48]. The fractional dose was calculated by dividing the planned dose distribution by the total number of fractions. Three different sets of simulations were carried out, accounting for a) only anatomical deformations, b) only set-up uncertainties, and c) a combination of both.

3.2.3.1 Anatomical deformations

To simulate one treatment, we produced one systematic DVF and 30 random DVFs (five DVFs from each of the six weekly models) using Equation 3.1. Each fraction was simulated using the combination of the same systematic and a different random DVF. This was repeated for each fraction of the treatment. Because all calculations were performed in the average anatomy, this allowed the same set of simulated DVFs to be used for all patients, to provide a 'paired' comparison.

3.2.3.2 Set-up uncertainties

Set-up uncertainties were also split into systematic uncertainties which are constant throughout a treatment and random uncertainties which are different for each fraction. These values were drawn from Gaussian distributions with standard deviations of Σ for systematic and σ for random uncertainties. Note that Σ and σ differ for the x , y , and z directions. For each fraction, the OAR was translated according to the sum of the sampled systematic and random errors. Similar to the deformations, the same systematic and random errors were used for each patient.

In order to obtain the population mean, and the standard deviations for the systematic and random set-up uncertainties for the dataset (\mathbf{M} , Σ and σ respectively), we followed the standard methodology described in [118]. The set-up error of each CBCT was calculated as the difference in the centre of mass position between the PTV in the pCT with the couch corrections applied and the pCT adjusted by the rigid registrations (the starting point for

deformable registration). To account for missing CBCTs, the set-up value from the CBCT closest in time was taken (nearest neighbour approach). M , Σ and σ were then calculated as in [118] using the training patients. The set-up uncertainties obtained in the training patients were also used for the validation patients.

3.2.3.3 Anatomical deformations and set-up uncertainties:

For simulations combining both anatomical deformations and set-up uncertainties, the same set of DVFs and set-up translations were used as when simulating treatments considering these uncertainties individually. For each fraction, the simulated DVF was first applied to each organ and then they were translated by the set-up error.

For all sets of simulations (considering anatomical deformations, set-up uncertainties and a combination of both), the dose to each voxel in the translated/deformed OARs was accumulated across all fractions to produce a simulated dose distribution for the treatment. Next, relevant dose-volume histogram (DVH) parameters for the OARs were computed from each distribution. This process was repeated 1000 times per patient. We calculated the difference between the planned and the simulated dose for the minimum dose to the primary CTV and for the DVH parameters used in planning the OARs at our institution (maximum dose to spinal cord and brainstem, mean dose to larynx, oral cavity and each of the parotids). Simulations were performed on both the training and the validation patient datasets.

3.2.3.4 Dose accumulation of actual treatment

A dose accumulation for the actual treatment was calculated using the DVFs and set-up errors using all actual CBCTs for each patient. If there was no CBCT for a given fraction, the nearest CBCT in the same week was used.

3.2.3.5 Volume change

To see how much the OAR volumes changed throughout treatment, the contours were first mapped into the average patient geometry and then warped by a DVF. The volume of the warped contour was then compared to the planned contour in the average geometry. This was done for both the DVFs coming from the CBCTs from the patients and DVFs from 100 simulated treatments produced by the model. The results from this are presented in Supplement 3.A.3.

3.3 Results

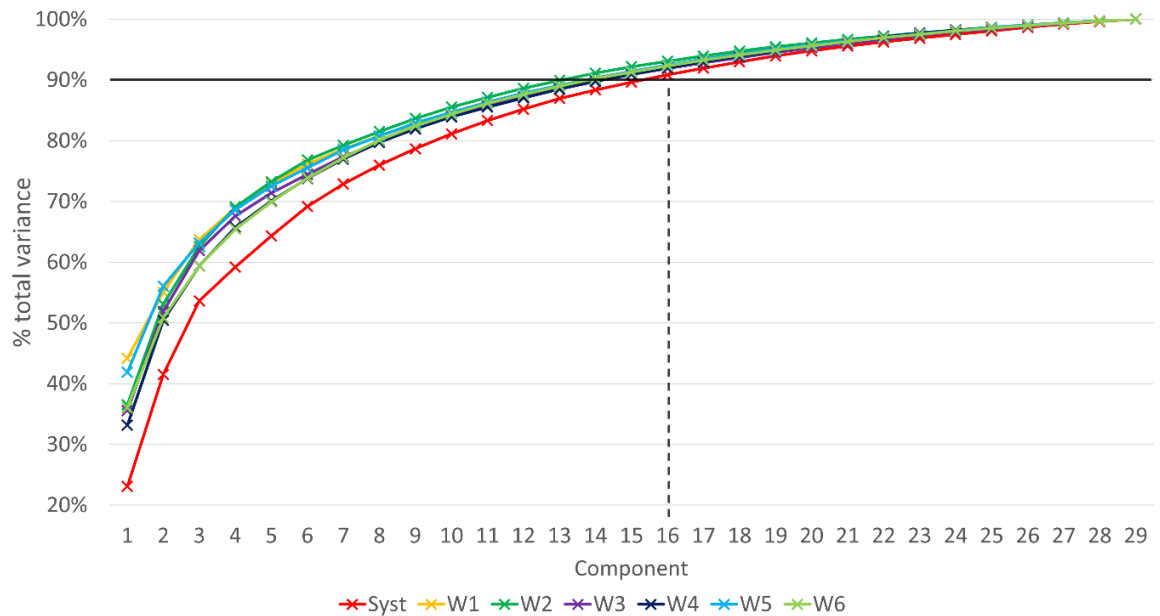


Figure 3.2: Cumulative contributions of the different components to the total variance for each weekly PCA model. The different PCA models are shown in different colours. The solid grey line indicates 90% of the total variance and the dashed grey line shows that all PCA models reach at least 90% the total variance by component 16.

3.3.1 PCA-based deformation models

The cumulative normalised variance for each PCA component in each of the models is shown in Figure 3.2. By component 16, around 90% of the total variation was included in all models.

3.3.1.1 Model evaluation

The model evaluation (Figure 3.3) shows that each of the weekly models had similar residual errors (both local and global). For the systematic model however, the training patients had lower residual errors than the weekly models while the validation patients had much higher residual errors. For all models, the region of the largest local residual error was around the oropharynx. Across all patients and models, the mean M_{res}^{90} was 2.2 mm.

For all models, except the systematic model when evaluated against the validation patients, the percentage of the Z-score values below 3 is ~ 99.7 (Table 3.2), suggesting that the DVFs the models were evaluated against are likely to be produced by the model. For the systematic model when evaluated against the validation patients, only 6% of the Z-scores were below 3, meaning it is unlikely these DVFs would be produced from the model.

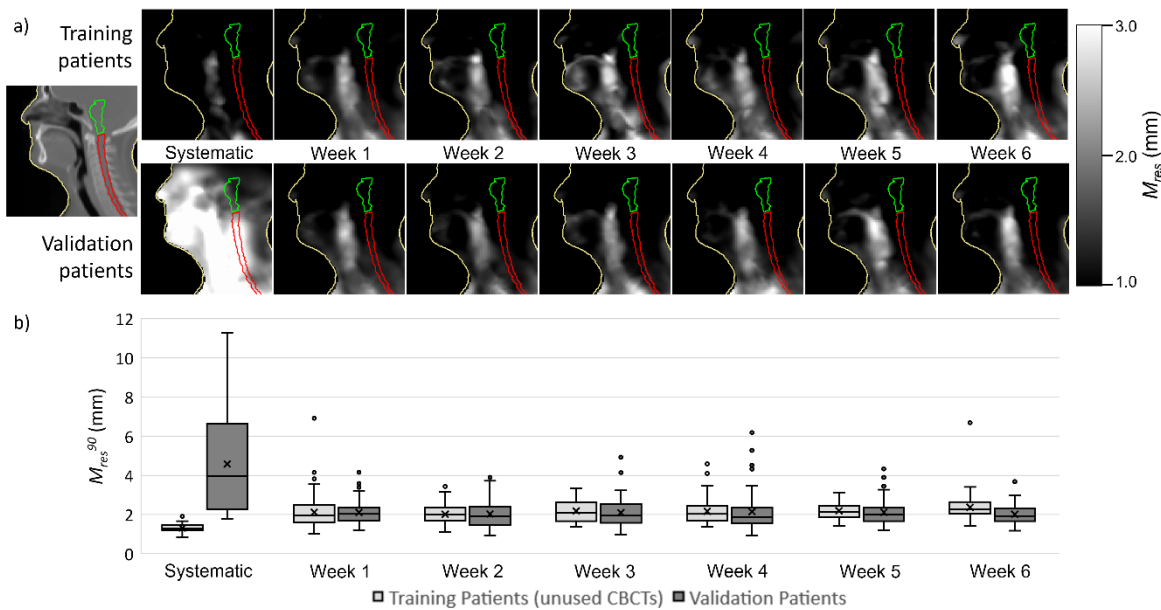


Figure 3.3: a) Average pCT and the mean of the M_{res} maps across all DVFs evaluated against for each of the models for the training and validation patients. The external contour and the spinal cord and brainstem are included on each plot to provide anatomical context. b) Box plots of M_{res}^{90} for each DVF used to evaluate the models against for training and validation patients.

	Systematic	W1	W2	W3	W4	W5	W6
Training	100	99.3	99.7	99.5	99.7	99.7	99.9
Validation	6.0	99.8	99.4	98.7	99.7	99.7	100

Table 3.2: The percentage of all the Z scores calculated for each model with values less than 3.

3.3.2 Set-up uncertainties

Table 3.3 shows the M , Σ and σ values calculated from the difference in the centre of mass position of the PTV for each of the CBCTs from the 30 training patients.

	Lateral	Longitudinal	Vertical
M	-0.02	0.02	-0.07
Σ	0.11	0.12	0.12
σ	0.16	0.14	0.18

Table 3.3: M , Σ and σ values from the 30 training patients.

3.3.3 Treatment simulations

Figure 3.4 shows the distribution of difference between each of the 1000 simulated treatment doses and the planned dose for all patients in the different error scenarios. For the brainstem, there was a median increase in dose for all scenarios considered, the largest being when both uncertainties were simulated together, with a median increase of 1.3 Gy

for the training patients and 1.1 Gy for the validation patients. The median dose change for the other organs were all <0.5 Gy for all scenarios except the CTV which had a median decrease to the minimum dose of 0.6 Gy when both uncertainties were considered.

Considering the standard deviation in the DVH parameters studied, for the minimum dose to the CTV and the mean dose to the larynx, the spread of results was larger when just anatomical deformations were considered than just set-up uncertainties. The standard deviation when both uncertainties were considered together was larger than just deformations for the CTV, but similar for the larynx. For all other organs considered, the standard deviations of the simulated results were similar when both uncertainties were considered together and when just set-up uncertainties were simulated, and these were larger than when just anatomical deformations were considered. For just anatomical deformations, the standard deviations were largest for the CTV (3.1 Gy and 2.9 Gy for training and validation patients respectively) and smallest for the oral cavity (0.4 Gy for both training and validation patients). For set-up uncertainties, the standard deviations were largest for the CTV (2.1 Gy and 2.4 Gy for training and validation patients respectively) and smallest for the larynx (0.8 Gy for both sets of patients). Across all scenarios, the standard deviation for the ipsilateral parotid was larger than the contralateral parotid by ~ 0.4 Gy.

For all scenarios, the 'real' values from the dose show a similar distribution to the simulated values for each DVH parameter, although the range is a bit smaller for most OARs. For both the brainstem and spinal cord in the validation patients, the median of the 'real' treatment values is lower than that of the simulated values when set-up uncertainties and a combination of both uncertainties were taken into consideration. Of course, it should be noted that the 'real' dose accumulated results contain one value per patient for each DVH parameter, whereas the simulated results each contain 1000 values per patient. For all organs and scenarios, the results for the training patients were similar to the validation patients.

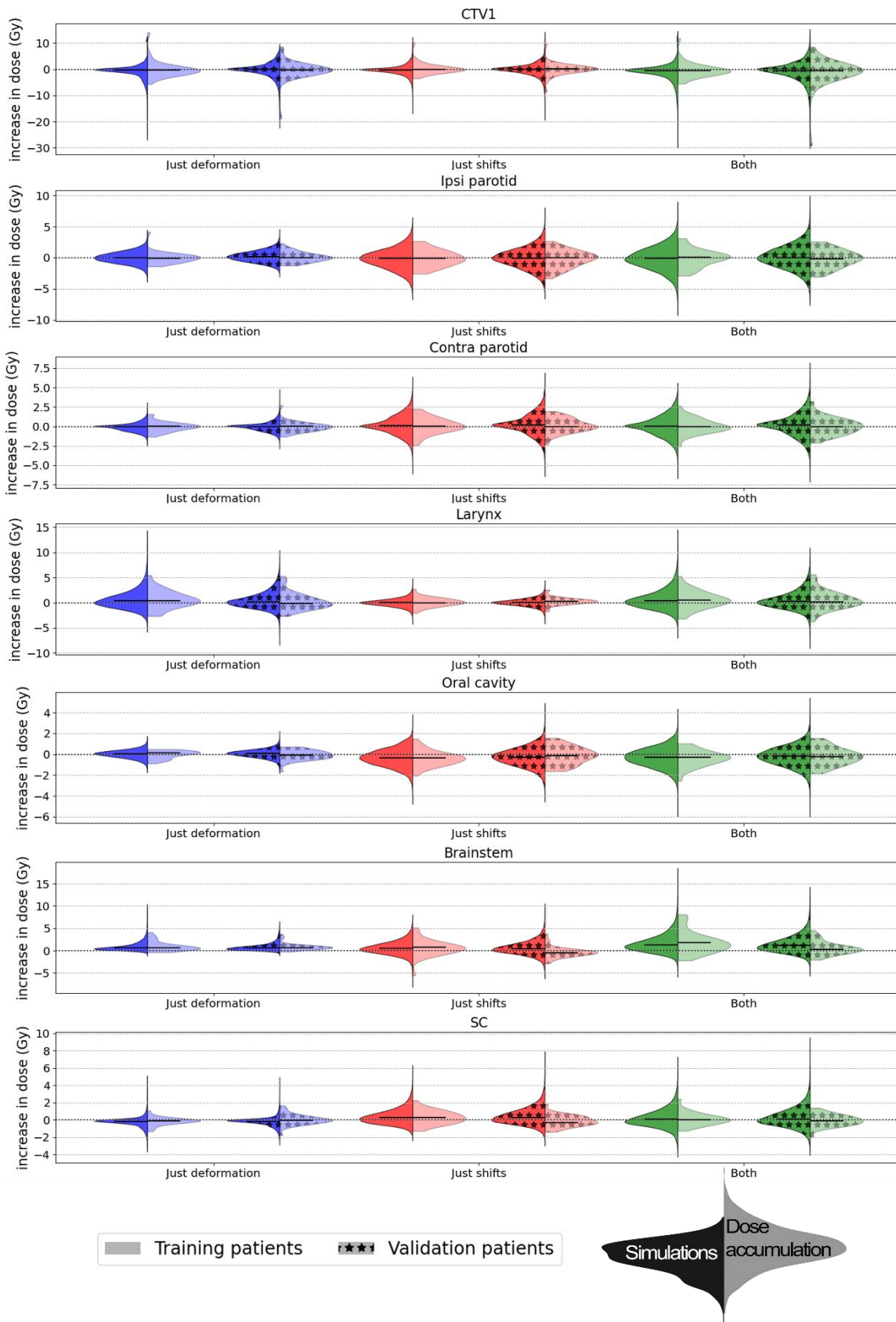


Figure 3.4: Violin plots showing the change in dose between the simulated treatments for all patients and the planned dose distribution for simulations considering just anatomical deformations (blue), just set-up uncertainties (red) and a combination of anatomical deformations and set-up uncertainties (green). The darker half (left) shows the simulated results and the lighter (right) the dose accumulation from actual CBCTs. The median is represented by a black line on each violin.

3.4 Discussion

We have proposed a novel method for modelling time-varying anatomical deformations, based on weekly population-based PCA models. Our methodology explicitly splits the deformations into systematic and weekly random variations. The effect of these deformations on the OAR dose was evaluated in combination with set-up uncertainties in the training patients as well as in unseen validation patients.

We found that for most organs used for planning (except the larynx and the primary CTV), the effect of anatomical deformations was smaller than set-up uncertainties, based on the off-line protocol in use at the time our data were collected. We expect that with modern treatment techniques where daily image guidance is implemented, patient set-up uncertainty will decrease. Consequently, uncertainties due to anatomical deformations will become more important.

The main novelty of our work is the inclusion of time-variation, allowing us to model weight loss and organ shrinkage. To generate the weekly models, we used just one CBCT per patient per week. We did, however, achieve very good performance. Our results agree with similar work done by Badawi *et al.* [68] who looked at comparing a PCA model using just one scan a week compared with all three scans they had available for each week in lung tumours. They concluded that using just one scan per week had a minimal impact on the residual testing error. Their method was different from ours in the fact that they created just a single PCA model that included all time points, whereas we separated ours out into weekly models to be able to account for time-varying effects. Also, their work was done using individual patient models, whereas ours were population-based.

Our work explicitly incorporated both systematic and random deformations separately. This is an important distinction as systematic variations originate from the arbitrary deformations ‘frozen’ in the pCT that affects the whole treatment course, because the pCT is used to plan the treatment. This separating of the systematic and random variations has been done before for prostate and lung patients [70], [100], but in those studies the random variations across the whole treatment course were modelled together and so time-variation was not considered.

We used PCA to create our models, and this method assumes that anatomical deformations are normally distributed around a non-zero mean position. This is a commonly used assumption in the literature for producing synthetic anatomical deformations [66], [67]. Our method required the displacements for each voxel within the patient's body as an input for the PCA models and this method for PCA has been used a number of times previously [67], [74], [101]. Other studies use the differences in the surface shape contours of specific organs as the input to the PCA models [65], [68], [69]. We chose the entire DVF method as we were aiming at simulating the impact of anatomical deformations for dose mapping and accumulation, so displacement information for each voxel was required. Additionally, by constructing one PCA model for the region of interest, deformations of multiple organs and the surrounding normal tissue can be modelled simultaneously, and their interactions are maintained.

In our simulations, delineation uncertainties were not taken into account. It has been shown that delineation uncertainties have a dosimetric effect on OARs [119]–[121], but that this effect decreases when evaluated with set-up uncertainties [122]. A future extension of this work would be to include delineation uncertainties in our analysis and compare the effect they have when combined with both set-up uncertainties and anatomical deformations.

To derive a population-based model, all our evaluations were performed in an average patient geometry. Therefore, our methodology requires the transfer of the DVFs and a contour for each OAR to the average patient anatomy, introducing uncertainty. Registrations were performed intra-patient, to measure deformations, and inter-patient to bring the deformation information to the average geometry to create the PCA models. All registrations were carefully visually assessed, but no quantitative measure of the uncertainty was taken. Veiga *et al.* [108] found the mean distance transform between manually delineated and deformed surfaces in H&N when using NiftyReg to be 0.3 ± 0.4 mm when averaged across all patients and structures analysed. Any displacement due to imperfect intra-patient registrations would show up as patient deformations and be included in the PCA model. However, the anatomical deformations we intended to capture within the model were assumed to be larger than the registration uncertainties. For

example, Barker *et al.* [9] found the median shift of the centre of mass of the parotids to be 3.1 mm medially.

For our models, we have used the first 16 components which correspond to roughly 90% of the total variance for each model. The final 10% in the last components is most likely to be dominated by registration uncertainties or CT artefacts as opposed to the main modes of deformation present. Several studies have found that between 2-5 principal components are enough to cover 90% of the variation present [65], [66], [68], [101]. However, these studies created intra-patient models, i.e., one model per patient, rather than a population-based model. Tsiamas *et al.* [71] compared models for individual patients with a population-based model and found that more components were necessary to cover the same percentage of variations for the population-based model as opposed to individual patient models. This is because each patient will experience different variations, meaning there is more variation to capture in population-based model and the overall variations will be more complex so more components are necessary. Budiarto *et al.* [69] produced a population-based PCA model for prostate and found that 15 principal components covered 90% of the variation, which is similar the 16 required in our case. To reach 95% of the total variance for their population-based prostate PCA models, Vile [70] found that 11 components were needed for the systematic deformations and 33 for the random deformations. However, their random model included a total of 210 input scans, significantly more than the 30 used in our study.

For the random weekly models, we found that the residual errors were similar between weeks and for both training and validation patients. The residual errors were smaller for the systematic model when evaluated against the training patients, but larger when considering the validation patients. This same effect was seen by Vile [70] for their PCA model created on a population of prostate patients. This is expected as the systematic errors remain the same throughout treatment and so the model will have captured some of the specific systematic changes for the training patients. The validation patients however will have slightly different systematic errors that were not captured by the model. This is likely due to differences between each patient and the 'average patient'. For example, validation patient 2 was larger than the average patient and had a particularly high M_{res}^{90} when evaluating the systematic DVF against the systematic model (Supplement 3.A.4).

Changing the number of training patients from 10 through to 30 (Supplement 3.A.2) caused the residual errors to decrease for the validation patients, especially for the systematic model. This suggests that perhaps increasing the number of training patients beyond 30 could reduce these systematic residual errors even more.

The region that had the largest residual errors in all our models was the oropharyngeal region. This could be due to the fact that this region experiences a large amount of random variations, even within a single patient, and so it is difficult for the models to accurately capture the full range of changes present. It could also be because this is the region that contains the primary CTV and is the high dose region for this set of patients. The exact size, shape and location of the primary CTV will vary between patients, as will the exact response of the tumour. This means that it can be difficult to capture all these possible changes in a single population-based model.

Across all simulations, the results for the validation patients were similar to those of the training patients and both were generally in good agreement with the dose accumulated from the DVFs from the actual CBCTs. For most DVH parameters, the range of the simulated results was larger than that of the dose accumulated results, although this is likely due to the larger number of samples present in the simulated results – 1000 treatment courses were simulated for each patient, whereas there will only be one dose accumulated result for each patient. This means that the models we have created can be used to simulate deformations on completely new patients without requiring any additional information. I.e., these models can be used in the planning process, without having to wait for multiple scans of the patient – the model can be used for new patients immediately.

Our simulations gave a mean volume change (Supplement 3.A.3) of 13.9% for the ipsilateral parotid for the training patients dataset by the end of treatment which is less than the average of $26 \pm 11\%$ found in the literature [55]. This could mean that our models were not picking up the full extent of the shrinking of the parotids, possibly due to the fact that different patients will experience different amounts of shrinkage or that we had a small sample size (30 patients). For most OARs, an increase in volume was seen in both the real and simulated DVFs between the pCT and week 1. This is likely due to the difference in quality of the CBCT and the pCT, making it challenging for the registration algorithm to perfectly match the soft tissue boundaries of some of the OARs. This can lead to the DVFs

seeming to produce an overall increase in volume from the pCT that is not actually there. This could also explain why the mean volume change of the parotids was smaller than that seen in literature.

Our method of modelling time-dependent anatomical deformations and running treatment simulations could potentially be used to prospectively evaluate treatment plans and identify the cases where anatomical deformations may have a large effect on the delivered dose. This can help in an adaptive radiotherapy programme, by predicting at the start of treatment whether adaptation may be required at some point during treatment. This ability at an early stage to predict the need for adaptation and replanning can help with resource allocation, e.g., allocating time to replan the treatment or time on the scanners.

This framework can also be used to add error bars to treatment planning. This could be useful to help focus the attention of the planners on areas where deformations could have a large impact on the delivered dose, meaning a plan can be created that is not only good in the nominal situation but also holds up under anatomic variations. Ultimately, these PCA models could also potentially be used to directly account for the effect of deformations by incorporating this uncertainty into the treatment planning process, e.g. by using probabilistic planning [99].

3.5 Conclusion

We have proposed a novel method to model time-varying anatomical deformations and have evaluated the effect of these deformations and set-up uncertainties on clinical H&N cancer plans. This method predicts plausible deformations based on a single CT scan and can, for instance, be used a priori to identify patients where adaptation may be beneficial.

3.6 Acknowledgements

Marcel van Herk was supported by NIHR Manchester Biomedical Research Centre. This work was supported by Cancer Research UK via funding to the Cancer Research Manchester Centre [C147/A25254].

Appendix 3A

3.A Supplementary material

3.A.1 Non-rigid registrations

The non-rigid registrations between the pCT and the CBCTs were performed using NiftyReg with a number of steps and different parameters. Firstly, a mask of the CBCT was created with a threshold of 100, using in-house software, so that the registrations were only focused within the patient's body. Then, the deformable registrations were performed using the Niftyreg's `reg_f3d.exe` with the pCT as the floating image, the CBCT as the reference image and the CBCT mask as the reference mask. The specified parameters used were:

```
-ln 5 -lp 4 -be 0.001 -smooR 1 -smooF 1 -lnc 20 -maxit 1000 -jl 0.0001
```

The `-ln 5 -lp 4` meant that there were 5 levels to the registration but only the first 4 were used – effectively down-sampling the image. The `smooR` and `smooF` were used because in some of the CBCTs later on in the treatment, there was a gap between the skin and the mask where the patient had lost weight. This was causing challenges in the registration as it was trying to match the mask and not the skin as was required to get the patient's deformation. Adding in the smoothing helped reduce the impact of this.

The next step was to then invert the deformation field to get the deformations with the pCT as the reference. This was done using NiftyReg's `reg_transform.exe` with `-invDef` using the pCT as the target reference.

3.A.2 Different patient numbers for training

To see the effect of training the model with fewer patients, we created a set of the systematic and random weekly PCA models using a random subset of 10 and 20 training patients. This was repeated for a total of 5 times with different random subsets of patients. For each of these models, the set of M_{res}^{90} values were calculated using just the validation set of patients, and these values were plotted as box plots seen in Figure 3.A.1.

It can be seen that for all the models, increasing the number of training patients decreased the size of the residual errors. This effect was largest for the systematic PCA model.

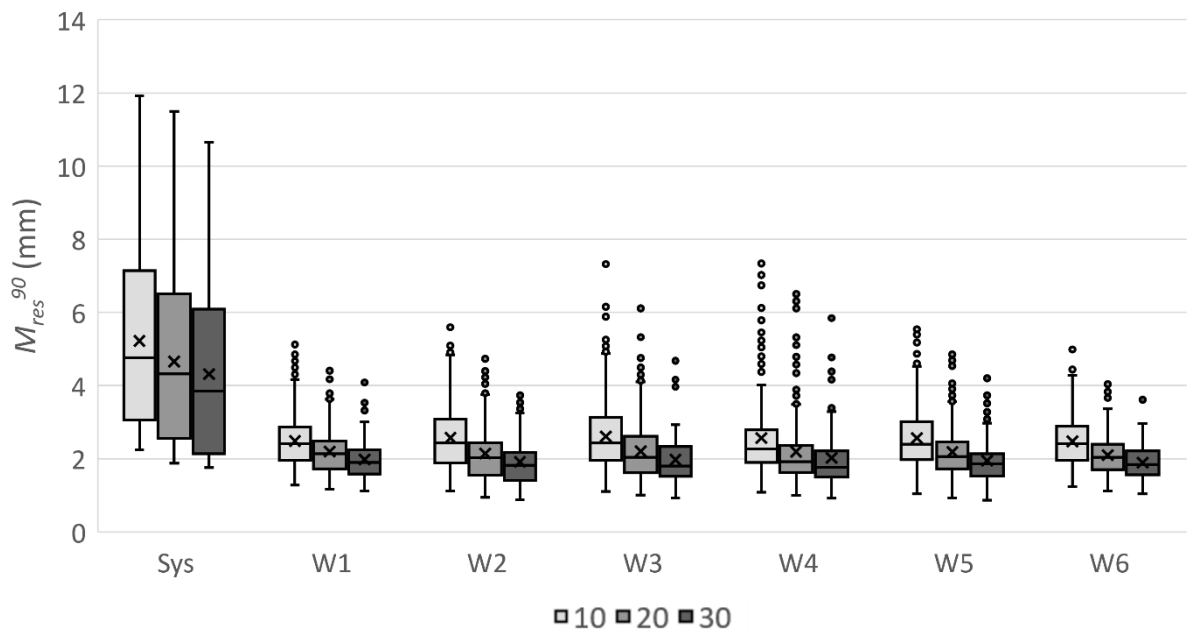


Figure 3.A.1: Box plots showing the M_{res}^{90} values for each DVF from the validation patients used to evaluate the different PCA models against 5 sets of PCA models trained on random subsets of 10 training patients, and 5 set on random subsets of 20 training patient. These are shown alongside the model trained on all 30 training patients.

3.A.3 Volume change of OARs

Figure 3.A.2 shows the simulated mean volume change in each OAR across the treatment weeks. Looking at the mean for both the real and simulated DVFs, for both parotids and the oral cavity, the volume change decreased as the weeks progress, although for the oral cavity the volume was always higher than that in the pCT. By week 6, the mean ipsilateral parotid volume had decreased by 14.1% in the CBCTs used for model creation and 13.9% in the simulated treatments compared to the volume in the pCT. For the contralateral parotid, these numbers were slightly smaller with a decrease of 13.5% for the real DVFs and 12.2% from the PCA models.

The mean volumes of the spinal cord, larynx and brainstem remained roughly similar throughout the weeks, although there was a slight increase seen in the brainstem volume. Both brainstem and spinal cord saw an increase in volume compared to the pCT and the larynx showed a decrease.

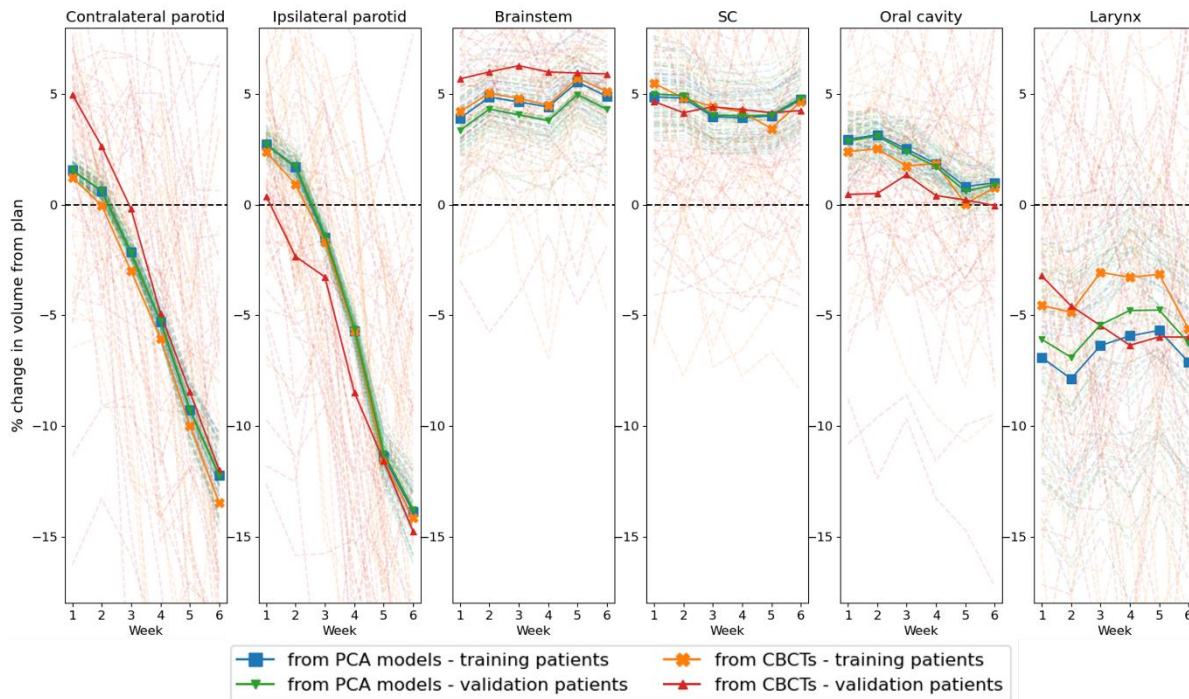


Figure 3.A.2: Volume change in OARs from the training and validation patients. Volume changes from the DVFs from the CBCTs are shown in orange (training) and red (validation) and from 100 simulated treatments from the PCA models in blue (training) and green (validation). The dotted lines show the volume change for each individual patient/simulated treatment and the darker, solid lines show the mean across all patients/treatments. Note: some of the dotted lines for individual patients/treatments extend outside the range of the axes on the plot.

3.A.4 Validation patient 2

Validation patient 2 was larger than the average patient and had an M_{res}^{90} of 7.7 mm. This patient was also used in the analysis of Chapter 2, see Figure 2.A.10.

Chapter 4

Can anatomical deformations in head and neck cancer be accounted for as set-up uncertainties?

Authors

Jennifer Robbins¹, Marcel van Herk¹, Andrew Green² and Eliana Vásquez Osorio¹

Affiliations

1. Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
2. European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom

Author contributions

I designed the experiments together with the co-authors. I took over development of the probabilistic treatment planning plugin from Ian Hinder (software engineer), which was used to create the probabilistic plans. I wrote the scripts in RayStation and created all the radiotherapy plans that were used in this paper. I wrote the lua code used to run the treatment simulations and performed all the analysis. I produced the figures and wrote the manuscript which was reviewed by the co-authors.

Abstract

Background: Anatomical deformations can cause uncertainties in the dose delivered to a patient during radiotherapy. However, tools to directly account for these variations in the planning process are lacking. The aim of this study was to see if anatomical deformations in head and neck (H&N) cancer patients can be accounted for by treating them as easier-to-model set-up uncertainties, which can be dealt with using several tools.

Methods: Margin, robust and probabilistic plans for different uncertainty settings corresponding to 1 mm, 3 mm and 5 mm margins were created for 19 H&N cancer patients. For each plan, 1000 treatment courses were simulated accounting for time-varying anatomical deformations, using a previously developed method. The 90th percentile of the D98 of the clinical target volume (CTV) was calculated for the set of simulated treatments, along with the 90th percentile of the dose metrics for each of the organs at risk (OARs) used to create the plans.

Results: The 90th percentile of the CTV D98 increased as the uncertainty settings increased for all planning approaches and was highest in the margin plans. On average, this value reached the accepted 95% of the prescribed dose for the 5 mm plans for all planning approaches and the 3 mm plans for the margin and robust approaches. OAR doses were seen to increase as the uncertainty setting increased and were also highest for the margin plans. In general, all simulations saw an increase in the 90th percentile of the OAR doses from the planned dose under deformations, but there was large variation between individual patients. Margin plans were generally more robust to deformations than robust and probabilistic plans, meaning that care should be taken when introducing such new techniques.

Conclusions: In general, plans that were robust to anatomical deformations in H&N could be produced by treating these variations as 3-5 mm set-up uncertainties. However, all simulations saw an increase in OAR dose from the planned dose, supporting the use of adaptive radiotherapy in cases where the OAR doses are close to tolerance.

4.1. Introduction

Radiotherapy is planned on a single computed tomography (CT) scan but is typically delivered in a large number of fractions spread across several weeks. This introduces uncertainties in the delivered dose as there may be changes in the patient geometry. These changes are due to differences in patient set-up positioning, organ motion within the patient and organ shape changes. As a result, a lower dose than intended could be delivered to the target, or a higher dose to surrounding organs at risk (OARs). This is a particular problem in head and neck (H&N) cancer patients as the target is often very close to OARs and there are changes to the patient's anatomy throughout treatment, e.g., weight loss, parotid shrinking and changes in neck positioning [8], [9], [11], [12], [14]–[19]. Methods have been developed to help improve the robustness of radiotherapy plans in face of these uncertainties and to help mitigate the effects they have on the dose delivered to the patient.

In particular for the target volumes, the most common method to account for geometrical uncertainties is to expand the clinical target volume (CTV) by a margin to create a planning target volume (PTV). The PTV is then used as the target for the plan optimisation, with the idea being that even if the CTV does change position, as long as it is within the PTV it will receive the intended dose. Recipes for the size of these margins have been developed [75], [76] based on the standard deviations of the systematic uncertainties (that remain the same throughout the whole treatment course) and random uncertainties (which are different for each fraction).

Besides margins, other planning approaches have been developed to account for uncertainties. One such method is robust planning where a number of discrete error scenarios are sampled and the plan is optimised based on the worst-case of the considered scenarios [83]. Another way of accounting for uncertainties is by using probabilistic planning where the plan is optimised based on the distribution of the cost function for each of a set of sampled scenarios in combination with the probability of that scenario occurring [36]. Probabilistic planning explicitly distinguishes between systematic and random uncertainties, while robust planning does not. For photon treatments, these methods are most commonly used to account for rigid translations of the target structures [32]–[37], [41].

Translations are the easiest type of uncertainty to account for because they are simple to model. However, with the improvement of techniques such as image guidance, set-up uncertainties are being reduced and therefore the uncertainties due to deformations are becoming more relevant. There have been many studies looking at modelling anatomical deformations [65], [68], [69], [71], [72], and the effect they can have on the delivered dose [13], [66], [67].

In Chapter 3 of this thesis, we showed that the dosimetric effect of anatomical deformations in H&N is similar to or smaller than that of set-up uncertainties for most OARs considered. However, no planning tools currently exist to explicitly account for anatomical deformations in H&N. As such, the aim of this study is to see if it is possible to create plans that are robust to anatomical deformations using tools intended to deal with set-up uncertainties. For this, we created margin, robust and probabilistic H&N plans using different uncertainty settings. We then compared their robustness to simulated anatomical deformations.

4.2. Materials and methods

4.2.1 Patient data

Variable	Value	Number of patients
Site	Anterior tongue	2
	Hypopharynx	2
	Neck nodes	5
	Oropharynx	9
	Supraglottis	1
T stage	2	7
	3	4
	4	3
	unknown	5
N stage	0	3
	1	2
	2	13
	unknown	1
Sex	Male	16
	Female	3

Table 4.1: Patient characteristics.

For this study, the planning CT (pCT) and delineations from 19 H&N cancer patients treated at our institution between 2014-2018 were used (Table 4.1). These data were collected from the ukCAT distributed learning database (ethics approval from the UK North West - Haydock Research Ethics Committee, reference number 17/NW/0060, local consent ref. 2018-018).

4.2.2 Plan creation

A total of nine plans were created for each patient, using three different planning approaches (margin-based, robustly optimised and probabilistic) and three different uncertainty settings (1 mm, 3 mm and 5 mm). Plans were created in RayStation (v6.99) following a class solution used at our institution, with a prescribed dose to the primary CTV of 66 Gy in 30 fractions. A 3x3x3 mm dose grid was used. Table 4.2 shows the OAR objectives used for planning. Target objectives were adapted between planning approaches, but all other objectives remained unchanged.

OAR	Objective
Brainstem	Maximum dose < 45.0 Gy
Brainstem + 5 mm (PRV)	Maximum dose < 52.0 Gy
Larynx	Mean dose < 38.0 Gy
Left/right parotids	Mean dose < 22.0 Gy
Oral cavity	Mean dose < 37.5 Gy
Spinal cord + 5 mm (PRV)	Maximum dose < 43.0 Gy

Table 4.2: OAR planning objectives.

For the margin-based plans, the target objectives were based on a PTV which was derived by expanding the CTV by the uncertainty setting for that plan. The robustly optimised plans used the CTV as the target and the robustness setting used was equal to the uncertainty setting for that plan. For each uncertainty setting, a total of seven scenarios were sampled, which was the default for RayStation v6.99's robust optimisation. Probabilistic plans were created using a probabilistic treatment planning plugin, similar to that used by Bohoslavsky *et al.* [36]. Details on the probabilistic planning algorithm can be found in Supplement 4.A.1.

All three probabilistic plans for each patient were optimised based on the CTV, using a confidence level of 90% and a standard deviation of 1.4 mm for the random uncertainties

[54]. The standard deviation of the systematic uncertainties was altered between the plans to change the uncertainty setting, using values of 0.0 mm, 0.8 mm and 1.6 mm, equivalent to a 1 mm, 3 mm and 5 mm margin respectively according to the van Herk margin formula [76].

4.2.3 Treatment simulations

For each plan, 1000 treatments simulations were run using the time-dependent PCA models for anatomical deformations developed in Chapter 3, created from a population of 30 oropharyngeal cancer patients. As these PCA models were population-based, the pCT of each patient had to be first registered to the average patient geometry. These non-rigid registrations were performed using NiftyReg [61], with a bending energy parameter of 0.0005. The dose distribution of the patient was then mapped into the average patient geometry and extended outside the patient's body, to improve the validity of the assumption of dose-shift invariance [41], [48].

The treatment simulations were run following the method detailed in Section 3.2.3.1 of this thesis. In short, this involved generating one simulated DVF from the systematic PCA model and five simulated DVFs from each of the random weekly PCA models. At each fraction, the systematic DVF and one of the random DVFs were combined and the patient anatomy deformed accordingly. The dose at each fraction was accumulated to get an overall simulated treatment dose. This was repeated for each plan of each patient.

For each simulated treatment, the minimum dose across 98% of the CTV (i.e., the D98) was calculated and used as a measure of target coverage. A plan was said to be robust to anatomical deformations if at least 90% of the simulated treatments achieved at least 95% of the prescribed dose (62.7 Gy) in the D98. We denote the 90th percentile of D98 as $D98^{90}$. Regarding the OARs, the mean dose to the larynx, oral cavity and each of the parotids was calculated along with the maximum dose to the spinal cord and the brainstem for each simulated treatment.

To estimate the equivalent uncertainty setting required to reach the desired robustness of 95% of the prescribed dose for $D98^{90}$, we used a linear interpolation on the mean of the $D98^{90}$ across all 19 patients. This interpolation was done by assuming a linear relationship for $D98^{90}$ between the 1 mm and 3 mm plans and between the 3 mm and 5mm plans. The

point between these plans at which the mean $D98^{90}$ reached 95% of the prescribed dose was considered an estimate of the equivalent uncertainty setting (see the stars marked in Figure 4.1). This equivalent uncertainty setting was then used to obtain the mean 90th percentile of each OAR dose metric for each planning approach at the required robustness. This allowed us to compare how well each approach could spare OARs when dealing with anatomical deformations.

4.3. Results

For the CTV, as expected, all plans for all patients showed a decrease in the dose from the planned D98 to the $D98^{90}$. As the uncertainty setting increased from 1 mm to 5 mm, $D98^{90}$ increased and the difference between $D98^{90}$ and the planned D98 got smaller (Figure 4.1). Using the 1 mm setting, both $D98^{90}$ and the difference between $D98^{90}$ and the planned D98 were similar for all three planning approaches, with a $D98^{90}$ of 62.3 Gy which was a decrease of 1.9 Gy from the planned D98. The margin plans were seen to have the highest $D98^{90}$ and the smallest decrease of $D98^{90}$ from the planned D98. Conversely, probabilistic plans had the lowest $D98^{90}$ and the largest decrease from the planned D98. Considering the mean $D98^{90}$ across all patients, for the probabilistic plans, the 5 mm setting was sufficient to be robust to anatomical deformations, while for margin and robust planning the 3 mm plans were robust.

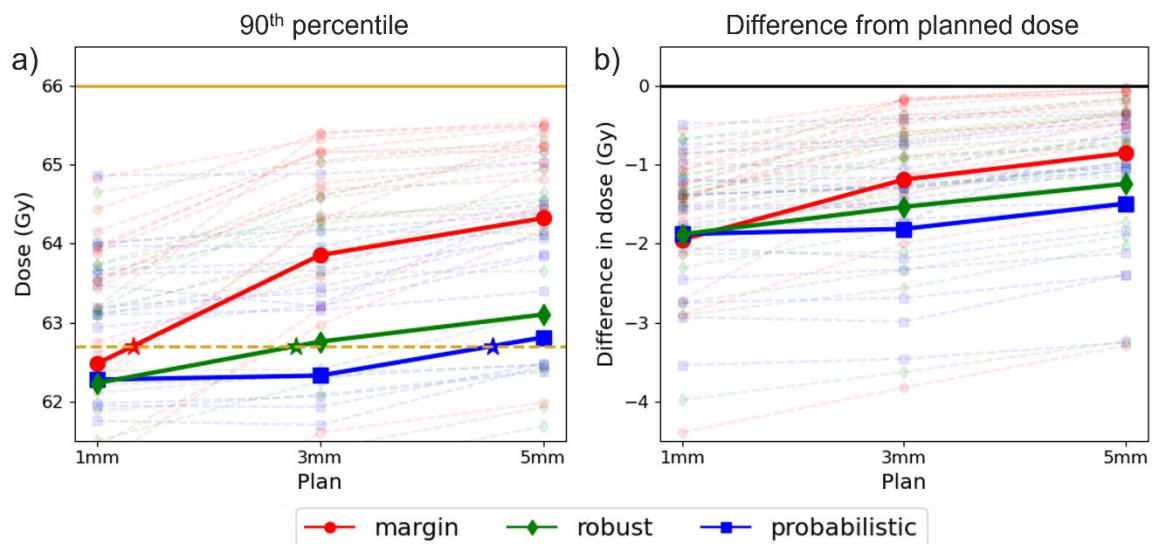


Figure 4.1: a) $D98^{90}$ and b) the difference between the planned D98 and $D98^{90}$ for the CTV for each plan. The margin plans are plotted in red, the robust in green and the probabilistic plans in blue. For each plot, the solid lines are the mean across all patients, and individual patient results are shown as faint dotted lines. The yellow dotted line shows 95% of the prescribed dose (62.7 Gy) and the solid yellow line shown the prescribed dose (66 Gy). The stars in plot a) show the point at which the $D98^{90}$ reaches 95% of the prescribed dose for each planning approach. Note: data for some individual patients are outside the dose range of the plots.

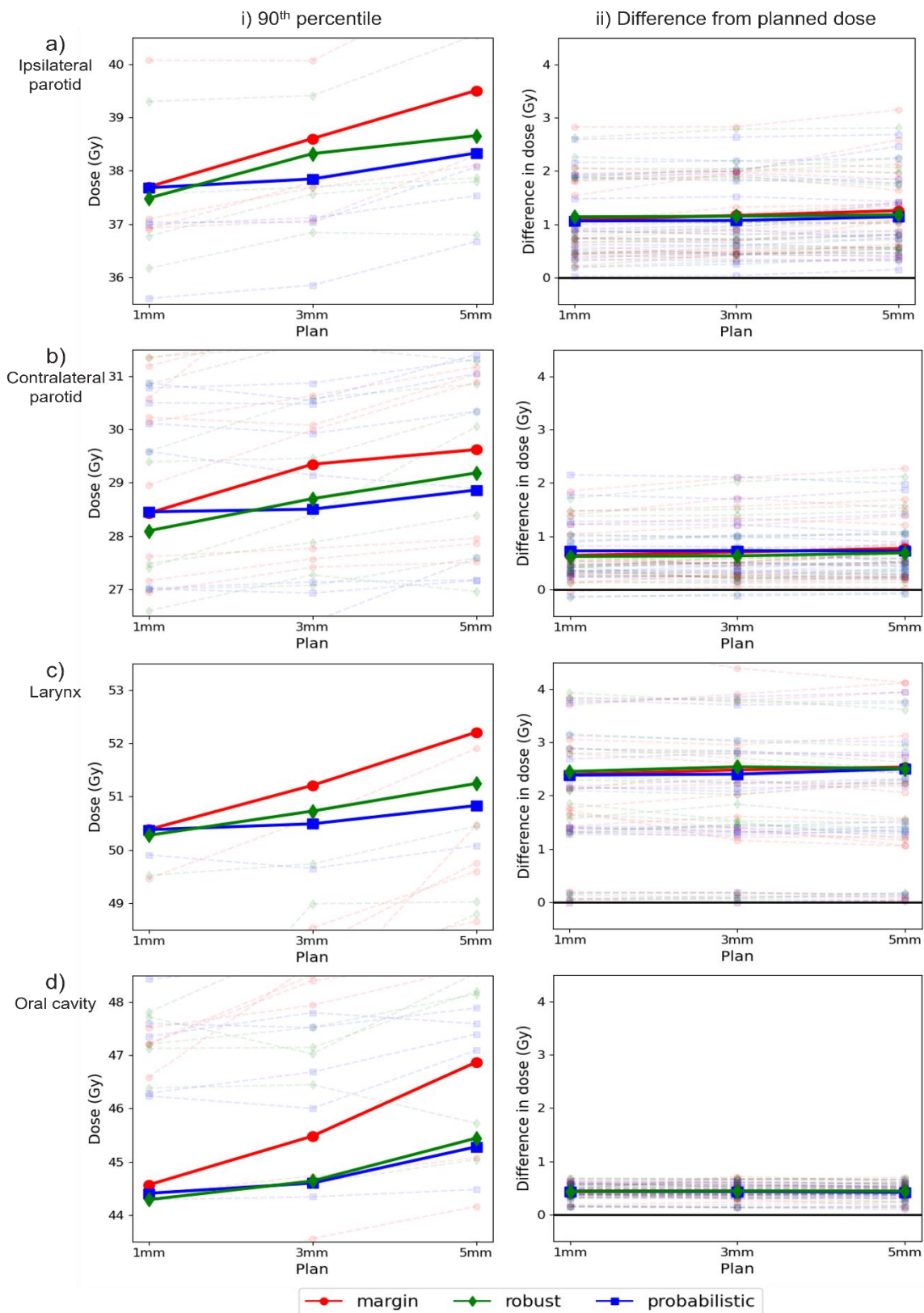


Figure 4.2: i) the 90th percentile and ii) the difference between the planned value and 90th percentile of the mean dose under deformations for a) the ipsilateral parotid, b) the contra lateral parotid, c) the larynx and d) the oral cavity for each plan. The margin plans are plotted in red, the robust in green and the probabilistic plans in blue. For each plot, the solid lines are the mean across all patients, and individual patient results are shown as faint dotted lines. Note: data for some individual patients are outside the dose range of the plots; each plot has the same dose range but is centred on a different dose.

For both parotids, the larynx and the oral cavity, as the uncertainty setting increased, the 90th percentile of the mean dose under anatomical deformations increased for all planning approaches (Figure 4.2 i). In general, the 90th percentile was highest for the margin plans for each of these OARs. When considering the difference between the 90th percentile of the mean dose under deformations and the planned mean dose to these OARs, in general, there was little difference between the different plans (Figure 4.2 ii). Note that the dose difference was calculated between the 90th percentile of the mean dose under deformations and the planned mean dose to these OARs, in general, there was little difference between the different plans (Figure 4.2 ii). Note that the dose difference was calculated between the 90th percentile and the planned dose for *each individual plan*. The fact that there was little difference between the different plans, means the planned OAR dose must have followed the same trends as the 90th percentile of the OAR dose. The mean increase from the planned mean dose to the 90th percentile of the mean dose under deformations was 1.1 Gy and 0.7 Gy for the ipsilateral and contralateral parotids respectively, 2.5 Gy for the larynx, and 0.4 Gy for the oral cavity.

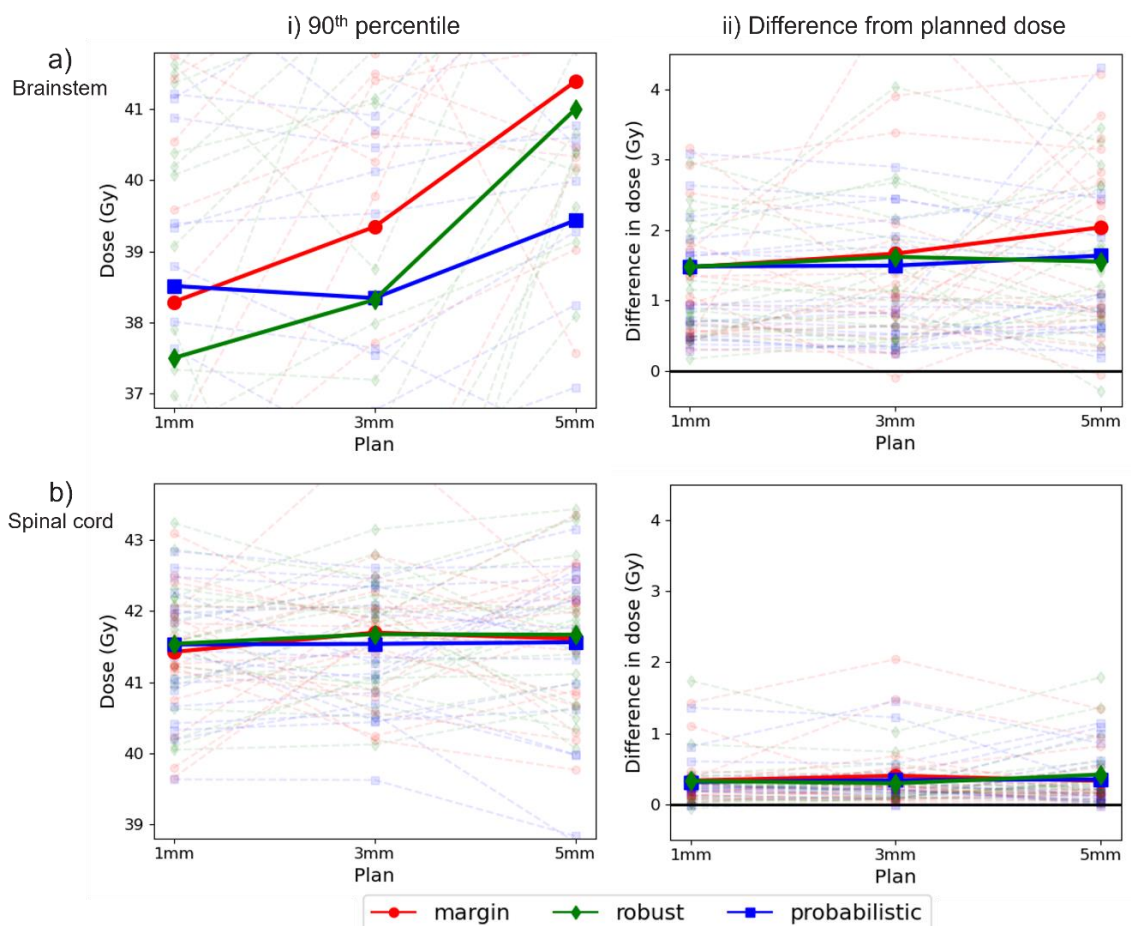


Figure 4.3: i) the 90th percentile and ii) the difference between the planned value and 90th percentile of the maximum dose under deformations for a) the brainstem and b) the spinal cord for each plan. The margin plans are plotted in red, the robust in green and the probabilistic plans in blue. For each plot, the solid lines are the mean across all patients, and individual patient results are shown as faint dotted lines. Note: data for some individual patients are outside the dose range of the plots; each plot has the same dose range but is centred on a different dose.

On average, the 90th percentile of the maximum dose to the brainstem was highest for the 5 mm margin plan and lowest for the 1 mm robust plan, see Figure 4.3. In general, the 90th percentile of the maximum dose under deformations increased as the uncertainty setting increased. The 5 mm margin plan also had the highest increase of the 90th percentile of the maximum dose under deformations to the brainstem at 2.0 Gy, while all other plans were similar with an increase between 1.4-1.7 Gy.

There was little difference in both the 90th percentile of the maximum dose under deformations for the spinal cord and the difference between that and the planned maximum dose between the different plans, with an average value of 41.6 Gy and 0.3 Gy respectively.

For all OARs considered, there was a large variation between different patients in the 90th percentile of the dose metric under deformations and its difference from the planned dose.

	Margin	Robust	Probabilistic
Uncertainty setting (mm)	1.3	2.8	4.5

Table 4.3: Interpolated uncertainty setting required for each planning approach so the mean CTV D_{98}^{90} is 95% of the prescribed dose (shown by the stars in Figure 4.1).

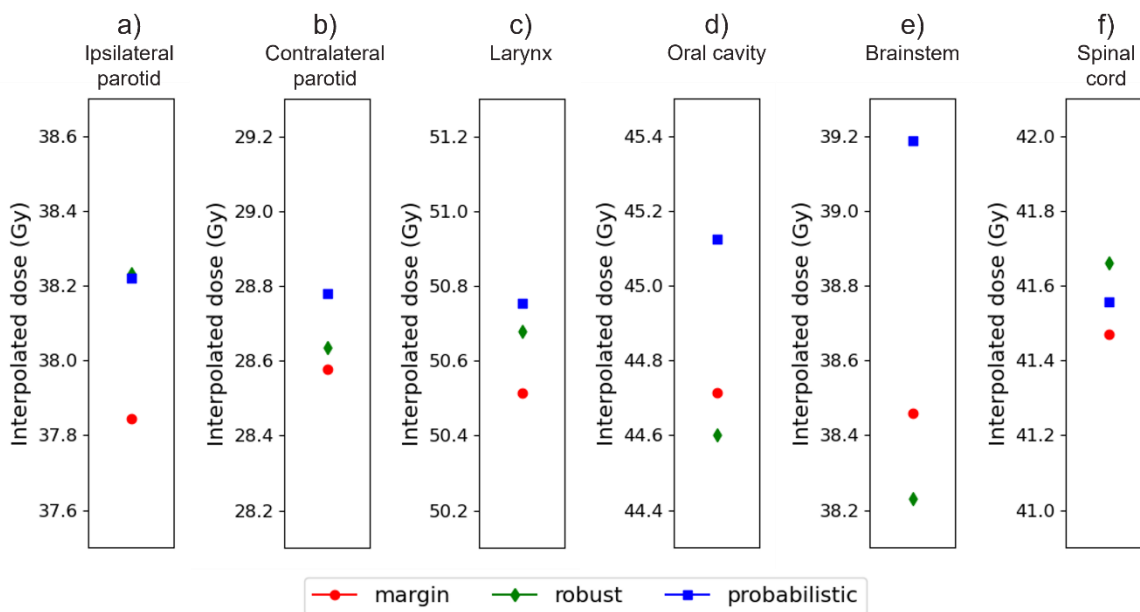


Figure 4.4: Interpolated mean 90th percentile of the dose metric to each of the OARs, for the uncertainty setting required for the desired robustness for each planning approach (Table 4.3). The margin plans are plotted in red, the robust in green and the probabilistic plans in blue. Note: each plot has the same range but is centred on a different dose.

Table 4.3 gives the uncertainty setting required to reach the desired level of robustness for each planning approach (shown by the stars in Figure 4.1), and Figure 4.4 shows the

interpolated OAR dose metrics for this uncertainty setting. It can be seen that for all OARs except the oral cavity and the brainstem, the margin plan gave the lowest OAR dose. The probabilistic plan gave the highest dose to all OARs except the spinal cord and ipsilateral parotid. The largest difference in OAR dose between different planning approaches was seen in the brainstem, with a difference of 1.0 Gy.

4.4. Discussion

In this study we have compared the robustness to anatomical deformations of margin-based, robustly optimised and probabilistic plans with different uncertainty settings for 19 H&N cancer patients. For all planning approaches, as the uncertainty setting increased from 1 mm through to 5 mm, the CTV D_{98}^{90} under deformations got higher and the decrease between the planned dose and D_{98}^{90} reduced. However, the OAR doses were also seen to increase as the uncertainty setting increased. In general, for each uncertainty setting, the margin plans had the highest D_{98}^{90} but also had the highest OAR doses. Averaged across the 19 patients, for the margin and robust planning approaches the 3 mm plans were sufficiently robust and for the probabilistic approach the 5 mm plans were sufficiently robust. To reach the desired level of robustness, the margin approach required the smallest uncertainty setting and this gave the lowest OAR doses compared to the robust and probabilistic approaches.

While there have been no other studies explicitly accounting for anatomical deformations by treating them as set-up uncertainties, there have been numerous studies using both robust and probabilistic planning approaches similar to ours to simply account for set-up uncertainties [32]–[37], [41]. When accounting for anatomical deformations, our results show that both robust and probabilistic planning can deliver lower OAR doses than margin-based planning for the same uncertainty setting (Figure 4.2 and Figure 4.3). This has been seen in other studies using these planning approaches to account for set-up uncertainties in H&N [34], [41]. However, in both those studies, plans were adapted to ensure a similar level of target coverage between the margin-based plans and either the robust or probabilistic plans. Wagenaar *et al.* [34] renormalised both their robustly optimised and margin-based plans to ensure a specific level of target coverage and Fontanarosa *et al.* [41] adjusted the weights of the objective functions to ensure that target coverage was the

same between the two planning approaches. While we did not renormalise our plans to ensure the same level of robustness, we used linear interpolation to find the uncertainty setting for each planning approach that gives the desired level of robustness. We found that for the same level of plan robustness, margin plans had the lowest OAR doses. However, the difference in OAR dose between the planning approaches at the desired level of robustness was small (Figure 4.4).

Other studies have gone beyond set-up uncertainties and used probabilistic methods to directly account for anatomical deformations [86], [99]. Baum *et al.* [86] created coverage probability maps from the positions of the prostate, bladder and rectum on five pre-treatment CT scans and used these in their plan optimisation. Tilly *et al.* [99] assumed a single fraction treatment and sampled 100 possible deformations from their PCA model for cervical cancer patients and optimised based on the worst 10% of these scenarios. They found an increase in target robustness and a decrease in OAR doses in their probabilistically optimised plans compared to the margin-based plan. However, both of these methods do not explicitly distinguish between random and systematic uncertainties. The latter method is also computationally expensive and more challenging than simply treating the anatomical deformations as set-up uncertainties.

PCA models of anatomical deformations in H&N have been shown to have the largest residual error around the oropharyngeal region (Chapter 2), with the PCA models used in this study having a mean global residual error of 2.2 mm (Section 3.3.1.1). This means that around the oropharyngeal region, the dose under deformations could be poorly simulated for some patients. Additionally, our models were trained and validated in Chapter 3 on oropharyngeal patients, but 10 of the patients used for this study did not have oropharyngeal tumours. These patients were included in this study to assess the effect of anatomical deformations in other H&N subsites, but the actual deformations for these patients could be slightly different than simulated.

As the anatomical deformations used in this study were population-based, the dose distribution for each patient had to be first mapped to the average patient geometry from the PCA models. This was done by registering the pCT of each patient to the average pCT, which will not be perfect and can introduce errors. However, these errors will be the same

for all simulated treatments and all plans for each individual patient, and so the different plans for the patient can still be compared.

Our results show that the simulated maximum dose to the brainstem increased from the planned dose when anatomical deformations were taken into consideration. While the brainstem itself is not expected to deform, the maximum dose is generally located at the bottom of the brainstem where it connects to the spinal cord, and so changes in flexion of the spinal cord could result in this increase in dose.

The plans used in this study were created using a class solution from our institution. These plans were not individualised for each patient and so may not represent the most optimal plan possible for the patient for the given planning approach and uncertainty setting. The class solution was chosen to avoid bias and ensure a fair comparison between the planning approaches.

For the probabilistic plans in this study, only the standard deviation of the systematic uncertainties was changed for the different uncertainty settings. The standard deviation of the random uncertainties remained at 1.4 mm [54]. Changing the size of the systematic uncertainties changes which scenarios are sampled by the optimiser, whereas the random uncertainties have no effect on which scenarios are sampled. Just changing the size of the systematic uncertainties meant that the differences between the different uncertainty settings were more similar to those used in robust planning, where changing the uncertainty setting also changes the scenarios sampled in the optimisation.

While on average the 5 mm plans for all three planning approaches and the 3 mm margin and robust plans were robust to anatomical deformations in terms of target coverage, there was a lot of variation between different patients. This means that it would have to be determined on a patient-by-patient basis as to which planning approach and uncertainty setting best accounts for the anatomical deformations, and whether any of these plans are robust enough.

Robust and probabilistic planning required uncertainty settings that were ~ 2 and ~ 3.5 times larger than those of margin-based planning respectively to reach the desired level of plan robustness. This means that the margin-based plans were better at accounting for anatomical deformations than the robust and probabilistic plans. We hypothesise that

margin plans may be more robust to deformations because the shape of the CTV is smoothed by the PTV expansion, while the original CTV contours may have sharp corners that are more easily underdosed under deformations. This raises the need of further research on the impact of CTV shape on plan robustness.

In general, for all planning approaches and uncertainty settings, OAR doses were seen to increase from the planned dose and the target dose was seen to decrease. This means that anatomical deformations should not be ignored in the planning process and plans should be evaluated to see how well they account for these changes. So, while in general treating anatomical deformations as set-up uncertainties can produce plans that are robust in terms of target coverage, care still needs to be taken to ensure OARs are not overdosed. For example, in cases where an OAR dose is close to the clinical goal, adaptive radiotherapy could be considered to stop the OAR being overdosed.

4.5. Conclusion

Margin-based plans were seen to be the most robust to anatomical deformations in terms of target coverage, but they had the highest OAR doses for the same uncertainty settings. Similarly, increasing the uncertainty setting increased the robustness but also the OAR dose. In general, anatomical deformations in H&N could be accounted for by treating them as set-up uncertainties. However, the robust and probabilistic plans were more sensitive to deformations, and this should be taken into account when replacing margin plans by robust or probabilistic planning techniques that only use translational error scenarios. For some patients, these plans were not robust and so alternative ways of accounting for the anatomical deformations should be considered in these cases, e.g., directly accounting for the anatomical deformations in the plan optimisation process or by adaptive radiotherapy.

4.6 Acknowledgements

We would like to acknowledge Ian Hinder for his work integrating the probabilistic treatment planning plugin into RayStation. Marcel van Herk was supported by NIHR Manchester Biomedical Research Centre. This work was supported by Cancer Research UK via funding to the Cancer Research Manchester Centre [C147/A25254].

Appendix 4A

4.A Supplementary material

4.A.1 Probabilistic planning algorithm

The probabilistic planning algorithm directly accounts for set-up uncertainties and starts by creating separate 3D Gaussian error kernels for both random and systematic uncertainties. The standard deviations of the error kernels are Σ and σ for the systematic and random uncertainties respectively, and can be different in each of the x , y and z directions. Both error kernels have the same resolution as the dose grid (3x3x3 mm in our case) and are truncated so that any values with a probability of less than 1% of the maximum value are removed, and the remaining values renormalised to make their sum equal to 1.

To account for random uncertainties, the dose distribution is convolved by the random error kernel for each voxel. Systematic uncertainties are accounted for by translating the blurred dose distribution, assuming dose-shift invariance. Each point in the systematic error kernel is sampled and the total cost function is computed for each scenario. These scenarios are then sorted by ascending cost function and weighted by the probability of that scenario occurring. The weighted cost functions are then summed until the cumulative probability of all scenarios in the sum reaches a specified confidence level, in this case 90%. This truncated weighted sum is the total cost that is then optimised.

Blank page

Chapter 5

The impact of target volume shape on plan robustness with different planning approaches

Authors

Jennifer Robbins¹, Marcel van Herk¹, Andrew Green² and Eliana Vásquez Osorio¹

Affiliations

1. Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
2. European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom

Author contributions

I designed the experiments together with the co-authors. I designed the cylindrical phantom in ImSimQA. I created the simple shapes in RayStation and selected the most appropriate for analysis. Together with EVO, I selected the real CTV shapes to use, aiming at getting a large range of shape variability. I took over development of the probabilistic treatment planning plugin from Ian Hinder (software engineer), which was used to create the probabilistic plans. I wrote the scripts in RayStation and created all the radiotherapy plans that were used in this paper. I wrote the lua code used to run treatment simulations on the plans and performed all the analysis. I produced all the figures and wrote the manuscript which was then reviewed by all the co-authors.

Abstract

Background: In radiotherapy, margins are commonly used to account for uncertainties, but robust and probabilistic planning present alternative approaches for this purpose. The aim of this study was to compare the robustness to set-up uncertainties of margin-based, robustly optimised and probabilistic plans for different clinical target volume (CTV) shapes.

Methods: Margin, robust and probabilistic plans were created on a cylindrical phantom for 12 different CTV shapes: a sphere, a cube and 10 CTVs extracted from head and neck cancer patients. The sphericity of each CTV shape was used as a measure of its complexity. For each plan, the conformity index (*CI*) was calculated and then 1000 treatments were simulated to measure the impact of set-up uncertainties. For each simulated treatment the minimum CTV dose was calculated.

Results: For all shapes, the margin plan was seen to have the best robustness, with 90% of the simulated treatments having a minimum CTV dose well above 95% of the prescribed dose. However, the margin plan also had the poorest *CI*. For both the robust and probabilistic plans, about half of the CTVs did not achieve target coverage, including the cube. For each planning approach, as the sphericity of the CTV decreased, target coverage dropped and the *CI* worsened. Both robust and probabilistic plans struggled to achieve adequate coverage on the 'corners' of the more complex CTVs.

Conclusions: Margin plans achieved the highest target coverage but also had the worst *CI*. As target sphericity decreased, plans were less robust for all three planning approaches. Care needs to be taken when planning on CTVs with low sphericity, especially when using robust or probabilistic planning as they may underdose the 'corners' of more complex CTV shapes.

5.1 Introduction

In radiotherapy, organ motion and differences in patient set-up positioning cause uncertainties to the delivered dose. These uncertainties are typically split into systematic errors, which remain the same throughout the treatment course, and random errors, which are different for each fraction. Such uncertainties are generally accounted for by extending the clinical target volume (CTV) by a margin to the planning target volume (PTV). A common

methodology to calculate the margin is to use the van Herk margin formula [76]. However, margin recipes were derived for targets with limited curvature which is unrealistic in head and neck (H&N) cancer patients and are based on idealised dose distributions which may not actually be realistically achievable. Gordon *et al.* [81] showed that prostate plans created using these margin recipes were able to tolerate larger set-up uncertainties than expected and as such were too conservative. Additionally, Zheng *et al.* [79] calculated that CTVs with different sizes and shapes would require different margin expansions to ensure that the CTV is within the PTV 95% of the time. As an alternative to PTV margins, methods such as robust planning and probabilistic planning optimise on the CTV and take the uncertainties into account in the optimisation process. These methods have two advantages over margins: 1) the robustness of the plan can be explicitly traded off with organ at risk (OAR) dose, and 2) the shaping of the dose distribution is taken into account in the robustness evaluation.

Robust planning works by calculating the cost function for a set of discrete error scenarios that are assumed to be equally likely to occur, and aims to optimise the worst-case of these scenarios. This method is often used to account for range uncertainties as well as set-up uncertainties for proton therapy [24]–[31] but its use has also been studied for set-up uncertainties in photon treatments for glioblastoma [32], lung [33], head and neck (H&N) [34] and breast [35].

In probabilistic planning, both systematic and random uncertainties are directly taken into account and dealt with differently by the optimiser. Random errors are accounted for by blurring the evaluated dose distribution while systematic errors are sampled discretely like in robust planning. The cost function for each sampled scenario is weighted by the probability of that scenario occurring, and the probability distribution of these functions across all scenarios is optimised to reach the given planning constraint at a pre-set probability. Studies on the effectiveness of probabilistic planning to account for set-up uncertainties have been performed in prostate [36]–[38], spinal cord [39], [40] and H&N [41] patients, generally showing better OAR sparing for the same probability of target coverage.

However, to the best of the author's knowledge, no study so far evaluated the direct effect of target shape complexity on the effectiveness of the different planning approaches. The

aim of this study was to compare margin-based plans with robustly and probabilistically optimised plans under realistic set-up uncertainties for complex target shapes as encountered in H&N radiotherapy.

5.2 Materials and methods

For this study, a uniform cylindrical phantom (diameter 15 cm and length 20 cm) was used. 12 separate CTV shapes were used, consisting of a sphere (4.8 cm diameter), a cube (4.8 cm length) and 10 actual primary CTV shapes from H&N cancer patients (Table 5.1). All shapes were centred on the centre of the cylindrical phantom and were treated as the CTV. The sphericity of each shape was calculated as a measure of its complexity using the following equation:

$$sphericity = \frac{\sqrt[3]{36\pi V^2}}{A}, \quad (5.1)$$

where V is the volume of the shape and A is the surface area. This measure is 1 for perfect sphere, and <1 for more complex shapes.

Three VMAT plans were made in a research version of RayStation (version 6.99) for each shape using margin-based planning, robust planning and probabilistic planning with a prescribed dose of 66 Gy in 30 fractions, using a 3x3x3 mm dose grid. The plans were kept simple, using just a minimum dose constraint for the target and both maximum and uniform target objectives along with a dose fall-off objective outside the target (Table 5.2). The only region of interest for this study was the target shape – no explicit OARs were included, with the rest of the phantom being considered as normal tissue.

Both the robustly optimised plans and the probabilistic plans used the CTV as the target, while the margin-based plans used a PTV that was derived by isotropically expanding the CTV by 4 mm. For the target objectives and constraints, an isotropic robustness setting of 4 mm was used for the robust plans.

A probabilistic treatment planning plugin was used to create the probabilistic plans, similar to the one used by Bohoslavsky *et al.* in [36], using a confidence level of 90%. Details on the probabilistic planning algorithm can be found in Supplement 4.A.1 in Chapter 4. In the probabilistic plans, standard deviations of the systematic (Σ) and random (σ) uncertainties were used of 1.2 mm and 1.4 mm respectively [54], equivalent to a 4 mm margin according

to the van Herk margin formula [76]. The dose fall-off objective remained unchanged across the planning approaches.













Shape	3D Rendering	Volume (cm ³)	Surface area (cm ²)	Sphericity
Sphere		57.28	72.20	1.00
Cube		105.94	132.02	0.82
CTV1		53.24	77.91	0.88
CTV2		69.19	93.86	0.87
CTV3		74.19	110.53	0.77
CTV4		81.72	121.08	0.75
CTV5		50.04	91.17	0.72
CTV6		61.41	107.60	0.70
CTV7		53.83	104.24	0.66
CTV8		86.54	143.58	0.66
CTV9		91.54	159.09	0.62
CTV10		58.89	128.59	0.57

Table 5.1: Shape, volume, surface area and sphericity of each target shape. The real CTV shapes are ordered by sphericity, with CTV1 being the most spherical.

Function	Parameters	Weight
Minimum target dose	62.7 Gy (95% prescribed dose)	Constraint
Maximum target dose	69.3 Gy (105% prescribed dose)	50
Uniform target dose	66 Gy	10
Dose fall-off	62.7 Gy to 55 Gy in 5mm from PTV edge	5

Table 5.2: Planning cost functions and their weights.

To compare the plans, several evaluations were performed. First, the conformity index (CI) was calculated to assess the spread of the high dose region, according to the Radiation Therapy Oncology Group guidelines [123]. This is calculated by:

$$CI = \frac{V_{95\%}}{V_{target}}, \quad (5.2)$$

where $V_{95\%}$ is the volume enclosed by the 95% isodose surface and V_{target} is the volume of the PTV. Values between 1-2 are considered to be acceptable, with 1 being the ideal but unreachable value [123]. The mean distances between the surface of the CTV and the 95%, 50% and 10% isodose surface were next computed to evaluate the spread of the high, medium and low dose regions within each plan, and these results are presented in Supplement 5.A.1. The minimum distance between the CTV and the 95% isodose surface was also calculated.

The target coverage of the plans was evaluated using an external error scenario simulator, similar to the software described in [124]. Here, 1000 treatments were simulated for each plan, taking random and systematic set-up uncertainties into account. The systematic and random set-up uncertainties were drawn from Gaussian distributions with standard deviations of $\Sigma = 1.2$ mm and $\sigma = 1.4$ mm respectively. For each simulation, the minimum cumulative dose to the CTV shape, CTV_{min} , and the D99 (minimum dose across 99% of the CTV) was recorded and summarised in a probability distribution. The 90th percentile of CTV_{min} was plotted against the minimum distance between the CTV and the 95% isodose surface for each plan. The 90th percentile of CTV_{min} for each plan was also plotted against the sphericity of the target shape to investigate whether there was a link between target coverage and shape complexity.

5.3 Results

Figure 5.1 shows the distribution of CTV_{min} for the simulated treatments. Given that the van Herk margin formula was used, all plans were expected to have 90% probability of achieving at least 95% of the prescribed dose (62.7 Gy in this case). For all target shapes, the 90th percentile of CTV_{min} was highest for the margin-based optimisations where it was well above the expected 95% for all shapes, indicating that the margin plans were overly robust. The robust and probabilistic planning approaches only reached the threshold of 95% of the prescribed dose to at least 90% of the simulated treatments for half of the plans: the sphere and five real CTVs.

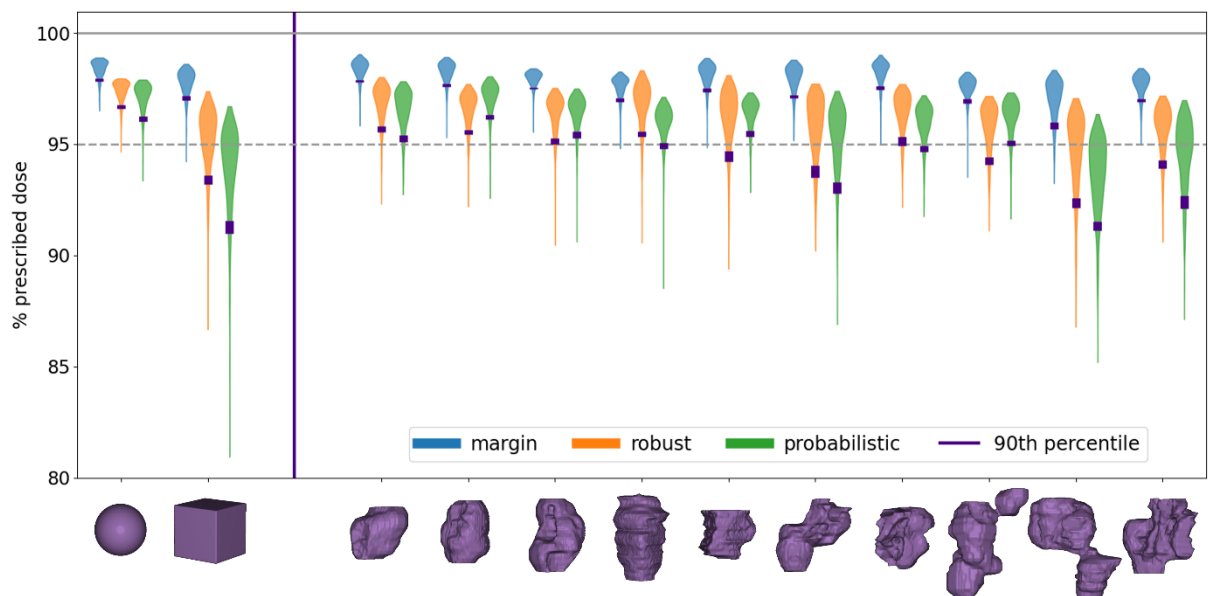


Figure 5.1: Violin plots showing the distributions of CTV_{min} for the different planning approaches and different target shapes. Plans for the sphere and cube are on the left, separated from the real H&N CTV shapes. The x-axis is labelled with a 3D rendering of the target shape used to create the plans. For each shape, the margin plan is shown in blue, the robust in orange and the probabilistic in green. The line on each violin represents the 90th percentile of the distribution (± 16 treatments corresponding to a 95% confidence level according to binomial statistics), i.e., 90% of the simulated treatments have values above the line.

For the D99 (Figure 5.2), the 90th percentile was above 95% of the prescribed dose for all plans, and highest for the margin plans. Only the probabilistic plan for the cube and CTVs 6 (👉), 9 (👉) and 10 (👉) along with CTV9 (👉) for the robust plans had any simulated treatment values below 95% of the prescribed dose.

Figure 5.3 shows there was a strong correlation between the minimum distance between the CTV and the 95% isodose surface and CTV_{min} – as one increased so did the other. The margin plans were seen to have the largest distance and CTV_{min} values, while probabilistic

plans had lower values. For each planning approach, the sphere had the highest values for both the minimum distance and CTV_{min} , apart from CTV2 (●) in the probabilistic plans. A CTV_{min} of 95% of the prescribed dose corresponded to a minimum distance between the CTV and 95% isodose surface of ~2.6 mm.

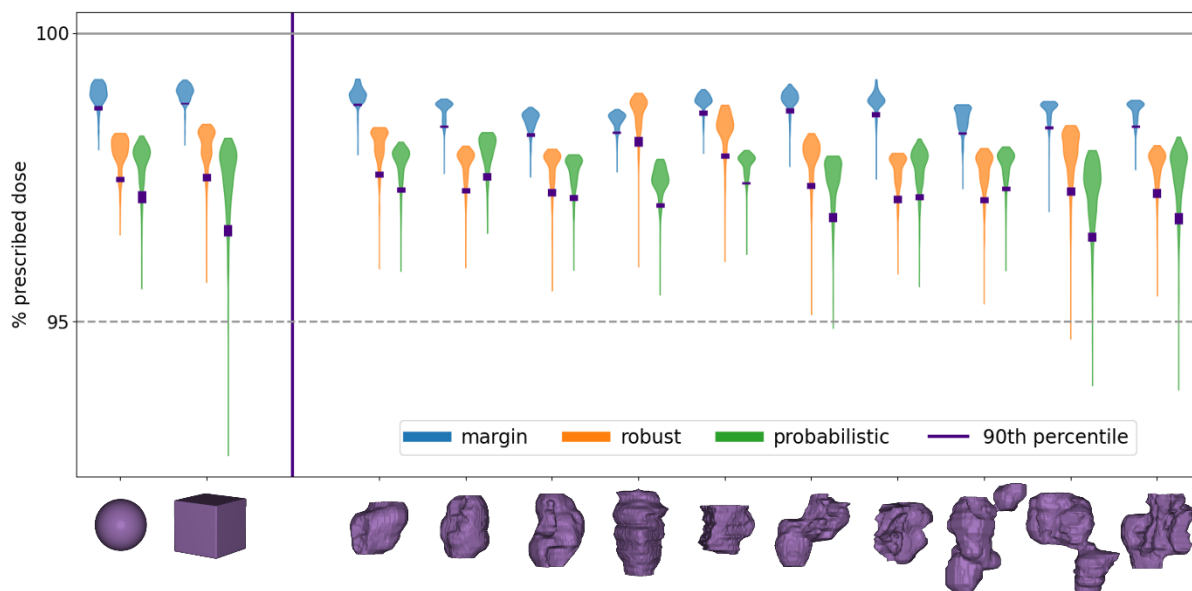


Figure 5.2: Violin plots showing the distributions for the D99 (minimum dose across 99% of the CTV) for the different planning approaches and each different target shape. Plans for the sphere and cube are on the left, separated off from the real H&N CTV shapes. The x-axis is labelled with a 3D rendering of the target shape used to create the plans. For each shape, the margin plan is shown in blue, the robust in orange and the probabilistic in green. The line on each violin represents the 90th percentile of the distribution (± 16 treatments corresponding to a 95% confidence level according to binomial statistics), i.e., 90% of the simulated treatments have values above the line.

Figure 5.4 shows the CI for the different plans, which ranged from 1.07 in the probabilistic plan for CTV9 (●) to 1.45 in the margin plan for CTV5 (●). For each shape, the margin plans had the poorest CI with a mean of 1.38 and the robust and probabilistic plans were similar, with means of 1.17 and 1.15 respectively.

As the sphericity of the shape decreased, there was a general trend for the 90th percentile of CTV_{min} to decrease for each planning approach (Figure 5.5). This same trend was seen for the D99. There was a general trend that shapes with a lower sphericity had a higher CI for each planning approach.

For the sphere, the 95% isodose surface covered the whole PTV in each plan. For the cube, the 95% isodose surface covered the whole PTV for the margin plan, whereas for the robust and probabilistic plans, it did not reach all of the voxels at the corners of the PTV (Figure 5.6).

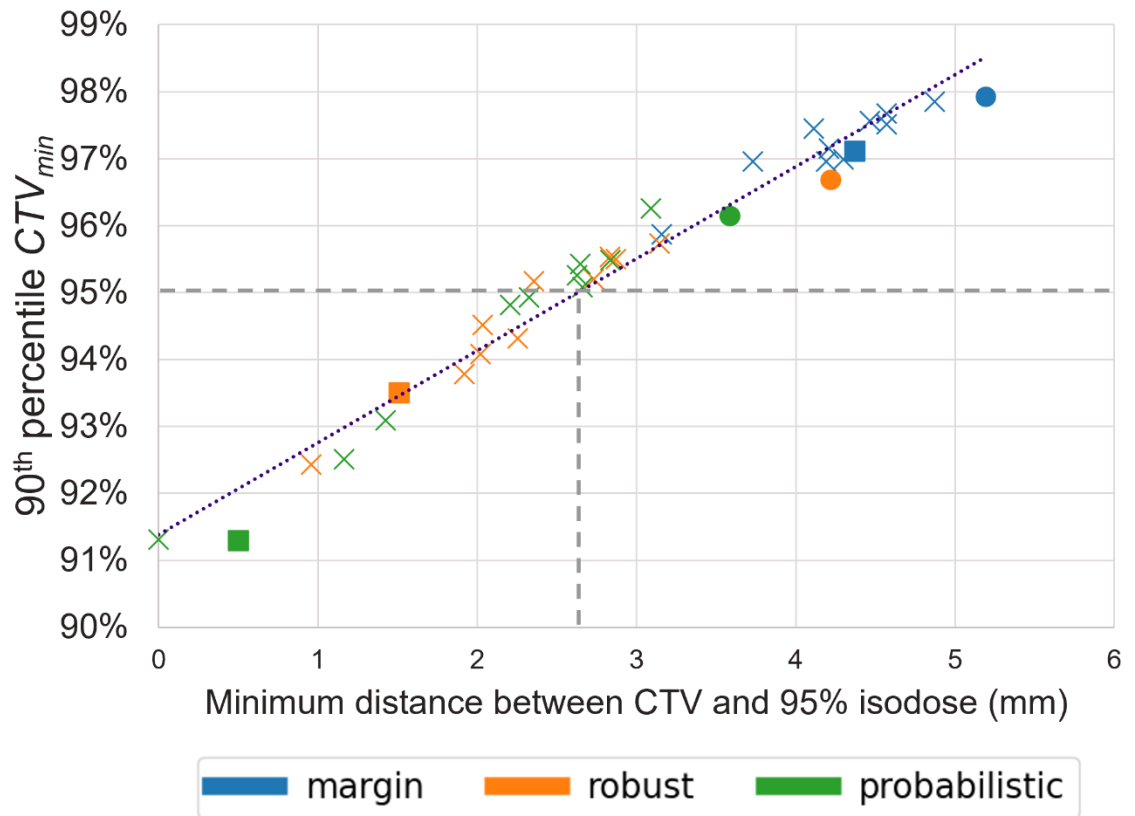


Figure 5.3: Scatter plots showing the relationship between the 90th percentile of CTV_{min} and the minimum distance between the CTV and 95% isodose surface. The 'x's represent the real H&N CTV shapes and the spherical and cubic targets are represented by a circle and square respectively. The margin plans are shown in blue, the robust in orange and the probabilistic in green. A trendline is shown in purple.

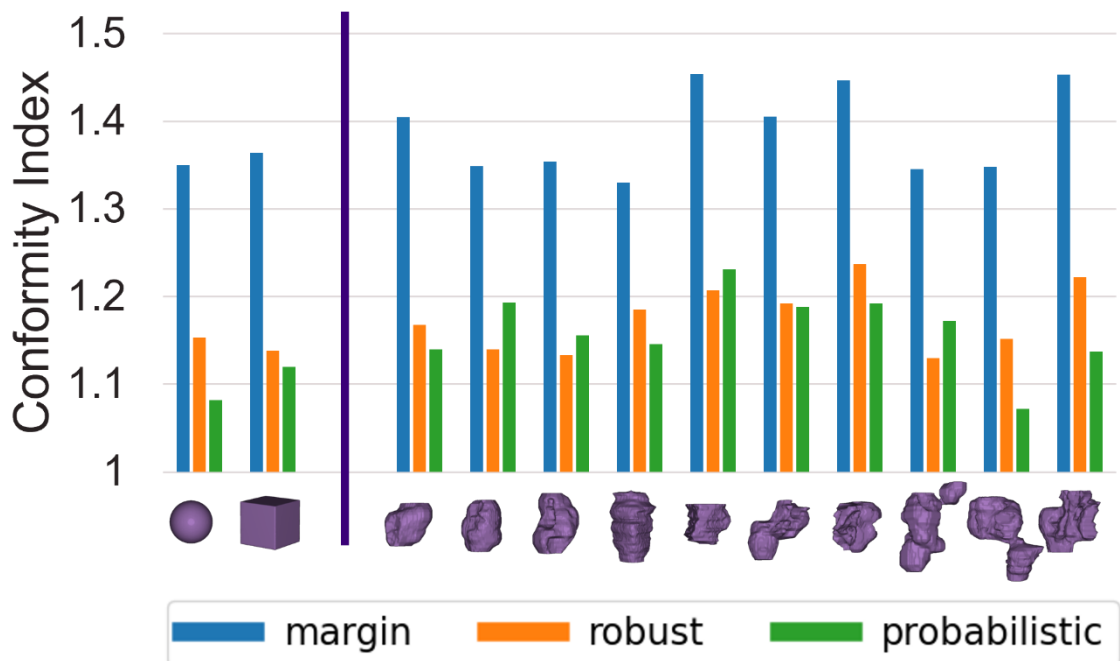


Figure 5.4: Bar charts showing the CI for the different plans. Plans for the sphere and cube are on the left, separated off from the real H&N CTV shapes. The x-axis is labelled with a 3D rendering of the target shape used to create the plans. For each shape, the margin plan is shown in blue, the robust in orange and the probabilistic in green.

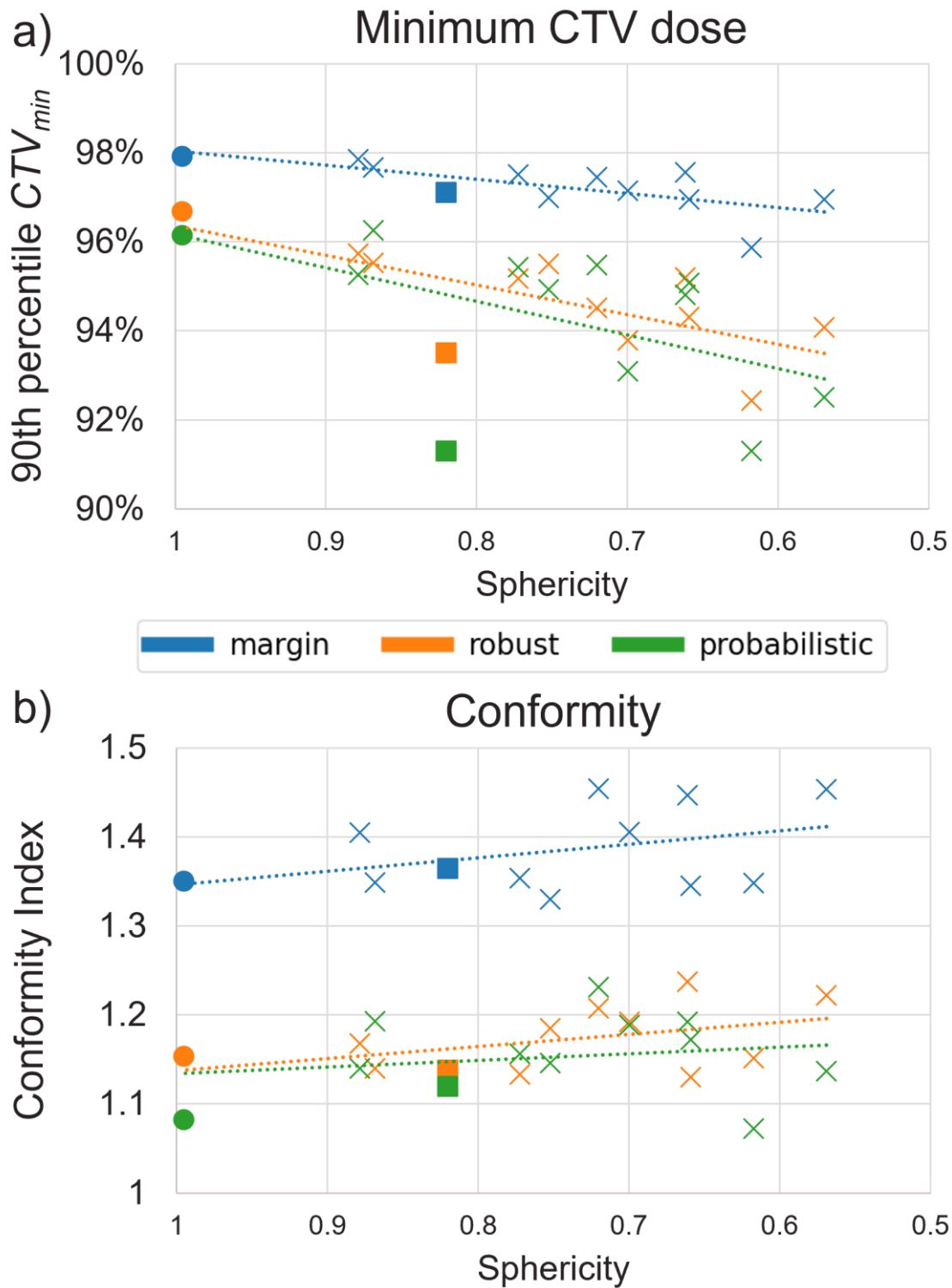


Figure 5.5: Scatter plots showing the relationship between CTV sphericity and a) the 90th percentile of CTV_{min} and b) CI for each plan. The 'x's represent the real H&N CTV shapes and the spherical and cubic targets are represented by a circle and square respectively. The margin plans are shown in blue, the robust in orange and the probabilistic in green. A trendline is shown for each of the planning approaches.

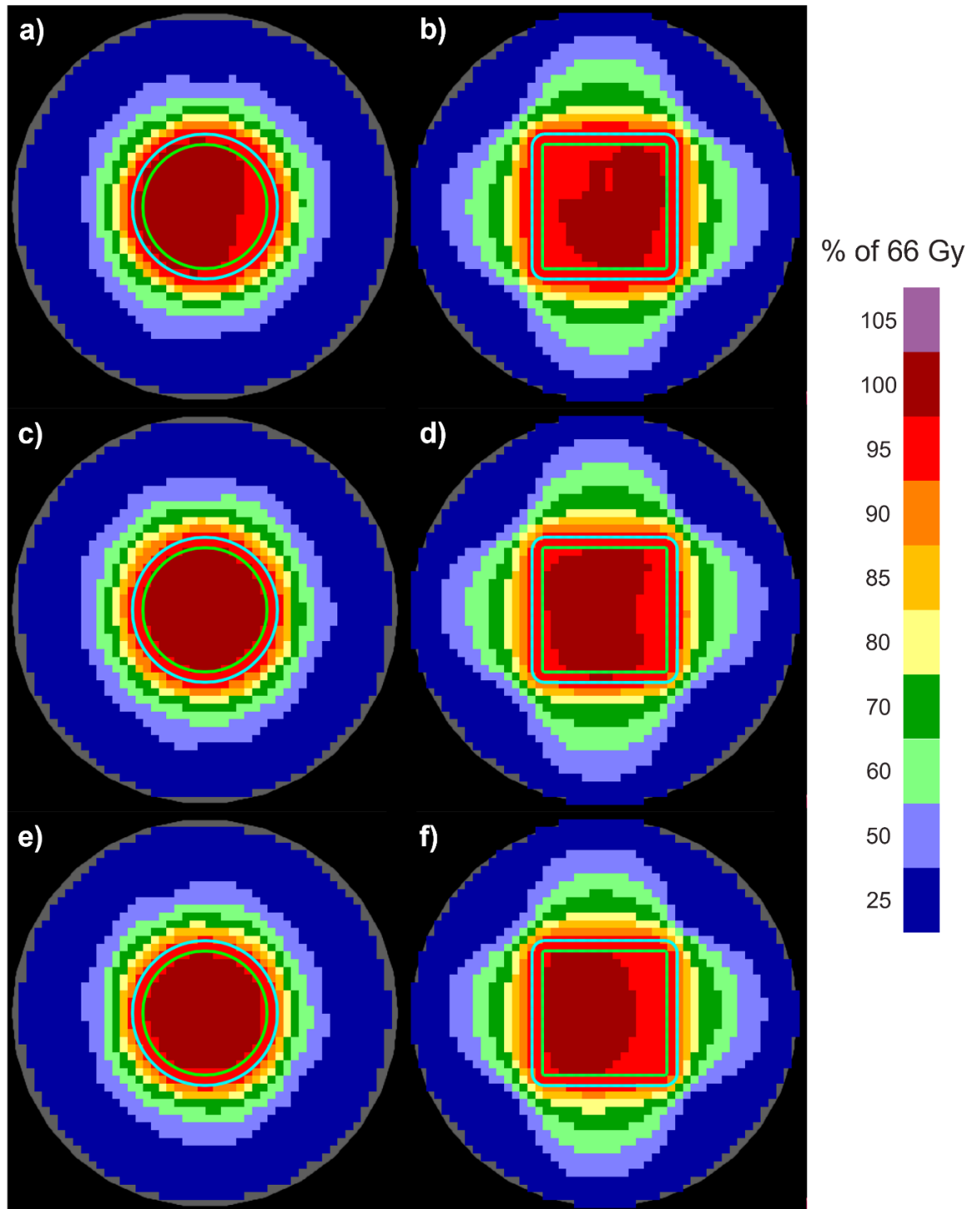


Figure 5.6: The dose distributions on the axial slice through the centre of the cylindrical phantom for the plans for the spherical and cubic target shapes. a-b) show the margin plan, c-d) the robust plans and e-f) the probabilistic plans. The CTV is delineated in green on each plan and the PTV (the CTV with a 4 mm expansion) is delineated in blue.

5.4 Discussion

In this study, we have compared simple plans for different CTV shapes using margin, robust and probabilistic planning approaches. We investigated the effect of set-up uncertainties on target dose by running treatment simulations, and also looked at conformity. Our results

have shown that as the minimum distance between the CTV and the 95% isodose surface increased, the 90th percentile of the minimum dose to the CTV also increased. The margin plans were seen to have both a higher robustness and larger dose spreads than the robust and probabilistic plans. Plans for the sphere were the most robust within each planning approach. We found that all plans achieved adequate coverage to 99% of the CTV but some of the robust and probabilistic plans had the 90th percentile of CTV_{min} significantly below the intended 95% of the prescribed dose. The margin plans were seen to have the highest target dose but also the poorest CI , while both the probabilistic and robust plans had a better CI and lower target dose.

There was a general trend that the lower the sphericity of the target shape, the lower the CTV_{min} , and the higher the CI . This implies that plans get 'worse' as the CTV gets less spherical – there is a lower dose to the target and the high dose region spreads further outside the target, into the normal tissue. This suggests that for CTVs with low sphericity, more care needs to be taken into the planning process to ensure a good quality plan, than for more spherical targets. To the best of the knowledge of the authors, no other studies have looked into the effect of target shape complexity on the efficacy of different planning algorithms.

There have been some studies looking at the effect of differences in CTV shape, size and tissue density on required margins. Zheng *et al.* [79] calculated the required margins for different size spheres, cylinders and concave regions within a CTV. They concluded that different size margins would be required for the different shapes and sizes of CTV to account for the random component of the set-up uncertainties. Witte *et al.* [78] also concluded that both the target size and tissue density affected the required margin expansion to account for random error. However, these margin expansions were calculated mathematically and were based on idealised dose distributions which may not actually be realistic or deliverable. Actual plans were not created and evaluated for the different shapes/sizes/tissue densities considered in these studies.

Numerous studies have compared either robust planning or probabilistic planning to conventional margin-based plans under set-up uncertainties for a range of sites [25], [27], [31], [32], [36]–[38], [41]. In general, these results have found that robust and probabilistic planning can help increase OAR sparing while keeping a similar level of target coverage.

These studies were conducted using real patient geometries, and the effect of different target shape complexity has not been evaluated.

The standard deviations of the systematic and random uncertainties used in the treatment simulations and probabilistic planning were 1.2 mm and 1.4 mm respectively. Using the van Herk margin formula [76], this should equate to a margin of 4 mm, which was the value used for the PTV expansion and the robustness setting for the robust plans. For a spherical target, this formula was derived under the assumption that the 95% isodose surface should extend by this margin beyond the CTV (following the PTV for conventional margin plans). In our plans for the spherical target, the mean distance between the surface of the sphere and the 95% isodose surface was 6.8 mm, 5.3 mm and 4.7 mm for the margin, robust and probabilistic plans respectively (Supplement 5.A.1). This is larger than expected for the margin and robust plans, and on par for the probabilistic plans. This shows that the margin-based plan extends the 95% isodose surface further than the formula recommends, potentially due to a lack of degrees of freedom for plan optimisation. This behaviour is not just seen in the sphere: the mean across all the shapes was 7.1 mm, 5.5 mm and 5.2 mm for the margin, robust and probabilistic planning approaches respectively. This effect has been noted before in both prostate [81] and H&N [49]. Both studies noted that the plans were able to tolerate larger set-up uncertainties than expected for the size of margins, suggesting that margin-based planning could be more conservative than intended. Gordon *et al.* [81] conclude that this is likely due to the planning system effectively creating an additional dosimetric margin around the target.

This over-cautiousness in the margin plans was also seen in the results for CTV_{min} for the simulated treatments. The van Herk margin formula was derived to ensure that a minimum of 95% of the prescribed dose to the target was achieved at least 90% of the time. All margin plans exceeded this aimed robustness.

In general, it was also seen that as the sphericity of the shape reduced, the 90th percentile of CTV_{min} reduced. This is especially noticeable in the cube. Looking at Figure 5.6, it can be seen that the robust and probabilistic plans were ‘cutting off the corners’ with the high dose. Although the real CTV shapes do not always have easily identifiable ‘corners’ as such, it is likely that this is also what was happening in those shapes with lower sphericity.

Looking at the shapes that have the lowest minimum CTV dose, CTVs 6 (👉), 9 (👉) and 10 (👉), these shapes are seen to have the most obvious 'corners' that are fairly square like, and it was these 'corners' where the minimum dose under uncertainties was below the expected 95% of the prescribed dose (Supplement 5.A.2). This explains why CTV_{min} was so low for these shapes while the D99 was within the accepted range – the majority of the target shape was covered by the high dose, but a few of the 'corner' voxels were being underdosed, resulting in a lower absolute minimum dose.

The reason for this 'cutting off corners' is possibly due to the way the scenario sampling works for the robust and probabilistic plans. The default behaviour for the robust planning algorithm in RayStation (v6.99) for this size of uncertainty only samples the nominal scenario (with no errors) and scenarios in the six cardinal directions, not any combinations of them, i.e., along the diagonals. The probabilistic planning algorithm weights the different scenarios sampled by their probability. The scenarios along the diagonals have the lowest probability of occurring so will be given less emphasis by the optimiser. This could explain why the corners of the cube (and other shapes with lower sphericity) are being cut off and the 95% isodose surface does not extend as far out as it perhaps should, explaining the lower CTV_{min} values for both robust and probabilistic plans.

Another reason for this could be the way the planning system represents the target shapes and OARs while planning. For each planning structure, every voxel in the dose grid is assigned a value from 0-1 depending on the percentage of that voxel that is within the specified structure. Voxels fully outside a given structure have a value of 0, voxels fully inside will have a value of 1 and voxels along the surface will have a value that is the fractional volume of the voxel that is within the structure. This representation of the structure used for the planning process is fairly rough and does not maintain the exact shape of the contour. This will not matter too much where the contour is smooth, but where there are 'corners', this representation is not perfect. Expanding the CTV by a specified margin to a PTV has the effect of smoothing out some of the 'corners' of the CTV. This means the imperfections in this way of representing planning structures will likely have less impact on margin planning which is based on the 'smoother' PTV contour than on robust or probabilistic planning which use the more 'jagged' CTV contour as the input.

Comparing the different planning approaches, for all shapes, the margin plan had the highest dose to the target in both CTV_{min} and the D99. However, the spread of the high dose region was largest for the margin plans, meaning there would be a higher dose to the normal tissue than for the robust and probabilistic plans. As always with planning, there is a trade-off between getting a higher dose to the target and limiting the dose to the rest of the patient. As to which of these planning approaches would be ‘better’, this would depend entirely on the scenario, and a clinical decision would have to be made as to whether it was worth sacrificing getting a high dose to a few voxels on the ‘corners’ of the CTV to spare the normal tissue from receiving as high a dose.

In order to create the plans for each shape, only a minimum dose constraint and both maximum and uniform target objective functions were used, with a dose fall off function. While these functions were based on realistic values, these plans were massively simplified due to the lack of specific OARs. In practise, there are typically many OARs that are considered in H&N patients which will have an effect on how well target coverage can be achieved and how far dose can be allowed to spread outside the target. However, these simplified plans can give an idea of how well the different planning approaches can perform on different shapes under optimal conditions. It stands to reason that a planning approach that struggles to achieve the required target coverage in this simplified scenario is likely to also struggle in a real patient with non-uniform tissue densities and OARs that have limits on how much dose should be applied.

An extension of this work would be to include some OARs to the cylindrical phantom, e.g., a simplified version of the brainstem, spinal cord or parotids, and re-run the plans with dose objectives on these to investigate whether target shape has an effect on the ability of the planning approaches to spare OARs as well as achieving coverage. However, this is beyond the current scope of this paper.

This study only looks at 12 different shapes – a sphere, a cube and then 10 real CTV shapes from H&N cancer patients. This is enough to show that there is a correlation between sphericity of the shape and both the minimum dose to the target and conformity. Future work could be done to investigate this link further by using a larger sample of different shapes with a wide range of sphericity and features. This could also be extended to consider

other measures for the complexity of the shape, e.g., deviation from the convex hull or the complexity measure suggested by Brinkhoff *et al.* [125].

5.5 Conclusion

All plans achieved 95% of the prescribed dose to the majority of the target, however some of the robust and probabilistic plans were seen to underdose a small portion (<1%) of the CTV. This effect was larger in CTV shapes with lower sphericity, likely due to the plans underdosing the 'corners' of the shape. However, the probabilistic and robust plans had better conformity than the margin plans and were closer to the intended robustness. It was seen that as the sphericity of the target decreased, in general plans performed worse for all three planning approaches. This suggests more care needs to be taken when planning on CTVs with low sphericity, especially when using probabilistic or robust planning.

5.6 Acknowledgements

We would like to acknowledge Ian Hinder for his work integrating the probabilistic treatment planning plugin into RayStation. Marcel van Herk was supported by NIHR Manchester Biomedical Research Centre. This work was supported by Cancer Research UK via funding to the Cancer Research Manchester Centre [C147/A25254].

Appendix 5A

5.A Supplementary material

5.A.1 High, medium and low dose spread

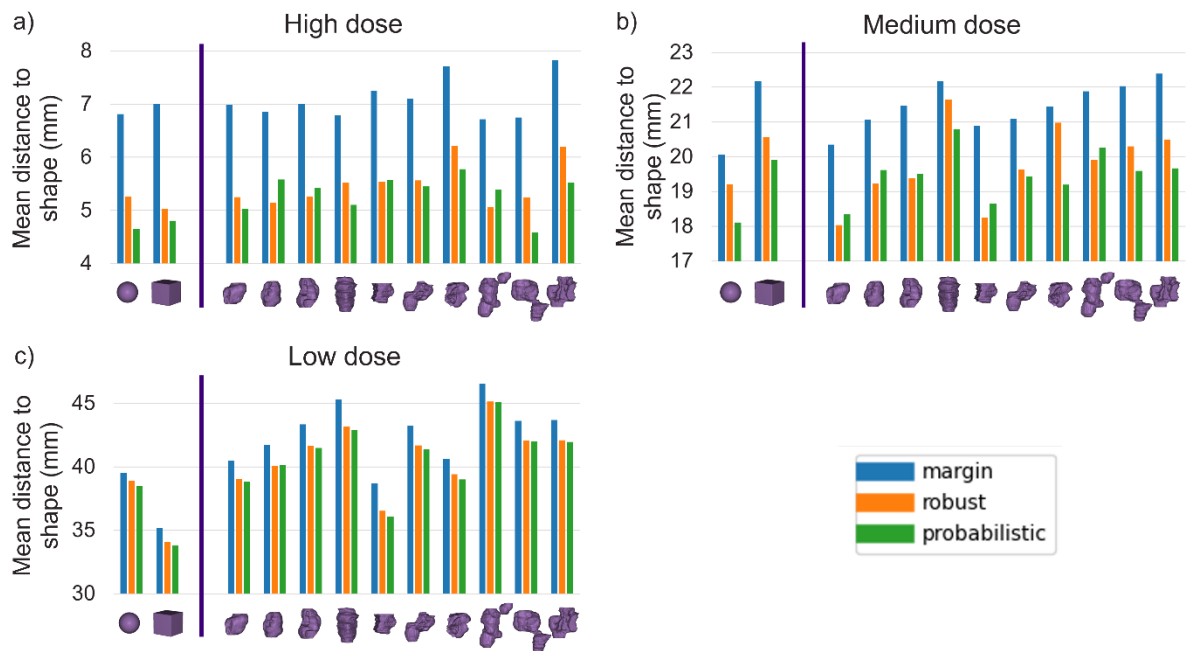


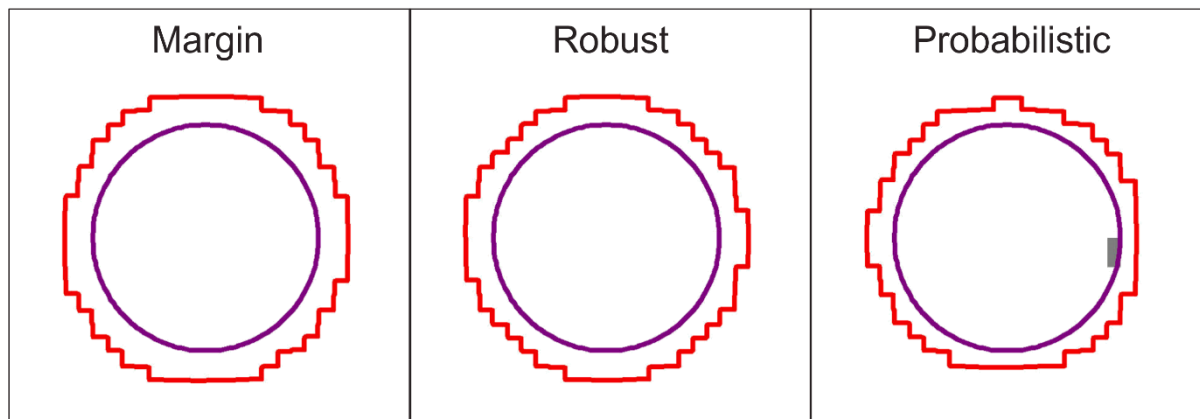
Figure 5.A.1: Bar charts showing the mean distance between the CTV surface and the a) 95% isodose surface, b) 50% isodose surface and c) the 10% isodose surface. Plans for the sphere and cube are on the left, separated off from the real H&N CTV shapes. The x-axis is labelled with a 3D rendering of the target shape used to create the plans. The margin plans are shown in blue, the robust in orange and the probabilistic in green.

Looking at the spread of the high, medium and low dose regions for each plan by considering the distance between the target shape and the 95%, 50% and 10% isodose surfaces (Figure 5.A.1), it can be seen that the margin plans had the largest dose spread. For the mean distance between the 95% isodose surface, the mean across all shapes was 7.1 mm, 5.5 mm and 5.2 mm for the margin, robust and probabilistic plans respectively. For the medium dose region, the means were 21.4 mm, 19.8 mm, 19.4 mm for the margin, robust and probabilistic plans respectively and for the low dose spread, 41.7 mm, 40.2 and 40.0 mm.

For each planning approach, there was a general trend that as the sphericity of the shape decreased, the high, medium and low dose regions spread further.

5.A.2 Isodose contours

Sphere ●



Cube ■

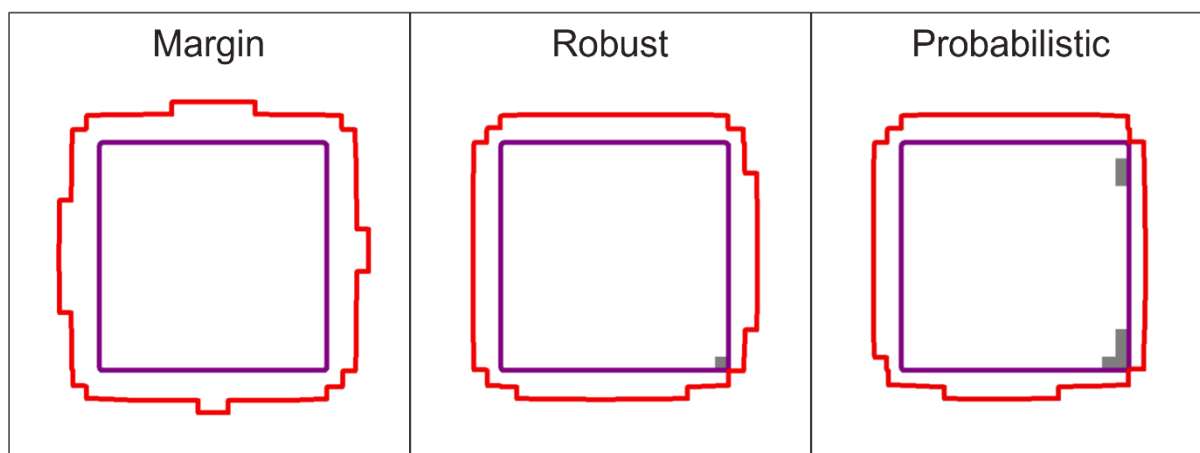
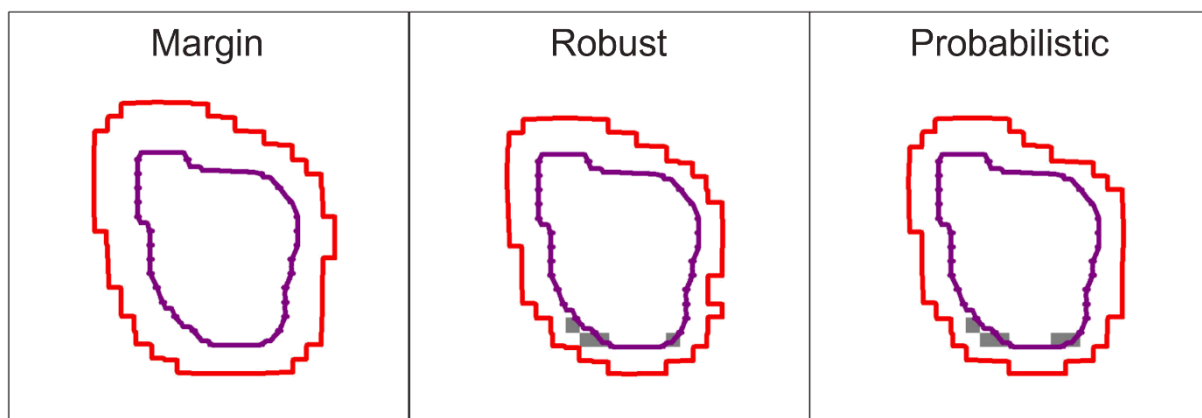


Figure 5.A.2: The CTV (purple) and the 95% isodose contours (red) for a single slice of each plan. The same slice is depicted for plans made using the three planning approaches for a given CTV shape, but different slices are shown for each shape. The highlighted regions show voxels within the CTV that had a final simulated treatment dose of less than 95% of the prescribed dose in any of 250 simulated treatments. Note: a voxel was considered within the CTV if any portion of the voxel fell within the 3D CTV contour – it may appear to be outside the contour in the figures as they are just 2D cross sections of the overall image.

CTV1



CTV2

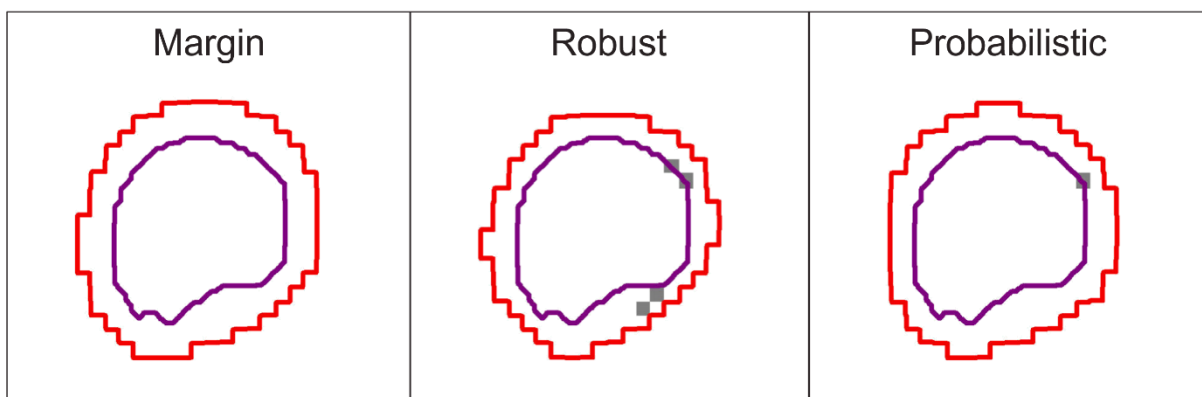
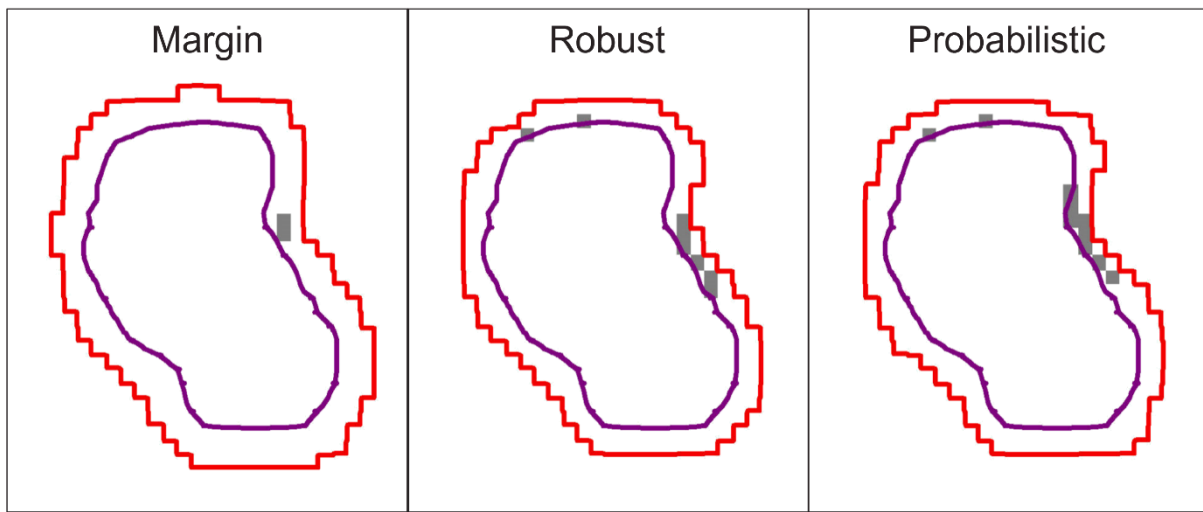


Figure 5.A.2 continued: The CTV (purple) and the 95% isodose contours (red) for a single slice of each plan. The same slice is depicted for plans made using the three planning approaches for a given CTV shape, but different slices are shown for each shape. The highlighted regions show voxels within the CTV that had a final simulated treatment dose of less than 95% of the prescribed dose in any of 250 simulated treatments. Note: a voxel was considered within the CTV if any portion of the voxel fell within the 3D CTV contour – it may appear to be outside the contour in the figures as they are just 2D cross sections of the overall image.

CTV3



CTV4

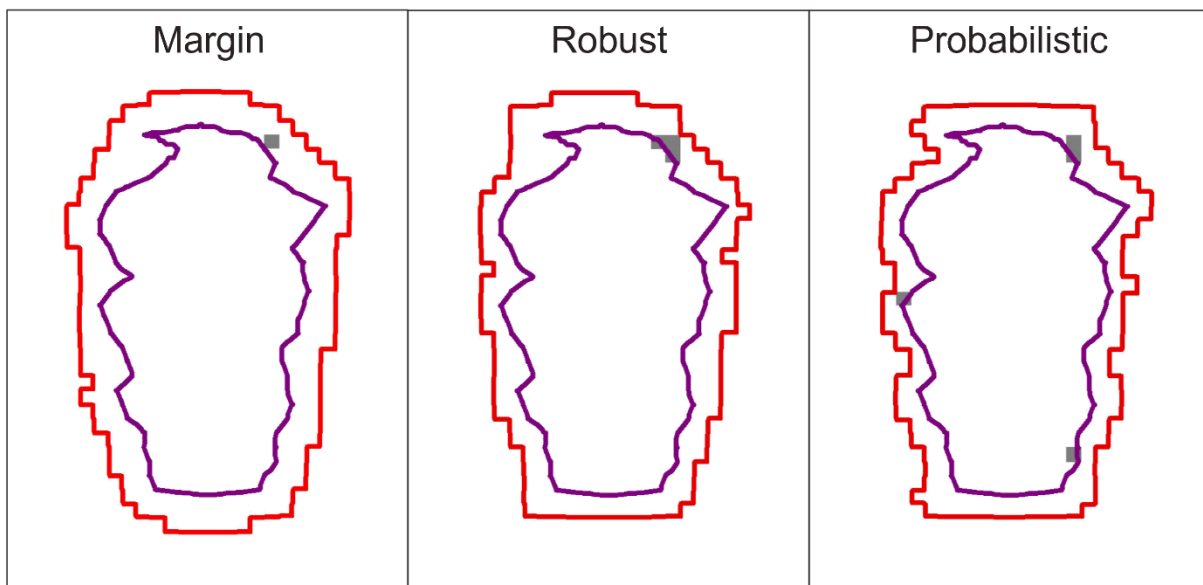
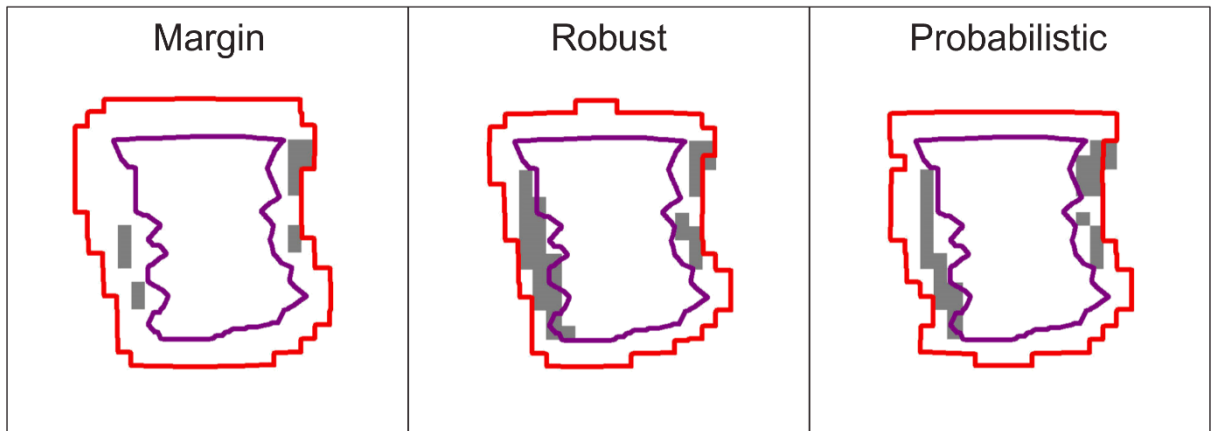


Figure 5.A.2 continued: The CTV (purple) and the 95% isodose contours (red) for a single slice of each plan. The same slice is depicted for plans made using the three planning approaches for a given CTV shape, but different slices are shown for each shape. The highlighted regions show voxels within the CTV that had a final simulated treatment dose of less than 95% of the prescribed dose in any of 250 simulated treatments. Note: a voxel was considered within the CTV if any portion of the voxel fell within the 3D CTV contour – it may appear to be outside the contour in the figures as they are just 2D cross sections of the overall image.

CTV5



CTV6

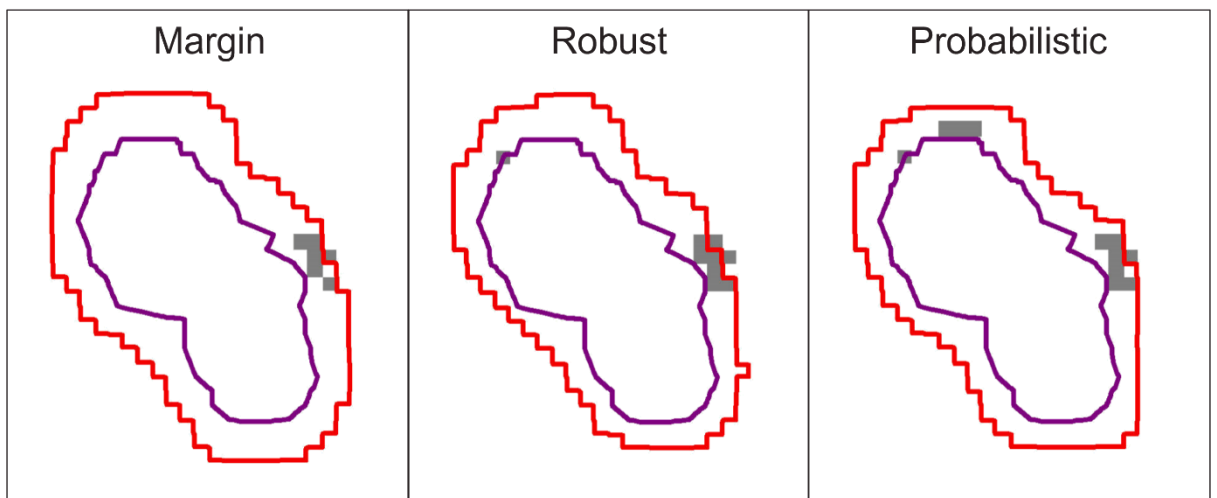
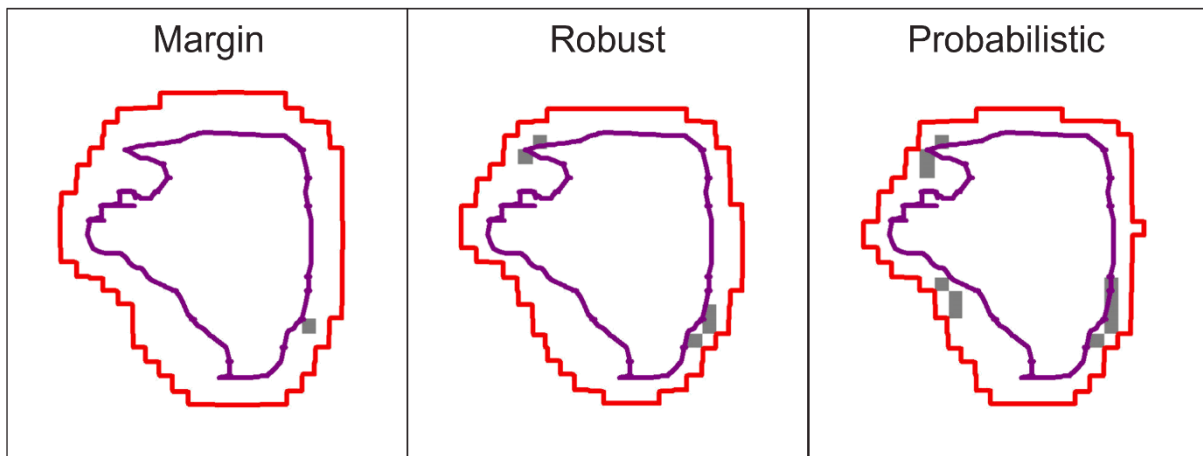


Figure 5.A.2 continued: The CTV (purple) and the 95% isodose contours (red) for a single slice of each plan. The same slice is depicted for plans made using the three planning approaches for a given CTV shape, but different slices are shown for each shape. The highlighted regions show voxels within the CTV that had a final simulated treatment dose of less than 95% of the prescribed dose in any of 250 simulated treatments. Note: a voxel was considered within the CTV if any portion of the voxel fell within the 3D CTV contour – it may appear to be outside the contour in the figures as they are just 2D cross sections of the overall image.

CTV7



CTV8

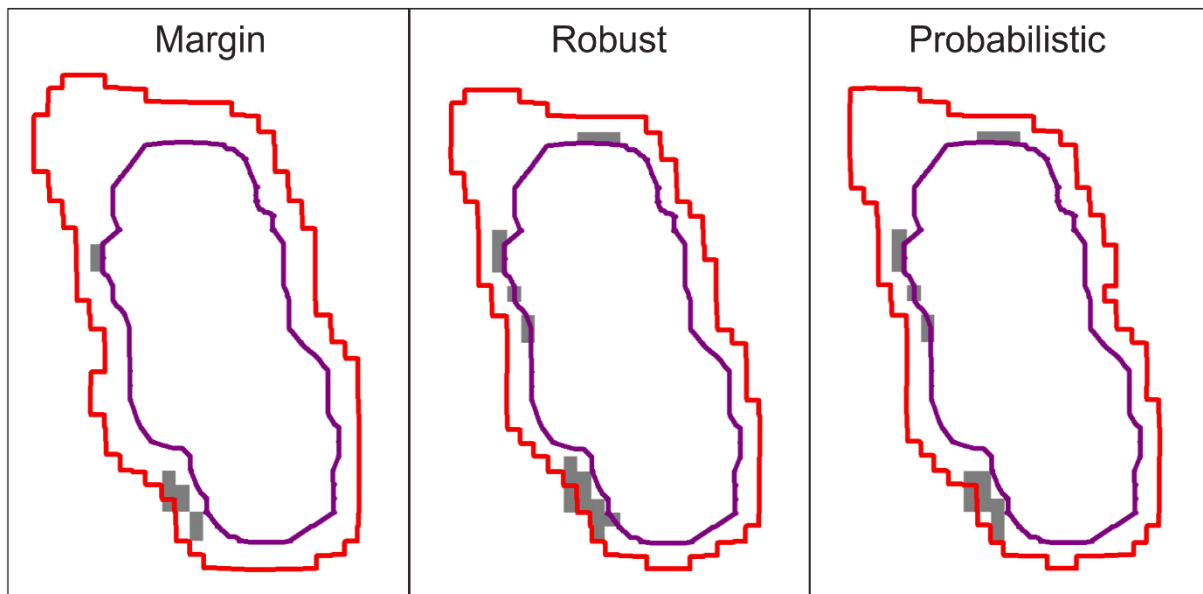
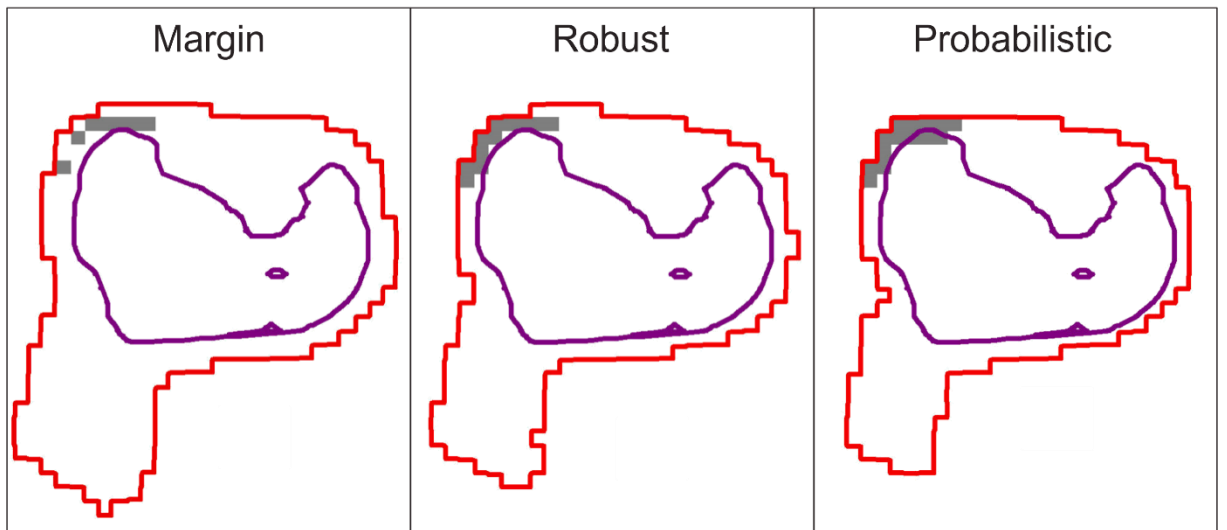


Figure 5.A.2 continued: The CTV (purple) and the 95% isodose contours (red) for a single slice of each plan. The same slice is depicted for the three planning approaches of a given CTV shape, but different slices are shown for each shape. The highlighted regions show voxels within the CTV that had a final simulated treatment dose of less than 95% of the prescribed dose in any of 250 simulated treatments. Note: a voxel was considered within the CTV if any portion of the voxel fell within the 3D CTV contour – it may appear to be outside the contour in the figures as they are just 2D cross sections of the overall image.

CTV9



CTV10

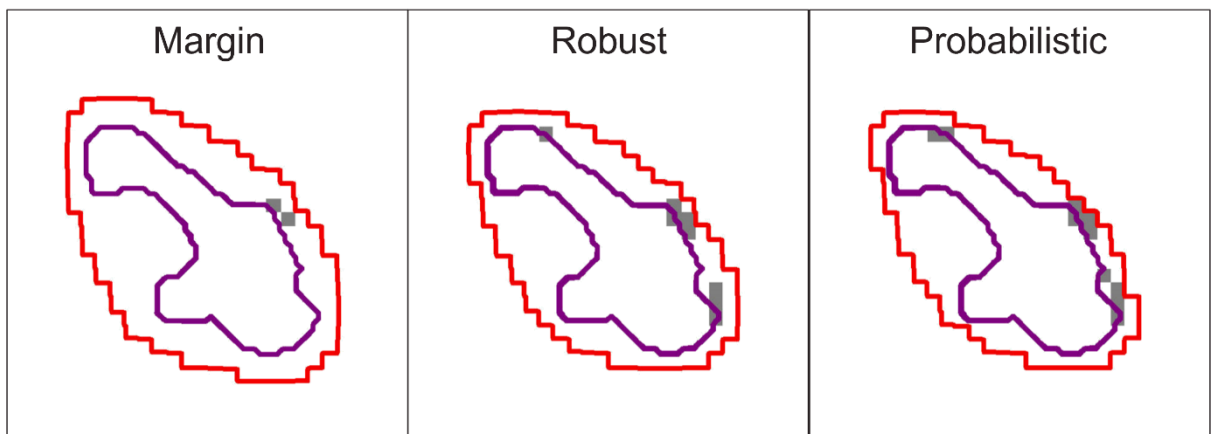


Figure 5.A.2 continued: The CTV (purple) and the 95% isodose contours (red) for a single slice of each plan. The same slice is depicted for the three planning approaches of a given CTV shape, but different slices are shown for each shape. The highlighted regions show voxels within the CTV that had a final simulated treatment dose of less than 95% of the prescribed dose in any of 250 simulated treatments. Note: a voxel was considered within the CTV if any portion of the voxel fell within the 3D CTV contour – it may appear to be outside the contour in the figures as they are just 2D cross sections of the overall image.

Blank page

Chapter 6

Discussion

The overall aim of this thesis was to understand geometrical uncertainties in H&N radiotherapy; how to model them, how to evaluate their impact and how to make plans that are robust against them. Uncertainties can cause underdosing of the target, making the treatment less effective or overdosing OARs, increasing the risk and severity of side effects for the patient. For H&N cancer patients, these side effects include xerostomia (dry mouth) and dysphagia (difficulty swallowing), which can have a large impact on the quality of life for the patient [5]–[7]. As such, it is important to improve the models of the uncertainties and assess the robustness of different plans to these variations. To achieve the overall aim of this thesis, a set of specific aims were defined:

1. Create and evaluate a time-dependent H&N anatomical deformation model including random and systematic components.
2. Evaluate the impact of set-up uncertainties and anatomical deformations on the dose to the target and OARs.
3. Compare different planning approaches to deal with set-up uncertainties and anatomical deformations.
4. Evaluate the impact of target shape on plan robustness.

The first aim was addressed by Chapters 2 and 3. Chapter 2 developed a method to evaluate the ability of PCA models to represent unseen changes in the patient or population. This chapter highlighted the fact that the oropharyngeal region in H&N cancer patients is particularly challenging to model using PCA. Chapter 3 created a set of time-dependent PCA models to describe both the systematic and random anatomical deformations occurring in a population of H&N cancer patients.

Chapter 3 went on to address the second aim by evaluating the dosimetric effect of both anatomical deformations and set-up uncertainties on the clinical plans for a set of H&N cancer patients. The results show that for most OARs, the anatomical deformations had a similar or smaller impact than set-up uncertainties.

This finding was then used to address the third aim in Chapter 4, investigating whether different planning approaches which were intended to account for set-up uncertainties could also account for anatomical deformations. This was done by creating margin, robust and probabilistic plans accounting for differing levels of set-up uncertainties. Treatment simulations for these plans were run, using the anatomical deformation models from Chapter 3. A 5 mm uncertainty setting was found to be sufficiently robust to anatomical deformations for the probabilistic plans, while a 3 mm uncertainty setting was sufficiently robust for the margin and robust plans. The third aim was also addressed in Chapter 5 where simple margin, robust and probabilistic plans were created for a variety of CTV shapes and the dosimetric effect of set-up uncertainties on these plans was evaluated. Here, margin plans were found to be the most robust to set-up uncertainties.

The final aim was addressed by Chapter 5, which investigated whether there was any link between plan robustness and the complexity of the target shape. The results show that plan robustness decreased as the CTV got less spherical, and both robust and probabilistic planning approaches struggled to reach a high enough dose to the target for more complex CTVs as the 'corners' of these shapes were underdosed. This implies that care should be taken when using robust and probabilistic planning to ensure an adequate dose to the target is reached for complex target shapes.

6.1 Comparison to literature

There have been many studies using PCA to model anatomical deformations. Some considered just the displacement of the surface contour of specific organs [65], [68], [69], while others considered the displacement of each voxel within the scan [67], [70], [74], [100], [101], as was done in Chapters 2 and 3. While the inputs for these two implementations are slightly different, the overall method for modelling deformations is the same, although depending on the intended application, one implementation may be preferable to the other. For example, considering the displacement of each individual voxel is more useful when the model is intended to be used for dose mapping and accumulation, as it was in Chapter 3.

Before PCA models can be used, they need to be first evaluated to see how well they represent actual deformations seen in a patient or population. Similar evaluation strategies

to the generalisability method proposed in Chapter 2 have been used previously in prostate [65], [69], [70] and lung [68], [100] patients, but they mainly evaluated the PCA model against the input DVFs instead of against unseen data. However, Badawi *et al.* [68] did create models using just a subset of their input data, which they evaluated against the unused DVFs in order to determine whether it was enough to use just one input scan per week of data or whether three per week were needed. Budiarto *et al.* [69] evaluated their population-based model on DVFs from three unseen patients, but not on any unseen DVFs from their training patients. Badawi *et al.* [68] and Söhn *et al.* [65] plotted heat maps of the local residual errors, similar to the M_{res} maps shown in Chapter 2.

Both Szeto *et al.* [100] and Vile [70] used a LOOCV, similar to the model robustness method proposed in Chapter 2. However, neither study evaluated residual errors locally; they just considered the scan as a whole.

While these studies have used similar methods to those proposed in Chapter 2, there are some key differences. Besides their lack of evaluation on unused DVFs to determine whether their models represented the full range of deformations seen within the patient or population of training patients, none of these studies considered the Z-score of the weights required to generate the closest DVF to the reference DVF that can be produced from the model. These Z-scores are important to ensure that the model can accurately describe not only the direction of the deformations but also their size. The Z-scores also give a value for the likelihood of the closest DVF being produced from the PCA model, which gives an indication as to whether the DVF being evaluated against is within the same distribution as the input DVFs. Also, none of these other studies investigated the sensitivity of the model by considering the effect adding Gaussian noise to the DVF had on the residual error.

Thornqvist *et al.* [103] developed a method for evaluating PCA models for anatomical deformations. They created 5,000-10,000 different patient geometries from their model and used these to create a coverage probability map for the different organs. From this, they compared specific iso-probability contours with the actual contours delineated on the different scans. They then used the sensitivity and precision of the iso-probability contours to evaluate the model.

In H&N, there have been a couple of studies using PCA to model anatomical deformations [71], [72], however no evaluation of the residual errors of the PCA models were conducted. While Chetvertkov *et al.* [72] looked at using their PCA model to see if they could predict time trends for the main modes of deformation, there was no explicit time-variation built into their models. In contrast, our population-based deformation model in Chapter 3 explicitly accounted for time by creating separate PCA models for each of the weeks of treatment. Pakela *et al.* [126] also explicitly built in a time-dependence to their H&N anatomical deformation models, however this was done by modelling the changes as either a Markov process or by using a quantum mechanics approach as opposed to PCA. None of these other H&N models were used to assess the dosimetric effect of the modelled changes on the patient, as was done in Chapter 3. This is an important step, because it is the actual effect of the deformations on the delivered dose that can determine how well they are currently accounted for. Importantly, we are the first to explicitly differentiate systematic and random components of anatomical deformations in H&N cancer, in a similar way to Vile [70] in the prostate. Szeto *et al.* [100] also modelled the random and systematic parts separately for their model of lung deformations, however, they used the voxel-wise standard deviation of all DVFs for each patient as the inputs for their model for the random variations. This distinction between the systematic and random errors is important as they will have different effects on the delivered dose, and so need to be handled differently when accounting for these uncertainties [76].

Along with conventional margin-based planning, robust and probabilistic planning have both been studied to account for set-up uncertainties in various sites [32], [33], [35]–[37], including H&N [34], [41]. These approaches have been shown to reduce OAR dose compared to margin plans while keeping a similar or improved level of target coverage. While there were no specific OARs considered in the plans in Chapter 5, the high dose region spread further into the normal tissue for the margin plans, suggesting that it might be harder to spare OARs if they had been considered. The robust and probabilistic plans were also seen to have a lower dose to the target than the margin plans which is not always the case in other studies conducted into robust and probabilistic planning. The reason for this discrepancy could be that other studies either re-normalise the plans once optimised to ensure a specific level of target coverage [34], or they adjust the margins or the weights

for the objective functions to ensure a similar level of target coverage between the two plans they are comparing [37], [41]. Neither of these was done for the studies in Chapters 4 and 5 to guarantee a fair comparison between the optimisation approaches.

We also found that the margin plans in Chapter 5 were overly robust, with the 90th percentile probability level of the minimum dose to the CTV being higher than 95% of the prescribed dose. This over-robustness of the margin plans has been seen before in clinical plans for prostate [81] and H&N [49] patients, suggesting that perhaps margins could be reduced to help spare OARs. This also agrees with the work in Chapter 4 which showed that the margin plans required a much smaller uncertainty setting than robust and probabilistic plans to reach the desired level of robustness to anatomical deformations.

The work in Chapter 5 investigated whether the complexity of the CTV shape affects the robustness of the plan. Witte *et al.* [78] investigated whether the size and tissue density of the target had any effect on the required margin size, and Zheng *et al.* [79] considered the effect of target shape as well as size on the necessary margins. Both these studies concluded that these factors would affect the size of margin expansion required to account for the random component of the uncertainties. However, both these studies evaluated mathematical derivations for margin recipes and actual plans were not created and analysed. To the best of the author's knowledge, there have been no other investigations into the effect of target shape on either robust or probabilistic planning, and we are the first to perform planning studies to investigate the effect of shape complexity on plan robustness.

The probabilistic planning algorithm used in Chapters 4 and 5 was the same implementation (but in a different treatment planning system) as the one used by Bohoslavsky *et al.* [36] and Fontanarosa *et al.* [41], although rotational uncertainties were not included in our study. In Chapters 4 and 5, all objective functions were kept the same between the margin and probabilistic plans, with the only difference being that target objectives were switched from PTV to CTV and were made probabilistic. For Chapter 4, the objective functions came from a class solution used at our institution, including minimum, maximum, uniform and minimum DVH target objectives, and Chapter 5 used minimum, maximum and uniform target objectives. However, Fontanarosa *et al.* [41] only used a minimum CTV dose objective function, and discarded any other objective functions relating

to the PTV. This was done to avoid possible problems due to the fact that each objective function is calculated independently, meaning that different error scenarios could be combined in the different cost functions for the same target. This is a similar concept to using an objective-wise vs composite worst-case optimisation in robust planning. We chose to follow the workflow used in the clinic at our institution as closely as possible, where multiple target objectives are used and therefore effects from this correlation problem will have been included in our results. In future, this impact of potential correlation problems could be investigated by adapting the probabilistic planning algorithm to treat all probabilistic objectives as a single overarching function and sample all the errors together and then comparing it to sampling each objective function individually.

Our implementation of probabilistic planning samples the errors at each point in the systematic error kernel (which has the same resolution as the dose grid), and then weights all the sampled scenarios by the probability of the scenario occurring. However, Moore *et al.* [38] randomly sampled systematic errors from a Gaussian distribution, and then summed the cost function of each scenario without explicitly weighting it by the probability. Their method works because the random sampling from the Gaussian distribution already takes the probability of the scenario occurring into account – with more likely scenarios being sampled more often. However, this technique requires a larger number of scenarios to be sampled to represent the probability distribution and would therefore take a very long time.

6.2 Limitations and future work

The PCA models in Chapters 2, 3 and 4 relied on DVFs produced by deformable image registration. While no formal quantitative evaluation of the registration uncertainties was performed, they will not be perfect. For instance, considering the volume changes of the OARs using the anatomical deformation model (Supplement 3.A.3), there was an overall systematic increase in volume seen in the brainstem, spinal cord and oral cavity and a decrease in the larynx, likely to do with differences in the CBCT image quality and CT image quality affecting the registrations. Veiga *et al.* [108] found that regions of low quality within the CBCT had larger variations in the registrations between different algorithms. Using data with better quality of on-treatment scans, or perhaps different deformable registration

algorithms could perhaps reduce registration uncertainties. For example, Veiga *et al.* [108] suggest that using a registration algorithm parameterized by a stationary velocity field could produce DVFs with more desirable physical properties.

Uncertainties will be present in both the intra- and inter-patient registrations. For the intra-patient registrations, these will cause uncertainties in the length and orientation of each vector within the DVF, and these uncertainties will be different for each scan. This will have the overall effect of adding noise to the models. The inter-patient registrations only effect the mapping of the DVFs into the average geometry. Any uncertainties in these registrations will manifest as uncertainties in the position of each vector in the DVF. This effect will remain the same for all DVFs from the same patient. These uncertainties will have the overall effect of blurring the model. Figure 6.1 shows a representation of the effect of these registration uncertainties for a vector within a DVF.

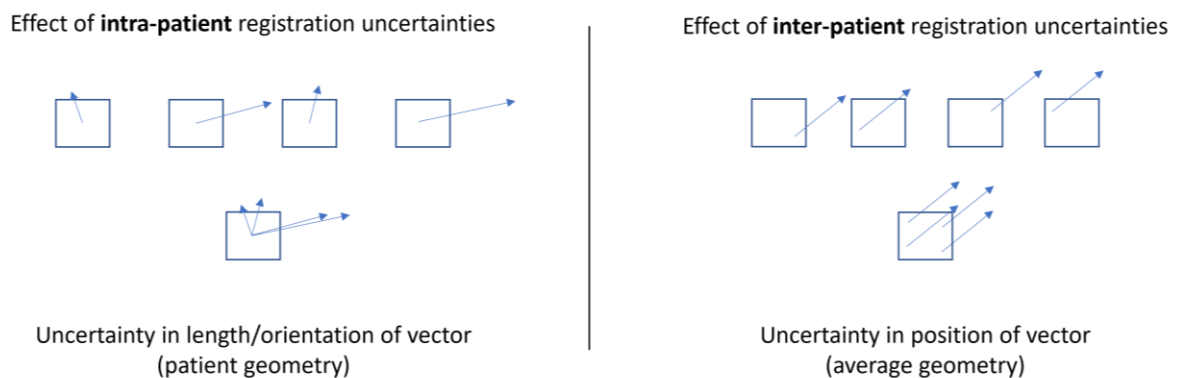


Figure 6.1: a representation of the effect of intra- (left) and inter-patient (right) registration uncertainties.

One way to get an idea of the regions of higher registration uncertainty is to take the average of all scans registered to a common reference scan [127]. Regions that are blurred or ‘fuzzy’ in this average scan are regions where the registrations were less consistent between the different scans. For each patient, an average of all the CBCTs registered to the pCT was calculated and an example of this for patient 4 is shown in Figure 6.2. Along with the average registered CBCT, a single registered CBCT is shown, to give an idea of the sharpness and quality of the individual scan for comparison.

For the inter-patient registrations, the average pCT is already the mean of each patient’s pCT registered to the common reference frame. This is shown in Figure 6.3, along with examples of 2 individual patient’s pCTs registered to the average geometry.

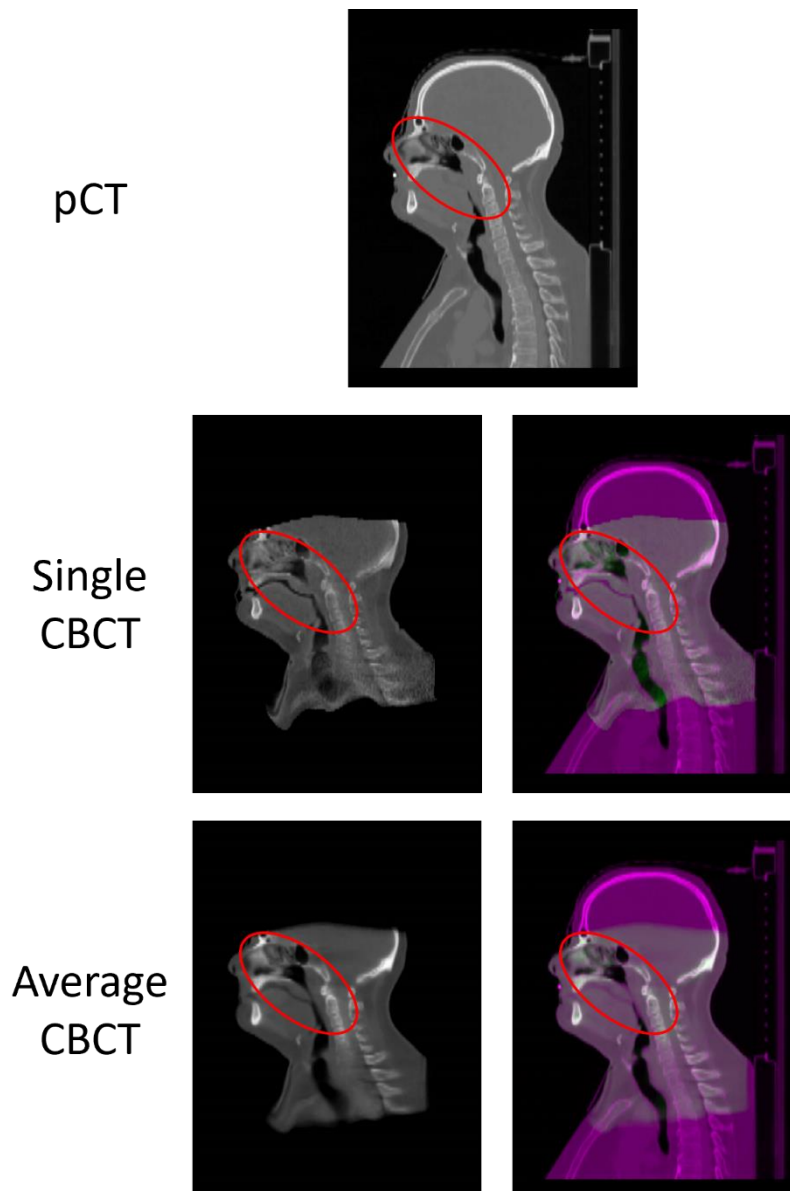


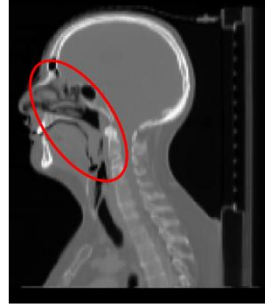
Figure 6.2: The pCT of patient 4 and the results of the intra-patient registrations for a single CBCT (middle row) and the mean of all CBCTs (bottom row). On the right, the CBCT and average CBCT (green) are superimposed over the pCT (pink). The region circled in red shows the area where the registrations were seen to struggle.

For both the intra- and inter-patient registrations, the areas of least consistency were shown to be around the air-tissue interfaces, especially in the oropharynx and the nasopharynx (see circled regions in Figures 6.2 and 6.3). This region is also the region that was found to have the largest residual error in the PCA models. It is possible that the registration uncertainties are contributing to the larger residual error, by blurring/adding noise to the PCA models, making it difficult to pick out the actual anatomical deformations taking place in this region.

Average
pCT



Patient
8



Patient
10



Figure 6.3: The average pCT and the results of the inter-patient registrations for patient 8 (middle row) and patient 10 (bottom row). On the right, the patient pCT (green) is superimposed with the average pCT (pink). The region circled in red shows the area where the registrations were seen to struggle.

In this thesis, the DVFs from the intra-patient registrations are mapped to the average patient geometry using the following equation

$$T_{pat_to_pop} \cdot T_{CBCT_to_pCT}, \quad (6.1)$$

where $T_{pat_to_pop}$ is the transformation from the patient pCT to the average pCT and $T_{CBCT_to_pCT}$ is the transformation from the CBCT to the pCT in the patient geometry. However, this only repositions the vectors in the new geometry; the length and orientation of the vectors remains unchanged. To include and changes in orientation/length of the vectors, the DVFs must also be transformed by the inverse of the transformation between the patient pCT to the average pCT [128], $T_{pat_to_pop}^{-1}$

$$T_{pat_to_pop} \cdot T_{CBCT_to_pCT} \cdot T_{pat_to_pop}^{-1}. \quad (6.2)$$

Adding this extra transformation ensures that the vectors are rescaled and reorientated accordingly to account for deformations between the patient and the average geometry. For example, if a patient is much larger than the average geometry, the vectors from the intra-patient deformations might be expected to be larger than average and so when warping to the average geometry, the DVFs from CBCT to pCT need to be rescaled to account for this.

To get an idea to the difference adding this extra transformation would make to the results in this thesis, one of the intra-patient DVFs from the final week of treatment for validation patient 2 was transformed into the average geometry using Equation 6.2. The population-based model from Chapter 2 was then evaluated against this DVF and the M_{res} values compared to those done previously, using Equation 6.1.

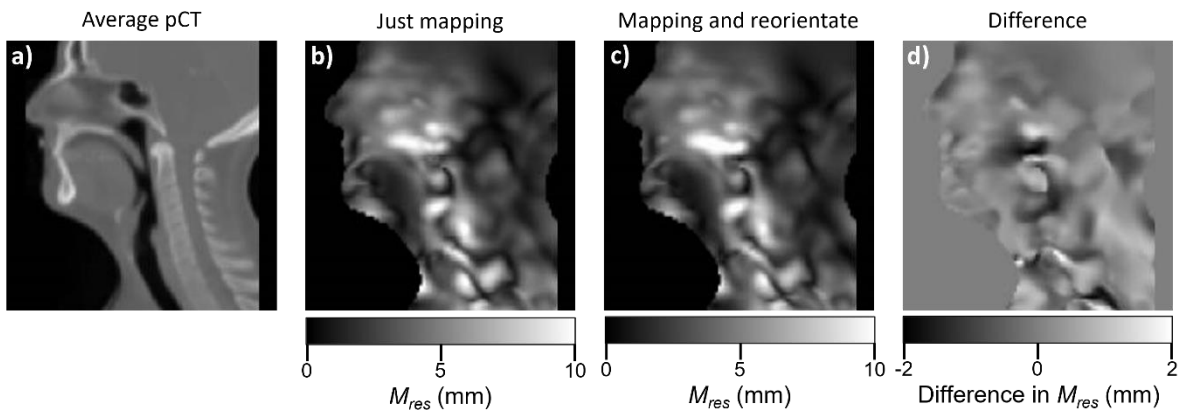


Figure 6.4: a) average planning CT. The M_{res} values for the population-based model in Chapter 2 evaluated against a DVF from week 6 for validation patient 2 where the DVFs were mapped into the average geometry by b) just mapping geometry and c) mapping and re-orientating the DVFs from the patient geometry. d) the difference in M_{res} values between the two methods.

Figure 6.4 shows the difference in the M_{res} values for a DVF from validation patient 2 between just mapping the vectors into the average patient geometry (Equation 6.1) and both mapping and reorientating the vectors (Equation 6.2). Across all voxels in the DVFs, the difference ranged from -3.6 to 3.2 mm, with a standard deviation of 0.4 mm. The difference between the M_{res}^{90} values was 0.08 mm.

The anatomical deformation models presented in Chapter 3 were population-based. This has the advantage that it can be used based on the first CT of a new patient prior to taking any additional scans. However, it is likely that not all patients experience the same deformations so the model could be more accurate for some patients than others. One

possible extension of this work could be to split patients into groups based on expected deformations, e.g., patients that are likely to experience large amounts of weight loss, or patients that may only have minimal parotid shrinking. Separate models could then be used for the different patient groups, and criteria should be developed to assign a new patient to the best fitting group. If a patient already has a few scans, the PCA evaluation method proposed in Chapter 2 could be used to identify which model has the lowest residual error and therefore best represents the likely deformations for that patient.

One limitation of the anatomical deformation models presented in Chapter 3 is the fact that the residual error was fairly high for the systematic model when evaluated against the validation patients, with a mean M_{res}^{90} of 4.6 mm. This could potentially be improved by training the model on more patients, as an increase in training patients was shown to decrease the residual errors (Supplement 3.A.2). For all models, the residual error was seen to be largest around the oropharynx. It is unlikely that increasing the number of training patients will completely eliminate the error in this region, as shown in Chapter 2. This is likely to be a limitation of PCA itself and models need to be carefully interpreted when used in this region. In Chapter 3, anatomical deformations were seen to have a larger effect than set-up uncertainties on the CTV and larynx dose. However, these are both within the region with higher residual error on the models, and so further work needs to be conducted to see if this is a real effect, or to do with the limitations of the modelling method itself.

To capture the long-term time trends, separate weekly PCA models were created. These models were independent from each other and so any treatment simulations would not take into account previous fractions when simulating the following fraction. This means that progressive changes may not be captured smoothly. For example, in one treatment simulation, a simulated DVF during week 1 may include a particularly large amount of parotid shrinking, while the following DVF simulated from the week 2 model may include experience a smaller level of shrinking. In reality, this is unlikely to happen, and the long-term changes are progressive throughout the treatment course. However, in this thesis, the models are not used to consider individual treatment courses. Rather, the models are used to look at the distribution of end points after multiple (1000) treatment courses are run. As such, while individual treatments themselves may experience unusual progressive

changes, this averages out over the 1000 treatments per patient and produces something much smoother and more realistic (see Figure 3.A.2).

However, in applications where simulating individual treatment courses would be required, these models could be adapted. This could be done by creating the weekly models using the residual deformation compared to the previous week. The models for week 1 would remain the same, but the input for subsequent models would be the residual DVF between the DVF from that week and the previous week. To run a simulated treatment course, a DVF would be simulated from the PCA model for week 1, as before. For each subsequent week, a residual DVF would be simulated from that weekly residual model, and this would be summed to the DVF from the previous week. This would ensure that any of the long-term changes were progressive and take into account the previous fraction.

The plans created in Chapter 4 were based on a class solution used at our institution and were not adapted for each specific patient. This means that these might not be the best possible plans for each patient. An extension to this study would be to have an experienced planner create a plan for the patient using say the 3 mm PTV margins, and then use the objectives and weights from this personalised plan for the other plans for that patient. This would ensure that the plans were of better quality and more realistic to the plans that might be clinically used.

The work in Chapter 5 massively simplified the challenge of planning radiotherapy for H&N patients. This was conducted in a cylindrical phantom, of uniform density, with only a single, central target and no specific OARs. While this gives an idea to the efficacy of the different planning approaches under ideal and simplified conditions, it does not give a realistic view of the complexity of H&N plans with multiple OARs near the target. A future direction of this work would be to extend this by approximating OARs within the phantom, e.g., the parotids or spinal cord, and see how target sphericity and planning approach affect OAR sparing. In addition, this analysis could also be run on plans created in a real patient geometry.

The probabilistic planning approach used in this thesis was stochastic optimisation and to the best of the author's knowledge, this approach has not been used before to directly account for anatomical deformations. There have been a couple studies using alternative

probabilistic planning approaches that have incorporated anatomical deformations into their plan optimisation [86], [99]. Baum *et al.* [86] used five pre-treatment CTs to create probability maps for the CTV, bladder and rectum for each of their prostate patients, which were then used in their CP optimisation. Tilly *et al.* [99] assumed single fraction treatments and randomly sampled 100 systematic errors from their PCA deformation model for cervical cancer patients for each iteration of the CVaR optimisation. Ideally, however, both systematic and random deformations would be incorporated into probabilistic planning and accounted for differently and neither of these studies do that.

My proposal to probabilistically account for anatomical deformations would be adapt current stochastic optimisation methods, using the systematic and weekly random PCA models developed in Chapter 3 as inputs.

- Firstly, the mean and components from each of the PCA models would need to be mapped from the average patient geometry to the geometry of the patient being planned.
- Each component l of a PCA model has a corresponding variance (λ_l) and the eigenvector of that component should be multiplied by $\sqrt{\lambda_l}$ to give the standard deviation of the displacement in each of the 3 directions for each voxel. This can then be used to create Gaussian error kernels for each voxel in each component of the different models.
- Considering random errors, the Gaussian error kernels for each component of the weekly PCA model should be convolved to create a total error kernel for each voxel for that week of treatment. If needed, a separate dose blurring could be performed for each weekly model. However, it is likely that they can be convolved to create a single blurred dose which could be used for random errors without too much loss in accuracy.
- To account for the systematic deformations, scenarios need to be sampled from the systematic PCA model. The number of sampled scenarios will depend on the desired accuracy of the model and any time limitations as more samples will take more time. For each component, regular steps should be sampled, e.g., sample every $0.5\sqrt{\lambda_l}$ (so multiply the eigenvector by $0, \pm 0.5\sqrt{\lambda_l}, \pm\sqrt{\lambda_l}$, etc.). For each component,

the probability of that scenario occurring will be the probability of selecting that scaling factor from a Gaussian distribution with standard deviation $\sqrt{\lambda_l}$. Different combinations of these steps across the different components should be sampled, by summing the scaled eigenvector of each component with the mean. For example, to sample the scenario where each component is scaled by the corresponding $\sqrt{\lambda_l}$, sum the weighted eigenvalues and the mean to get the overall DVF for that scenario. The probability of each scenario occurring will be the probability of that scaling factor for each component multiplied together. The sampled scenarios can be truncated to only include those that have a probability of at least 1% of the maximum probability, and then the probabilities of the remaining scenarios renormalised so the cumulative probability is 1 [36].

There are likely to be challenges in this implementation due to the large number of scenarios that will need to be sampled, which makes it much more computationally expensive than conventional margin-based planning. The run-time could be reduced by sampling fewer scenarios, e.g., only sampling every $\sqrt{\lambda_l}$ rather than every $0.5\sqrt{\lambda_l}$ for the above method for accounting for anatomical deformations. However, this will reduce the accuracy of the method and so a balance needs to be struck to ensure that enough scenarios are sampled to account for the full range of uncertainties present but that the optimisation runs in a reasonable amount of time. Exactly where this balance will be will depend on the specific situation and the requirement of the planner.

One potential way to reduce the computational time is by following the work of Moore *et al.* [38] where they initially started with 8 samples and ran the optimisation. Then, the number of samples is doubled, and the optimisation continued from the previous end point, until it had optimised with 128 samples. This means that a large number of scenarios was sampled overall, and so included in the final optimisation, but time was saved by not doing this at every single step in the optimisation. This will be faster than sampling the final number of scenarios at each iteration, however, it may make the system more unstable or perhaps more likely to get stuck in a local minimum and not find the overall optimal solution.

6.3 Clinical relevance

The aim of this thesis was to understand geometrical uncertainties in H&N radiotherapy; how to model them, how to evaluate their impact and how to make plans that are robust against them. The work in Chapters 3 and 4 has shown that anatomical deformations have an effect on the delivered dose to the patient and so should be taken into account when creating a plan. We have shown that, on average, treating anatomical deformations as set-up uncertainties allows the creation of robust plans, but this is not the case for every single patient. This is likely because different patients will have different tumour locations and slightly different geometries and so the planned dose distribution will be different for each patient and will therefore be affected differently by anatomical deformations. Ideally, a plan evaluation tool such as the one developed in this thesis should be integrated into the treatment planning system to show the planner the dosimetric effect of anatomical deformations and set-up errors. Such a tool would allow the planner to focus on regions that are most likely to be affected by anatomical deformations and allow them to improve the plan robustness and reduce the risk of overdosing OARs. The PCA evaluation method developed in Chapter 2 could also be used to highlight regions where the deformation models are less accurate and inform the planners of regions where more care needs to be taken. For example, if robust planning is being used, the robustness settings could be higher in the region where the PCA deformation models are less accurate.

The work in Chapter 5 has shown that when set-up uncertainties are taken into consideration, robust and probabilistic planning need to be used with care. While both methods have the potential to reduce dose to the normal tissue and possibly OARs, this can come at the cost of target coverage. This trade-off becomes more important for more complex target shapes, as the robust and probabilistic plans failed to achieve adequate target coverage on a few voxels in the 'corners' of the CTV. In these situations, a clinical decision would have to be taken to decide whether it is more important to deliver a high dose to every single voxel in the CTV, or if it is worth using robust or probabilistic planning and sacrificing those few corner voxels in order to reduce dose to the surrounding healthy tissue and OARs. In addition, robust and probabilistic planning approaches were seen to be less robust for anatomical deformations.

The work in Chapter 5 has also shown that margin plans in H&N may be over conservative and achieve more than the required level of target coverage when set-up uncertainties are taken into account in the plan evaluation. Similarly, the margin-based plans required a smaller uncertainty setting than robust or probabilistic plans to account for anatomical deformations. This means that potentially margins could be reduced which could help OAR sparing and reduce normal tissue dose, thus reducing the risk or severity of possible side effects for the patient.

In future work, planning systems should be developed that anticipate anatomical deformations and directly account for them in the planning process, i.e., using the method described in this discussion.

Chapter 7

Conclusions

In this thesis, a number of studies were presented with an overall aim to understand geometrical uncertainties in H&N radiotherapy; how to model them, how to evaluate their impact and how to make plans that are robust against them. The main findings in this thesis are:

- PCA models of anatomical deformations in H&N cancer patients struggle to represent the deformations seen around the oropharynx.
- For most organs, the dosimetric effect of anatomical deformations in oropharyngeal cancer patients is similar to or smaller than the effect of set-up uncertainties, except for the in the primary CTV and the larynx.
- On average, anatomical deformations can be accounted for by treating them as set-up uncertainties for all planning approaches using 3-5 mm uncertainty settings, but this is not the case for every single patient.
- For all planning approaches, increasing the uncertainty setting increases the plan robustness but also the OAR dose.
- Margin plans are more robust to set-up uncertainties and anatomical deformations than robust or probabilistic plans, but they spread the high dose region further beyond the CTV.
- When accounting for set-up uncertainties, robust and probabilistic plans do not achieve an adequate target dose for a few voxels on the 'corners' of more complex CTVs.
- When set-up uncertainties are taken into consideration, as the sphericity of the CTV decreases, plans get less robust for margin, robust and probabilistic planning.

Collectively, the work presented in this thesis indicate that both set-up uncertainties and anatomical deformations have dosimetric effects on H&N cancer patients and that current planning methods cannot fully account for these uncertainties in all patients. While robust and probabilistic planning can help reduce OAR dose, care needs to be taken when utilising them to ensure adequate target coverage is met, especially for more complex CTVs.

Blank page

Chapter 8

Publications and Presentations

8.1 Publications

8.1.1 Peer-reviewed

- R. Argota-Perez, **J. Robbins**, A. Green, M. van Herk, S. Korreman, and E. Vásquez-Osorio, “Evaluating principal component analysis models for representing anatomical changes in head and neck radiotherapy,” *Phys. Imaging Radiat. Oncol.*, vol. 22, pp. 13–19, Apr. 2022.

8.1.2 Non peer-reviewed

- **J. Robbins**, R. Argota-Perez, A. Green, M. van Herk, S. Korreman, and E. Vasquez Osorio, “OC-0363 Evaluation of how well a PCA model represents anatomical variations during H&N radiation treatment,” *Radiotherapy and Oncology*, vol. 161, pp. S267–S269, Aug. 2021.
- **J. Robbins**, E. Vásquez Osorio, A. Green, and M. van Herk, “PH-0045 Comparing robustness of margin and robustly optimised plans to anatomical deformations in H&N,” *Radiotherapy and Oncology*, vol. 161, pp. S19–S20, Aug. 2021.
- **J. Robbins**, M. Van Herk, A. Green, B. Eiben, A. McPartlin, and E. Vásquez Osorio, “PO-1650: Evaluating plan robustness for organ deformation and set-up uncertainties in head and neck cancer,” *Radiotherapy and Oncology*, vol. 152, pp. S905–S906, Nov. 2020.
- **J. Robbins**, E. Vásquez Osorio, A. Green, A. McWilliam, A. McPartlin, and M. Van Herk, “EP-1816 A robustness comparison of margin based and robust plans for head and neck VMAT patients,” *Radiotherapy and Oncology*, vol. 133, pp. S984–S985, Apr. 2019.

8.1.3 In preparation

- **J. Robbins**, M. Van Herk, B. Eiben, A. Green and E. Vásquez Osorio, “Probabilistic evaluation of plan quality for time-dependent anatomical deformations in head and neck cancer patients.”

- **J. Robbins**, M. Van Herk, A. Green and E. Vásquez Osorio, “The impact of target volume shape on plan robustness with different planning approaches.”
- **J. Robbins**, M. Van Herk, A. Green and E. Vásquez Osorio, “Can anatomical deformations in head and neck cancer be accounted for as set-up uncertainties?”

8.2 Presentations

- ESTRO 2022, *Is robust planning safe? A comparison of margin and robust photon plans on different target shapes*, May 2022, Copenhagen, **mini-oral presentation**.
- ESTRO 2021 Hybrid Meeting, *Evaluation of how well a PCA model represents anatomical variations during H&N radiation treatment*, August 2021, Virtual, **oral presentation**.
- ESTRO 2021 Hybrid Meeting, *Comparing robustness of margin and robustly optimised plans to anatomical deformations in H&N*, August 2021, Virtual, **mini-oral presentation**.
- ICCR 19: The 19th International Conference on the Use of Computers in Radiation Therapy, *Modelling the effect of time varying organ deformations in head and neck cancer using a PCA model*, June 2019, Montreal, **oral presentation**.
- ESTRO 2020 Virtual Meeting, *Evaluating plan robustness for organ deformation and set-up uncertainties in head and neck cancer*, November 2020, Virtual, **poster**.
- ESTRO 2019, *A robustness comparison of margin based and robust plans for head and neck VMAT patients*, April 2019, Milan, **poster**.

Bibliography

- [1] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, "The ever-increasing importance of cancer as a leading cause of premature death worldwide," *Cancer*, vol. 127, no. 16, pp. 3029–3030, Aug. 2021.
- [2] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, May 2021.
- [3] J. Van Loon, J. Grutters, and F. Macbeth, "Evaluation of novel radiotherapy technologies: what evidence is needed to assess their clinical and cost effectiveness, and how should we get it?," *Lancet Oncol.*, vol. 13, no. 4, pp. e169–e177, Apr. 2012.
- [4] J. M. Borrás *et al.*, "The impact of cancer incidence and stage on optimal utilization of radiotherapy: Methodology of a population based analysis by the ESTRO-HERO project," *Radiother. Oncol.*, vol. 116, no. 1, pp. 45–50, Jul. 2015.
- [5] P. Strojanc *et al.*, "Treatment of late sequelae after radiotherapy for head and neck cancer," *Cancer Treat. Rev.*, vol. 59, p. 79, Sep. 2017.
- [6] V. Dilalla, G. Chaput, T. Williams, and K. Sultanem, "Radiotherapy side effects: integrating a survivorship clinical lens to better serve patients," *Curr. Oncol.*, vol. 27, no. 2, pp. 107–112, 2020.
- [7] S. A. Bhide and C. M. Nutting, "Advances in radiotherapy for head and neck cancer," *Oral Oncol.*, vol. 46, no. 6, pp. 439–441, Jun. 2010.
- [8] Z. Stauch *et al.*, "An evaluation of adaptive planning by assessing the dosimetric impact of weight loss throughout the course of radiotherapy in bilateral treatment of head and neck cancer patients," *Med. Dosim.*, vol. 45, no. 1, pp. 52–59, Mar. 2020.
- [9] J. L. Barker *et al.*, "Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 59, no. 4, pp. 960–970, Jul. 2004.

- [10] E. K. Hansen, M. K. Bucci, J. M. Quivey, V. Weinberg, and P. Xia, "Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 2, pp. 355–362, Feb. 2006.
- [11] H. C. Y. Cheng *et al.*, "A prospective study on volumetric and dosimetric changes during intensity-modulated radiotherapy for nasopharyngeal carcinoma patients," *Radiother. Oncol.*, vol. 104, no. 3, pp. 317–323, Sep. 2012.
- [12] X. Wang *et al.*, "Anatomic and dosimetric changes during the treatment course of intensity-modulated radiotherapy for locally advanced nasopharyngeal carcinoma," *Med. Dosim.*, vol. 35, no. 2, pp. 151–157, Jun. 2010.
- [13] C. Han, Y. J. Chen, A. Liu, T. E. Schultheiss, and J. Y. C. Wong, "Actual Dose Variation of Parotid Glands and Spinal Cord for Nasopharyngeal Cancer Patients During Radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 70, no. 4, pp. 1256–1262, Mar. 2008.
- [14] J. L. Robar *et al.*, "Spatial and Dosimetric Variability of Organs at Risk in Head-and-Neck Intensity-Modulated Radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, no. 4, pp. 1121–1130, Jul. 2007.
- [15] S. A. Bhide *et al.*, "Weekly Volume and Dosimetric Changes During Chemoradiotherapy With Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Observational Study," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 76, no. 5, pp. 1360–1368, Apr. 2010.
- [16] A. Fiorentino *et al.*, "Parotid gland volumetric changes during intensity-modulated radiotherapy in head and neck cancer," *Br. J. Radiol.*, vol. 85, no. 1018, pp. 1415–1419, Oct. 2012.
- [17] C. Lee *et al.*, "Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration," *Radiother. Oncol.*, vol. 89, no. 1, pp. 81–88, Oct. 2008.
- [18] E. M. Vásquez Osorio, M. S. Hoogeman, A. Al-Mamgani, D. N. Teguh, P. C. Levendag, and B. J. M. Heijmen, "Local Anatomic Changes in Parotid and Submandibular Glands During Radiotherapy for Oropharynx Cancer and Correlation With Dose, Studied in

Detail With Nonrigid Registration," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 70, no. 3, pp. 875–882, 2008.

- [19] P. H. Ahn *et al.*, "Random Positional Variation Among the Skull, Mandible, and Cervical Spine With Treatment Progression During Head-and-Neck Radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 73, no. 2, pp. 626–633, Feb. 2009.
- [20] H. Tachibana, K. Motegi, and S. Moriya, "Impact of shoulder deformation on volumetric modulated arc therapy doses for head and neck cancer," *Phys. Med.*, vol. 53, pp. 118–128, Sep. 2018.
- [21] E. Neubauer *et al.*, "Assessment of shoulder position variation and its impact on IMRT and VMAT doses for head and neck cancer," *Radiat. Oncol.*, vol. 7, no. 1, Feb. 2012.
- [22] C. Chen, X. Lin, J. Pan, Z. Fei, L. Chen, and P. Bai, "Is it necessary to repeat CT imaging and replanning during the course of intensity-modulated radiation therapy for locoregionally advanced nasopharyngeal carcinoma?," *Jpn. J. Radiol.*, vol. 31, no. 9, pp. 593–599, Sep. 2013.
- [23] M. Beltran *et al.*, "Dose variations in tumor volumes and organs at risk during IMRT for head-and-neck cancer," *J. Appl. Clin. Med. Phys.*, vol. 13, no. 6, pp. 101–111, Nov. 2012.
- [24] S. van de Water, I. van Dam, D. R. Schaart, A. Al-Mamgani, B. J. M. Heijmen, and M. S. Hoogeman, "The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probability in intensity-modulated proton therapy for oropharyngeal cancer patients," *Radiother. Oncol.*, vol. 120, no. 1, pp. 56–62, 2016.
- [25] W. Liu, S. J. Frank, X. Li, Y. Li, R. X. Zhu, and R. Mohan, "PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization," *Med. Phys.*, vol. 40, no. 2, 2013.
- [26] W. Liu, Y. Li, X. Li, W. Cao, and X. Zhang, "Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques," *Med. Phys.*, vol. 39, no. 6, pp. 3089–3101, 2012.

- [27] W. Liu *et al.*, “Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers,” *Med. Phys.*, vol. 40, no. 5, 2013.
- [28] W. Liu *et al.*, “Impact of Respiratory Motion on Worst-Case-Scenario Optimized Intensity-Modulated Proton Therapy for Lung Cancers,” *Pract. Radiat. Oncol.*, vol. 5, no. 2, pp. 77–86, 2015.
- [29] H. Li *et al.*, “Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients,” *Radiother. Oncol.*, vol. 114, no. 3, pp. 367–372, 2015.
- [30] M. Stuschke, A. Kaiser, C. Pöttgen, W. Lübcke, and J. Farr, “Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to intensity modulated photon plans,” *Radiother. Oncol.*, vol. 104, no. 1, pp. 45–51, 2012.
- [31] W. Liu, X. Zhang, Y. Li, and R. Mohan, “Robust optimization of intensity modulated proton therapy,” *Med. Phys.*, vol. 39, no. 2, pp. 1079–1091, 2012.
- [32] A. K. Exeli *et al.*, “Cerebral cortex dose sparing for glioblastoma patients: IMRT versus robust treatment planning,” *Radiat. Oncol.*, vol. 13, no. 1, pp. 4–11, 2018.
- [33] X. Zhang *et al.*, “Robust optimization in lung treatment plans accounting for geometric uncertainty,” *J. Appl. Clin. Med. Phys.*, 2018.
- [34] D. Wagenaar, R. G. J. Kierkels, J. Free, J. A. Langendijk, S. Both, and E. W. Korevaar, “Composite minimax robust optimization of VMAT improves target coverage and reduces non-target dose in head and neck cancer patients,” *Radiother. Oncol.*, vol. 136, pp. 71–77, Jul. 2019.
- [35] C. A. Jensen, A. M. A. Roa, M. Johansen, J. Å. Lund, and J. Frengen, “Robustness of VMAT and 3DCRT plans toward setup errors in radiation therapy of locally advanced left-sided breast cancer with DIBH,” *Phys. Medica*, vol. 45, pp. 12–18, Jan. 2018.
- [36] R. Bohoslavsky, M. G. Witte, T. M. Janssen, and M. Van Herk, “Probabilistic objective functions for margin-less IMRT planning,” *Phys. Med. Biol.*, vol. 58, no. 11, pp. 3563–3580, 2013.

- [37] M. G. Witte, J. Van Der Geer, C. Schneider, J. V. Lebesque, M. Alber, and M. Van Herk, "IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP," *Med. Phys.*, vol. 34, no. 9, pp. 3544–3555, 2007.
- [38] J. A. Moore, J. J. Gordon, M. Anscher, J. Silva, and J. V. Siebers, "Comparisons of treatment optimization directly incorporating systematic patient setup uncertainty with a margin-based approach," *Med. Phys.*, vol. 39, no. 2, pp. 1102–1111, 2012.
- [39] J. Unkelbach, T. Bortfeld, B. C. Martin, and M. Soukup, "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning," *Med. Phys.*, vol. 36, no. 1, pp. 149–163, 2009.
- [40] A. Trofimov, J. Unkelbach, T. DeLaney, and T. Bortfeld, "Visualization of a variety of possible dosimetric outcomes in radiation therapy using dose-volume histogram bands," *Pract. Radiat. Oncol.*, vol. 2, no. 3, pp. 164–171, 2012.
- [41] D. Fontanarosa *et al.*, "An in silico comparison between margin-based and probabilistic target-planning approaches in head and neck cancer patients," *Radiother. Oncol.*, vol. 109, no. 3, pp. 430–436, 2013.
- [42] J. Unkelbach *et al.*, "Robust radiotherapy planning," *Phys. Med. Biol.*, vol. 63, no. 22, pp. 1–46, Nov. 2018.
- [43] M. Sharma, E. Weiss, and J. V. Siebers, "Dose deformation-invariance in adaptive prostate radiation therapy: Implication for treatment simulations," *Radiother. Oncol.*, vol. 105, no. 2, pp. 207–213, Nov. 2012.
- [44] A. J. Lomax, "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: The potential effects of inter-fraction and inter-field motions," *Phys. Med. Biol.*, vol. 53, no. 4, pp. 1027–1042, 2008.
- [45] D. Pflugfelder, J. J. Wilkens, and U. Oelfke, "Worst case optimization: A method to account for uncertainties in the optimization of intensity modulated proton therapy," *Phys. Med. Biol.*, vol. 53, no. 6, pp. 1689–1700, 2008.
- [46] J. de Leon *et al.*, "Analysis of data to Advance Personalised Therapy with MR-Linac (ADAPT-MRL)," *Clin. Transl. Radiat. Oncol.*, vol. 31, p. 64, Nov. 2021.

- [47] J. J. W. Lagendijk, B. W. Raaymakers, and M. van Vulpen, "The Magnetic Resonance Imaging–Linac System," *Semin. Radiat. Oncol.*, vol. 24, no. 3, pp. 207–209, Jul. 2014.
- [48] T. Craig, J. Battista, and J. Van Dyk, "Limitations of a convolution method for modeling geometric uncertainties in radiation therapy. I. The effect of shift invariance," *Med. Phys.*, vol. 30, no. 8, pp. 2001–2011, Aug. 2003.
- [49] J. V. Siebers, P. J. Keall, Q. Wu, J. F. Williamson, and R. K. Schmidt-Ullrich, "Effect of patient setup errors on simultaneously integrated boost head and neck IMRT treatment plans," *Int. J. Radiat. Oncol.*, vol. 63, no. 2, pp. 422–433, Oct. 2005.
- [50] M. van Herk, P. Remeijer, and J. V Lebesque, "Inclusion of geometric uncertainties in treatment plan evaluation," *Int. J. Radiat. Oncol.*, vol. 52, no. 5, pp. 1407–1422, Apr. 2002.
- [51] J. Leong, "Implementation of random positioning error in computerised radiation treatment planning systems as a result of fractionation," *Phys. Med. Biol*, vol. 32, no. 3, pp. 327–334, 1987.
- [52] S. D. Mccarter and W. A. Beckham, "Evaluation of the validity of a convolution method for incorporating tumour movement and set-up variations into the radiotherapy treatment planning system," *Phys. Med. Biol*, vol. 45, pp. 923–931, 2000.
- [53] J. J. Gordon and J. V. Siebers, "Convolution method and CTV-to-PTV margins for finite fractions and small systematic errors," *Phys. Med. Biol.*, vol. 52, no. 7, pp. 1967–1990, Apr. 2007.
- [54] S. van Kranen, S. van Beek, C. Rasch, M. van Herk, and J. J. Sonke, "Setup Uncertainties of Anatomical Sub-Regions in Head-and-Neck Cancer Patients After Offline CBCT Guidance," *Int. J. Radiat. Oncol.*, vol. 73, no. 5, pp. 1566–1573, Apr. 2009.
- [55] C. L. Brouwer, R. J. H. M. Steenbakkers, J. A. Langendijk, and N. M. Sijtsema, "Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help?," *Radiotherapy and Oncology*, vol. 115,

- no. 3. Elsevier Ireland Ltd, pp. 285–294, 01-Jun-2015.
- [56] W. R. Crum, T. Hartkens, and D. L. G. Hill, “Non-rigid image registration: theory and practice,” *Br. J. Radiol.*, vol. 77, pp. 140–153, Jan. 2014.
- [57] O. Zachariadis *et al.*, “Accelerating B-spline interpolation on GPUs: Application to medical image registration,” *Comput. Methods Programs Biomed.*, vol. 193, p. 105431, Sep. 2020.
- [58] J. Shackelford, N. Kandasamy, and G. Sharp, “Unimodal B-Spline Registration,” *High Perform. Deform. Image Regist. Algorithms Manycore Process.*, pp. 13–43, 2013.
- [59] K. K. Brock, S. Mutic, T. R. McNutt, H. Li, and M. L. Kessler, “Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132,” *Med. Phys.*, vol. 44, no. 7, pp. e43–e76, Jul. 2017.
- [60] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, “Iconic feature based nonrigid registration: the PASHA algorithm,” *Comput. Vis. Image Underst.*, vol. 89, no. 2–3, pp. 272–298, Feb. 2003.
- [61] M. Modat *et al.*, “Fast free-form deformation using graphics processing units,” *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 278–284, Jun. 2010.
- [62] G. S. Mageras *et al.*, “A method of incorporating organ motion uncertainties into three-dimensional conformal treatment plans,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 35, no. 2, pp. 333–342, 1996.
- [63] M. S. Hoogeman, M. Van Herk, D. Yan, L. J. Boersma, P. C. M. Koper, and J. V. Lebesque, “A model to simulate day-to-day variations in rectum shape,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 54, no. 2, pp. 615–625, 2002.
- [64] C. Lorenz and N. Krahnstöver, “Generation of point-based 3D statistical shape models for anatomical objects,” *Comput. Vis. Image Underst.*, vol. 77, no. 2, pp. 175–191, 2000.
- [65] M. Söhn, M. Birkner, D. Yan, and M. Alber, “Modelling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and

- evaluation," *Phys. Med. Biol.*, vol. 50, no. 24, pp. 5893–5908, 2005.
- [66] M. Söhn, B. Sobotta, and M. Alber, "Dosimetric treatment course simulation based on a statistical model of deformable organ motion," *Phys. Med. Biol.*, vol. 57, no. 12, pp. 3693–3709, 2012.
- [67] S. Thörnqvist *et al.*, "Treatment simulations with a statistical deformable motion model to evaluate margins for multiple targets in radiotherapy for high-risk prostate cancer," *Radiother. Oncol.*, vol. 109, no. 3, pp. 344–349, 2013.
- [68] A. M. Badawi, E. Weiss, W. C. Sleeman, C. Yan, and G. D. Hugo, "Optimizing principal component models for representing interfraction variation in lung cancer radiotherapy," *Med. Phys.*, vol. 37, no. 9, pp. 5080–5091, 2010.
- [69] E. Budiarto *et al.*, "A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases," *Phys. Med. Biol.*, vol. 56, no. 4, pp. 1045–1061, 2011.
- [70] D. J. Vile, "Statistical modeling of interfractional tissue deformation and its application in radiation therapy planning," Virginia Commonwealth University, 2015.
- [71] P. Tsiamas *et al.*, "Principal component analysis modeling of Head-and-Neck anatomy using daily Cone Beam-CT images," *Med. Phys.*, vol. 45, no. 12, pp. 5366–5375, 2018.
- [72] M. A. Chetvertkov *et al.*, "Use of regularized principal component analysis to model anatomical changes during head and neck radiation therapy for treatment adaptation and response assessment," *Med. Phys.*, vol. 43, no. 10, pp. 5307–5319, Oct. 2016.
- [73] I. T. Jolliffe, *Principal Component Analysis*, Second Edi. Springer, 2002.
- [74] R. Li *et al.*, "On a PCA-based lung motion model," *Phys Med Biol*, vol. 56, no. 18, pp. 6009–6030, 2011.
- [75] J. C. Stroom, H. C. J. De Boer, H. Huizenga, and A. G. Visser, "Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 43, no. 4, pp. 905–919, 1999.

- [76] 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," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, no. 4, pp. 1121–1135, 2000.
- [77] A. McKenzie, M. Van Herk, and B. Mijnheer, "Margins for geometric uncertainty around organs at risk in radiotherapy," *Radiother. Oncol.*, vol. 62, no. 3, pp. 299–307, 2002.
- [78] M. G. Witte, J. Van Der Geer, C. Schneider, J. V. Lebesque, and M. Van Herk, "The effects of target size and tissue density on the minimum margin required for random errors," *Med. Phys.*, vol. 31, no. 11, pp. 3068–3079, Nov. 2004.
- [79] B. Zheng, Z. Huang, and J. Li, "The Effects of the Shape and Size of the Clinical Target Volume on the Planning Target Volume Margin," *PLoS One*, vol. 9, no. 10, p. e109244, Oct. 2014.
- [80] J. Gordon and J. Siebers, "Evaluation of Dosimetric Margins in Prostate IMRT Treatment Plans," *Med. Phys.*, vol. 35, no. 2, pp. 569–575, 2008.
- [81] J. J. Gordon, A. J. Crimaldi, M. Hagan, J. Moore, and J. V. Siebers, "Evaluation of clinical margins via simulation of patient setup errors in prostate IMRT treatment plans," *Med. Phys.*, vol. 34, no. 1, pp. 202–214, Jan. 2007.
- [82] J. V. Lebesque and R. B. Keus, "The simultaneous boost technique: the concept of relative normalized total dose," *Radiother. Oncol.*, vol. 22, no. 1, pp. 45–55, 1991.
- [83] A. Fredriksson, A. Forsgren, and B. Hårdemark, "Minimax optimization for handling range and setup uncertainties in proton therapy," *Med. Phys.*, vol. 38, no. 3, pp. 1672–1684, 2011.
- [84] A. Fredriksson and R. Bokrantz, "A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning," *Med. Phys.*, vol. 41, no. 8, 2014.
- [85] W. Chen *et al.*, "Including robustness in multi-criteria optimization for intensity-modulated proton therapy," *Phys. Med. Biol.*, vol. 57, no. 3, pp. 591–608, 2012.
- [86] C. Baum, M. Alber, M. Birkner, and F. Nüsslin, "Robust treatment planning for

intensity modulated radiotherapy of prostate cancer based on coverage probabilities," *Radiother. Oncol.*, vol. 78, no. 1, pp. 27–35, 2006.

- [87] A. Fredriksson, "A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization," *Med. Phys.*, vol. 39, no. 8, pp. 5169–5181, 2012.
- [88] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, no. 3, pp. 21–41, 2000.
- [89] Y. Li *et al.*, "Selective robust optimization: A new intensity-modulated proton therapy optimization strategy," *Med. Phys.*, vol. 42, no. 8, pp. 4840–4847, 2015.
- [90] F. Albertini, E. B. Hug, and A. J. Lomax, "Is it necessary to plan with safety margins for actively scanned proton therapy?," *Phys. Med. Biol.*, vol. 56, no. 14, pp. 4399–4413, 2011.
- [91] H. A. Loebner *et al.*, "Development of a Monte Carlo based robustness calculation and evaluation tool," *Med. Phys.*, 2022.
- [92] S. Van Der Voort, S. Van De Water, Z. Perkó, B. Heijmen, D. Lathouwers, and M. Hoogeman, "Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 95, no. 1, pp. 163–170, 2016.
- [93] S. E. McGowan, F. Albertini, S. J. Thomas, and A. J. Lomax, "Defining robustness protocols: A method to include and evaluate robustness in clinical plans," *Phys. Med. Biol.*, vol. 60, no. 7, pp. 2671–2684, 2015.
- [94] M. Wedenberg, C. Beltran, A. Mairani, and M. Alber, "Advanced Treatment Planning," *Med. Phys.*, vol. 45, no. 11, pp. 1011–1023, 2018.
- [95] H. E. Morgan and D. J. Sher, "Adaptive radiotherapy for head and neck cancer," *Cancers Head Neck* 2020 51, vol. 5, no. 1, pp. 1–16, Jan. 2020.
- [96] C. Kurz *et al.*, "Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images," *Radiat. Oncol.*, vol. 11, no. 1, pp. 1–9, Apr. 2016.

- [97] N. Deiter, F. Chu, N. Lenards, A. Hunzeker, K. Lang, and D. Mundy, "Evaluation of replanning in intensity-modulated proton therapy for oropharyngeal cancer: Factors influencing plan robustness," *Med. Dosim.*, vol. 45, no. 4, pp. 384–392, Dec. 2020.
- [98] N. T. C. Fung, W. M. Hung, C. K. Sze, M. C. H. Lee, and W. T. Ng, "Automatic segmentation for adaptive planning in nasopharyngeal carcinoma IMRT: Time, geometrical, and dosimetric analysis," *Med. Dosim.*, vol. 45, no. 1, pp. 60–65, Mar. 2020.
- [99] D. Tilly, Å. Holm, E. Grusell, and A. Ahnesjö, "Probabilistic optimization of dose coverage in radiotherapy," *Phys. Imaging Radiat. Oncol.*, vol. 10, no. March, pp. 1–6, 2019.
- [100] Y. Z. Szeto, M. G. Witte, M. van Herk, and J. J. Sonke, "A population based statistical model for daily geometric variations in the thorax," *Radiother. Oncol.*, vol. 123, no. 1, pp. 99–105, 2017.
- [101] Q. Zhang *et al.*, "A patient-specific respiratory model of anatomical motion for radiation treatment planning," *Med. Phys.*, vol. 34, no. 12, pp. 4772–4781, 2007.
- [102] S. Dhou, J. Lewis, W. Cai, D. Ionascu, and C. Williams, "Quantifying day-to-day variations in 4DCBCT-based PCA motion models," *Biomed. Phys. Eng. Express*, vol. 6, no. 3, p. 035020, Apr. 2020.
- [103] S. Thörnqvist *et al.*, "Adaptive radiotherapy in locally advanced prostate cancer using a statistical deformable motion model," *Acta Oncol. (Madr.)*, vol. 52, no. 7, pp. 1423–1429, 2013.
- [104] O. Weistrand and S. Svensson, "The ANACONDA algorithm for deformable image registration in radiotherapy," *Med. Phys.*, vol. 42, no. 1, pp. 40–53, Jan. 2015.
- [105] M. Modat, D. M. Cash, P. Daga, G. P. Winston, J. S. Duncan, and S. Ourselin, "Global image registration using a symmetric block-matching approach," *J. Med. Imaging*, vol. 1, no. 2, p. 024003, Sep. 2014.
- [106] B. C. Lowekamp, D. T. Chen, L. Ibáñez, D. Blezek, and H. J. Johnson, "The Design of SimpleITK," 2013.

- [107] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” 2011.
- [108] C. Veiga *et al.*, “Toward adaptive radiotherapy for head and neck patients: Uncertainties in dose warping due to the choice of deformable registration algorithm,” *Med. Phys.*, vol. 42, no. 2, pp. 760–769, Feb. 2015.
- [109] O. Hamming-Vrieze *et al.*, “Impact of setup and range uncertainties on TCP and NTCP following VMAT or IMPT of oropharyngeal cancer patients,” *Phys. Med. Biol.*, vol. 64, no. 9, p. 095001, Apr. 2019.
- [110] T. Bostel *et al.*, “Dosimetric impact of interfractional variations in prostate cancer radiotherapy—implications for imaging frequency and treatment adaptation,” *Front. Oncol.*, vol. 9, no. SEP, Sep. 2019.
- [111] R. A. Rozendaal, B. J. Mijnheer, O. Hamming-Vrieze, A. Mans, and M. Van Herk, “Impact of daily anatomical changes on EPID-based in vivo dosimetry of VMAT treatments of head-and-neck cancer,” *Radiother. Oncol.*, vol. 116, no. 1, pp. 70–74, Jul. 2015.
- [112] D. Yan, D. A. Jaffray, and J. W. Wong, “A Model to Accumulate Fractionated Dose in a Deforming Organ,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 44, no. 3, pp. 665–675, 1999.
- [113] E. Fontenla, C. A. Pelizzari, J. C. Roeske, and G. T. Y. Chen, “Numerical analysis of a model of organ motion using serial imaging measurements from prostate radiotherapy,” *Phys. Med. Biol.*, vol. 46, no. 9, pp. 2337–2358, 2001.
- [114] E. Fontenla, C. A. Pelizzari, J. C. Roeske, and G. T. Y. Chen, “Using serial imaging data to model variabilities in organ position and shape during radiotherapy,” *Phys. Med. Biol.*, vol. 46, no. 9, pp. 2317–2336, 2001.
- [115] T. E. Marchant, C. J. Moore, C. G. Rowbottom, R. I. MacKay, and P. C. Williams, “Shading correction algorithm for improvement of cone-beam CT images in radiotherapy,” *Phys. Med. Biol.*, vol. 53, no. 20, p. 5719, Sep. 2008.
- [116] M. Rosu and G. D. Hugo, “Advances in 4D Radiation Therapy for Managing Respiration: Part II – 4D Treatment Planning,” *Z. Med. Phys.*, vol. 22, no. 4, p. 272,

Dec. 2012.

- [117] E. W. Korevaar *et al.*, “Practical robustness evaluation in radiotherapy – A photon and proton-proof alternative to PTV-based plan evaluation,” *Radiother. Oncol.*, vol. 141, pp. 267–274, Dec. 2019.
- [118] M. Van Herk, “Errors and Margins in Radiotherapy,” *Semin. Radiat. Oncol.*, vol. 14, no. 1, pp. 52–64, 2004.
- [119] A. McWilliam, W. Beasley, and C. G. Rowbottom, “Relationship between geometric and dosimetric accuracy of auto-contouring in head and neck VMAT treatment planning,” *Radiother. Oncol.*, vol. 115, p. S489, Apr. 2015.
- [120] H. Nourzadeh, W. T. Watkins, M. Ahmed, C. Hui, D. Schlesinger, and J. V. Siebers, “Clinical adequacy assessment of autocontours for prostate IMRT with meaningful endpoints:,” *Med. Phys.*, vol. 44, no. 4, pp. 1525–1537, Apr. 2017.
- [121] S. Martin *et al.*, “Impact of target volume segmentation accuracy and variability on treatment planning for 4D-CT-based non-small cell lung cancer radiotherapy,” *Acta Oncol. (Madr.)*, vol. 54, no. 3, pp. 322–332, Mar. 2015.
- [122] E. Aliotta, H. Nourzadeh, and J. Siebers, “Quantifying the dosimetric impact of organ-at-risk delineation variability in head and neck radiation therapy in the context of patient setup uncertainty,” *Phys. Med. Biol.*, vol. 64, no. 13, p. 135020, Jul. 2019.
- [123] E. Shaw *et al.*, “Radiation Therapy Oncology Group: radiosurgery quality assurance guidelines,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 27, pp. 1231–1239, 1993.
- [124] J. Robbins, E. Vasquez Osorio, A. Green, A. McWilliam, A. McPartlin, and M. Van Herk, “EP-1816 A robustness comparison of margin based and robust plans for head and neck VMAT patients,” *Radiother. Oncol.*, vol. 133, pp. S984–S985, Apr. 2019.
- [125] T. Brinkhoff, H. Kriegel, R. Schneider, and A. Braun, “Measuring the Complexity of Polygonal Objects,” *ACM-GIS*, vol. 109, 1995.
- [126] J. M. Pakela, M. M. Matuszak, R. K. Ten Haken, D. L. McShan, and I. El Naqa, “Dynamic stochastic deep learning approaches for predicting geometric changes in head and neck cancer,” *Phys. Med. Biol.*, vol. 66, no. 22, p. 225006, Nov. 2021.

- [127] G. Palma, S. Monti, and L. Cella, "Voxel-based analysis in radiation oncology: A methodological cookbook," *Phys. Med.*, vol. 69, pp. 192–204, Jan. 2020.
- [128] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels, "Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration," *IEEE Trans. Med. Imaging*, vol. 30, no. 2, pp. 251–265, Feb. 2011.